POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Matematica

Tesi di Laurea Magistrale

Stochastic models of Chemical Reaction Networks: multiscale approximations and convergence

Relatore:
Prof.ssa Paola Siri

Candidato:
Claudio Del Sole

Correlatori:
Prof. Enrico Bibbona
Prof. Franco Fagnola

Anno Accademico 2019-2020

Contents

Introduction

Chemical reaction networks are mathematical models widely used to describe the dynamical behaviour of systems in biology, epidemiology, chemistry. In such models, individual units (e.g. molecules), belonging to different groups (e.g. chemical species), interact with each other according to specific laws, named reactions, which may be represented in a graph. Reactions are usually modeled as stochastic counting processes, so that the number of molecules in the system for each species is described by a continuous time Markov chain.

When the number of molecules is very high, species dynamics may be suitably described in terms of chemical concentrations, and the stochastic model is well approximated by a deterministic, continuous dynamical system: this approach is known in literature as classical scaling, and dates back to 1970s [14].

More recently, the study of chemical reaction systems in cellular biology has renewed the interest in stochastic models. Indeed, these systems may involve species with vastly different number of molecules or reactions whose propensities vary over several orders of magnitude, so that the continuous approximation does not provide a satisfactory description. In order to properly characterize such complex phenomena, rescaled versions of the original models may be analysed on different time-scales, and a comprehensive theory of convergence for chemical reaction networks, extending the classical one, is widely developed in literature [1] [2] [13].

In some specific cases, the intrinsic discreteness of the model cannot be disregarded. For instance, in signal transduction processes, a single molecule may trigger a biochemical cascade which causes a transition in the cell state. An example of a chemical reaction system displaying such peculiar behaviour was suggested by Togashi and Kaneko [17]: in this model, biochemical cascades are driven by fast autocatalytic reactions, while inflows and outflows of single species happen at much slower rates.

For this class of systems, the above-mentioned theory of convergence cannot always be applied, since many of the underlying assumptions fail. The main criticalities arise when the time-scale is accelerated so that trajectories of the stochastic process describing species concentrations display sharp peaks or rapid switches to different stable states, both induced by the fast autocatalytic reactions and corresponding to failed or completed transitions, respectively. In particular, such peculiar features prevent the model to converge to any limit on the Skorohod space of cadlag functions equipped with the classical Skorohod topology [16].

An alternative sequential topology on the Skorohod space was proposed by Jakubowski in order to weaken the Skorohod topology while preserving its useful

properties [9] [11]. Although being non-metrizable, many of the fundamental results of the classical theory of weak convergence on metric spaces hold true, provided a stronger notion of convergence for probability measures is adopted. In this topology, sharp peaks corresponding to failed transitions coalesce and cancel out: therefore, it seems a promising candidate with respect to which convergence of the Togashi-Kaneko model may be verified.

In the first chapter of this thesis, the classical theory of chemical reaction networks is analysed in detail, as well as its most recent developments in the direction of a multiscale approximation approach. In particular, the main results concerning convergence on different time-scales are stated and discussed.

The second chapter is devoted to the theory of weak convergence in metric and non-metric spaces [4], with focus on the Skorohod space of cadlag functions. A new topology on this space, devised by Jakubowski, is presented, together with a stronger notion of convergence in distribution, based on the Skorohod Representation Theorem, which extends the usual definition to non-metric spaces [10].

In the last chapter, a simplified version of the Togashi-Kaneko example is studied in the framework of multiscale approximation theory, and convergence to non-degenerate limit models is verified, whenever possible [3]. Moreover, some useful properties of this model are formally derived: in particular, it is shown that, under suitable conditions, autocatalytic cascades may be separated from inflow and outflow reactions, in the sense that no inflows and outflows happen during an autocatalytic cascade.

As for convergence at faster time-scales, it is verified that this version of the Togashi-Kaneko model does not converge to a naturally arising limiting model with respect to the Skorohod topology; specifically, the corresponding sequence of probability measures on the Skorohod space is not relatively compact. On the contrary, uniform tightness of such sequence is proved to hold in the weaker Jakubowski topology, implying in turn convergence for subsequences. In order to obtain convergence of the entire sequence to the above-mentioned candidate limit, convergence of the finite-dimensional distribution has still to be proved.

