The dynamics of stochastic mono-molecular reaction systems in stochastic environments

Cappelletti, Daniele; Pal Majumder, Abhishek; Wiuf, Carsten

Published in:
Stochastic Processes and Their Applications

DOI:
10.1016/j.spa.2021.03.010

Publication date:
2021

Document version
Early version, also known as pre-print

Citation for published version (APA):
The dynamics of stochastic mono-molecular reaction systems in stochastic environments

Daniele Cappelletti, Abhishek Pal Majumder, Carsten Wiuf

July 20, 2022

Keywords: Markov chain, Markov modulated process, Stochastic recurrence equation, Gene regulation, Mono-molecular reaction network, Stationary distribution.

Abstract

We study the stochastic dynamics of a system of interacting species in a stochastic environment by means of a continuous-time Markov chain with transition rates depending on the state of the environment. Models of gene regulation in systems biology take this form. We characterise the finite-time distribution of the Markov chain, provide conditions for ergodicity, and characterise the stationary distribution (when it exists) as a mixture of Poisson distributions. The mixture measure is uniquely identified as the law of a fixed point of a stochastic recurrence equation. This recursion is crucial for statistical computation of moments and other distributional features.

1 Introduction

Reaction networks are widely used in system biology to describe the evolution of interacting molecular species. Though the nature of interaction is motivated by examples of (bio)chemical reactions, similar models are considered in genetics [15], epidemiology [28], and ecology [17]. In recent years, stochastic models of reaction networks, based on continuous time Markov chains (CTMCs), have become fashionable, as the counts of molecular species in experimental settings might be low or fluctuate considerably over time. In addition to this, a system of interacting species might itself be embedded into a stochastic environment that affects the reaction propensities and the availability of resources.

A particular focus of stochastic reaction network theory has been to understand the long term behaviour of a model. For a class of reaction networks that has a ‘complex balanced’ equilibrium in a deterministic sense, the stationary distribution has a Poisson product-form [1, 2, 9], akin to results in queuing theory. Another important class of reaction networks, closely related to complex balanced networks and also considered in this paper, is that of mono-molecular reaction networks [16, 20]. For this class, not only the stationary distribution but also the finite-time distributions can be determined. Specifically, the latter takes the form of a convolution of Multinomial and Poisson distributions with intensities evolving in time according to an ODE system (which itself might have time-dependent rate constants). If the rate constants are independent of time, then the intensities of the stationary distribution are determined by the equilibrium of the ODE system.

In this work we study the evolution of a stochastic mono-molecular reaction network in a stochastic environment. In the biological context the stochastic environment might itself be a reaction network, however, we assume a more general setting. As an example, consider the reaction network

\[ 0 \xrightarrow{\kappa_1} E_1, \quad 0 \xrightarrow{\kappa_2} E_2, \quad E_1 \xrightarrow{\kappa_3} E_1 + S, \quad E_2 + S \xrightarrow{\kappa_4} E_2. \]
The first two reactions form a fluctuating stochastic environment, where the molecules (resources) $E_1$ and $E_2$ continuously are produced and degraded. If $E_1$ and/or $E_2$ are present they catalyse the production of the substrate $S$ and its subsequent degradation. The availability of $E_1, E_2$ is determined by the first two reversible reactions which occur independently of the number of substrate molecules $S$. The production rate of $S$ depends linearly on the number of $E_1$ molecules, while the degradation of $S$ depends linearly on the number of $E_2$ molecules as well as the number of substrate molecules. Biologically, it is an example of a non-mutual symbiotic relationship (parasitism) where the presence of the parasite species (here $S$) depends on the presence of the host species (here $E_1, E_2$), but not vice versa.

We interpret a stochastic reaction network in a stochastic environment (to be defined in Section 2) as a Markov process $\{(X(t), Z(t)) : t \geq 0\}$ on a joint state space $\Gamma \times \mathbb{Z}_{\geq 0}^d$. The marginal process $\{X(t) : t \geq 0\}$ is itself a Markov process on the state space $\Gamma$ and constitutes the environment. The process counting the number of molecules of the species in the reaction network is $\{Z(t) : t \geq 0\}$ with state space $\mathbb{Z}_{\geq 0}^d$. Recent work has been done to understand (exponential) ergodicity in similar, though different, settings [10, 32, 33], where the environmental process is a switching process between regimes.

In the setting of a stochastic reaction network in a stochastic environment, we derive the finite-time distributions of $Z(t)$, conditional on the trajectories $X(s), 0 \leq s \leq t$ (in fact, this result does not rely on the Markovian property of the environment, as discussed in Remark 5.1). Specifically, we show that the finite-time distributions are obtained as a convolution of multinomial and Poisson distributions, similar to the mono-molecular reaction networks in a constant or time-dependent environment discussed above. Furthermore, we give conditions under which the joint process $\{(X(t), Z(t)) : t \geq 0\}$ is ergodic in terms of ergodicity of the environmental process $\{X(t) : t \geq 0\}$ and structural conditions on the process $\{Z(t) : t \geq 0\}$. In recent work [18], an approach based on linear Lyapunov functions provides sufficient conditions for the existence and uniqueness of a stationary distribution of large class of stochastic reaction networks. It is noteworthy, that these results remain inconclusive for simple examples in our setting.

An explicit characterisation of the stationary distribution is however not available. In Theorem 6.8 we interpret the stationary distribution of the mono-molecular species as a mixture of Poisson distributions, thus providing an example of a Poisson representation [17][Section 7.7]. We show that the mixing measure appears as the solution of a random affine equation (also called a stochastic recurrence equation [8]). Stochastic recurrence equations have been studied in other contexts, see for example [4]. In our case, it is quite remarkable that the randomness of the environment is reflected in the solution of the stochastic recurrence equation only through path-wise functionals (for simple cases they are of integral forms). We demonstrate the usefulness of the stochastic recurrence equation by providing a simulation scheme to simulate from the stationary distribution of the process $\{Z(t) : t \geq 0\}$ given the state of the environment.

This paper is organised as follows. In Section 2 we provide background and notation, and introduce the processes we are interested in formally. In Section 3 an example is given, where the exact finite-time distribution of the counting process $\{Z(t) : t \geq 0\}$ can be computed. We generalise the example in Section 5 and discuss the long time behaviour in Section 6. A simulation scheme using the invariant measure is provided in Section 7 and examples are given in Section 8.

## 2 Background and definitions

In the following, $\mathbb{Z}, \mathbb{Z}_{\geq 0}$, and $\mathbb{Z}_{\geq 0}$ denote the integers, the non-negative integers, and the positive integers, respectively. Similarly, $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ denote the real and the non-negative real numbers. For any $u \in \mathbb{R}$, $[u] = \max\{m \in \mathbb{Z} : m \leq u\}$ denotes the largest integer smaller than or equal to $u$. We denote the $i$-th unit vector in $\mathbb{R}^n$ by $e_i$ and let $e = \sum_{i=1}^n e_i$ be the vector whose entries are all one. If $u, v \in \mathbb{R}^n$, we write $u \leq v$ or $u \geq v$ if the inequality holds component-wise. Moreover, define

$$u! = \prod_{i=1}^n u_i! \quad \text{and} \quad u^v = \prod_{i=1}^n u_i^{v_i}.$$  

The $n \times n$ identity matrix is denoted by $I_n$. If $A = (a_{ij})_{n \times n}$ is a real $n_1 \times n_2$ matrix, then $A^\top$ denotes the transpose matrix, and $\|A\|_1 = \max_{j=1}^{n_2} \sum_{i=1}^{n_1} |a_{ij}|$ denotes the $L_1$-matrix norm of $A$. In particular, if
v ∈ \mathbb{R}^n \text{ then } \|v\|_1 = \sum_{i=1}^n |v_i|$. For square matrices we define
\[
\prod_{i=m_1}^{m_2} A_i = \begin{cases} A_{m_2}A_{m_2-1}\ldots A_{m_1} & \text{if } m_1 \leq m_2, \\
A_{m_2}A_{m_2+1}\ldots A_{m_1} & \text{otherwise.}
\end{cases}
\]
The extremes $m_1$ and $m_2$ are potentially infinite. Note the matrices are ordered from upper to lower index.

For a random variable $W$, $\mathcal{L}(W)$ denotes the law of $W$. Given two random variables $W_1$ and $W_2$, defined on the same probability space, we write $W_1 \parallel W_2$ if they are independent, and $W_1 \sim W_2$ if $\mathcal{L}(W_1) = \mathcal{L}(W_2)$. Given two probability distributions $\mathcal{L}_1$ and $\mathcal{L}_2$ on $\mathbb{R}^n$ (or a subset thereof), their convolution is denoted by $\mathcal{L}_1 * \mathcal{L}_2$. Let
\[
\bigotimes_{i=1}^n \mathcal{L}_i = \mathcal{L}_1 \ast \cdots \ast \mathcal{L}_n.
\]

Convergence in distribution is denoted $W_i \xrightarrow{d} W$ for $i \to \infty$, where $i \in \mathbb{Z}_{\geq 0}$ or $i \in \mathbb{R}_{\geq 0}$. A stochastic process $\{W(t) : t \geq 0\}$ with values in a metric space $H$ is tight if for $\varepsilon > 0$ there exists a compact set $M_\varepsilon \subseteq H$ such that $\sup_{t \in \mathbb{R}_{\geq 0}} P(W(t) \notin M_\varepsilon) \leq \varepsilon$. A continuity set $A$ of a random variable $W$ is a measurable set such that $P(W \in \partial A) = 0$, where $\partial A$ denotes the boundary of the set $A$.

Let $P_2(x, \gamma) : \mathcal{A} \times \mathcal{F} \to [0, 1]$, $t > 0$, be the Markov transition kernel of a CTMC $\{X(t) : t \geq 0\}$ on a measurable space $(\mathcal{A}, \mathcal{F})$. If all states communicate with each other, the chain is said to be irreducible. A stationary distribution for $\{X(t) : t \geq 0\}$ is a probability measure $\pi$ on $(\mathcal{A}, \mathcal{F})$ satisfying $\int_{\mathcal{A}} \int_{\mathcal{F}} P_2(x, dy) d\pi(x) = \pi(B)$ for all $B \in \mathcal{F}$ and $t > 0$. The process $\{X(t) : t \geq 0\}$ is said to be ergodic for every initial condition if for any $x \in \mathcal{A}$ there exists a probability measure $\pi_x$ on $(\mathcal{A}, \mathcal{F})$ such that $\pi_x(A) = \lim_{t \to \infty} P_2(x, A)$ for all $A \in \mathcal{F}$. The probability measures $\pi_x$ are necessarily stationary distributions. If $\pi_x$ is independent of $x$ then the process is said to be ergodic. If the state space is irreducible and the process is ergodic then $\pi_x$ is the unique stationary distribution on $(\mathcal{A}, \mathcal{F})$.

Next we introduce notation for probability distributions and random variables that will appear in various results. $\text{Exp}(\mu)$ with $\mu > 0$ denotes an exponential random variable with mean $\mu$. For $m, n \in \mathbb{Z}_{\geq 0}$, and $p = (p_1, \ldots, p_n) \in [0, 1]^n$ with $\sum_{j=1}^n p_j \leq 1$, $\text{Multi}(m, p)$ denotes an $\{x \in \mathbb{Z}_{\geq 0}^n : \sum_{i=1}^n x_i \leq m\}$-valued random variable such that
\[
P(\text{Multi}(m, p) = (i_1, \ldots, i_n)) = \frac{m!}{i_1! \ldots i_n!(m - \sum_{j=1}^n i_j)!} p_1^{i_1} \cdots p_n^{i_n} \left(1 - \sum_{j=1}^n p_j\right)^{m - \sum_{j=1}^n i_j}.
\]
Note that the latter is not the usual notation for a multinomial random variable and the multinomial coefficient. In particular, if $n = 1$, then $\text{Multi}(m, p)$ is a binomial random variable, $\text{Bin}(m, p)$. For $m \in \mathbb{Z}_{\geq 0}$, $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{R}_{\geq 0}^n$, we let $\text{Pois}(\mu)$ denote an $\mathbb{Z}_{\geq 0}^n$-valued random variable such that
\[
P(\text{Pois}(\mu) = (i_1, \ldots, i_n)) = \prod_{j=1}^n e^{-\mu_j} \frac{\mu_j^{i_j}}{i_j!}.
\]
Namely, $\text{Pois}(\mu)$ is distributed as $n$ independent Poisson variables with non-negative rates.

## 2.1 Stochastic reaction systems with stochastic environments

A reaction network with species set $S = \{S_1, \ldots, S_d\}$ is a set of reactions $\mathcal{R} = \{y_1 \to y'_1, y_2 \to y'_2, \ldots, y_k \to y'_k\}$, where $y_r = \sum_{j=1}^d \gamma_{rj} S_j$, $y'_r = \sum_{j=1}^d \gamma'_{rj} S_j$ are linear non-negative integer combinations of species. The left hand side of a reaction is called the reactant, the right hand side, the product, both are complexes, and $\xi_r = y'_r - y_r \in \mathbb{R}^d$ is the reaction vector, the net gain of species in a reaction. We assume that all species take part in at least one reaction and that the reactant and product sides are never identical.

The evolution of the species counts is usually modelled as a homogeneous CTMC $\{Z(t) : t \geq 0\}$ with state space $\mathbb{Z}_{\geq 0}^d$ [3, 13]. Here, we consider a generalisation of the standard setting, assuming the transition rates of $\{Z(t) : t \geq 0\}$ change over time in a stochastic way.
Definition 2.1. A stochastic reaction system with stochastic environment is a triple \( (\mathcal{R}, \Lambda, \{X(t): t \geq 0\}) \) such that:

- \( \mathcal{R} \) is a reaction network,
- \( \{X(t): t \geq 0\} \) is a homogeneous CTMC with irreducible discrete state space \( \Gamma \). Furthermore, we assume \( \{X(t): t \geq 0\} \) is non-explosive (or regular), implying that the number of jumps in a bounded time interval is almost surely (a.s.) finite [26].
- \( \Lambda = (\lambda_1, \ldots, \lambda_k) \) is a vector of \( k \) reaction rate functions, referred to as a kinetics, such that \( \lambda_r: \Gamma \times \mathbb{R}_{\geq 0}^d \to \mathbb{R}_{\geq 0} \) is not constantly zero, and \( \lambda_r(x, z) > 0 \) only if \( z \geq y_r, r = 1, \ldots, k \).

The evolution of the species counts of \( \mathcal{S} \) is modelled by the stochastic process \( \{Z(t): t \geq 0\} \), such that \( \{(X(t), Z(t)): t \geq 0\} \) is a homogeneous CTMC with state space \( \Gamma \times \mathbb{Z}_{\geq 0}^d \), and transition rate function satisfying the following. Let \( (x, z), (x', z') \in \Gamma \times \mathbb{Z}_{\geq 0}^d \).

1. The transition rate from \( (x, z) \) to \( (x', z') = (x, z + \xi) \) is

\[
\sum_{r: y_r \to y'_r \in \mathcal{R}} \lambda_r(x, z).
\]

2. The transition rate from \( (x, z) \) to \( (x', z') = (x', z) \) equals the transition rate of \( \{X(t): t \geq 0\} \) from \( x \) to \( x' \), and thus does not depend on \( z \).

3. All other transition rates are zero.

While we assume that \( \{X(t): t \geq 0\} \) is non-explosive, we do not require the same for \( \{(X(t), Z(t)): t \geq 0\} \). In case an explosion occurs for \( \{(X(t), Z(t)): t \geq 0\} \), then infinitely many jumps occur for \( \{Z(t): t \geq 0\} \) in a finite time. In such a case, we understand the process as living in a “cemetery space" \( \Gamma \times \Delta \) from the explosion time on. However, while an example of explosive stochastic reaction system with stochastic environment can be seen in Example 4.2, models satisfying Assumption 1 (which are the main object of our analysis) are in fact non-explosive.

By definition \( \{X(t): t \geq 0\} \) and \( \{(X(t), Z(t)): t \geq 0\} \) are CTMCs, but not necessarily \( \{Z(t): t \geq 0\} \). The requirements on the kinetics ensure \( \mathbb{Z}_{\geq 0}^d \) is invariant for \( \{Z(t): t \geq 0\} \).

If the rate functions \( \lambda_r \) do not depend on the stochastic environment \( \{X(t): t \geq 0\} \), that is, \( \lambda_r(x, z) = \lambda_r(z) \) for \( 1 \leq r \leq k \), then we simply refer to \( (\mathcal{R}, \Lambda) \) as a stochastic reaction system. This definition coincides with standard terminology [3, 13].

Let \( (F_s^X)_{s \geq 0} \) denote the filtration of \( \{X(t): t \geq 0\} \). Then, it follows from the above definition that for \( 0 < s \leq t \) the random variable \( X(t) \) is conditionally independent of \( Z(s) \) given \( F_s^X \), that is,

\[
X(t) \perp Z(s) \bigg| F_s^X \quad \text{for} \quad s \leq t.
\]  

The property (2.1) can be recast as \( \mathcal{L}(X(t)|X(s), Z(s)) = \mathcal{L}(X(t)|X(s)) \). Models with a similar conditional structure are studied in [6].

Following [24], we might write

\[
Z(t) = Z(0) + \sum_{r=1}^{k} N_r \left( \int_0^t \lambda_r(X(s), Z(s)) \, ds \right) \xi_r,
\]

where \( \{N_r: r = 1, \ldots, k\} \) is a set of i.i.d. unit-rate Poisson processes, independent of \( \{X(t): t \geq 0\} \), and \( Z(0) \) and \( Z(0) \) is the initial state at time zero. An equivalent description of \( \{Z(t): t \geq 0\} \) is the following: Conditioned on the path of \( \{X(t): t \geq 0\} \), \( X(s) = x(s) \) for all \( s \in [0, t] \), the process \( \{Z(t): t \geq 0\} \) is a non-homogeneous CTMC

\[
Z(t) = Z(0) + \sum_{r=1}^{k} N_r \left( \int_0^t \lambda_r(x(s), Z(s)) \, ds \right) \xi_r, \quad \text{a.s.}
\]  

(2.2)
A common choice of kinetics is stochastic mass-action kinetics, which in our setting takes the form

$$\lambda_r(x, z) = \kappa_r(x) \frac{z!}{(z - y_r)!} 1_{\{x \in \mathbb{Z}_{\geq 0}^i: x \geq y_r\}}(z), \quad r = 1, \ldots, k,$$

for some functions $\kappa_r: \Gamma \rightarrow \mathbb{R}_{\geq 0}$, $r = 1, \ldots, k$. Mass-action kinetics corresponds to the hypothesis that the system is well-stirred such that the propensity of a reaction is proportional to the number of ordered ways the molecules in the reactant can be chosen. A stochastic reaction system (with stochastic environment) with mass-action kinetics is called a stochastic mass-action system (with stochastic environment).

**Example 2.1.** A typical biological situation is that of a system of non-mutual symbiotic interactions where some species evolve conditionally on the presence of other species, the stochastic environment. As an example, consider the reaction network

$$0 \xrightarrow{\lambda_1} mS_3, \quad S_2 \xrightarrow{\lambda_3} S_2 + mS_3, \quad S_1 + S_3 \xrightarrow{\lambda_6} S_1,$$

where $m$ is an integer. The set of reactions is enumerated according to the index of $\lambda_r$, written above (or below) the arrow of $y_r \rightarrow y'_r$. Assume the species counts evolve according to a stochastic mass-action system with constant environment, that is, with transition rates

$$\lambda_1(w) = \kappa_1, \quad \lambda_2(w) = \kappa_2w_1, \quad \lambda_3(w) = \kappa_3,$$

$$\lambda_4(w) = \kappa_4w_2, \quad \lambda_5(w) = \kappa_5w_2, \quad \lambda_6(w) = \kappa_6w_1w_3,$$

with $\kappa_r > 0$ and $w = (w_1, w_2, w_3) \in \mathbb{Z}_{\geq 0}^3$. Let \{Y(t): t \geq 0\} be the associated CTMC.