Numerical simulations have been carried out using MATLAB®; the code developed for this purpose can be found in the annex to this thesis.

Chapter 1

Chemical Reaction Networks

1.1 Mathematical modeling

The mathematical formulation of a chemical reaction system includes two parts: a *reaction network* and a choice of *dynamics* [1, ch. 2].

The reaction network is the static component of the model and consists of three sets:

- $\mathcal{S}$, the set of *species*, the chemical components whose number or concentration we are interested in modeling dynamically;
- $\mathcal{C}$, the set of *complexes*, which are linear and non-negative combinations of species that describe how species can interact;
- $\mathcal{R}$, the set of *reactions*, which are ordered pairs of complexes describing how to convert a complex to another.

In this work, elements of these three sets are referred to consistently with the following notations.

- Species $S_1, \ldots, S_n$ are represented as capital letters.
- Complexes $C_1, \ldots, C_m$ are more conveniently identified with non-negative column vectors, whose i-th entry is the number of components of species i constituting the complex:

$$C_j = \sum_i (z_j)_i S_i = z_j \cdot S, \qquad z_j \in \mathbb{N}_0^n. \tag{1.1}$$

  Note that the *null complex*, i.e. the complex corresponding to a null vector, is an admissible complex, usually represented as $\emptyset$.

- Reactions $R_1, \ldots, R_l$ are represented as ordered pairs of non-negative column vectors (i.e. complexes),

$$R_k = (\nu_k, \nu_k'), \qquad \nu_k, \nu_k' \in \mathcal{C}, \tag{1.2}$$

and usually depicted as two complexes connected by an arrow

$$\sum_i (\nu_k)_i S_i \to \sum_i (\nu'_k)_i S_i, \qquad \nu_k, \nu'_k \in \mathcal{C},$$

or simply (with a slight abuse of notation)

$$\nu_k \to \nu'_k, \qquad \nu_k, \nu'_k \in \mathcal{C}.$$

The first element of the pair (complex ν_k) is sometimes referred to as the *substrate* of the reaction, while the second element (complex ν'_k) is called the *product*.

In case a reaction can occur both ways (i.e. the reaction is reversible), a double arrow is used:

$$\sum_i (\nu_k)_i S_i \rightleftharpoons \sum_i (\nu'_k)_i S_i, \qquad \nu_k, \nu'_k \in \mathcal{C}. \tag{1.3}$$

Consistently with the definition given above of a reaction as an ordered pair, this writing actually implies two different reactions, usually identified as the *forward* reaction $\nu_k \to \nu'_k$ and the *reverse* reaction $\nu'_k \to \nu_k$.

Definition 1.1. A chemical reaction network (CRN) is a triple $\{ \mathcal{S}, \mathcal{C}, \mathcal{R} \}$, where

- $\mathcal{S} = \{ S_i \}_{i=1}^n$ is the set of species,
- $\mathcal{C} = \{ C_j \}_{j=1}^m = \{ z_j \in \mathbb{N}_0^n \}_{j=1}^m$ is the set of complexes,
- $\mathcal{R} = \{ R_k \}_{k=1}^l = \{ \nu_k \to \nu'_k \colon \nu_k, \nu'_k \in \mathcal{C} \}_{k=1}^l$ is the set of reactions.

Chemical reaction networks have a natural graph representation, in which complexes indentify with the vertices and reactions describe the edges: the resulting directed graph is referred to as the *complex graph*. When introducing a CRN, it is common to forgo an explicit description of the triple $\{ \mathcal{S}, \mathcal{C}, \mathcal{R} \}$ and simply illustrate the complex graph implied by the reaction network.

Example 1.2 (Enzyme catalysis)**.** In biochemistry, one of the best-known models of enzyme kinetics is due to Michaelis and Menten (1913), describing the catalytic action of an enzyme. In particular, an enzyme E binds to a substrate S, forming an enzyme-substrate complex ES, and catalyses the convertion of the substrate into a product P, after which the product and the enzyme are released. Moreover, production and degradation of the enzyme occur.