The number of $S_3$ molecules does not affect the numbers of $S_1$ and $S_2$ molecules. Therefore, we might regard the latter as constituting a stochastic environment, which affects the production and degradation of $S_3$. Namely, consider \{X(t): t \geq 0\} = \{Y_1(\cdot), Y_2(\cdot)\} with state space $\Gamma = \mathbb{Z}_{\geq 0}^3$ and \{Z(t): t \geq 0\} = $Y_3(\cdot)$. In this case, the reaction graph associated with \{Z(t): t \geq 0\} is simply

$$0 \xrightarrow{\tilde{\lambda}_1} mS_3, \quad S_3 \xrightarrow{\tilde{\lambda}_2} 0, \quad \tilde{\lambda}_1(x, z) = \kappa_5x_2, \quad \tilde{\lambda}_2(x, z) = \kappa_6x_1z.$$

The process \{X(t): t \geq 0\} does not have to be confined to non-negative vectors. For example, rescaled species counts could be considered in the spirit of multiscale analysis [22]. Moreover, even negative states of \{X(t): t \geq 0\} could be considered to model the effect of temperature on the transition rates of \{Z(t): t \geq 0\}.

**Remark 2.1.** In the example, the rate of consumption of $S_3$ depends linearly on the number of $S_3$ molecules. This particular situation with linear transition rates will be important later. In general, consider a species $S$ that is consumed as $S \rightarrow y'$ with linear transition rate, and where $y'$ is a complex not involving $S$ (consumption might here be degradation as in $S \rightarrow 0$ or conversion/transformation where $S$ is converted into other molecules). Then, one might construct the process \{Z(t): t \geq 0\} such that the present molecules of $S$ are consumed conditionally independent of each other, given \{X(t): t \geq 0\}. In fact, the reaction rate function of $S \rightarrow y'$ is of the form $\lambda(X(t), z) = \kappa(X(t))z_i$ (assuming $S$ is the $i$-th species). This rate might be constructed from $z_i$ independent exponentially distributed random variables with rate $\kappa(X(t))$.

### 3 A case study

We begin our formal analysis with an example, essentially expanding Example 2.1. We show that the distribution of $Z(t)$ can be explicitly computed in terms of the path \{X(s): 0 \leq s \leq t\}. This can be accomplished in general assuming the reaction rate functions are linear in $Z(t)$, see Section 4.

Consider a stochastic reaction system, \{(R, \Lambda, \{X(t): t \geq 0\})\},

$$0 \xrightarrow{\lambda_1} mS, \quad S \xrightarrow{\lambda_2} 0, \quad \lambda_1(x, z) = \kappa_1(x), \quad \lambda_2(x, z) = \kappa_2(x)z,$$

for fixed $m \in \mathbb{Z}_{\geq 0}$. The distribution of $Z(t)$, given $X^t$, is a convolution of simple distributions. Specifically, Proposition 3.1 shows that the variable $Z(t)$ is distributed as a sum $N_0 + N_1 + \ldots + mN_m$ of conditionally
independent random variables, given $\mathcal{F}_t^X$. Here $N_0$ counts the number of $S$ molecules that were present at time 0 and survived until time $t$, and $N_i, \ i = 1, \ldots, m$, counts the number of birth events of $m$ $S$ molecules, where $i$ of them survive until time $t$. Note that this conditional independence is expected by Remark 2.1.

A similar proof of Proposition 3.1 (and of the more general Proposition 5.3) for the case $m = 1$ can be found in [20, proof of Theorem 1]. A weaker version of [20, proof of Theorem 1] is stated in this paper as Theorem 5.1. The subtle difference is that a deterministic (though not necessarily constant) environment $\{X(t) : t \geq 0\}$ is considered in [20], and in Proposition 3.1 (and later, in Proposition 5.3) we condition on $\mathcal{F}_t^X$. We provide here an independent proof to illustrate the general idea used in the example.

**Proposition 3.1.** For the model described above, the conditional distribution of $Z(t)$ given $\mathcal{F}_t^X$, for $t > 0$, is

$$\mathcal{L} \left( \text{Bin} \left( Z(0), e^{-\int_0^t \nu_2(X(s)) ds} \right) \bigg| \mathcal{F}_t^X \right) \ast \left( \bigotimes_{i=1}^m \mathcal{L}(iN_i|\mathcal{F}_t^X) \right),$$

where $N_i, \ i = 1, \ldots, m$, are Poisson random variables with mean

$$(m) \int_0^t \nu_1(X(u)) e^{-\int_0^u \nu_2(X(s)) ds} (1 - e^{-\int_u^t \nu_2(X(s)) ds})^{m-i} du, \quad i = 1, \ldots, m. \quad (3.1)$$

**Proof.** We begin with the case $m = 1$. We show that $\mathcal{L}(Z(t)|\mathcal{F}_t^X)$ is

$$\mathcal{L} \left( \text{Bin} \left( Z(0), e^{-\int_0^t \nu_2(X(s)) ds} \right) \bigg| \mathcal{F}_t^X \right) \ast \mathcal{L} \left( \text{Pois} \left( \int_0^t \nu_1(X(u)) e^{-\int_0^u \nu_2(X(s)) ds} du \right) \bigg| \mathcal{F}_t^X \right). \quad (3.2)$$

Conditioned on $\mathcal{F}_t^X$, the process $\{Z(s) : 0 \leq s \leq t\}$ evolves as a non-homogeneous CTMC in the sense of (2.2). Moreover, the degradation events of the $S$ molecules are conditionally independent of each other, see Remark 2.1. Specifically, the time until degradation of an $S$ molecule is exponentially distributed with time dependent rate. Hence, given $\mathcal{F}_t^X$, the number of $S$ molecules that survived until time $t$ among the $Z(0)$ molecules present at time 0 has distribution

$$\mathcal{L} \left( \text{Bin} \left( Z(0), e^{-\int_0^t \nu_2(X(s)) ds} \right) \bigg| \mathcal{F}_t^X \right),$$

which accounts for the first term in (3.2). We will next show that given $\mathcal{F}_t^X$, the total number of surviving $S$ molecules at time $t$ out of those born in $(0,t]$ has distribution

$$\mathcal{L} \left( \text{Pois} \left( \int_0^t \nu_1(X(u)) e^{-\int_0^u \nu_2(X(s)) ds} du \right) \bigg| \mathcal{F}_t^X \right). \quad (3.3)$$

Let the total number of birth instances of $S$ molecules in $[0,t]$ be $B_t$, which conditioned on $\mathcal{F}_t^X$ follows a Poisson distribution with mean $\lambda_t = \int_0^t \nu_1(X(u)) du$. Birth instances conditioned on $B_t$ and $\mathcal{F}_t^X$ can be thought of as independent realisations from the density

$$\frac{\nu_1(X(\cdot))}{\int_0^t \nu_1(X(u)) du}. \quad (3.4)$$

Each of these particles survives until time $t$ with conditional probability

$$P_t = \frac{\int_0^t \nu_1(X(u)) e^{-\int_0^u \nu_2(X(s)) ds} du}{\int_0^t \nu_1(X(u)) du},$$

where the exponential term is the probability of survival given a molecule is born at time $u$. Multiplying $P_t$ with $\lambda_t$ proves (3.3) by thinning of Poisson processes, and in turn (3.2) because the $S$ molecules present at time 0 and those born afterwards behave conditionally independently given $\mathcal{F}_t^X$, see Check: Remark 2.1.

We generalise (3.2) to arbitrary $m > 1$. The fate of the initial $Z(0)$ molecules is described similarly to the case $m = 1$. For each birth occurrence (that is, a firing of the reaction $0 \rightarrow mS$) there are $m$ molecules of $S$ being born and among them either 0, 1, ... or $m$ will survive until time $t$. Given $\mathcal{F}_t^X$, and again due
to the conditional independence of the behaviour of the molecules, the probability that exactly \( j \) out of the \( m \) molecules of \( S \) that are born at time \( u \in (0, t] \) survive until time \( t \) is

\[
P_u^t(j) = \binom{m}{j} e^{-j \int_0^t \kappa_2(X(s)) \, ds} \left(1 - e^{-\int_0^t \kappa_2(X(s)) \, ds}\right)^{m-j}.
\]

Let the unordered birth instances of \( S \) molecules in \((0, t]\) be \( U_1, \ldots, U_{B_t} \). Define \( Y(t) = \sum_{i=1}^{B_t} Y_i \), where \( Y_i \) is the number of molecules surviving up to time \( t \) among those born at time \( U_i \in (0, t]. \) Note that the random variables \( Y_i \) are conditionally independent and identically distributed given \( F_t^X \) and \( B_t \), as they are not ordered and the molecules of \( S \) degrade independently on each other, given \( F_t^X \) (see Remark 2.1). Hence, the Laplace transform of the surviving \( S \) molecules that are born in \((0, t]\) is

\[
E \left( e^{-\gamma Y(t)} \middle| F_t^X \right) = \sum_{r=0}^{\infty} P(B_t = r \middle| F_t^X) \prod_{i=1}^{r} E \left( e^{-\gamma Y_i(t)} \middle| B_i = r, F_t^X \right) \\
= \sum_{r=0}^{\infty} e^{-\int_0^t \kappa_1(X(u)) \, du} \frac{\left(\int_0^t \kappa_1(X(u)) \, du\right)^r}{r!} \left[ \int_0^t \left(\frac{\kappa_1(X(u))}{\int_0^t \kappa_1(X(s)) \, ds} \sum_{j=0}^{m} P_u^t(j) e^{-\gamma j} \right) \, du \right]^r \\
= e^{-\int_0^t \kappa_1(X(u)) \, du} e^{\sum_{j=0}^{m} \int_0^t \kappa_1(X(u)) P_u^t(j) e^{-\gamma j} \, du} \\
= e^{-\sum_{j=0}^{m} (1-e^{-\gamma j}) \int_0^t \kappa_1(X(u)) P_u^t(j) \, du}, \tag{3.5}
\]

where for the last equation we use \( P_u^t(0) = 1 - \sum_{j=1}^{m} P_u^t(j) \). Hence, \( \mathcal{L}(Y(t) \middle| F_t^X) = \mathcal{L}(\sum_{i=1}^{m} \mathcal{L}(iN_i \middle| F_t^X)) \), where \( N_i, i = 1, \ldots, m, \) are Poisson random variables with mean \( \{3.1\} \). The proof is concluded by

\[
\mathcal{L}(Z(t) \middle| F_t^X) = \mathcal{L} \left( \text{Bin} \left( Z(0), e^{-\int_0^t \kappa_2(X(s)) \, ds} \right) \middle| F_t^X \right) * \mathcal{L}(Y(t) \middle| F_t^X).
\]

### 4 General case of study

The main aim is to study ergodicity (for any initial condition) of a stochastic reaction system with stochastic environment, assuming that \( \{X(t) : t \geq 0\} \) is an ergodic process. To do so, we consider a specific family of models for which the stationary distribution can be characterised explicitly. To motivate the necessity of further assumptions, we give two examples. The first is a transient reaction system, regardless of the ergodicity of the process \( \{X(t) : t \geq 0\} \). The second is an example of an explosive reaction system with stochastic ergodic environment.

**Example 4.1.** Assume the chain \( \{X(t) : t \geq 0\} \) has two states, denoted by 0 and 1. This is often the case in genetic models, where \( X(t) \) denotes whether a gene is active at time \( t \) or not. Let \( q_{10} \) be the transition rate from 1 to 0 and \( q_{01} \) the transition rate from 0 to 1.

Consider the following stochastic mass-action system with stochastic environment:

\[
S_1 + S_2 \xrightleftharpoons[1]{\lambda_1} S_3, \quad S_1 + S_3 \xrightleftharpoons[3]{\lambda_2} S_2, \quad 0 \xrightarrow[1]{\lambda_3} S_1,
\]

with

\[
\lambda_1(x, z) = \kappa_1 x z_1 z_2, \quad \lambda_2(x, z) = \kappa_2 (1-x) z_1 z_3 \quad \text{and} \quad \lambda_3(x, z) = \kappa_3,
\]

for positive constants \( \kappa_1, \kappa_2 \) and \( \kappa_3 \). The total amount of molecules of \( S_2 \) and \( S_3 \) is conserved. Assume that

\[
Z_2(t) + Z_3(t) = b > 0.
\]

When \( X(t) = 1 \), degradation of an \( S_1 \) molecule consumes a molecule of \( S_2 \), which is not recreated because \( \lambda_2(1, z) = 0 \). Hence, when \( X(t) = 1 \) at most \( b \) molecules of \( S_1 \) can be degraded. Similarly, \( \lambda_1(0, z) = 0 \) implies that at most \( b \) molecules of \( S_1 \) can be degraded when \( X(t) = 0 \). Hence, by using renewal theory (see
for example [11, Theorem 3.3]) and by setting up the renewal intervals in between two consecutive visits of \( \{ X(t) : t \geq 0 \} \) to 1, we have that

\[
\lim_{t \to \infty} \frac{Z_1(t)}{t} = \lim_{t \to \infty} \frac{Z_1^+(t) - Z_1^-(t)}{t} \geq \kappa_3 - 2b \frac{q_{10}q_{01}}{q_{10} + q_{01}}, \quad \text{a.s.,}
\]

where \( Z_1^+(t) \) and \( Z_1^-(t) \) denote the number of \( S_1 \) molecules created and degraded by time \( t \), respectively. It follows that for any choice of parameters such that the latter is strictly positive, the model is transient.

An interesting observation is the following. Consider the stochastic mass-action system with fixed environment

\[
S_1 + S_2 \xrightarrow{\lambda_1} S_3, \quad S_1 + S_3 \xrightarrow{\lambda_2} S_2, \quad 0 \xrightarrow{\lambda_3} S_1,
\]

where

\[
\tilde{\lambda}_1(z) = \tilde{\kappa}_1 z_1 z_2, \quad \tilde{\lambda}_2(z) = \tilde{\kappa}_2 z_1 z_3, \quad \tilde{\lambda}_3(z) = \tilde{\kappa}_3,
\]

with \( \tilde{\kappa}_1, \tilde{\kappa}_2, \tilde{\kappa}_3 \) positive. For all parameter values, the function \( V(z) = z \) is a Lyapunov function, hence the process is always ergodic. Therefore, the stochastic fluctuations of the environment allow for a behaviour that is not possible in the constant environment. Hence, the example suggests that the effect of the stochastic environment is not merely averaged out over time in some way, but might influence the dynamics substantially.

**Example 4.2.** Consider the following stochastic mass-action system with stochastic environment:

\[
2S \xrightarrow{\frac{\lambda_1}{\lambda_2}} 3S, \quad \lambda_1(x, z) = \kappa_1(x)z(z - 1), \quad \lambda_2(x, z) = \kappa_2(x)z(z - 1)(z - 2).
\]

Assume the process \( \{ X(t) : t \geq 0 \} \) is ergodic, and that there exists a recurrent state \( x' \) such that \( \kappa_1(x') \neq 0 \) and \( \kappa_2(x') = 0 \). In this case, with probability one, \( \{ X(t) : t \geq 0 \} \) hits \( x' \) infinitely often. When this occurs, the number of \( S \) molecules evolves according to the stochastic mass-action system with constant environment

\[
2S \xrightarrow{\tilde{\lambda}_1} 3S,
\]

and \( \tilde{\lambda}_1(z) = \kappa_1(x')z(z - 1) \). Specifically, the species count of \( S \) evolves according to a pure birth process with (almost) quadratic birth rate. It follows that whenever \( \{ X(t) : t \geq 0 \} \) hits \( x' \), \( \{ Z(t) : t \geq 0 \} \) has a positive probability of exploding. Hence it will eventually explode with probability one.

### 4.1 Mono-molecular reaction networks

We introduce a family of stochastic reaction systems with stochastic environment.

**Assumption 1.** \( (\mathcal{R}, \Lambda, \{ X(t) : t \geq 0 \}) \) is a stochastic mass-action system with stochastic environment, such that each reaction takes one of the following forms:

**Production:** \( 0 \to m_j S_j \),  
**Conversion:** \( S_i \to S_j \),  
**Degradation:** \( S_i \to 0 \),

for \( m = (m_1, \ldots, m_d) \in \mathbb{Z}_0^d \), and \( 1 \leq i, j \leq d \). For notational convenience, we set \( m_j = 0 \) if there is not a reaction \( 0 \to \beta S_j \) for any \( \beta > 0 \).

The stochastic reaction system in Section 3 satisfies Assumption 1, while those of Example 4.1 and Example 4.2 do not. Under Assumption 1, let \( \tilde{\lambda}_{ij} : \Gamma \to \mathbb{R}_{\geq 0}^d \), for \( 0 \leq i, j \leq d \), be the functions such that:

- \( \kappa_k(x, z) = \tilde{\kappa}_{0j}(x), z \in \mathbb{Z}^n, \) if \( 0 \to m_j S_j \) is the kth reaction of \( \mathcal{R} \).
- \( \kappa_k(x, z) = \tilde{\kappa}_{ij}(x)z_i, z \in \mathbb{Z}^n, \) if \( S_i \to S_j \) is the kth reaction of \( \mathcal{R} \).
- \( \kappa_k(x, z) = \tilde{\kappa}_{i0}(x)z_i, z \in \mathbb{Z}^n, \) if \( S_i \to 0 \) is the kth reaction of \( \mathcal{R} \).
- \( \tilde{\lambda}_{ij} \) is zero for all other choices of indices \( 0 \leq i, j \leq d \).
We further define the function $A: \Gamma \to \mathbb{R}^{d\times d}$ as $A(\cdot) = (a_{ij}(\cdot))_{1 \leq i,j \leq d}$ with
\[
a_{ij}(x) = \begin{cases} 
\bar{\lambda}_j(x) & \text{for } i \neq j, \\
-\sum_{k \neq i} \bar{\lambda}_k(x) - \tilde{\lambda}_{i0}(x) & \text{for } i = j,
\end{cases}
\] (4.2)
and the function $B: \Gamma \to \mathbb{R}^d_{\geq 0}$ as $B(\cdot) = (b_i(\cdot))_{1 \leq i \leq d}$ with
\[
b_i(x) = \tilde{\lambda}_{i0}(x).
\] (4.3)
For $x \in \Gamma$, $A(x)$ has non-positive column sums and non-negative off-diagonal elements. Hence it may be considered the transpose of a sub-generator matrix of a Markov chain.

We introduce the composite functions $A_X: \mathbb{R}^d_{\geq 0} \to \mathbb{R}^{d\times d}$ and $B_X: \mathbb{R}^d_{\geq 0} \to \mathbb{R}^d$ as $A_X(t) = A(X(t))$ and $B_X(t) = B(X(t))$, $t \geq 0$, respectively, and note that they only depend on the Markov process through $X(t)$.

### 4.2 The matrix $\Phi(\cdot)$

Given a function $M: \mathbb{R}_{\geq 0} \to \mathbb{R}^{n\times n}$ and a non-singular matrix $H_0 \in \mathbb{R}^{n\times n}$, a fundamental matrix solution of
\[
H(0) = H_0 \quad \text{and} \quad \frac{d}{dt} H(t) = M(t)H(t) \quad \text{for all} \quad t > 0,
\] (4.4)
is a function $H: \mathbb{R}_{\geq 0} \to \mathbb{R}^{n\times n}$ that is non-singular for all $t \in \mathbb{R}_{>0}$ and solves (4.4), where the derivative is intended component-wise (for details, see [31, Section 1.10]).

For our purpose, we define $\{\Phi(t): t \geq 0\}$ as the $\mathbb{R}^{d\times d}$-valued stochastic process that solves
\[
\Phi(0) = I_d \quad \text{and} \quad \frac{d}{dt} \Phi(t) = A_X(t)\Phi(t) \quad \text{for all} \quad t > 0
\] (4.5)
path-wise. Note that $\Phi$ is well-defined with probability one on the event that the non-explosive process $\{X(t): t \geq 0\}$ jumps finitely many times in finite time. The precise form of $\Phi(t)$ (including almost sure existence and uniqueness) is given in Lemma 4.1. Intuitively, (4.5) might be thought of as a time-varying Kolmogorov forward equation, as described below. As a consequence, a formal interpretation of the quantity $\Phi$ can be derived, see Lemma 5.2.