This model may be represented schematically via its complex graph:

$$S + E \rightleftharpoons ES \to P + E, \qquad E \rightleftharpoons \emptyset. \tag{1.4}$$

This representation formally corresponds to the reaction network described by the sets

$$\mathcal{S} = \{ S, E, ES, P \}, \qquad \mathcal{C} = \{ S + E,\ ES,\ P + E,\ E,\ \emptyset \},$$
$$\mathcal{R} = \{ S + E \rightleftharpoons ES,\ ES \to P + E,\ E \rightleftharpoons \emptyset \}.$$

Note that the set $\mathcal{R}$ of reactions is here written in a compact form, as it actually contains five elements, since reversible reactions have to be counted twice. Moreover, this example illustrates the role of the null complex $\emptyset$: it is used to model inflows and outflows of species.

The *state* of a chemical reaction system is described by a column vector $X \in \mathbb{N}_0^n$, whose i-th entry gives the number of molecules of species S_i present in the system. Each time a reaction occurs, the state of the system is updated accordingly: in particular, whenever reaction

$$R_k : \nu_k \to \nu_k'$$

occurs, species contained in substrate complex ν_k are converted into species contained in product complex ν_k', and therefore the state X is updated summing the column vector

$$\zeta_k := \nu_k' - \nu_k \ \in \mathbb{Z}^n, \tag{1.5}$$

which is referred to as the *reaction vector*.

Definition 1.3. The *stoichiometric matrix* of a chemical reaction network $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ is the matrix $S \in \mathbb{Z}^{n\times l}$ containing reaction vectors $\{\zeta_k\}_{k=1}^{l}$ as columns:

$$S = \begin{bmatrix} \zeta_1 & \dots & \zeta_l \end{bmatrix}. \tag{1.6}$$

Other two matrices are useful to describe a chemical reaction network [18]:

- the *complex stoichiometric matrix* $Z \in \mathbb{Z}^{n\times m}$ is the matrix containing vectors $\{z_j\}_{j=1}^{m}$, defined in (1.1), as columns:

$$Z = \begin{bmatrix} z_1 & \dots & z_m \end{bmatrix}. \tag{1.7}$$

- the matrix $B \in \mathbb{Z}^{m\times l}$ is the incidence matrix of the complex graph, i.e. the matrix whose entries are given by

$$B_{jk} = \begin{cases} -1 & \text{if } C_j = \nu_k \\ +1 & \text{if } C_j = \nu_k' \\ 0 & \text{otherwise} \end{cases} \tag{1.8}$$

The fundamental relation connecting these matrices is expressed as

$$S = ZB. \tag{1.9}$$

Example 1.4 (Enzyme catalysis)**.** Recall that the Michealis-Menten model for enzyme kinetics, described in Example 1.2, is represented by the complex graph

$$S + E \rightleftharpoons ES \to P + E, \qquad E \rightleftharpoons \emptyset.$$

Therefore, adopting the labeling order introduced in the previous Example, the corresponding stoichiometric matrix S is

$$S = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \in \mathbb{Z}^{4\times 5}.$$

Moreover, complex stoichiometric matrix Z and incidence matrix B are given by

$$Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \in \mathbb{N}^{4\times 5}, \qquad B = \begin{bmatrix} -1 & +1 & 0 & 0 & 0 \\ +1 & -1 & -1 & 0 & 0 \\ 0 & 0 & +1 & 0 & 0 \\ 0 & 0 & 0 & -1 & +1 \\ 0 & 0 & 0 & +1 & -1 \end{bmatrix} \in \mathbb{N}^{5\times 5}.$$

One can easily check that the relation $S = ZB$ holds true.

Remark 1.5. Matrices Z (complex stoichiometric matrix) and B (incidence matrix) completely determine a CRN, except for species and reactions relabeling. Indeed, the incidence matrix fully characterizes the complex graph, while the complex stoichiometric matrix captures the composition of complexes (i.e. the vertices of the complex graph) in terms of chemical species.