To uncover the interpretation of $\Phi$ and how it carries over to the random case, we consider the situation where $\{X(t): t \geq 0\}$ is deterministic and constant. Hence assume $A_X = A_X(t)$ is independent of time. In this case, $\{Z(t): t \geq 0\}$ is a mass-action system in the standard sense: The transition rates are constant and as a consequence the process is a time-homogeneous CTMC. If Assumption 1 holds, then the molecules in the system evolve independently of each other, in the sense of Remark 2.1. In particular, the species type of a specific molecule evolves as a CTMC with state space $\{S_1, \ldots, S_d, 0\}$, where 0 is an absorbing state denoting degradation. The transition matrix for this process is
\[
Q = \begin{pmatrix}
A_X^\top & -A_X^\top e \\
0 & 0
\end{pmatrix}.
\]
Hence, if $P(t)$ is the matrix of transition probabilities between the states $S_1, \ldots, S_d$, such that the entry $P^{(i,j)}(t)$ is the probability of a single molecule being transformed from state $S_i$ to state $S_j$ over time $t$, then by Kolmogorov’s backward equation,
\[
\frac{d}{dt} P(t) = P(t)A_X^\top
\]
(note that once the process has ‘leaked’ to 0 it cannot return, so $P(t)$ describes the dynamics before absorption). It follows that $P(t)$ is the transpose matrix of $\Phi(t)$. The same holds even if the matrix $A_X$ changes over time as seen in [20], and even for stochastic environment as we will see in Lemma 5.2.

In this spirit the matrix $\Phi(\cdot)$ has been used to describe state probabilities for deterministically changing environments [20] and for networks of $M/M/\infty$ queues with time-varying stochastic rates [12, 27], which [20] might be considered a special case of. We state a version of these results in Theorem 5.1 below.
Furthermore, we define \( \{W(t) : t \geq 0\} \) as the \( \mathbb{R}^d_{\geq 0} \)-valued stochastic process solving (path-wise)
\[
W(0) = 0 \quad \text{and} \quad \frac{d}{dt}W(t) = AX(t)W(t) + BX(t) \quad \text{for all} \quad t > 0.
\]
Note that the term \( BX(t) \) accounts for the molecules entering the system over time. These molecules are subsequently transformed according to \( AX(t) \). By Lemma 11.3 in the Appendix (Section 11.4) we have
\[
W(t) = \int_0^t \Phi(u,t)BX(u)\,du \quad \text{for all} \quad t \in \mathbb{R}_{\geq 0},
\]
where the integral is intended component-wise, and
\[
\Phi(u,t) = \Phi(t)\Phi^{-1}(u), \quad \text{for} \quad t \geq u \geq 0.
\]
Clearly, \( \Phi(0,t) = \Phi(t) \). Moreover, \( \Phi(u,u+\cdot) \) is the fundamental matrix solution solving
\[
\Phi(u,u) = I_d \quad \text{and} \quad \frac{d}{dt}\Phi(u,u+t) = AX(u+s)\Phi(u,u+t) \quad \text{for all} \quad t > 0.
\]
Furthermore, the matrix \( \Phi(u,t) \) fulfills the equality
\[
\Phi(u,t) = \prod_{i=0}^{k-1} \Phi(t_i,t_{i+1}) = \Phi(t_{k-1},t_k)\ldots \Phi(t_0,t_1),
\]
for any \( 0 \leq u = t_0 \leq t_1 \leq \ldots \leq t_k = t \).

**Lemma 4.1.** Almost surely, a unique fundamental matrix solution \( \Phi \) of (4.5) exists. Specifically, let \( T_0 = 0 \) and \( T_i, \ i \geq 1 \), denote the \( i \)-th jump time of \( \{X(t) : t \geq 0\} \). For \( s \in [T_i,T_{i+1}) \), let \( A_i = AX(s) \). Then,
\[
\Phi(t) = e^{A_N(t-T_N)} \prod_{i=0}^{N(t)-1} e^{A_i(T_{i+1}-T_i)}, \quad t \geq 0, \ a.s.,
\]
where \( N(t) = \sup\{i : T_i \leq t\} \) is a.s. finite. As a consequence, for \( t \geq u \), almost surely
\[
\Phi(u,t) = \begin{cases} 
  e^{A_N(t-T_N)} \left( \prod_{i=0}^{N(t)-1} e^{A_i(T_{i+1}-T_i)} \right) e^{A_N(u)(T_N(u)-u)} & \text{for} \quad N(u) + 1 \leq N(t),
  \\
  e^{A_N(u)(t-u)} & \text{for} \quad N(u) = N(t).
\end{cases}
\]

**Proof.** By assumption \( \{X(t) : t \geq 0\} \) is non-explosive, hence \( N(t) \) is a.s. finite. Observe that on the interval \( [T_i,T_{i+1}) \) the matrix \( AX(\cdot) \) is the constant matrix \( A_i \) for all \( i \). Hence, the unique solution to the Kolmogorov’s forward equation \( \Phi'(t) = A_i\Phi(t) \) for \( t \in [T_i,T_{i+1}) \) is \( \Phi(t) = e^{A_i(t-T_i)}\Phi(T_i) \). The result follows by induction on \( i \geq 0 \) and by the fact that \( \Phi(0) \) is the identity matrix. Non-singularity of \( \Phi(t) \) for \( t \geq 0 \) follows from the fact that exponential matrices are non-singular. Finally, equation (4.10) follows from \( \Phi(u,t) = \Phi(t)\Phi^{-1}(u) \) and from (4.9) applied to both \( \Phi(t) \) and \( \Phi^{-1}(u) \).

**5 Finite-time distribution**

The goal of this section is to describe the distribution of \( Z(t) \), given \( \mathcal{F}_t^X \), by means of \( (\Phi(t), W(t)) \). We start with a similar result from \( [20] \) when \( \{X(t) : t \geq 0\} \) is deterministic.

**Theorem 5.1.** Suppose Assumption 1 holds and that \( 0 \leq m_j \leq 1 \) for \( 1 \leq j \leq d \). Moreover, assume the process \( \{X(t) : t \geq 0\} \) is deterministic, that is, there exists a function \( x : \mathbb{R}_{\geq 0} \rightarrow \Gamma \) such that
\[
X(t) = x(t) \quad \text{for all} \quad t \geq 0,
\]
\( a.s. \). Then, for \( t \geq 0 \),
\[
\mathcal{L}(Z(t)) = \left( \bigotimes_{i=1}^d \mathcal{L}(\text{Multi}(Z_i(0),\Phi(t)e_i)) \right) \ast \mathcal{L}(\text{Pois}(W(t))),
\]
Theorem 1 in [20] is stronger than the theorem above as it allows for rates to change continuously in time, rather than by jumps in $\Gamma$. The weaker version proposed in Theorem 5.1 is sufficient for our purposes. In Remark 5.1, we discuss an extension to [20, Theorem 1] which is a consequence of our results.

We next provide a probabilistic interpretation of $\Phi(u, t)$.

**Lemma 5.2.** Suppose Assumption 1 holds and let $\Phi(\cdot)$ be as in (4.5). For $1 \leq i, j \leq d$ and $t \geq u \geq 0$, let $P_X^{(i,j)}(u, t)$ be the probability, given $\mathcal{F}_t^X$, that a molecule of species $S_i$, present at time $u$, is a molecule of species $S_j$ at time $t$. Then, for $1 \leq i, j \leq d$ and $t \geq u \geq 0$,

(a) $P_X^{(i,j)}(u, t)$ is the $(j, i)$-th entry of $\Phi(u, t)$,

(b) the probability given $\mathcal{F}_t^X$ that a molecule of species $S_i$, present at time $u$, eventually is degraded by time $t$ is

$$1 - \sum_{j=1}^d P_X^{(i,j)}(u, t) = 1 - e^T \Phi(u, t) e_i = 1 - \|\Phi(u, t) e_i\|_1,$$

(c) $\|\Phi(u, t) e_i\|_1 \leq 1$.

**Proof.** To prove (a), consider a molecule of species $S_i$ that is present at time $u$, and let $F(t)$ denote its type (a species or 0 if it is degraded) at time $u + t$, for $t \geq 0$.

Define $\tilde{X}(t) = X(t + u)$ for $t \geq 0$. Consider a modification of the model where production is not allowed (that is, $\lambda_{ii}$ is the null function for all $1 \leq i \leq d$). Denote by $\{\tilde{Z}(t): s \geq 0\}$ the process associated with this model, and let $\{\tilde{X}(t): t \geq 0\}$ be its stochastic environment.

Assuming $\tilde{Z}(0) = e_i$, the process $\{\tilde{Z}(t): t \geq 0\}$ is distributed as the process $\{F(t): t \geq 0\}$, and the dependence on the environment is the same for the two processes, conditionally on $(\tilde{X}(0), \tilde{Z}(0))$. Hence, it follows from Theorem 5.1 and (4.7) that $L(\tilde{Z}(t)|\mathcal{F}_t^X) = L(\text{Multi}(1, \Phi(u, t) e_i)|\mathcal{F}_u^X)$ for $t \geq 0$. Hence, given $\mathcal{F}_t^X$, then $P_X^{(i,j)}(u, t) = P(F(t) = S_j|\mathcal{F}_t^X) = P(\tilde{Z}(t) = e_j|\mathcal{F}_t^X)$ is the $(j, i)$-th entry of $\Phi(u, t)$, and part (a) is proven.

The proof is then concluded as (b) is a straightforward consequence of (a), and (c) follows from (b). □

The next proposition provides a multi-dimensional version of Proposition 3.1. The distribution in (5.1) below has two main components. The first component is similar to that of Proposition 3.1 and represents the fate of the molecules present at time zero, sorted by initial species type at time zero. The second component, represents the fate of the molecules born after time zero. If $m_j$ molecules of species $S_j$ are born, then a number of them die before time $t$. The remaining molecules might change species type by transformation over time. This is recorded in the vector $\nu$ in (5.1) below.

The overall structure of the distribution resembles that in Proposition 3.1, though it is more complex due to the fact that there might be many species types and not just one, and that the molecules might be transformed over time. Compared to the deterministic result in [20], we allow $m_j \geq 0$ to be arbitrary and not restricted to the values 0, 1.

For any $n_1, n_2 \in \mathbb{Z}_{\geq 0}$, let $\Theta_{n_1, n_2}$ be the set of $n_2$-tuples $\{(\nu_1, \ldots, \nu_{n_2}) \in \mathbb{Z}_{\geq 0}^{n_2} : 1 \leq \sum_{j=1}^{n_2} \nu_j \leq n_1\}$. We refer to $\nu \in \Theta_{n_1, n_2}$ as a configuration. If $\Theta_{m_j, d} = \emptyset$ for some $j$ in (5.1) below, then we take the corresponding law to be a degenerate distribution at $0 \in \mathbb{Z}_{\geq 0}^d$.

**Proposition 5.3.** Suppose Assumption 1 holds and let $\Phi(\cdot)$ be as in (4.5). Then, for any fixed $t \geq 0$, the conditional distribution of $Z(t)$ given $\mathcal{F}_t^X$ is

$$\left( \frac{d}{i=1} L(\text{Multi}(Z_i(0), \Phi(t) e_i)|\mathcal{F}_t^X) \right) \ast \left( \frac{d}{j=1} \otimes_{\nu \in \Theta_{m_j, d}} L(\nu N_{\nu j}(t)|\mathcal{F}_t^X) \right),$$

where the initial condition $Z(0) = (Z_1(0), \ldots, Z_d(0))$ is the molecular counts at time zero,

$$N_{\nu j}(t) \sim \text{Pois} \left( \int_0^t \tilde{\lambda}_{0j}(X(u)) g_X^{\nu}(\nu, m_j) \, du \right),$$

and $g_X^{\nu}(\nu, m_j)$ is the probability, given $\mathcal{F}_t^X$, that the multinomial random variable $\text{Multi}(m_j, \Phi(u) e_j)$ takes the value $\nu$, given $\mathcal{F}_t^X$ ($\nu N_{\nu j}$ is the vector $\nu$ multiplied by the number $N_{\nu j}$).
Proof. We begin with the case $0 \leq m_j \leq 1$ for all $j = 1, \ldots, d$. Given $F^X_t$, it follows from Theorem 5.1 that

$$
\mathcal{L}(Z(t) \mid F^X_t) = \left( \bigotimes_{i=1}^d \mathcal{L}(\text{Multi}(Z_i(t), \Phi(t)e_i) \mid F^X_t) \right) * \mathcal{L}\left(\text{Pois}(W(t)) \mid F^X_t\right).
$$

(5.2)

If $m_j = 0$, then $\Theta_{m_j,d} = \emptyset$. If $m_j = 1$, then $\Theta_{m_j,d} = \{ e_i \mid i = 1, \ldots, d \}$ and $g^X_{u,t}(e_i, m_j) = e_i^\top \Phi(u,t)e_j$. It follows that

$$
\sum_{j=1}^d \lambda_{0j}(X(u)) g^X_{u,t}(e_i, m_j) = e_i^\top \Phi(u,t) B_X(u).
$$

Using the definition of $W(t)$ in (4.6) then (5.1) is the same as (5.2), and the result holds.

The multinomial terms in (5.2) represent the degradation of the molecules present at time 0, and the Poisson term represents the number of the molecules born after time 0 which survived up to time $t$.

We next generalise the result to arbitrary $(m_1, \ldots, m_d) \in \mathbb{Z}_{\geq 0}^d$. First, we note that the argument for the multinomial part in (5.2), representing the distribution of the number of molecules at time $t$ that survived from the initial $Z(0)$ molecules at time 0, does not depend on $(m_1, \ldots, m_d)$, hence it carries over to the general case.

Let $V_t$ be the vector counting the number of molecules, born in $(0, t]$ and alive at time $t$, by their type at time $t$. The random variable $V_t$ is a sum $\sum_{j=1}^d V^j_t$, where $V^j_t \in \mathbb{Z}_{\geq 0}$ counts the molecules that are produced by the reaction $0 \to m_j S_j$ (if present) and subsequently transformed by conversion into other molecules and/or degraded. Let $B^j_t$ be the number of times the reaction $0 \to m_j S_j$ occurred in $[0, t]$, if this is present. Observe that the random variables $V^j_t$, $1 \leq j \leq d$, are conditionally independent given $F^X_t$, since they are generated by independent processes of production, and each molecule degrades or is transformed independently of the others, given $F^X_t$ (Remark 2.1).

By conditioning on the time of birth events in the computation of the Laplace transform of $V^j_t$, similarly to the argument applied to show (3.5), we have that for every $\gamma \in \mathbb{Z}_{\geq 0}^d$

$$
E\left( e^{-\gamma^\top V^j_t} \mid F^X_t \right) = \sum_{r=0}^{\infty} P(B^j_t = r \mid F^X_t) \prod_{i=1}^r E \left( e^{-\gamma^\top V^i_t} \mid B^i_t = r, F^X_t \right)
$$

$$
= \sum_{r=0}^{\infty} e^{-\int_0^t \lambda_{ij}(X(u)) du} \left( \int_0^t \lambda_{ij}(X(u)) du \right)^r \prod_{i=1}^r \left[ \int_0^t \left( \int_0^t \lambda_{ij}(X(s)) ds - \sum_{\nu \in \Theta_{m_j,d}} g^X_{u,t}(\nu, m_j) e^{-\gamma^\top \nu} \right) du \right]\nu.
$$

(5.3)

where $g^X_{u,t}(\nu, m_j)$ is the probability given $F^X_t$ that $m_j$ molecules of $S_j$ born at time $u$ have become $\nu_i$ molecules of $S_i$, $1 \leq i \leq d$, at time $t$, where $\nu = (\nu_1, \ldots, \nu_d)$. By Lemma 5.2(a), $g^X_{u,t}(\nu, m_j)$ is the probability given $F^X_t$ that Multi$(m_j, \Phi(u,t)e_j)$ takes the value $\nu$. It follows that

$$
\mathcal{L}(V^j_t \mid F^X_t) = \bigotimes_{\nu \in \Theta_{m_j,d}} \mathcal{L}(\nu \text{ Pois} \left( \int_0^t \lambda_{ij}(X(u)) g^X_{u,t}(\nu, m_j) du \right) \mid F^X_t),
$$

(5.4)

where each Poisson variable counts how often the configuration $\nu$ appears at time $t$. Using the conditional independence of the variables $V^j_t$, the Poisson part in the convolution (5.1) follows from (5.3), and the proposition is proved.

Note that if $0 \leq m_j \leq 1$ for all $j = 1, \ldots, d$, then the Poisson mixture in (5.1) simplifies to

$$
\bigotimes_{j=1}^d \bigotimes_{\nu \in \Theta_{m_j,d}} \mathcal{L}(\nu \text{ Pois}(t) \mid F^X_t) = \mathcal{L}(\text{Pois}(W(t)) \mid F^X_t),
$$

which is the form it takes in Theorem 5.1.
Remark 5.1. Propositions 3.1 and 5.3 hold under more general conditions. Firstly, the reaction rates $\lambda_r$ may depend directly on time $t$, that is, be of the form $\lambda_r(x, z, t)$, rather than just $\lambda_r(x, z)$. The requirements for this generalisation are minimal: for the current proofs to hold we only need the matrix-valued process $\{\Phi(s): 0 \leq s \leq t\}$ and the integrals of $\lambda_0(x(S))$ over $[0, t]$ to be well-defined and almost-surely finite. A sufficient condition is that $\lambda_r(x, z, \cdot)$ is Lipschitz in $[0, t]$ for all $x, z$. Moreover, this generalisation would imply a more general version of [20, Theorem 1], with $m_j \geq 0$, and not confined to the values $0, 1$.

Furthermore, the proofs of Propositions 3.1 and 5.3 are valid without the assumption that $\{X(t): t \geq 0\}$ is Markovian, as long as (2.1) holds. This is because the dynamics of the process $\{X(t): t \geq 0\}$ itself are not important once we condition on $F_t^X$. We may further allow the state space $\Gamma$ to be continuous, and $\{X(t): t \geq 0\}$ to be anything other than a jump process, as long as $\{\Phi(s): 0 \leq s \leq t\}$ and the integrals of $\lambda_0(x(S))$ are well-defined and almost-surely finite, as discussed above.

Aforementioned remark on possible generalisations cannot be considered in the following sections without substantially changing the results. As a matter of fact, the following sections rely strongly on the regeneration properties of $\{X(t): t \geq 0\}$ and, consequently, of $\{\Phi(t): t \geq 0\}$.

6 The stationary distribution and ergodicity

In this section we study the long term behaviour of the processes fulfilling Assumption 1. To motivate this further consider a process $\{X(t): t \geq 0\}$ with state space $\{0, 1\}$ and the following simple reaction system:

$$0 \xrightarrow{\lambda_1} S, \quad \lambda_1(x, z) = 1, \quad \lambda_2(x, z) = xz.$$ 

The species $S$ is constantly produced while degradation at time $t$ depends on whether $X(t)$ is 1 or 0. Intuitively, if the process $\{X(t): t \geq 0\}$ takes the value zero for a minuscule fraction of time one might expect a transient behaviour of $\{Z(t): t \geq 0\}$. However, this is not the case. The model falls in the category of Section 3 with $m = 1$, $\kappa_1(x) = 1$, and $\kappa_2(x) = x$. If $Z(0) = 0$, then the distribution of $Z(t)$ is

$$\mathcal{L}(Z(t)|F_t^X) = \mathcal{L} \left( \text{Pois} \left( \int_0^t e^{-s} X(s) ds \right) \right) \bigg| F_t^X.$$ 

If $\{X(t): t \geq 0\}$ is ergodic, then the intensity of the Poisson distribution converges in distribution as $t \to \infty$. Consequently, it follows from Levy’s continuity theorem [11], that $Z(t)$ also converges in distribution and $Z(t)$ cannot be transient. It is also a consequence of Theorem 6.1 below.

6.1 Structural conditions for ergodicity

To rigorously analyse this stability phenomenon we give graphical conditions under which the models satisfying Assumption 1 are ergodic for every initial condition.

Definition 6.1. Suppose Assumption 1 holds for a stochastic reaction system with stochastic environment $(\mathcal{R}, \Lambda, \{X(t): t \geq 0\})$, and let $S_i, S_j \in \mathcal{S}$. We say that

- $S_j$ is obtainable from $S_i$, denoted $S_i \rightsquigarrow S_j$, if $S_i = S_j$, or there exists a sequence of reactions $S_{i_k} \to S_{i_{k+1}}$, $1 \leq k \leq n - 1$, with $i_1 = i$ and $i_n = j$,

- $S_i$ is properly produced, denoted $0 \rightsquigarrow S_i$, if there exists $0 \to m_i, S_i$ and a sequence of reactions $S_{i_k} \to S_{i_{k+1}}$, $1 \leq k \leq n - 1$, with $i_n = i$,

- $S_i$ is properly degraded, denoted $S_i \rightsquigarrow 0$, if for every species $S_j$ obtainable from $S_i$, there is a sequence of reactions $S_{i_k} \to S_{i_{k+1}}$, $1 \leq k \leq n - 1$, with $i_1 = j$, and such that $S_{i_n} \to 0$.

Furthermore, we write $S_i \not\rightsquigarrow S_j$, $0 \not\rightsquigarrow S_i$, and $S_i \not\rightsquigarrow 0$ if the relation does not hold.

Assumption 2. The process $\{X(t): t \geq 0\}$ is ergodic with stationary distribution $\pi$. Furthermore, assume

(a) every properly produced species is properly degraded,
(b) the expectation of \( \tilde{\lambda}_0(\cdot) \) with respect to \( \pi \) is finite for any \( 1 \leq j \leq d \), that is

\[
\sum_{x \in \Gamma} \tilde{\lambda}_0(x) \pi(x) \leq \sum_{x \in \Gamma} \|B(x)\|_1 \pi(x) < +\infty.
\]

The process \( \{(X(t),Z(t)) : t \geq 0\} \) can only be ergodic for every initial condition if the properly produced species are also properly degraded. Hence, the essential restriction is (b).

Under Assumptions 1 and 2, the relation \( \leadsto \) induces a partition of the species set \( \mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_h \cup \mathcal{S}_T \cup \mathcal{S}_P \), \( h \in \mathbb{Z}_{\geq 0} \), where \( \mathcal{S}_i \) consists of the species in a closed strongly connected component of the graph associated to \( \leadsto \), not containing 0, the set \( \mathcal{S}_P \) consists of the properly produced species (hence also properly degraded species), and \( \mathcal{S}_T \) consists of the remaining (transient) species. The transient species are either properly degraded (but not produced) or converted into other species in one or more connected components, potentially both.

As an example, consider the reaction network 0 → 2S1, S1 → S2 → 0, S3 → S4 → S5, S5 → S4. The partition induced by \( \leadsto \) is \( \mathcal{S} = \{S_4,S_5\} \cup \{S_1,S_2\} \cup \{S_3\} \) with \( \mathcal{S}_P = \{S_1,S_2\} \) and \( \mathcal{S}_T = \{S_3\} \).

The first main result is the following.

**Theorem 6.1.** Consider a stochastic reaction system with stochastic environment, fulfilling Assumptions 1 and 2. Then \( \{(X(t),Z(t)) : t \geq 0\} \) is ergodic for every initial condition.

By ‘ergodic for every initial condition’ we understand that given an initial condition of the process there exists of a unique limiting distribution as \( t \to \infty \), see Section 2. The support of this distribution does not have to be a unique irreducible component, though. For example, the reaction network \( S_1 \to S_2 \) and \( S_1 \to S_3 \) (with constant environment) is ergodic for every initial condition, but the limiting distribution is not concentrated on an irreducible component. Since \( \Gamma \) is irreducible by assumption, the possible division of \( \Gamma \times \mathbb{Z}_{\geq 0}^d \) is induced by the structure of the reaction network.

We prove the theorem by making use of properties of the fundamental matrix solution \( \Phi(\cdot) \). The proof is in Section 6.4 and draws on material in Section 6.2 and Lemma 6.4.

### 6.2 Connecting Assumption 2 with properties of \( \Phi(\cdot) \)

Here we highlight some connections between properties of the matrix \( \Phi(\cdot) \) and Assumptions 1 and 2.

If the state space \( \Gamma \) consists of just one element then the matrices \( A_X \) and \( B_X \) are constant, deterministic matrices, and the environment is not stochastic but constant. Moreover, if further \( m_j \in \{0,1\} \) for all \( 1 \leq j \leq d \), then the process \( \{Z(t) : t \geq 0\} \) is a CTMC associated with a mono-molecular stochastic reaction network in the sense of [20]. Questions about ergodicity and the form of the stationary distribution have been fully answered in this case [20].

Therefore, assume \( \Gamma \) has at least two elements and let \( x \in \Gamma \). Define \( \tau_0^x \geq 0 \) to be the first time \( \{X(t) : t \geq 0\} \) hits the state \( x \), and let \( T_0^x \) denote the time spent in \( x \) during this first visit. If \( X(0) = x \), then \( \tau_0^x = 0 \). Recursively define for \( k \geq 1 \),

\[
\tau_k^x = \inf\{t > \tau_{k-1}^x + T_{k-1}^x : X(t) = x\}, \quad \text{and} \quad T_k^x = \inf\{t > \tau_k^x : X(t) \neq x\} - \tau_k^x.
\]

The time between the \( (k-1) \)-th and \( k \)-th visits to \( x \) is \( \tau_k^x - \tau_{k-1}^x \) for \( k \geq 1 \). By the strong Markov property the random variables \( \tau_k^x - \tau_{k-1}^x \), \( k \geq 1 \), are independent and identically distributed. Moreover, \( \tau_0^x \) is independent of \( \tau_k^x - \tau_{k-1}^x \) for all \( k \geq 1 \). If \( \{X(t) : t \geq 0\} \) is ergodic then the waiting times \( \tau_k^x \), \( k \geq 1 \), have positive finite expectation (recall, \( \Gamma \) is irreducible). If \( X(0) \neq x \), then the same holds for \( \tau_0^x \); otherwise \( \tau_0^x = 0 \).

We state here technical implications of Assumptions 1 and 2 in terms of \( \Phi(\cdot) \), which we will make use of later on. Specifically, we will introduce the quantity \( \alpha \), which is interpreted as the number of times the process \( \{X(t) : t \geq 0\} \) needs to come back to a given state \( x \) before all reactions of \( \{Z(t) : t \geq 0\} \) have had a positive probability of occurring. For further intuition on the role of \( \alpha \), see Example 6.1.

**Lemma 6.2.** Suppose Assumptions 1 and 2 hold and that \( \Gamma \) contains at least two elements. Then, there exists \( \alpha \in \mathbb{Z}_{\geq 0} \) such that for all \( x \in \Gamma \),

\[
\alpha
\]
(a) \( E\left( \max_{1 \leq i \leq d} \| \Phi(\tau_{0}^{x}, \tau_{i}^{x})e_{i} \|_{1} < 1, \right) \) < 1,

(b) \( E\left( \min_{1 \leq i,j \leq d} \Phi_{ij}(\tau_{0}^{x}, \tau_{i}^{x}) > 0, \right) \)

Proof. (a) Let \( S_{i_{1}}, S_{i_{2}}, \ldots, S_{i_{d}}, \) be the properly degraded species. By Lemma 5.2(b) we have \( \| \Phi(\tau_{k}^{x}, \tau_{k+k}^{x})e_{i} \|_{1} \leq 1 \) for any \( k, k' \in \mathbb{Z}_{\geq 0}, 1 \leq i \leq d, \) and \( x \in \Gamma. \) Hence it suffices to show that there exists \( \alpha \in \mathbb{Z}_{\geq 0} \) such that for all \( x \in \Gamma, \)

\[
P\left( \max_{1 \leq \ell \leq d} \| \Phi(\tau_{0}^{x}, \tau_{\ell}^{x})e_{\ell} \|_{1} < 1 \right) > 0, \tag{6.1}
\]

For any \( 1 \leq \ell \leq \hat{d}, \) there exists a directed path of reactions \( S_{i_{\ell}} = S_{j_{0}} \rightarrow S_{j_{1}} \rightarrow \cdots \rightarrow S_{j_{\ell-1}} \rightarrow 0 (= S_{i_{\ell}} \) for convenience) with \( \alpha_{\ell} \in \mathbb{Z}_{\geq 0}. \) For any \( 0 \leq k \leq \alpha_{\ell} - 1, \) let \( x_{k} \in \Gamma \) be such that \( \lambda_{j_{k},j_{k+1}}(x_{k}) > 0, \) and \( x \in \Gamma. \) Since \( \Gamma \) is irreducible, with positive probability the process \( \{X(t): t \geq 0\} \) visits the states \( x_{0}, x_{1}, \ldots, x_{\alpha_{\ell} - 1} \) in that order, between time \( \tau_{k}^{x} \) and \( \tau_{\alpha_{\ell}}^{x} \) (if there is more than one visit to \( x \) between visits to \( x_{k-1} \) and \( x_{k} \), then the path could be contracted). Hence, for any \( 1 \leq \ell \leq \hat{d} \) and any \( x \in \Gamma, \) there is positive probability that a molecule of species \( S_{i_{\ell}}, \) present at time \( \tau_{0}^{x}, \) is degraded by time \( \tau_{\alpha_{\ell}}^{x}. \) Therefore, it follows from Lemma 5.2(a) that

\[
P(\| \Phi(\tau_{0}^{x}, \tau_{\ell}^{x})e_{\ell} \|_{1} < 1) > 0 \quad \text{for all} \quad \alpha \geq \alpha_{\ell}, \text{ and all} \quad x \in \Gamma.
\]

To prove (6.1), we need to show that molecules of different properly degraded species can all be degraded by the same return time to \( x \) with positive probability. Define

\[
\tilde{\alpha} = \sum_{1 \leq \ell \leq \hat{d}} \alpha_{\ell}.
\]

Assume that a molecule of every properly degraded species is present at time \( \tau_{\tilde{\lambda}}^{x}. \) With positive probability, the molecule of species \( S_{i_{\ell}} \) is degraded by time \( \tau_{\alpha_{\ell}}^{x}, \) while the other molecules do not change. Then, with positive probability the molecule of species \( S_{i_{\ell}} \) is degraded by time \( \tau_{\alpha_{\ell},+\alpha_{\ell}}^{x}, \) while the molecules of the other species do not change, and so on. In conclusion, with positive probability all the molecules of properly degraded species are degraded by time \( \tau_{\alpha_{\ell}}^{x}, \) and (6.1) holds for any \( \alpha \geq \tilde{\alpha}. \)

(b) Assume \( S_{i} \sim S_{j}. \) By a similar argument as before, there exists \( \alpha_{ij} \in \mathbb{Z}_{\geq 0} \) such that a molecule of species \( S_{i}, \) present at time \( \tau_{\alpha_{ij}}^{x}, \) is transformed into a molecule of species \( S_{j} \) by time \( \tau_{\alpha_{ij}}^{x}, \) with positive probability, for any \( x \in \Gamma. \) Similarly as before, by choosing

\[
\tilde{\alpha} = \sum_{1 \leq \ell,j \leq \hat{d}} \alpha_{ij} \quad \text{for} \quad S_{i} \sim S_{j},
\]

then the desired holds for all \( \alpha \geq \tilde{\alpha}. \) Hence, (a) and (b) hold at the same time for \( \alpha \geq \max(\tilde{\alpha}, \tilde{\alpha}). \)

To gain intuition about the role played by \( \alpha \) in Lemma 6.2, we give an example of a mass-action system with stochastic environment and find an integer \( \alpha \) satisfying Lemma 6.2(a).

**Example 6.1.** Let \( \{X(t): t \geq 0\} \) be a Markov process with two states \( 0 \) and \( 1 \) such that \( q_{10} \) is the transition rate from \( 1 \) to \( 0, \) and \( q_{01} \) that from \( 0 \) to \( 1. \) Consider the following stochastic mass-action system with stochastic environment:

\[
0 \xrightarrow{\lambda_{1}} S_{1} \xrightarrow{\lambda_{2}} S_{2} \xrightarrow{\lambda_{3}} S_{3} \xrightarrow{\lambda_{4}} 0
\]

with transition rates

\[
\lambda_{1}(x, z) = \kappa_{1} \quad \lambda_{2}(x, z) = \kappa_{2}xz_{1} \quad \lambda_{3}(x, z) = \kappa_{3}(1-x)z_{2} \quad \lambda_{4}(x, z) = \kappa_{4}xz_{3}
\]

for positive constants \( \kappa_{1}, \kappa_{2}, \kappa_{3}, \) and \( \kappa_{4}. \) Assumptions 1 and 2 hold, and all species are properly produced and properly degraded. Hence Lemma 6.2 applies. The conclusions of the lemma hold for \( \alpha = 2, \) but not for \( \alpha = 1. \) It follows from the simple observation that a molecule of species \( S_{1}, \) present at time 0, can only
be removed from the system after the process \( \{X(t) : t \geq 0\} \) has visited the state 1 twice: \( S_1 \) can only be transformed into \( S_2 \) when \( X(t) = 1 \), \( S_2 \) can only be transformed into \( S_3 \) when \( X(t) = 0 \), and finally \( S_3 \) can be degraded only when \( X(t) = 1 \). Using Lemma 5.2, it follows that

\[
\|\Phi(\tau^1_0, \tau^1_1)e_i\|_1 = \sum_{j=1}^{3} P_{X}^{(i,j)}(\tau^1_0, \tau^1_1) = 1 \quad \text{a.s.}, \quad \|\Phi(\tau^1_0, \tau^1_2)e_i\|_1 = \sum_{j=1}^{3} P_{X}^{(i,j)}(\tau^1_0, \tau^1_2) < 1 \quad \text{a.s.}
\]

We next state another technical implication which will be useful in developing the results.

**Lemma 6.3.** Suppose Assumptions 1 and 2 hold and that \( \Gamma \) contains at least two elements. Then,

\[
E\left( \int_{\tau^1_0}^{\tau^1_2} \|\Phi(u, \tau^1_1)B_X(u)\|_1 \, du \right) < +\infty.
\]

**Proof.** It follows from Lemma 5.2(c) that for any \( x \in \Gamma \)

\[
E\left( \int_{\tau^1_0}^{\tau^1_2} \|\Phi(u, \tau^1_1)B_X(u)\|_1 \, du \right) \leq E\left( \int_{\tau^1_0}^{\tau^1_2} \|B_X(u)\|_1 \, du \right).
\]

Hence, it suffices to show that the expectation of the right-hand side is bounded. By the definition of \( B_X(\cdot) \), it suffices to show that

\[
E\left( \int_{\tau^1_0}^{\tau^1_2} \tilde{\lambda}_{0i}(X(u)) \, du \right) < \infty
\]

for all \( i \) such that \( m_i \neq 0 \). By ergodicity of \( \{X(t) : t \geq 0\} \) and non-negativity of \( \tilde{\lambda}_{0i}(\cdot) \),

\[
\lim_{t \to \infty} \frac{1}{t} \int_{0}^{t} \tilde{\lambda}_{0i}(X(u)) \, du = \sum_{x \in \Gamma} \tilde{\lambda}_{0i}(x)\pi(x) \quad \text{a.s.},
\]

where \( \pi \) denotes the stationary distribution of \( \{X(t) : t \geq 0\} \). Since \( \Gamma \) is irreducible and contains at least two elements, \( \lim_{n \to \infty} \tau^x_n = \infty \) a.s. It follows that

\[
\lim_{n \to \infty} \frac{1}{\tau^x_n} \int_{0}^{\tau^x_n} \tilde{\lambda}_{0i}(X(u)) \, du = \sum_{x \in \Gamma} \tilde{\lambda}_{0i}(x)\pi(x) \quad \text{a.s.}
\]

Since \( \tau^x_0 < \infty \) a.s., then by the strong law of large numbers, it holds with probability one that

\[
\lim_{n \to \infty} \frac{1}{\tau^x_n} \int_{0}^{\tau^x_n} \tilde{\lambda}_{0i}(X(u)) \, du = \lim_{n \to \infty} \frac{1}{\tau^x_n - \tau^x_0} \int_{\tau^x_0}^{\tau^x_n} \tilde{\lambda}_{0i}(X(u)) \, du
\]

\[
= \lim_{n \to \infty} \frac{n}{\sum_{k=1}^{n} (\tau^x_k - \tau^x_{k-1})} \frac{1}{n} \sum_{k=1}^{n} \int_{\tau^x_{k-1}}^{\tau^x_k} \tilde{\lambda}_{0i}(X(u)) \, du
\]

\[
= \frac{1}{E(\tau^x_1 - \tau^x_0)} E\left( \int_{\tau^x_0}^{\tau^x_2} \tilde{\lambda}_{0i}(X(u)) \, du \right).
\]

Consequently

\[
E\left( \int_{\tau^x_0}^{\tau^x_2} \tilde{\lambda}_{0i}(X(u)) \, du \right) = E(\tau^x_1 - \tau^x_0) \sum_{x \in \Gamma} \tilde{\lambda}_{0i}(x)\pi(x).
\]

The proof is concluded by the fact that the right-hand side is bounded by Assumption 2. \[ \square \]
6.3 A recurrence relation

A random variable $V$ with values in $\mathbb{R}^n$ satisfies a Stochastic Recurrence Equation (SRE) characterised by a pair of random variables $(Q_1, Q_2)$ with values in $\mathbb{R}^{n \times n} \times \mathbb{R}^m$ if the following holds

$$ V \sim Q_1 V + Q_2, \quad V \parallel (Q_1, Q_2). \quad (6.2) $$

Furthermore, if the law of every random variable satisfying (6.2) is the same, then this law is said to be the weakly unique solution of (6.2). This type of equation appears in modelling perpetuities which are applicable in financial mathematics [8]. General existence and uniqueness conditions for the law of $V$ are given in [8], where other applications are also mentioned.

Assume that $\Gamma$ has at least two elements, and define

$$ n_t^x = \sup \{ n: \tau_n^x \leq t \}, \quad \text{and} \quad \tau_t^x = \tau_{n_t^x}^x, \quad (6.3) $$

where $n_t^x$ is the number of times $\{ X(t): t \geq 0 \}$ visits $x \in \Gamma$ before time $t$. Note that if $\Gamma$ has at least two elements and $\{ X(t): t \geq 0 \}$ is ergodic (as by Assumption 2) and irreducible (as by Definition 2.1), then $\{ X(t): t \geq 0 \}$ will visit $x$ infinitely many times almost surely [26]. For $k \in \mathbb{Z}_{\geq 0}$, define

$$ C_k^x = \Phi(\tau_{k+1}^x), \quad \text{and} \quad D_k^x = \int_{\tau_k^x}^{\tau_{k+1}^x} \Phi(u, \tau_{k+1}^x)B_X(u)\,du. \quad (6.4) $$

By the strong Markov property, $(C_k^x, D_k^x)_{k \geq 0}$, is a sequence of independent, identically distributed random variables. For notational convenience, let

$$ C_{-1}^x = \Phi(0, \tau_0^x), \quad \text{and} \quad D_{-1}^x = \int_0^{\tau_0^x} \Phi(u, \tau_0^x)B_X(u)\,du. $$

Then, using (4.8), we obtain

$$ \Phi(t) = \Phi(\tau_t^x(t), t) \prod_{k=-1}^{n_t^x-1} C_k^x, \quad (6.5) $$

$$ W(t) = \int_{\tau_t^x(t)}^t \Phi(u, \tau_t^x)B_X(u)\,du + \Phi(\tau_t^x, t) \left( D_{n_t^x-1} + \sum_{k=1}^{n_t^x-2} \left( \prod_{i=k}^{n_t^x-1} C_i^x \right) D_k^x \right). \quad (6.6) $$

The proof of the next two lemmas can be found in the Appendix (Sections 11.1 and 11.2). In the lemma below the order of the matrices in the product of $\Phi_x^\infty$ and $W_x^\infty$ is reversed compared to (6.5) and (6.6). In general we cannot assure the existence of the random variables $\prod_{k=0}^\infty C_k^x$ and $\sum_{k=1}^\infty (\prod_{i=k}^\infty C_i^x)D_k^x$, intended as strong convergence limits. More details can be found in the proof of Lemma 6.4 (Section 11.1), and from [7].