On the other hand, the stoichiometric matrix S alone is not enough to describe a CRN: for example, the following networks

$$(1) \quad A + B \rightleftharpoons 2B, \qquad\qquad (2) \quad A \rightleftharpoons B,$$

are both characterized by the same stoichiometric matrix

$$S = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \in \mathbb{Z}^{2\times 2},$$

but determine different chemical reaction systems. Indeed, the stoichiometric matrix only captures *net* changes in the system state caused by each reaction (as described by reaction vectors), and therefore does not consider those species present in both the substrate and the product of a reaction, which are, in a sense, "crossed out".

Chemical reactions such as $A + B \rightleftharpoons 2B$ are known in chemistry as *autocatalytic* reactions: chemical species B is both a substrate and a product for the reaction, and acts as a catalyst. The effect of such autocatalytic reaction on the state of the system is the same of the simpler reaction $A \rightleftharpoons B$. However, as is known from chemistry, the presence of a catalyst considerably speeds up the reaction, since it lowers the activation energy. Similarly, in the mathematical model, catalysts affect the definition of reaction rates: catalyzed reactions happen at much faster rates.

In conclusion, stioichiometric matrix S does not completely determine a CRN because it is "blind" to autocatalytic processes (and to catalysts in general).

Having defined a notion of reaction network as the static component of a chemical reaction system, one turns to the problem of modeling its dynamical behaviour. In particular, the *state* X of the system (a column vector representing the counts of the different species in $\mathcal{S}$) is now regarded as a function of time:

$$X = \{\, X(t) \in \mathbb{N}_0^n,\ t \geq 0 \,\}\,. \tag{1.10}$$

Consider a column vector $Y = \{\, Y(t) \,\}_{t\geq 0}$ whose k-th entry counts the number of times reaction R_k has occurred up to time $t \geq 0$

$$Y = \{\, Y(t) \in \mathbb{N}_0^l,\ t \geq 0 \,\}\,; \tag{1.11}$$

then, the state $X(t)$ of the system at time $t \geq 0$ is clearly given by

$$X(t) = X(0) + \sum_k \zeta_k Y_k(t) = X(0) + S\,Y(t), \qquad t \geq 0, \tag{1.12}$$

where $X(0)$ represents the initial state.

A choice of *dynamics* for a chemical reaction system consists in specifying the form of the counting vector Y.

In a stochastic setting, $Y = \{ Y(t) \}_{t \geq 0}$ is modeled as a multivariate counting process with intensity functions denoted by

$$\lambda_k \colon \mathbb{N}_0^n \mapsto \mathbb{R}_{\geq 0}, \qquad k = 1, \ldots, l;$$

the number of times reaction R_k has occurred up to time $t \geq 0$ is thus given by

$$Y_k(t) = \xi_k \left(\int_0^t \lambda_k(X(s)) ds \right), \qquad t \geq 0, \tag{1.13}$$

where ξ_k, $k = 1, \ldots, l$, are independent unit Poisson processes (for further details, see [1, ch. 1]).

With this definition of the counting vector, the stochastic process X describing the state of the system can be modeled as a continuous time Markov chain with state space $\mathbb{S} = \mathbb{N}_0^n$:

$$X(t) = X(0) + \sum_k \xi_k \left(\int_0^t \lambda_k(X(s)) ds \right) \zeta_k, \qquad t \geq 0. \tag{1.14}$$

Note that intensity functions do not depend on time s explicitely, but only through the state $X(s)$: therefore, X is a time-homogeneous Markov chain.

A specific form for the intensity functions is called a *kinetics*. A minimal requirement for a kinetics is to be *stoichiometrically admissible*, that is

$$\lambda_k(x) = 0 \qquad \text{if} \quad x < \nu_k, \qquad k = 1, \ldots, l, \tag{1.15}$$

where $x < \nu_k$ is to be understood component-wise. Stoichiometric admissibility ensures that a reaction cannot occur if molecules present in the system are not sufficient to produce the substrate, and guarantees that the process remains in $\mathbb{S} = \mathbb{N}_0^n$ for all time.