**Lemma 6.4.** Suppose Assumptions 1 and 2 hold, and $S_T = \emptyset$. Furthermore, assume $\Gamma$ is not a singleton set, and $X(0) = x \in \Gamma$. Then,

$$ \Phi_x^\infty = \prod_{k=\infty}^0 C_k^x \quad \text{and} \quad W_x^\infty = D_0^x + \sum_{k=1}^\infty \left( \prod_{i=k}^\infty C_i^x \right) D_k^x \quad (6.7) $$

exist a.s. Moreover,

$$ (\Phi(\tau_t^x(t)), W(\tau_t^x(t))) \overset{c}{\to} (\Phi_x^\infty, W_x^\infty) \quad \text{for} \quad t \to \infty, \quad (6.8) $$

$$ E(\|W_x^\infty\|_1) < \infty, \quad \text{and for all continuity sets} \ A \ \text{of} \ \Phi_x^\infty, W_x^\infty), $$

$$ \lim_{t \to \infty} P \left( (\Phi(\tau_t^x(t)), W(\tau_t^x(t)) \in A | X(t) = x \right) = P((\Phi_x^\infty, W_x^\infty) \in A). \quad (6.9) $$

The columns of the matrix $\Phi_x^\infty$ corresponding to the species in $S_i$, $i = 1, \ldots, h$, are identical.
As the transient species are never produced but are all eventually degraded or transformed into other species, the assumption $S_T = \emptyset$ is more of convenience than necessity, if our aim is to study the limit behaviour of $\{(X(t), Z(t)) : t \geq 0\}$ as $t$ tends to $\infty$. As a matter of fact, the entries of transient species are necessarily zero within every closed irreducible set, so including them would not affect the analysis of stationary distributions. However, considering transient species would unnecessarily complicate the form $\Phi_\infty^\varepsilon$ takes and the proof of Lemma 6.4, so we assume $S_T = \emptyset$ to simplify our presentation.

Lemma 6.5. Suppose Assumptions 1 and 2 hold and that $\Gamma$ contains at least two elements. Let $(C, D)$ be distributed as $(C_0^\varepsilon, D_0^\varepsilon)$, and independent of $\{X(t) : t \geq 0\}$. Then $(\Phi_\infty^\varepsilon, W_\infty^\varepsilon)$ is a solution to the SRE

$$
\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \sim \begin{pmatrix} C & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} + \begin{pmatrix} 0 \\ D \end{pmatrix}, \quad (V_1, V_2) \parallel (C, D),
$$

(6.10)

Moreover, $W_\infty^\varepsilon$ is the weakly unique solution to the bottom row of $V_2 \sim CDV_2 + D$ in (6.10).

6.4 Proof of Theorem 6.1

Before proving Theorem 6.1 we need some preliminary results. The following lemma relies on the characterisation of the distribution of $Z(t)$ for $t \geq 0$ given in Proposition 5.3. Recall that the state space of $\{X(t) : t \geq 0\}$ is irreducible, hence whether the process $\{W(t) : t \geq 0\}$ in (4.6) is tight or not holds regardless of the initial condition $X(0)$.

Lemma 6.6. Suppose Assumption 1 holds. Furthermore, assume the process $\{X(t) : t \geq 0\}$ is ergodic and let $\{W(t) : t \geq 0\}$ be as in (4.6). Then, $\{W(t) : t \geq 0\}$ is tight if and only if the process $\{(X(t), Z(t)) : t \geq 0\}$ is ergodic for every initial condition.

The proof of the lemma is in the Appendix (Section 11.3).

Lemma 6.7. Assume that $\{X(t) : t \geq 0\}$ is ergodic and that $\Gamma$ contains at least two elements. Let $x \in \Gamma$ and define $\tau^x(t)$ as in (6.3). Then, the process

$$\{W(t) - W(\tau^x(t)) : t \geq 0\}
$$

is tight.

Proof. The process $\{X(t) : t \geq 0\}$ is ergodic, hence tight. Moreover, the state space $\Gamma$ is by assumption discrete. It follows that for any $\varepsilon > 0$ there exists a finite set $\Upsilon_\varepsilon \subseteq \Gamma$ such that

$$\inf_{t \geq 0} P(X(t) \in \Upsilon_\varepsilon) > 1 - \frac{\varepsilon}{2}.
$$

By assumption, the process $\{X(t) : t \geq 0\}$ is non-explosive. Hence, since $\Upsilon_\varepsilon$ is finite, it follows that for any $T > 0$ (with the given $\varepsilon$) there exists a finite set $\bar{\Upsilon}_{\varepsilon, T} \subseteq \Gamma$ such that

$$\inf_{t \geq 0} P\left(X(u) \in \bar{\Upsilon}_{\varepsilon, T} \text{ for all } u \in [t, t + T]\right) > 1 - \varepsilon.
$$

Indeed, for any $x \in \Upsilon_\varepsilon$, there is a finite set $\Upsilon^x_{\varepsilon, T}$ such that the probability of $\{X(u) \in \Upsilon^x_{\varepsilon, T} \text{ for all } u \in [t, t + T]\}$, given $X(t) = x$, is at least $1 - \frac{\varepsilon}{2}$. Now take $\bar{\Upsilon}_{\varepsilon, T}$ to be the union of $\Upsilon^x_{\varepsilon, T}$ over $x \in \Upsilon_\varepsilon$. The finiteness of $\bar{\Upsilon}_{\varepsilon, T}$ in turn implies that

$$\inf_{t \geq 0} P\left(\|B_X(u)\|_1 \leq M_{\varepsilon, T} \text{ for all } u \in [t, t + T]\right) > 1 - \varepsilon,
$$

where

$$M_{\varepsilon, T} = \max_{\bar{x} \in \bar{\Upsilon}_{\varepsilon, T}} \|B(\bar{x})\|_1 < \infty \quad \text{a.s.}$$

18
By standard renewal theory the age process \( \{ t - \tau^x(t): t \geq 0 \} \) is ergodic, and the unique stationary distribution has density with respect to the Lebesgue measure \([11, \text{Theorem 3.11}]\)

\[
f^x(u) = \frac{P(\tau^x_1 \geq u)}{E(\tau^x_1)} \quad \text{for all} \quad u \geq 0.
\]

Hence, for all \( \varepsilon > 0 \) there exists \( M'_\varepsilon \) such that

\[
\inf_{t \geq 0} P\left(t - \tau^x(t) \leq M'_\varepsilon \right) > 1 - \frac{\varepsilon}{2}.
\]

By Lemma 5.2,

\[
\|W(t) - W(\tau^x(t))\|_1 = \left\| \int_{\tau^x(t)}^t \Phi(u,t)B_X(u) du \right\|_1 \\
\leq \int_{\tau^x(t)}^t \|B_X(u)\|_1 du \leq (t - \tau^x(t)) \cdot \sup_{u \in [\tau^x(t),t]} \|B_X(u)\|_1 \quad \text{a.s.}
\]

It follows that for any \( \varepsilon > 0 \),

\[
\inf_{t \in \mathbb{R}_d, \geq 0} P\left(\|W(t) - W(\tau^x(t))\|_1 \leq M'_\varepsilon \cdot M_e, M_t \right) > 1 - \varepsilon,
\]

which concludes the proof. \( \square \)

We are now ready to prove Theorem 6.1.

**Proof of Theorem 6.1.** First, assume that \( \Gamma \) has at least two elements. Due to Lemma 6.6, it is sufficient to prove that the process \( \{W(t): t \geq 0\} \) is tight for any initial condition \( X(0) \). Assume \( X(0) = x \in \Gamma \). Since

\[
W(t) = (W(t) - W(\tau^x(t))) + W(\tau^x(t)) \quad \text{for all} \quad t \geq 0,
\]

the process \( \{W(t): t \geq 0\} \) is tight because of Lemmas 6.4 and 6.7. Indeed, Lemma 6.7 gives tightness of \( \{W(t) - W(\tau^x(t))\|t \geq 0\} \), while Lemma 6.4 provides convergence in distribution of \( W(\tau^x(t)) \) to a random variable as \( t \to \infty \), hence \( \{W(\tau^x(t))\|t \geq 0\} \) is tight. The proof is then concluded for \( \Gamma \) containing at least two elements.

If \( \Gamma \) contains a single element \( x \), then we consider a modification of the model with \( \Gamma' = \{x_1, x_2\} \) and \( \{X'(t): t \geq 0\} \) being any irreducible CTMC on \( \Gamma' \). Since \( \Gamma' \) is finite, \( \{X'(t): t \geq 0\} \) is ergodic \([26]\). We define a stochastic reaction system with stochastic environment \( (\mathcal{R}, \mathcal{N}', \{X'(t): t \geq 0\}) \) by letting

\[
\lambda_r(x_1, z) = \lambda_r(x_2, z) = \lambda_r(x, z) \quad \text{for all} \quad z \in \mathbb{Z}^d_{\geq 0},
\]

and all reactions \( y_r \to y'_r \) of \( \mathcal{R} \). Then the distribution of \( \{Z'(t): t \geq 0\} \) is the same as that of \( \{Z(t): t \geq 0\} \). Specifically, \( \{(X(t), Z(t)): t \geq 0\} \) is ergodic for every initial condition if and only if \( \{(X'(t), Z'(t)): t \geq 0\} \) is ergodic for every initial condition, and since \( \Gamma' \) contains at least two elements the proof is concluded. \( \square \)

### 6.5 Characterisation of the stationary distribution

We characterise the explicit structure of the stationary distribution of models satisfying Assumptions 1 and 2 in the case \( m_i \in \{0, 1\} \) for \( 1 \leq i \leq d \). If \( d = 1 \), the model is a \( M/M/\infty \) queue with stochastic environment, and the stationary distribution has been characterised by a recursive equation under the assumption that \( \Gamma \) has finitely many states \([27]\).

For \( x \in \Gamma \), let \( q_x \) be the rate of the exponential holding time of \( \{X(t): t \geq 0\} \) in \( x \), and let \( G_x: \mathbb{R}_0 \to \mathbb{R}^d \) be the function defined as

\[
G_x(s) = \int_0^s e^{A(x)u} B(x) du = \sum_{n=0}^{\infty} \frac{1}{(n+1)!} A^n(x)B(x)s^{n+1}, \quad (6.11)
\]

where \( A(x) \) and \( B(x) \) are as given in (4.2) and (4.3), respectively.
Theorem 6.8. Suppose Assumptions 1 and 2 hold, that \( S_T = \emptyset \), and that \( \Gamma \) has at least two elements. Further, assume that \( m_i \in \{0, 1\} \) for \( 1 \leq i \leq d \). Then, \( \{(X(t), Z(t)) : t \geq 0\} \) is ergodic for every initial condition, and the irreducible closed sets are

\[
\Xi_{n_1, \ldots, n_h} = \big\{(x, z) \in \Gamma \times \mathbb{Z}_0^d : \sum_{j \in \mathcal{S}_i} z_j = n_i \text{ for } i = 1, \ldots, h\big\},
\]

with \( n_1, \ldots, n_h \in \mathbb{Z}_0 \). Moreover, the unique stationary distribution restricted to \( \Xi_{n_1, \ldots, n_h} \) is given by

\[
\hat{\pi}_{n_1, \ldots, n_h}(x, z) = \pi(x) \int_{\mathbb{R}_0^{d_1} \times \mathbb{R}_0^{d_2}} P\left(N(w) + \sum_{i=1}^h M_i(n_i, u \epsilon_j) = z\right) \mu_x(du, dw),
\tag{6.12}
\]

for any \( (x, z) \in \Xi_{n_1, \ldots, n_h} \), where \( N(w) \sim \text{Pois}(w) \) and \( M_i(k, v) \sim \text{Multi}(k, v) \), \( j \) is any function \( j : \{1, \ldots, h\} \to \{1, \ldots, d\} \) such that \( S_{j(i)} \in \mathcal{S}_i \) for all \( i \), \( \pi \) is the stationary distribution of \( \{X(t) : t \geq 0\} \) and \( \mu_x \) is the distribution of

\[
\left(e^{A(x)}U_x, \Phi_{x,x}^\infty, e^{A(x)}U_xW_x^\infty + G_x(U_x)\right),
\tag{6.13}
\]

with \( U_x \sim \text{Exp}(q_x) \) being independent of \( (\Phi_{x,x}^\infty, W_x^\infty) \).

Proof. Ergodicity for any initial condition follows from Theorem 6.1. The form of the irreducible sets follow from the definition of the partition of \( \mathcal{S} \).

Let \( (x, z) \in \Xi_{n_1, \ldots, n_h} \). To obtain the stationary distribution evaluated at \( (x, z) \) we analyse

\[
P(X(t) = x, Z(t) = z | X(0) = x, Z(0) = z) = P(X(t) = x | X(0) = x) P(Z(t) = z | Z(0) = z, X(0) = x, X(t) = x),
\tag{6.14}
\]

which converges to \( \hat{\pi}_{n_1, \ldots, n_h}(x, z) \) as \( t \to \infty \) by ergodicity for every initial condition. The first term in the product is a consequence of (2.1). By ergodicity of \( \{X(t) : t \geq 0\} \), to show that (6.14) converges to (6.12) as \( t \to \infty \), it suffices to show that

\[
\lim_{t \to \infty} P(Z(t) = z | Z(0) = z, X(0) = x, X(t) = x) = \int_{\mathbb{R}_0^{d_1} \times \mathbb{R}_0^{d_2}} P\left(N(w) + \sum_{i=1}^h M_i(n_i, u \epsilon_j) = z\right) \mu_x(du, dw).
\]

By Proposition 5.3,

\[
\mathcal{L}(Z(t) | Z(0) = z, X(0) = x, X(t) = x) = \mathcal{L}\left(N(W(t)) + \sum_{i=1}^d \tilde{M}_i(z_i, \Phi(t) e_i) \right | X(0) = x, X(t) = x),
\tag{6.15}
\]

where \( N(w) \sim \text{Pois}(w) \) and \( \tilde{M}_i(k, v) \sim \text{Multi}(k, v) \), \( i = 1, \ldots, d \), are independent random variables. Furthermore, on the event \( \{X(0) = x, X(t) = x\} \),

\[
(\Phi(t), W(t)) = \left(e^{A(x)(t-\tau^x(t))} \Phi(\tau^x(t)), G_x(t-\tau^x(t)) + e^{A(x)(t-\tau^x(t))} W(\tau^x(t))\right) \text{ a.s.},
\tag{6.16}
\]

where \( G_x \) is defined in (6.11). Indeed, note that

\[
G_x(t - \tau^x(t)) = \int_0^{t-\tau^x(t)} e^{A(x)u} B(x) du = \int_{\tau^x(t)}^t e^{A(x)(t-u)} B(x) du.
\]

Motivated by (6.15) and (6.16), we study the limit of the joint distribution

\[
\mathcal{L}(t - \tau^x(t), \Phi(\tau^x(t)), W(\tau^x(t)) | X(0) = x, X(t) = x)
\]

as \( t \to \infty \). For any \( u, s \in \mathbb{R}_{\geq 0} \) with \( u \leq s \), define the event

\[
B_{u,s} = \{ \text{there is no jump of } \{X(t) : t \geq 0\} \text{ in } (s - u, s]\}.
\]

20
Observe that for any continuity set $A \subseteq \mathbb{R}^{d \times d} \times \mathbb{R}^{d}$ of $(\Phi_{\infty}^{x}, W_{\infty}^{x})$, and for any $s \leq t$, the event 
\[
\{X(t) = x, t - \tau^{x}(t) > s, (\Phi(\tau^{x}(t)), W(\tau^{x}(t))) \in A, X(0) = x\}
\] 
coincides with the event 
\[
\{B_{s,t}, X(t - s) = x, (\Phi(\tau^{x}(t - s)), W(\tau^{x}(t - s))) \in A, X(0) = x\},
\] 
since if $t - \tau^{x}(t) > s$, one has $\tau^{x}(t - s) = \tau^{x}(t) - s$.

Using the above equality one gets
\[
P(t - \tau^{x}(t) > s, (\Phi(\tau^{x}(t)), W(\tau^{x}(t))) \in A | X(0) = x, X(t) = x) = P(B_{s,t} \mid X(0) = x, X(t - s) = x, (\Phi(\tau^{x}(t - s)), W(\tau^{x}(t - s))) \in A)
\]
\[
\times P((\Phi(\tau^{x}(t - s)), W(\tau^{x}(t - s))) \in A | X(0) = x, X(t - s) = x)
\]
\[
\times P(X(0) = x, X(t - s) = x) = e^{-q_{x}s},
\]
which leads to the expression
\[
P(t - \tau^{x}(t) > s, (\Phi(\tau^{x}(t)), W(\tau^{x}(t))) \in A | X(0) = x, X(t) = x) = e^{-q_{x}s}P((\Phi(\tau^{x}(t - s)), W(\tau^{x}(t - s))) \in A | X(0) = x, X(t - s) = x) P(X(0) = x, X(t - s) = x)
\]
\[
\times P(X(0) = x, X(t) = x).
\]

Note that the third term on the right-hand side of the above equality converges to 1 as $t \to \infty$. Regarding the second term, it follows from (6.9) in Lemma 6.4 that
\[
\lim_{t \to \infty} P((\Phi(\tau^{x}(t - s)), W(\tau^{x}(t - s))) \in A | X(0) = x, X(t - s) = x) = P((\Phi_{\infty}^{x}, W_{\infty}^{x}) \in A | X(0) = x)
\]
\[
= P((\Phi_{\infty}^{x}, W_{\infty}^{x}) \in A),
\]
using that $A$ is a continuity set for $(\Phi_{\infty}^{x}, W_{\infty}^{x})$. Note that the conditioning on $X(0) = x$ was dropped in the last equality because the random variables $\Phi_{\infty}^{x}$ and $W_{\infty}^{x}$ do not depend on $X(0)$ (see (6.7)), since $x$ is eventually visited almost surely for any initial condition. Hence, $t - \tau^{x}(t)$ and $(\Phi(\tau^{x}(t)), W(\tau^{x}(t)))$ are asymptotically conditionally independent given $\{X(0) = x, X(t) = x\}$, as $t$ goes to infinity. Moreover, conditioned on $\{X(0) = x, X(t) = x\}$, $t - \tau^{x}(t)$ converges in distribution to an exponential random variable with rate $q_{x}$. Now using that $\Phi(t)$ converges to a matrix with identical columns for the species in $S_{i}$, $i = 1, \ldots, h$ (Lemma 6.4, Levy’s continuity theorem [11], (6.15), and (6.16), it follows that (6.12) holds, which concludes the proof.

The condition $S_{T} = \emptyset$ is not strict. If some transient species are present at time zero then they will eventually be degraded or converted, but never produced. Hence at stationarity there are no transient species present. Moreover, if all species are properly produced and degraded, that is, if $S = S_{P}$, then the multinomial terms in (6.12) are absent and the stationary distribution of $\{(X(t), Z(t)) : t \geq 0\}$ takes a simpler form, as expressed in the following result.

**Corollary 6.9.** Suppose Assumptions 1 and 2 hold, that $S = S_{P}$, and that $\Gamma$ has at least two elements. Further, assume that $m_{i} \in \{0, 1\}$ for $1 \leq i \leq d$. Then, $\{(X(t), Z(t)) : t \geq 0\}$ is ergodic, and the unique stationary distribution is given by
\[
\tilde{\pi}(x, z) = \pi(x) \int_{\mathbb{Z}_{\geq 0}^{d}} P(\text{Pois}(w) = z) \mu_{x}(dw) \quad (6.18)
\]
for any $(x, z) \in \Gamma \times \mathbb{Z}_{\geq 0}^{d}$, where $\pi$ is the stationary distribution of $\{X(t) : t \geq 0\}$ and $\mu_{x}$ is the distribution of $e^{A(x)\Phi_{\infty}^{x}}W_{\infty}^{x} + G_{x}(U_{x})$, with $U_{x} \sim \text{Exp}(q_{x})$ being independent of $(\Phi_{\infty}^{x}, W_{\infty}^{x})$.  

21
Proof. Ergodicity of \( \{(X(t), Z(t)) : t \geq 0\} \) follows from Theorem 6.8 and from the assumption that \( S = S_P \). Furthermore, it follows from Lemma 6.2(a) that \( \|\Phi^x_{\infty}\|_1 = 0 \), hence (6.18) follows from (6.12) and (6.13) in Theorem 6.8. \( \square \)

Corollary 6.9 can be used to find an expression for the factorial moments, in particular, for the case \( d = 1 \) and \( m = 1 \). In this scenario, \( \{Z(t) : t \geq 0\} \) is distributed as an M/M/\( \infty \) queue with stochastic environment (with \( \Gamma \) potentially infinite).

**Corollary 6.10.** Suppose Assumptions 1 and 2 hold, that \( S = S_P \), and that \( \Gamma \) has at least two elements. Moreover, if \( m_i \in \{0, 1\} \) for all \( 1 \leq i \leq d \). Let \( \hat{\pi} \) be the stationary distribution of \( \{(X(t), Z(t)) : t \geq 0\} \) given in (6.18) and let \( (X, Z) \) be a random variable with distribution \( \hat{\pi} \). For \( q \in \mathbb{Z}_{\geq 0} \), denote the \( q \)-th factorial moment by

\[
m_q = E\left( \frac{Z!}{(Z-q)!} 1_{\{z \in \mathbb{Z}_{\geq 0}^d : z \geq q\}}(Z) \right),
\]

where \( z = \prod_{i=1}^d z_i! \). Then,

\[
m_q = \sum_{x \in \Gamma} \pi(x) \int_{\mathbb{R}_{\geq 0}^d} u^q \mu_x(du).
\] (6.19)

Moreover, if \( d = 1 \), then \( m_0 = 1 \), and \( m_q, q \in \mathbb{Z}_{>0} \), can be found iteratively by

\[
m_q = \frac{1}{1 - E\left( (C^x_0)^q \right)} \sum_{i=0}^{q-1} \binom{q}{i} E\left( (C^x_0)^i (\tilde{D}^x_0)^{q-i} \right) m_i
\]

where

\[
\tilde{D}^x_0 = \begin{cases} 
\tilde{\lambda}_0(x) (1 - C^x_0) (1 - e^{-\tilde{\lambda}_0(x) U_x}) + e^{-\tilde{\lambda}_0(x) U_x} D^x_0 & \text{if } \tilde{\lambda}_0(x) > 0, \\
(1 - C^x_0) U_x \tilde{\lambda}_0(x) + e^{-\tilde{\lambda}_0(x) U_x} D^x_0 & \text{if } \tilde{\lambda}_0(x) = 0,
\end{cases}
\]

and \( U_x \) is an Exponential random variable with rate \( q_x \), independent of \( (C^x_0, D^x_0) \).

Proof. Since \( E\left( \frac{Z!}{(Z-q)!} 1_{\{z \in \mathbb{Z}_{\geq 0}^d : z \geq q\}}(Z) \right) = u^q \) for \( u \in \mathbb{R}_{\geq 0}^n, q \in \mathbb{Z}_{\geq 0} \), and \( Z' \sim \text{Pois}(u) \), it holds that

\[
E\left( \frac{Z!}{(Z-q)!} 1_{\{z \in \mathbb{Z}_{\geq 0}^d : z \geq q\}}(Z) \right) = \sum_{x \in \Gamma} \pi(x) E\left( \frac{Z!}{(Z-q)!} 1_{\{z \in \mathbb{Z}_{\geq 0}^d : z \geq q\}}(Z) \bigg| X = x \right),
\]

which proves (6.19). Assume that \( d = 1 \). By Theorem 6.8, Lemma 6.5 and the remarks above the corollary, \( \mu_x \) is the weakly unique solution of the SRE (6.13),

\[
V \sim C^x_0 V + ((1 - C^x_0) G_x(U_x) + e^{-\tilde{\lambda}_0(x) U_x} D^x_0), \quad V \parallel (C^x_0, D^x_0). \] (6.20)

Note that

\[
G_x(U_x) = \int_0^{U_x} e^{-\tilde{\lambda}_0(x) u} \tilde{\lambda}_0(x) du = \begin{cases} 
\tilde{\lambda}_0(x) (1 - e^{-\tilde{\lambda}_0(x) U_x}) & \text{if } \tilde{\lambda}_0(x) > 0, \\
U_x \tilde{\lambda}_0(x) & \text{if } \tilde{\lambda}_0(x) = 0.
\end{cases}
\]

Hence, (6.20) becomes

\[
V \sim C^x_0 V + \tilde{D}^x_0, \quad V \parallel (C^x_0, \tilde{D}^x_0).
\]

It follows from the binomial theorem and the independence of \( V \) and \((C^x_0, \tilde{D}^x_0)\) that for \( q \in \mathbb{Z}_{>0} \),

\[
E(V^q) = E((C^x_0 V + \tilde{D}^x_0)^q) = E(V^q) E((C^x_0)^q) + \sum_{i=0}^{q-1} \binom{q}{i} E(V^i) E((C^x_0)^i (\tilde{D}^x_0)^{q-i}). \] (6.21)
By Lemma 6.2(a) we have \(E((C_0^x)^n) < 1\) for some positive \(\alpha\), which implies \(E((C_0^x)^q) < 1\) for \(q \geq 1\), as \(d = 1\) and \(C_0^x \leq 1\) a.s. Hence by (6.21) we obtain

\[
\int_0^\infty u^q \mu_x(du) = E(V^q) = \frac{1}{1-E((C_0^x)^q)} \sum_{i=0}^{q-1} \binom{q}{i} E(V^i) E((C_0^x)^i (\tilde{D}_d^0)^{q-i})
\]

\[
= \frac{1}{1-E((C_0^x)^q)} \sum_{i=0}^{q-1} \binom{q}{i} E((C_0^x)^i (\tilde{D}_d^0)^{q-i}) \int_0^\infty u^i \mu_x(du),
\]

and the proof is concluded by (6.19).

\[\square\]

7 Generating samples from the conditional stationary distribution

The theory provides a way to simulate observations from the long-term dynamics of a model under the conditions of Theorem 6.8, given that the environmental state is known. The goal of this section is to show how to simulate from the conditional stationary distribution

\[
\tilde{\pi}(. \mid x) = \frac{\pi(x, \cdot)}{\pi(x)}.
\]

This is often desirable in the setting of biochemistry, as typically only a portion of the reactants is observable. For simplicity, we assume \(S = S_P\), as in Corollary 6.9. In our setting, the observed variables might be considered the environment.

Let \((X, Z)\) denote a random variable with distribution \(\tilde{\pi}\). If it is possible to simulate from the distribution \(\pi\) of \(X\), then it would also be possible to simulate from the joint distribution of \((X, Z)\) by first simulating \(X\) according to \(\pi\) and then \(Z\) according to the aforementioned conditional distribution. Hence, according to Theorem 6.8, the stationary distribution is characterised by the distribution of \(W_x^\infty\). With a slight abuse of notation, we denote by \(\mu_x\) the distribution of \(e^{A(x)U_i} W_x^\infty + G_x(U_x)\), with \(U_x\) as in Theorem 6.8. Then, in order to simulate from \(\tilde{\pi}(\cdot \mid x)\), by Theorem 6.8 it suffices to generate a sample \(U\) from \(\mu_x\), at which point

\[
\tilde{\pi}(\cdot \mid x) = \mathcal{L}(\text{Pois}(U)).
\]

Here we show how to generate a sample with approximate distribution \(\mu_x\), using that \(\mu_x\) is the weakly unique solution of (6.13).

First, for some \(n \in \mathbb{N}\) we generate a sequence \((C_{0,i}^x, D_{0,i}^x)_{i=1,...,n}\) of \(n\) independent realisations of \((C_0^x, D_0^x)\) in (6.4). To this aim, it is sufficient to generate \(n\) independent samples of the process \(\{X(t) : t \geq 0\}\) with \(X(0) = x\), until the stopping time \(\tau^*_1\). Once \(n\) independent samples of \(\{X(t) : t \geq 0\}\) are available, the desired sequence \((C_{0,i}^x, D_{0,i}^x)_{i=1,...,n}\) can be obtained by means of (4.10). Generate an observation \(U_x\) from an exponential random variable with rate \(q_x\), independent of \((C_{0,i}^x, D_{0,i}^x)_{i=1,...,n}\). Let \(V_0 = 0\) and define recursively for \(1 \leq i \leq n\),

\[
V_i = C_{0,i}^x V_{i-1} + D_{0,i}^x.
\]

We take \(V^*_n = e^{A(x)U_n} V_n + G_x(U_x)\) as an approximate sample from \(\mu_x\). The following result gives an estimate of the error made with this approximation in terms of the Wasserstein metric on \((\mathbb{Z}_d^d, \| \cdot \|_1))

\[
W_1(\nu_1, \nu_2) = \inf_{(Y_1, Y_2) : \mathcal{L}(Y_1) = \nu_1, \mathcal{L}(Y_2) = \nu_2} E(\|Y_1 - Y_2\|_1),
\]

and describes how the error decays to 0 exponentially as \(n\) goes to infinity.

**Proposition 7.1.** Suppose Assumptions 1 and 2 hold. Furthermore, assume that \(S = S_P\) and that \(\Gamma\) is not a singleton set. Let \(\alpha\) be as in Lemma 6.2, and let

\[
\Psi = \prod_{k=0}^{\alpha-1} (C_k^x) = \Phi(\tau^*_0, \tau^*_\alpha).
\]

Then for \(n \geq 1\),
(a) \( E(\|V_n\|_1) < \infty, \quad M = \int_{\mathbb{R}^d_{\geq 0}}\|u\|_1\mu_x(du) < \infty, \quad r = -\log(E(\|\Psi\|_1)) > 0; \)

(b) \( W_1(\mathcal{L}(V_n^\ast), \mu_x) \leq E(\|W^\infty_x\|)e^{-\frac{r}{\alpha}n}. \)

Proof. Consider a copy of \( W^\infty_x \) which is independent of \( U_x \). The distribution of \( e^{A(x)U_x}W^\infty_x + G_x(U_x) \) is \( \mu_x \) by Corollary 6.9. Since \( E(\|\Psi\|_1) < 1 \) by Lemma 6.2(a) and by the assumption \( S = S_P \), in order to prove part (a) we only need to show that both \( e^{A(x)U_x}V_n + G_x(U_x) \) and \( e^{A(x)U_x}W^\infty_x + G_x(U_x) \) are integrable.

Let \( \hat{\tau} \sim U_x \) be the first time the state \( x \) is left by \( \{X(t): t \geq 0\} \), assuming that \( X(0) = x \). Then, \( e^{A(x)U_x} \sim \Psi(0, \hat{\tau}) \) by Lemma 4.1, and \( E(e^{A(x)U_x}) \leq 1 \) by Lemma 5.2(c). Hence,

\[
E(\|G_x(U_x)\|_1) \leq E\left(\int_{0}^{U_x} \|e^{A(x)u}B(x)\|_1 du\right) \leq E\left(\int_{0}^{U_x} \|B(x)\|_1 du\right) = E(U_x)\|B(x)\|_1 < \infty.
\]

Since \( U_x \| (V_n, W^\infty_x) \), to prove (a) it suffices to show that both \( W^\infty_x \) and \( V_n \) are integrable. The former is true by Lemma 6.4. For the latter, by using the recursive definition of \( V_n \), the strong Markov property and the fact that \( E(\|C^\ast_0\|_1) \leq 1 \) (by Lemma 5.2(c)),

\[
E(\|V_n\|_1) = E\left(\left\|D^\ast_{0,n} + \sum_{i=1}^{n-1} \left(\prod_{j=i+1}^{n} C^\ast_{0,i}\right)D^\ast_{0,i}\right\|_1 \right) \leq \sum_{i=1}^{\alpha} E(\|D^\ast_{0,i}\|_1) + \sum_{i=1}^{\alpha} \sum_{j=1}^{\frac{n}{\alpha}} E(\|D^\ast_0\|_1) \leq \frac{\alpha}{1 - E(\|\Psi\|_1)} < \infty.
\]

For part (b), extend the sequence \((C^\ast_0, D^\ast_{0,1})_{i=1,\ldots,n}\) to an infinite sequence of i.i.d. random variables, independent of \( U_x \). Then, define

\[
V = D^\ast_{0,1} + \sum_{j=2}^{\infty} \left(\prod_{i=1}^{j-1} C^\ast_{0,i}\right)D^\ast_{0,j} \sim W^\infty_x
\]

so that \( e^{A(x)U_x}V + G^\ast(U_x) \) is distributed as \( \mu_x \) by Corollary 6.9. Moreover, as shown in the proof of Lemma 6.4, by exchangeability

\[
V_n \sim \tilde{V}_n = D^\ast_{0,1} + \sum_{j=2}^{n} \left(\prod_{i=1}^{j-1} C^\ast_{0,i}\right)D^\ast_{0,j}.
\]

Then we have

\[
W_1(\mathcal{L}(V_n^\ast), \mu_x) \leq E(\|e^{A(x)U_x}V + G^\ast(U_x) - e^{A(x)U_x}\tilde{V}_n - G^\ast(U_x)\|_1) \leq E(\|e^{A(x)U_x}\|_1)E(\|V - \tilde{V}_n\|_1) \leq E(\|V - \tilde{V}_n\|_1).
\]

The proof is then concluded by observing that

\[
V - \tilde{V}_n = \sum_{j=n+1}^{\infty} \left(\prod_{i=1}^{n} C^\ast_{0,i}\right)D^\ast_{0,j} = \left(\prod_{i=n}^{\infty} C^\ast_{0,n-i}\right)\tilde{V}, \quad \tilde{V} \sim V,
\]

\[
E(\|\prod_{i=0}^{n} C^\ast_{0,n-i}\|_1) \leq E(\|\Psi\|_1)(n/\alpha) \leq E(\|\Psi\|_1)(n/\alpha)^{-1} = e^{-r(\alpha)n}, \text{ and } E(\|V\|) = E(\|W^\infty_x\|). \quad \square
\]

Example 7.1. Consider the following reaction network with species \( S_1, S_2, S_3 \) and reactions

\[
S_1 + S_3 \xrightarrow{\lambda_1} S_1, \quad S_2 \xrightarrow{\lambda_2} S_2 + S_3
\]

\[
0 \xrightarrow{\alpha_1} S_1 + 2S_2 \xrightarrow{\alpha_2} 2S_1 + 2S_2 \xrightarrow{\alpha_1} 0,
\]

taken with stochastic mass-action kinetics. We take the top line of reactions to be a stochastic reaction network with stochastic environment given by the the bottom line of the reactions, and rate functions

\[
\lambda_1(x, z) = \kappa_1 x_1 z, \quad \lambda_2(x, z) = \kappa_2 x_2,
\]

24
\[ \alpha_1(x) = a_1, \quad \alpha_2(x) = a_2x_1x_2^2, \quad \alpha_3(x) = a_3x_1^2x_2, \quad \alpha_4(x) = a_4x_1^2x_2, \]

where \( x = (x_1, x_2) \) denotes the counts of \( S_1, S_2 \), and \( z \) the counts of \( S_3 \). The stationary distribution of the environment is

\[ \pi(x_1, x_2) \propto \frac{b_1^{x_1} b_2^{x_2}}{x_1! x_2!}, \]

with \( b_1 = \left( \frac{a_1a_2}{a_3a_4} \right)^{\frac{1}{3}}, \quad b_2 = \left( \frac{a_1a_3}{a_4a_2} \right)^{\frac{1}{3}} \)

on an irreducible component \( \Gamma \subset \mathbb{Z}^2_{\geq 0} \) determined by \( x_1 + x_2 \mod 3 = k, \ k = 0, 1, 2 \ [2] \).

Assumptions 1 and 2 hold, and Lemma 6.2 holds with \( \alpha = 1 \). From Theorem 6.8, the conditional distribution becomes

\[ \tilde{\pi}(z|x_1, x_2) = \int_{\mathbb{R}_+} P(\text{Pois}(y) = z)\mu_x(dy) = \int_{\mathbb{R}_+} \frac{y^z}{z!}e^{-y}\mu_x(dy) \]

with \( x = (x_1, x_2) \). Furthermore, \( \mu_x = \mathcal{L}(e^{-\kappa_xU_x}V + G_x(U_x)) \), where \( U_x \) is an exponential random variable with rate \( q_x \) that is independent of \( V \), such that

\[ V \sim e^{-\int_0^{\tau_1} \kappa_1X_1(s)ds}V + \int_{\tau_1}^{\tau_2} \kappa_2X_2(u)e^{-\int_0^u \kappa_1X_1(s)ds}du, \]

\[ q_x = \alpha_1(x) + \alpha_2(x) + \alpha_3(x) + \alpha_4(x). \]

Using the iterative method of Section 7 with rate constants \( (\kappa_1, \kappa_2, a_1, a_2, a_3, a_4) = (3, 4, 1, 2, 8, 8) \), the realisations of \( \tilde{\pi}(z|x_1, x_2) \) are found in Figure 1 for two different values of \( z = (x_1, x_2) \) (with \( K \) being the iteration number as in Proposition 7.1). The mixture of the Poisson distribution changes with \( (x_1, x_2) \).

![Figure 1: Simulated densities with K = 20 for two different states of the environment.](image)

---

### 8 Gene Regulatory Networks

An important class of reaction networks is that of gene regulatory networks [19, 23, 30] that are building blocks of large biological networks and underlie many cellular processes [29]. Gene regulation is the process by which a gene is regulated by other molecules, generally known as transcription factors (TFs). Regulation controls the production of the gene product, a protein. An elementary version of a gene regulatory system allows the gene to be in two states, an activated state where protein is produced and an inactivated state where protein is not produced (or produced at a lower rate). The activation and inactivation is controlled by TFs with potential feedback mechanisms. Here we consider two models of a gene regulatory mechanism [21, 30] and show that the stationary distributions exist and can be exactly characterised.
The first is the following reaction network with stochastic mass-action,

\[ G \xrightarrow{\alpha_1}{\alpha_2} G', \quad G' \xrightarrow{\lambda_1} P + G', \quad P \xrightarrow{\lambda_2} \emptyset, \]

where the dynamics of \( G, G' \) constitute the environment, \( x = (x_1, x_2) \) denotes the counts of \( G, G' \) and \( z \) that of \( P \). Here, \( G \) denotes the inactive gene and \( G' \) the activated gene. While in state \( G' \) the gene produces protein \( P \). Hence, we have a stochastic reaction network with a stochastic environment described as

\[ 0 \xrightarrow{\lambda_1} P, \quad P \xrightarrow{\lambda_2} \emptyset. \]

Denote the evolution of the counts of \( G, G' \) by \((X_1(t), X_2(t)) : t \geq 0 \) and observe that the total count is conserved \( X_1(t) + X_2(t) = N \) for all \( t \geq 0 \). In [30], the authors consider \( N = 1 \) and found the stationary distribution using Kummer’s differential equation after a suitable change of variable. Biologically, \( N = 1 \) would be the case for haploid (or monoploid) cells with one chromosome, such as bacteria. However, many organisms are diploid (\( N = 2 \), like humans) or polyploid (\( N \geq 2 \), like many plants). Also, in particular in synthetic biology or experimental system biology, one might consider the situation in which the gene is located on a plasmid, a small chromosomal-like structure, with \( N \geq 1 \). If \( N > 1 \), then the framework of [30] will not work to characterise the stationary distribution.

To characterise the stationary distribution for general \( N \) we rely on Theorem 6.8. The stationary distribution of the counts of \( G, G' \) is a binomial distribution [2],

\[ \pi_N(x) = \binom{N}{x_1} \left( \frac{a_2}{a_1 + a_2} \right)^{x_1} \left( \frac{a_1}{a_1 + a_2} \right)^{x_2}, \quad x \in \Gamma_N = \{(x_1, x_2) \in \mathbb{Z}_+^2 : x_1 + x_2 = N\}. \]

The function \( G_x(s) \) is

\[ G_x(s) = \sum_{n=0}^{\infty} \frac{1}{(n+1)!} (-\kappa_2)^n \kappa_1^t (N - x_1) s^{n+1} = \frac{\kappa_1^t (N - x_1)}{\kappa_2} (1 - e^{-\kappa_2 s}) \quad \text{for } s \geq 0. \]

Let \( U_x \) be the exponential holding time in \( x \) with rate \( q_x = a_1x_1 + a_2(N - x_1) \). Then the stationary distribution \( \tilde{\pi}_N \) is

\[ \tilde{\pi}_N(x, z) = \pi_N(x) \int_{\mathbb{R}_+} P(\text{Pois}(y) = z) \mu_x(dy), \quad (x, z) \in \Gamma \times \mathbb{Z}_+, \]

(8.1)

where \( \mu_x \) is characterised in Theorem 6.8 as the law of \( V^* = e^{-\kappa_2 U_x} V + G_x(U_x) \), and \( V \) solves an explicit SRE.

The second example is a generalisation of the first and models a more complex mechanism adding a TF as a separate species [21]. Specifically, the system is

\[ G \xrightarrow{\alpha_1}{\alpha_2} G', \quad G \xrightarrow{\lambda_1} P + G \quad G' \xrightarrow{\lambda_1'} P + G', \quad P \xrightarrow{\lambda_2} \emptyset, \]

\[ P \xrightarrow{\lambda_3} P + TF, \quad TF \xrightarrow{\lambda_4} \emptyset, \]

\[ \alpha_1(x) = a_1x_1, \quad \alpha_2(x) = a_2x_2, \quad \lambda_1(x, y) = \kappa_1 x_2, \quad \lambda_1'(x, y) = \kappa_1' x_2, \]

\[ \lambda_2(x, y) = \kappa_2 y, \quad \lambda_3(x, y, z) = \kappa_3 y, \quad \lambda_4(x, y, z) = \kappa_4 z, \]

where \( x = (x_1, x_2) \) denotes the counts of \( G, G' \) the count of \( P \) and \( z \) that of \( TF \), and the rate functions dependence of \( x, y, z \) reflect how we will build up the network.

In this example, the gene product \( P \) can be produced from either version of the gene, but at different rates \( \lambda_1, \lambda_1' \). Subsequently, \( P \) is involved in the production of a TF. The TF does not affect the activation of the gene nor the production of \( P \). A still more complete model would consider a feedback mechanism, where the TF binds to the gene to enhance production of \( P \).
As before we might at first take the dynamics of \( G, G' \) to constitute the stochastic environment and

\[
0 \xrightarrow{\lambda_1 + \lambda'_1} P, \quad P \xrightarrow{\lambda_3} \emptyset
\]

to be a stochastic reaction network with stochastic environment. Using Theorem 6.8 we characterise the stationary distribution, but now \( \mu_x \) in (8.1) is the law of

\[
e^{-\kappa_2 U_x} V_1 + \frac{\kappa'_1 N + x_1(\kappa_1 - \kappa'_{1})}{\kappa_2} \left(1 - e^{-\kappa_2 U_x}\right),
\]

where the law of \( V_1 \) is the unique solution of an explicit SRE, and \( U_x \sim \text{Exp}(\alpha_1(x) + \alpha_2(x)) \).

At a second level, to understand the evolution of the TF, we might take the joint dynamics of \( G, G', P \) represented by \( (X_1(t), X_2(t), Y(t)) : t \geq 0 \) to constitute the stochastic environment, where \( Y(t) \) is the count of \( P \), and

\[
0 \xrightarrow{\lambda_4} TF, \quad TF \xrightarrow{\lambda_4} \emptyset
\]

to be a stochastic reaction network with stochastic environment. Using (6.12) the conditional probability is

\[
\hat{\pi}(x, y) = \int_{\mathbb{R}^+} P(\text{Pois}(u) = z) \mu_{(x, y)}(du).
\]

where \( \mu_{(x, y)} \) is the law of

\[
e^{-\kappa_4 U_{(x, y)}} V_2 + \frac{\mu_{(x, y)}}{\kappa_4} \left(1 - e^{-\kappa_4 U_{(x, y)}}\right),
\]

\( U_{(x, y)} \) is an exponential variable with rate \( q_x = \alpha_1(x) + \alpha_2(x) + \lambda_1(x, y) + \lambda'_1(x, y) + \lambda_2(x, y) \), and \( V_2 \) solves an explicit SRE.

9 Discussion

A class of stochastic reaction networks with stochastic environment is studied and their finite time distribution as well as their closed form stationary distribution is characterised (when it exists). In applications, the stochastic environment itself is a stochastic reaction network. We focused on mono-molecular reaction networks, networks where reaction rates depend linearly on the counts of the species. This had the advantage that the finite time distribution of the molecular counts can be characterised through the paths of the stochastic environment [20]. The long term dynamics can then be characterised using the finite time characterisation. To go beyond the mono-molecular cases, finite time distributions are generally not known, perhaps with [25] as the exception where the reversible reaction \( A + B \xleftrightarrow{\kappa} C \) is studied. However the method of [25] cannot be extended to the present case.

The linearity assumption plays an important role in controlling non-explosivity of the molecular counts. If reactions are not mono-molecular, then it appears we have less control of the behaviour of the system. In fact, we relied on the characterisation in [20] which in turn builds on a close relationship between the stochastic system and a corresponding deterministic (ODE) system. If the reaction network is not mono-molecular, then this relationship between stochastic and deterministic counterparts is generally lost. See also Example 4.1 and 4.2.

We end by listing some open directions for future research.

Consider feedback imposed by the protein in the gene activation step, assuming stochastic mass-action kinetics,

\[
G + P \xrightarrow{\alpha_1} G' + P, \quad G' \xrightarrow{\alpha_2} G', \quad G' \xrightarrow{\lambda'_1} P + G', \quad P \xrightarrow{\lambda_2} \emptyset,
\]

with \( \alpha_1(x) = a_1 x_1 \). Clearly, we cannot take \( G, G' \) to constitute the environment as the transformation of \( G \) to \( G' \) is catalysed by \( P \). The evolution of the counts of \( P \) can be stochastically upper-bounded by the reaction network

\[
\emptyset \xrightarrow{\tilde{\alpha}_1} P, \quad P \xrightarrow{\alpha} \emptyset,
\]

where \( \tilde{\alpha}_1(x) = Na_1 \) and \( N \) is the conserved amount of \( G, G' \). Hence the feedback network is ergodic. More involved frameworks are needed to characterise the stationary distribution in such cases.
A consequence of the Markov assumption (2.1) is reflected in the following identity
\[
P(X(t) = x | (X(s), Z(s)) = (x', z')) = P(X(t) = x) X(s) = x', \quad 0 < s < t,
\]
that significantly simplified the structure of the stationary distribution here. In [6], a reduction method for arbitrary reaction networks is considered based on a locally independence structure derived from a graph (the so-called Kinetic Independence Graph). The locally independence structure is a weaker assumption than assumption (2.1). It would be interesting to characterise stochastic stability of reaction networks having independence structures weaker than (2.1).

Generating a sample from \( \mu_x \) will be extremely cumbersome if the conditioning event \( \{X = x\} \) is rare as any excursion path \( (\tau_0^x, \tau_1^x) \) will be large. In such cases, other simulation techniques are required. Here importance sampling might be applicable. It would be of interest to develop techniques that improve sampling for the models considered here.

10 Acknowledgement

CW is supported by the Independent Research Fund Denmark. DC has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme grant agreement No. 743269 (CyberGenetics project). We are indebted to a reviewer who pointed out errors and imprecisions in the original version. The manuscript has greatly improved due to reviewer’s comments and suggestions.

11 Appendix

11.1 Proof of Lemma 6.4

Since \( \tau_0^x = 0 \), then \( C_{n-1}^x = 0 \) and \( D_{n-1}^x = 0 \). By (6.5) and (6.6),
\[
(\Phi(\tau^x(t)), W(\tau^x(t))) = \left( \prod_{k=0}^{n-1} C_k^x, D_{n-1}^x + \sum_{k=0}^{n-2} \left( \prod_{i=k+1}^{n-1} C_i^x \right) D_k^x \right)
\]
\[
= \left( C_0^x, D_0^x \right) \left( C_{n-2}^x, D_{n-2}^x + C_{n-1}^x D_{n-2}^x \right) \left( \prod_{i=0}^{n-1} C_i^x \right) D_0^x.
\]

It follows from the renewal property of \( \{X(t): t \geq 0\} \) that
\[
(\{C_0^x, D_0^x\}, (C_1^x, D_1^x), \ldots, (C_{n-1}^x, D_{n-1}^x), \{X(s): \tau_{n-1}^x \leq s \leq t\})
\]
is distributed as
\[
(\{C_{n-1}^x, D_{n-1}^x\}, (C_{n-2}^x, D_{n-2}^x), \ldots, (C_0^x, D_0^x), \{X(s): \tau_{n-1}^x \leq s \leq t\})
\]
for all \( t \geq 0 \). Hence,
\[
(\Phi(\tau^x(t)), W(\tau^x(t))) \sim (\hat{\Phi}^x(t), \hat{W}^x(t)),
\]
where
\[
\hat{\Phi}^x(t) = \prod_{k=n-1}^{0} C_k^x = C_0^x C_1^x \cdots C_{n-1}^x,
\]
\[
\hat{W}^x(t) = D_0^x + \sum_{k=1}^{n-1} \left( \prod_{i=k}^{n-1} C_i^x \right) D_k^x = D_0^x + C_0^x D_1^x + C_0^x C_1^x D_2^x + \cdots + \left( \prod_{i=n-2}^{0} C_i^x \right) D_{n-1}^x.
\]
are functions of \( \{X(s) : 0 \leq s \leq t\} \). Moreover, the distributional equality of (11.1) and (11.2) implies that (11.3) also holds conditionally on the sigma algebra of \( \{X(s) : \tau_{n_i}^x \leq s \leq t\} \), in particular on the event \( \{X(t) = x\} \).

We will show that \( (\Phi^x(t), \tilde{W}^x(t)) \) converges strongly to a random variable \((\Phi^x_\infty, \tilde{W}^x_\infty)\) as \( t \to \infty \), by proving it separately for the two entries, thus proving (6.7) and (6.8).

By assumption, \( \tilde{F}^x(t) \) (which is distributed as \( \Phi(\tau^x(t)) \)) is a block matrix with one block \( \Phi^x_j(t) \), \( 1 \leq i \leq h \), for each strongly connected component \( S_i \) of the partition of \( \mathcal{S} \), and one block for \( \mathcal{S}_0 \). Let \( \alpha \) be as in Lemma 6.2(b). Then with positive probability the matrices \( \tilde{F}^x_j(\tau^x_\infty) \), \( 1 \leq i \leq h \) (which are distributed as the block of \( \Phi(\tau^x_j) \) relative to \( S_i \)), have all entries positive. Let \( C^x_{i,j} \) be the block of \( C^x_j = \Phi_1(\tau^x_j, \tau^x_{j+1}) \) relative to \( S_i \). Note that for any \( j \) the product \( (C^x_{i,j,a+1})^\top (C^x_{i,j,a+2})^\top \ldots (C^x_{i,j,a+\ell})^\top \) is a stochastic matrix a.s. (for being product of stochastic matrices, see Lemma 5.2(a)) and has a positive probability of having all entries positive (for being distributed as the transpose of \( \tilde{F}^x_j(\tau^x_\infty) \)). Hence, by [7, Theorem 1.4], for any fixed \( 1 \leq i \leq h \) the limit
\[
M^x_{\infty} = \lim_{n \to \infty} \prod_{j=0}^n (C^x_{i,j,a+1})^\top (C^x_{i,j,a+2})^\top \ldots (C^x_{i,j,a+\ell})^\top
\]
eexists a.s. Likewise, the limit
\[
M^x_{\infty}(\ell) = \lim_{n \to \infty} \left( \prod_{j=0}^n (C^x_{i,j,a+1+\ell})^\top (C^x_{i,j,a+2+\ell})^\top \ldots (C^x_{i,j,a+\ell+\ell})^\top \right) (C^x_{i,\ell-1})^\top (C^x_{i,\ell-2})^\top \ldots (C^x_{i,0})^\top
\]
eexists a.s. for \( 1 \leq \ell \leq \alpha - 1 \). Note that \( M^x_{\infty} \) and \( M^x_{\infty}(\ell) \) are the limits of \( \alpha \) different subsequences of \( \prod_{j=0}^n C^x_{i,j} \), which define a partition of it. Hence, since \( \lim_{n \to \infty} \tilde{F}^x_1(t) = \lim_{n \to \infty} \prod_{j=0}^n C^x_{i,j} \), the matrices \( M^x_{\infty} \) and \( M^x_{\infty}(\ell) \), for \( 1 \leq \ell \leq \alpha - 1 \), are the accumulation points of \( \tilde{F}^x_j(t)^\top \). As a consequence, in order to prove that \( \tilde{F}^x(t) \) converges strongly, we only need to show that \( M^x_{\infty}(\ell) = M^x_\infty \) a.s. for all \( 1 \leq \ell \leq \alpha - 1 \). Note that to this aim we cannot use associativity of matrix multiplication in the infinite product \( \prod_{j=0}^\infty C^x_{i,j} \), as this does not hold in general (consider for example the infinite product of transition rate matrices of a periodic Markov chain with period \( \alpha \)). We use another technique: for convenience, define
\[
M^x_{n,h} = \prod_{j=0}^n (C^x_{i,j,k+1})^\top (C^x_{i,j,k+2})^\top \ldots (C^x_{i,j,k})^\top.
\]
If we had \( M^x_{\infty}(\ell) \neq M^x_{\infty} \) for some \( 1 \leq \ell \leq \alpha - 1 \), then \( \lim_{n \to \infty} M_{n,h+1} \) would not exist since the sequence would have at least two accumulation points \( M^x_{\infty} = \lim_{n \to \infty} M_{n,a+1} \) and \( M^x_{\infty}(\ell) = \lim_{n \to \infty} M_{n,a+\ell+1} \). However, similarly as before, \( (C^x_{i,j,a+1})^\top (C^x_{i,j,a+2})^\top \ldots (C^x_{i,j,a+\ell})^\top \) is a stochastic matrix a.s. and has a positive probability of having all entries positive (for being distributed as the transpose of \( \tilde{F}^x_j(\tau^x_{a+1}) \)). Then, by [7, Theorem 1.4],
\[
\lim_{n \to \infty} M_{n,a+1} = \lim_{n \to \infty} \prod_{j=0}^n (C^x_{i,j,a+j})^\top (C^x_{i,j,a+j+1})^\top \ldots (C^x_{i,j,a+j})^\top
\]
eexists a.s. Hence, \( \tilde{F}^x_j(t) \) converges strongly, and so does \( \tilde{F}^x(t) \). Specifically,
\[
\lim_{t \to \infty} \tilde{F}^x_1(t) = \Phi^x_\infty \quad \text{a.s.} \tag{11.4}
\]
It also follows from [7, Theorem 1.4] that in the limit the columns are identical.

We now study the convergence of \( \tilde{W}^x(t) \) and show the limit has finite expectation. By Lemma 5.2(a), all entries of the matrices \( C^x_n \) are non-negative. The same holds for the entries of the matrices \( D^x_n \). Finally, for \( 0 \leq t_1 \leq t_2 \) we have \( n^x_{t_1} \leq n^x_{t_2} \), therefore
\[
\tilde{W}^x(t_1) \leq \tilde{W}^x(t_2) \quad \text{for all} \quad 0 \leq t_1 < t_2,
\]

29
where the inequality holds component-wise. Hence, \( \lim_{t \to \infty} \hat{W}^x(t) = W^x_\infty \).

Equation (6.7) and (6.8) are thus proved. In order to prove \( E(W^K_\infty) < \infty \), define \( J \in \mathbb{Z}^{d \times d}_{\geq 0} \) as the diagonal matrix with entries

\[
J_{ij} = \begin{cases} 
1 & \text{if } i = j \text{ and } S_i \text{ is properly degraded,} \\
0 & \text{otherwise.}
\end{cases}
\]

Since \((B_X)_i = 0\) for any \(1 \leq i \leq d\) such that \(0 \neq S_i\), the corresponding columns of \(D^\xi_k, k \geq 0\), are zero. Consequently, we might write \(C^x_kJ\) in the definition of \(W^x_\infty\) rather than \(C^x_k\), as the zero columns progress through the product,

\[
W^x_\infty = D^x_0 + \sum_{k=1}^{\infty} \left( \prod_{i=k-1}^{0} C^x_{k-j}J \right) D^x_k
= D^x_0 + \sum_{k=1}^{\infty} \left( \prod_{i=k-1}^{0} C^x_{k-j} \right) \left( \sum_{j=1}^{\infty} \prod_{i=j-1}^{0} \Psi_i \right) \left( \sum_{k=1}^{\infty} \prod_{i=k-1}^{0} C^x_{k-j+i} J \right) D^x_{k+j},
\]

where

\[
\Psi_i = \prod_{k=0}^{i} (C^x_{\alpha_i+k}J) \sim \prod_{k=0}^{\infty} (C^x_{\alpha_i+k}J),
\]

and the entries of the latter are identical to those of \(\Phi(\tau^x_{\alpha_i}, \tau^x_{\alpha_i+1})\) for properly degraded species, and zero otherwise. By the monotone convergence theorem, the strong Markov property, and Lemma 5.2(c)

\[
E(\|W^x_\infty\|_1) \leq \alpha E(\|D^x_0\|_1) + \alpha \sum_{j=1}^{\infty} (E(\|\Psi_0\|_1))^j E(\|D^x_0\|_1).
\]

Finally, due to Lemma 6.2(a) we have \(E(\|\Psi_0\|_1) < 1\), and by Lemma 6.3 we have \(E(\|D^x_0\|_1) < \infty\). It follows that

\[
E(\|W^x_\infty\|_1) \leq \frac{E(\|D^x_0\|_1)}{1 - E(\|\Psi_0\|_1)} < \infty.
\]

In order to prove (6.9), note that for \(\eta > 0\) and any function \(\sigma: \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}\) with \(\lim_{t \to \infty} \sigma(t) = \infty\),

\[
\lim_{t \to \infty} P\left( \|\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t)) - (\Phi^x_\infty, W^x_\infty) \|_1 > \eta \bigg| X(t) = x \right) = 0.
\]

Indeed, by ergodicity of \(\{X(t): t \geq 0\}\) there exists \(T_\eta\) such that

\[
P(X(t) = x) > \pi(x) - \eta \quad \text{for all } t \geq T_\eta,
\]

and if \(\eta\) is smaller than \(\pi(x)\)

\[
0 < (\pi(x) - \eta) P\left( \|\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t)) - (\Phi^x_\infty, W^x_\infty) \|_1 > \eta \bigg| X(t) = x \right)
\]

\[
< P\left( \|\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t)) - (\Phi^x_\infty, W^x_\infty) \|_1 > \eta, X(t) = x \right)
\]

\[
\leq P\left( \|\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t)) - (\Phi^x_\infty, W^x_\infty) \|_1 > \eta \right).
\]

Equation (11.6) then follows from the strong convergence of \((\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t)))\), specifically from (11.4) and (11.5). Therefore, for any continuity set \(A\) of \((\Phi^x_\infty, W^x_\infty)\) we have

\[
\lim_{t \to \infty} P\left( (\Phi(x(t)), W(x(t))) \in A \bigg| X(t) = x \right) = \lim_{t \to \infty} P\left( (\hat{\Phi}^x(t), \hat{W}^x(t)) \in A \bigg| X(t) = x \right)
\]

\[
= \lim_{t \to \infty} P\left( (\Phi^x_\infty, W^x_\infty) \in A \bigg| X(t) = x \right)
\]

\[
= \lim_{t \to \infty} P\left( (\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t))) \in A \bigg| X(t) = x \right).
\]

30
where the first equality follows from (11.3), the second equality follows from (11.4) and (11.5), and the third equality follows from (11.6) for a general function \( \sigma \) diverging to \( \infty \). Note that (6.9) is proven if in the second line of (11.7) the conditioning is removed. The idea is to show that the conditioning does not play any role in the limit by using the arbitrariness of the function \( \sigma \) and the ergodicity of \( \{X(t) : t \geq 0\} \).

Fix \( 0 < \varepsilon < \pi(x) \), and let \( \Gamma_\varepsilon \subseteq \Gamma \) be a finite set of states such that

\[
\sum_{x' \notin \Gamma_\varepsilon} \pi(x') < \frac{\varepsilon}{2},
\]

where we use the convention that a sum over an empty set is zero. For any real numbers \( 0 \leq s \leq t \) define

\[
\theta_x(s, t) = \sup_{x' \in \Gamma_\varepsilon} \left| P(X(t) = x \mid X(s) = x') - P(X(t) = x) \right|.
\]

Since \( \Gamma_\varepsilon \) is a finite set, and by ergodicity of the CTMC \( \{X(t) : t \geq 0\} \) with the fixed initial condition \( X(0) = x \), we have that for all \( s \geq 0 \)

\[
\lim_{t \to \infty} \theta_x(s, t) = 0.
\]

(11.9)

A first consequence of (11.9) is that there exists \( T_\varepsilon \) such that \( \theta_x(0, t) < \varepsilon \) for all \( t \geq T_\varepsilon \). Hence, we can define the function \( \sigma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \) as

\[
\sigma(t) = \begin{cases} 0 & \text{if } t \leq T_\varepsilon, \\ \sup\{0 \leq s \leq t : \theta_x(s, t) < \varepsilon\} & \text{if } t > T_\varepsilon. \end{cases}
\]

It also follows from (11.9) that \( \sigma(t) \) goes to infinity as \( t \) goes to infinity. Hence, it follows from the ergodicity of \( \{X(t) : t \geq 0\} \) that by potentially increasing \( T_\varepsilon \) we might further assume

\[
\sum_{x' \in \Gamma_\varepsilon} \left| P(X(\sigma(t)) = x') - \pi(x') \right| < \frac{\varepsilon}{2} \text{ for all } t > T_\varepsilon,
\]

(11.10)

\[
\sum_{x' \in \Gamma_\varepsilon} \left| P(X(t) = x') - \pi(x') \right| < \varepsilon \text{ for all } t > T_\varepsilon.
\]

(11.11)

In particular, it follows from (11.8) and (11.10) that

\[
0 < \sum_{x' \notin \Gamma_\varepsilon} P(X(\sigma(t)) = x') < \varepsilon \text{ for all } t > T_\varepsilon.
\]

(11.12)

For all \( t > T_\varepsilon \), we have

\[
P\left( \left( \hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t)) \right) \in A \mid X(t) = x \right)
\geq \sum_{x' \in \Gamma_\varepsilon} P\left( \left( \hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t)) \right) \in A \mid X(\sigma(t)) = x', X(t) = x \right) P(X(\sigma(t)) = x') P(X(t) = x)
\]

\[
= \sum_{x' \in \Gamma_\varepsilon} P\left( \left( \hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t)) \right) \in A \mid X(\sigma(t)) = x' \right) P(X(t) = x \mid X(\sigma(t)) = x') P(X(\sigma(t)) = x') \frac{P(X(\sigma(t)) = x')}{P(X(t) = x)}
\]

\[
\geq \frac{\pi(x) - 2\varepsilon}{\pi(x) + \varepsilon} \sum_{x' \in \Gamma_\varepsilon} P\left( \left( \hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t)) \right) \in A \mid X(\sigma(t)) = x' \right) P(X(\sigma(t)) = x')
\]

\[
\geq \frac{\pi(x) - 2\varepsilon}{\pi(x) + \varepsilon} \left( P\left( \left( \hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t)) \right) \in A \right) - \varepsilon \right),
\]

(11.13)

where in the second line we conditioned on the possible values of \( X(\sigma(t)) \), in the third line we utilise the fact that \( \sigma(t) \leq t \) by definition, the Markov property of \( \{X(t) : t \geq 0\} \), and Bayes’ formula. In the forth line
we use the definition of \(\sigma(t)\) and (11.11), and in the last line we use (11.12). Similarly, we have

\[
\begin{align*}
P\left(\left(\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t))\right) \in A \mid X(t) = x\right) \\
\leq \sum_{x' \notin \mathcal{D}_x} P\left(\left(\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t))\right) \in A \mid X(\sigma(t)) = x', X(t) = x\right) P(X(\sigma(t)) = x') P(X(t) = x) \\
+ \sum_{x' \notin \mathcal{D}_x} P(X(\sigma(t)) = x' \mid X(t) = x) \\
= \sum_{x' \in \mathcal{D}_x} P\left(\left(\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t))\right) \in A \mid X(\sigma(t)) = x'\right) P(X(\sigma(t)) = x') P(X(t) = x) \\
+ \sum_{x' \notin \mathcal{D}_x} P(X(t) = x \mid X(\sigma(t)) = x') P(X(\sigma(t)) = x') \\
\leq \frac{\pi(x) + 2\varepsilon}{\pi(x) - \varepsilon} \left(\varepsilon + \sum_{x' \notin \mathcal{D}_x} P\left(\left(\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t))\right) \in A \mid X(\sigma(t)) = x'\right) P(X(\sigma(t)) = x') \right) \\
\leq \frac{\pi(x) + 2\varepsilon}{\pi(x) - \varepsilon} \left( P\left(\left(\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t))\right) \in A \right) + \varepsilon \right) \, ,
\end{align*}
\]

where for the first inequality we condition on the possible values of \(X(\sigma(t))\), for the equality afterwards we utilise the fact that \(\sigma(t) \leq t\), the Markov property of \(\{X(t) : t \geq 0\}\) and Bayes’ formula. For the consecutive inequality we use the definition of \(\sigma(t)\), (11.11) and (11.12), and in the last line we use the law of total probability. In conclusion, it follows from (11.13) and (11.14) and by the arbitrariness of \(0 < \varepsilon < \pi(x)\) that

\[
\lim_{t \to \infty} P\left(\left(\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t))\right) \in A \mid X(t) = x\right) = \lim_{t \to \infty} P\left(\left(\hat{\Phi}^x(\sigma(t)), \hat{W}^x(\sigma(t))\right) \in A \right)
\]

which in turn implies (6.9) by (11.4), (11.5) and (11.7). The proof is then completed. \(\square\)

### 11.2 Proof of Lemma 6.5

Since \((C, D)\) is distributed as any of the i.i.d. random variables \((C_k^x, D_k^x)\) and is independent of them, for any \(n\) we have

\[
\begin{align*}
(C, D, (C_k^x)_{k=0,\ldots,n}, (D_k^x)_{k=0,\ldots,n}) & \sim (C, D, (C_k^x)_{k=1,\ldots,n+1}, (D_k^x)_{k=1,\ldots,n+1}) \\
& \sim (C_0^x, D_0^x, (C_k^x)_{k=1,\ldots,n+1}, (D_k^x)_{k=1,\ldots,n+1})
\end{align*}
\]

Hence, it follows from the definition of \(\Phi_\infty\) and \(W_\infty\) in (6.7) that

\[
\begin{align*}
& \left( C\Phi_\infty^x \right) \left( D + CW_\infty^x \right) = \left( D + CD_\infty^x + C \lim_{n \to \infty} \prod_{k=n}^{0} C_k^x \right) \left( \prod_{k=1}^{n+1} C_k^x \right) D_k^x \\
& \sim \left( D_0^x + \lim_{n \to \infty} \prod_{k=n+1}^{0} C_k^x \right) \left( \prod_{k=1}^{n+1} C_k^x \right) D_k^x = \left( \Phi_\infty^x \right) \left( W_\infty^x \right)
\end{align*}
\]

Using [14][Theorem 2.1] the existence of the solution to \(V_2 \sim CV_2 + D\) can be deduced under the conditions

\[
\begin{align*}
\prod_{i=n}^{0} C_i \overset{a.s.}{\to} 0, \quad \text{and} \quad \prod_{i=n}^{0} C_i \overset{a.s.}{\to} 0 \quad \text{as} \quad n \to \infty,
\end{align*}
\]

where \((C_n, D_n)_{n \geq 0}\) are i.i.d. distributed as \((C, D) \sim (C_0^x, D_0^x)\). Weak uniqueness can be also derived from (11.15) as follows: assume \(V \sim CV + D\) and \(\tilde{V} \sim CV + \tilde{D}\) with \(V \sim \mu\) and \(\tilde{V} \sim \tilde{\mu}\), then \(V^{(n)} \sim \mu\) and
\( \widetilde{V}(n) \sim \mu \) for any natural \( n \), where

\[
V^{(n)} = \begin{cases} 
V & \text{if } n = 0 \\
C_{n-1}V^{(n-1)} + D_{n-1} & \text{otherwise}
\end{cases}
\quad \text{and} \quad
\widetilde{V}^{(n)} = \begin{cases} 
\widetilde{V} & \text{if } n = 0 \\
C_{n-1}\widetilde{V}^{(n-1)} + D_{n-1} & \text{otherwise}.
\end{cases}
\]

Since for any \( n \geq 1 \) we have \( V^{(n)} - \widetilde{V}^{(n)} = (\prod_{i=n-1}^{0} C_i)(V - \widetilde{V}) \) and the latter tends to 0 by (11.15), then necessarily \( \mu = \widetilde{\mu} \). To conclude the proof, it then suffices to prove (11.15). By (6.4) and Lemma 6.2(a), if all species are properly degraded then there exists \( \alpha \geq 1 \) such that \( E(\|\widetilde{C}_i\|_1) < 1 \), where \( \widetilde{C}_i = C_{i\alpha + \alpha - 1}C_{i\alpha + \alpha - 2} \cdots C_{i\alpha} \sim \Phi(\tau^*_{\alpha_0}, \tau^*_{\alpha}), \ i \geq 0 \). Note that for \( 1 \leq k \leq \alpha \) and \( n \geq 0 \),

\[
M_{k+n\alpha} = \left\| \left( \prod_{i=k-1+n\alpha}^{0} C_i \right) D_{k+n\alpha} \right\| \leq \left( \prod_{i=n-1}^{0} \|\widetilde{C}_i\|_1 \right) \|D_{k+n\alpha}\|_1.
\]

Applying Jensen’s inequality, \( E(\log(\|\widetilde{C}_i\|_1)) \leq \log E(\|\widetilde{C}_i\|_1) < 0 \). Using (6.4) and Lemma 6.3, we similarly find \( E(\log(\|D_n\|_1)) \leq \log E(\|D_n\|_1) < \infty \). Hence as a consequence of the strong law of large numbers,

\[
\limsup_{k+n\alpha \to \infty} \frac{1}{1+n\alpha} \log(M_{k+n\alpha}) \leq \limsup_{n \to \infty} \left( \frac{1}{n} \sum_{i=0}^{n-1} \log(\|\widetilde{C}_i\|_1) \right) + \limsup_{n \to \infty} \left( \frac{1}{n} \log(\|D_n\|_1) \right) < 0 \quad \text{a.s.,}
\]

proving the second assertion of (11.15) (the second term converges to 0 if \( E(\log(\|D_n\|_1)) \) is uniformly bounded from below and is otherwise \( \leq 0 \)). The first assertion of (11.15) follows similarly.

\[ \square \]

### 11.3 Proof of Lemma 6.6

We state a preliminary lemma first.

**Lemma 11.1.** Consider a stochastic process \( \{L(t) : t \geq 0\} \) on \( \mathbb{R}_{\geq 0} \), and let \( m \) be a positive real number. Let \( \{J(t) : t \geq 0\} \) be another stochastic process, such that \( J(t) \sim \text{Pois}(mL(t)) \) for any \( t \geq 0 \). Then, \( \{L(t) : t \geq 0\} \) is tight if and only if \( \{J(t) : t \geq 0\} \) is tight.

**Proof.** Assume that \( \{L(t) : t \geq 0\} \) is tight. Then, by definition, for all \( \varepsilon > 0 \) there exists \( M_{\varepsilon} \) such that

\[
\sup_{t \geq 0} P(L(t) > M_{\varepsilon}) \leq \varepsilon.
\]

For \( 0 < \varepsilon \leq 1 \) define

\[
\widetilde{M}_{\varepsilon} = 2m \frac{\varepsilon}{\varepsilon - \frac{\varepsilon}{2}} M_{\frac{\varepsilon}{2}}.
\]

We have

\[
\sup_{t \geq 0} P\left( J(t) > \widetilde{M}_{\varepsilon} \right) \leq \left( 1 - \frac{\varepsilon}{2} \right) P\left( \text{Pois}(mM_{\frac{\varepsilon}{2}}) > \widetilde{M}_{\varepsilon} \right) + \frac{\varepsilon}{2}
\]

\[
\leq \left( 1 - \frac{\varepsilon}{2} \right) \frac{mM_{\frac{\varepsilon}{2}}}{M_{\varepsilon}} + \frac{\varepsilon}{2} = \varepsilon,
\]

where the second inequality follows from Markov’s inequality. Hence, the process \( \{J(t) : t \geq 0\} \) is tight.

For the other direction, assume that \( \{L(t) : t \geq 0\} \) is not tight. Then, there exists \( \varepsilon > 0 \) such that for all \( M > 0 \) there exists \( t \geq 0 \) with \( P(L(t) > M) > \varepsilon \). Then, for all \( M > \frac{4}{m} \) there exists \( t \geq 0 \) with

\[
P\left( J(t) > mM - 2\sqrt{mM} \right) > \varepsilon P\left( \text{Pois}(mM) > mM - 2\sqrt{mM} \right)
\]

\[
> \varepsilon P\left( |\text{Pois}(mM) - mM| \leq 2\sqrt{mM} \right) \geq \frac{\varepsilon}{4},
\]

where we used Chebychev’s inequality in the last step. Therefore, the process \( \{J(t) : t \geq 0\} \) is neither tight, and the proof is concluded. \[ \square \]
Lemma 11.2. The process \( \{W(t); t \geq 0\} \) is tight if and only if the process \( \{Z(t); t \geq 0\} \) is tight.

Proof. First, it follows from the definition of tightness that \( \{W(t); t \geq 0\} \) is tight if and only if \( \{\|W(t)\|_1; t \geq 0\} \) is tight. By Lemma 5.2(a), we further have

\[
\|W(t)\|_1 = \left\| \int_0^t \Phi(u,t)B_X(u) \, du \right\|_1 = \int_0^t e^\top \Phi(u,t)B_X(u) \, du.
\]

Assume that \( \{W(t); t \geq 0\} \) is tight. Since \( Z(t) \in \mathbb{Z}_d^d \geq 0 \), we have \( \|Z(t)\|_1 = \sum_{j=1}^d Z_j(t) = e^\top Z(t) \). Using Proposition 5.3, the random variable \( \|Z(t)\|_1 \), conditioned on \( \mathcal{F}_t^X \), is stochastically bounded by

\[
\|Z(t)\|_1 \leq \|Z(0)\|_1 + m \sum_{j=1}^d \sum_{\nu \in \Theta_{m_j,d}} N_{\nu j}(t),
\]

where \( m = \max(m_j; j = 1, \ldots, d) \) and \( N_{\nu j} \) are independent random variables with

\[
N_{\nu j}(t) \sim \text{Pois} \left( \int_0^t \lambda_{0j}(X(u)) g_{\nu j}^X(\nu, m_j) \, du \right).
\]

By definition of configuration and of \( m \), we have \( \|\nu\|_1 \leq m \) for all \( \nu \in \bigcup_{j=1}^d \Theta_{m_j,d} \). Using the definition of \( g_{\nu j}^X(\nu, m_j) \) and independence (conditioned on \( \mathcal{F}_t^X \)) of the random variables \( N_{\nu j}(t) \) for \( \nu \in \Theta_{m_j,d}, 1 \leq j \leq d \), equation (11.16) reduces to

\[
\|Z(t)\|_1 \leq \|Z(0)\|_1 + m \text{Pois} \left( \sum_{j=1}^d \int_0^t \lambda_{0j}(X(u)) \left[ 1 - g_{\nu j}^X((0, \ldots, 0), m_j) \right] du \right)
\]

\[
\sim \|Z(0)\|_1 + m \text{Pois} \left( \sum_{j=1}^d \int_0^t \lambda_{0j}(X(u)) \left[ 1 - (1 - e^\top \Phi(u,t)e_j)^{m_j} \right] du \right)
\]

\[
\leq \|Z(0)\|_1 + m \text{Pois} \left( m \sum_{j=1}^d \int_0^t \lambda_{0j}(X(u)) e^\top \Phi(u,t)e_j du \right)
\]

\[
\sim \|Z(0)\|_1 + m \text{Pois} \left( m \int_0^t e^\top \Phi(u,t)B_X(u) \, du \right),
\]

by recalling the definition (4.3) of \( B_X(u) \). Furthermore, it is used that \( 1 - ka \leq (1 - a)^k \) for \( 0 \leq a \leq 1 \) and \( k \geq 1 \). Using Lemma 11.1 and the stochastic upper bound, tightness of \( \{Z(t); t \geq 0\} \) follows.

Conversely, assume that \( \{Z(t); t \geq 0\} \) is tight with the given \( m \). Then, the molecular count process \( \{Z'(t); t \geq 0\} \) corresponding to the process with \( m' = (\min(1, m_1), \ldots, \min(1, m_d)) \) in (4.1) is also tight. Indeed, \( \|Z'(t)\|_1 \) is stochastically bounded from above by \( \|Z(t)\|_1 \) for all \( t \geq 0 \), because

- the lifetime distribution of any molecule created in any of the two settings is the same as it does not depend on the choice of \( m \),
- every molecule that is created in any of the two settings evolves independently of the other molecules being present, given \( \{\mathcal{F}_t^X\}_{t \geq 0} \) (see Remark 2.1),
- in the process \( \{Z'(t); t \geq 0\} \) less molecules are created than in the process \( \{Z(t); t \geq 0\} \) (the two processes can be coupled such that the reactions \( 0 \rightarrow m_j S_j \) and \( 0 \rightarrow m'_j S_j \) occur at the same time).

The choice of \( m \) does not affect the quantities \( A_X(t) \) and \( B_X(t) \), hence the process \( \{W(t); t \geq 0\} \) is not affected either. It follows from Proposition 5.3 that

\[
\text{Pois} \left( \int_0^t e^\top \Phi(u,t)B_X(u) \, du \right) \sim e^\top \sum_{j=1}^d \sum_{\nu \in \Theta_{m'_j,d}} \nu N_{\nu j}(t)
\]

\[
\leq e^\top Z'(t) = \|Z'(t)\|_1.
\]
Finally, it follows from the tightness of \( \{Z'(t): t \geq 0\} \) and Lemma 11.1 that \( \{W(t): t \geq 0\} \) is tight, which concludes the proof.

We are now ready to prove Lemma 6.6.

**Proof of Lemma 6.6.** Assume that \( \{(X(t), Z(t)): t \geq 0\} \) is ergodic for every initial condition. Then \( \{Z(t): t \geq 0\} \) is tight, which by Lemma 11.2 implies that \( \{W(t): t \geq 0\} \) is tight.

Now assume that \( \{W(t): t \geq 0\} \) is tight. Then, by Lemma 11.2 the process \( \{(X(t), Z(t)): t \geq 0\} \) is tight as well. It follows that \( \{(X(t), Z(t)): t \geq 0\} \) is tight because \( \{X(t): t \geq 0\} \) is ergodic by assumption. Hence, by Prokhorov’s theorem, \( \{(X(t), Z(t)): t \geq 0\} \) is sequentially compact [5]. Since \( \{(X(t), Z(t)): t \geq 0\} \) is a CTMC, the accumulation point of \( \{(X(t), Z(t)): t \geq 0\} \) is unique, which implies that \( \{(X(t), Z(t)): t \geq 0\} \) is ergodic for every initial condition.

### 11.4 Additional results

**Lemma 11.3.** Suppose \( \Phi(t) \) is a fundamental matrix solution to \( x' = A(t)x \). The solution of \( \lambda'(t) = A(t)\lambda(t) + b(t) \) with initial condition \( \lambda(u) = \lambda_u\), is

\[
\lambda(t) = \Phi(t)\Phi^{-1}(u)\lambda_u + \int_u^t \Phi(t)\Phi^{-1}(s)b(s)ds, \quad t > u \geq 0.
\]

**Proof.** Using that \( \Phi'(t) = A(t)\Phi(t) \) by assumption, we find the derivative of (11.17) to be

\[
\lambda'(t) = \Phi(t)'\Phi^{-1}(u)\lambda_u + \Phi(t)\Phi^{-1}(t)b(t) + \int_u^t \Phi(t)'\Phi^{-1}(s)b(s)ds
\]

\[
= A(t)\Phi(t)\Phi^{-1}(u)\lambda_u + b(t) + \int_u^t A(t)\Phi(t)\Phi^{-1}(s)b(s)ds = A(t)\lambda(t) + b(t).
\]

**References**