Definition 1.6. *Mass action kinetics* is a kinetics which follows the law of mass action, i.e. a kinetics whose intesity function (also called *reaction rate*) for reaction R_k is

$$\lambda_k(x) = \kappa_k \prod_{i=1}^{n} \frac{x_i!}{(x_i - (\nu_k)_i)!} = \kappa_k \frac{x!}{(x - \nu_k)!}, \qquad k = 1, \ldots, l, \tag{1.16}$$

where constant κ_k is called *reaction rate constant*.

Mass action kinetics is stoichiometrically admissible and is the most common choice of kinetics for chemical reaction systems. Note that the rate is proportional to the number of distinct groups of molecules present in the system which can form the substrate of the reaction: this choice of kinetics reflects the assumption that the system is well-stirred.

Example 1.7 (Enzyme catalysis)**.** Consider once again the Michaelis-Menten model for enzyme kinetics, described in Example 1.2:

$$S + E \rightleftharpoons ES \to P + E, \qquad E \rightleftharpoons \emptyset.$$

Adopting the labeling introduced there, reaction rates can be expressed as

$$\lambda_1(x) = \kappa_1 x_S x_E, \qquad \lambda_2(x) = \kappa_2 x_{ES}, \qquad \lambda_3(x) = \kappa_3 x_{ES}$$
$$\lambda_4(x) = \kappa_4 x_E, \qquad \lambda_5(x) = \kappa_5.$$

Therefore, the state of the system at time $t \geq 0$ is described by the following equations:

$$X_S(t) = X_S(0) - \xi_1 \left(\int_0^t \kappa_1 X_S(s) X_E(s) ds \right) + \xi_2 \left(\int_0^t \kappa_2 X_{ES}(s) ds \right),$$

$$\begin{aligned} X_E(t) = X_E(0) &- \xi_1 \left(\int_0^t \kappa_1 X_S(s) X_E(s) ds \right) + \xi_2 \left(\int_0^t \kappa_2 X_{ES}(s) ds \right) \\ &+ \xi_3 \left(\int_0^t \kappa_3 X_{ES}(s) ds \right) - \xi_4 \left(\int_0^t \kappa_4 X_E(s) ds \right) + \xi_5 \left(\int_0^t \kappa_5 ds \right), \end{aligned}$$

$$\begin{aligned} X_{ES}(t) = X_{ES}(0) &+ \xi_1 \left(\int_0^t \kappa_1 X_S(s) X_E(s) ds \right) - \xi_2 \left(\int_0^t \kappa_2 X_{ES}(s) ds \right) \\ &- \xi_3 \left(\int_0^t \kappa_3 X_{ES}(s) ds \right), \end{aligned}$$

$$X_P(t) = X_P(0) + \xi_3 \left(\int_0^t \kappa_3 X_{ES}(s) ds \right).$$

1.2 Linear Algebra for CRNs

Matrices S (stoichiometric matrix), Z (complex stoichiometric matrix), and B (incidence matrix), defined above, capture geometrical properties of CRNs, which are *static* properties, i.e. indepedendent from the choice of dynamics.

Definition 1.8. A chemical reaction network $\{ \mathcal{S}, \mathcal{C}, \mathcal{R} \}$ is called *reversible* if each reaction is reversible, i.e.

$$\nu_k \to \nu'_k \in \mathcal{R} \quad \Longrightarrow \quad \nu'_k \to \nu_k \in \mathcal{R}, \tag{1.17}$$

or, more formally,

$$\forall k \ \exists h \quad \text{s.t.} \quad \nu_k = \nu'_h, \ \nu'_k = \nu_h.$$

For reversible networks, reactions can be represented in a simplified form: indeed, forward reaction $\nu_k \to \nu'_k$ and reverse reaction $\nu'_k \to \nu_k$ can be identified as the same reaction R_k, which is now represented as an unordered pair,

$$R_k = \{ \nu_k, \nu'_k \}, \qquad \nu_k, \nu'_k \in \mathcal{C}, \tag{1.18}$$

and depicted using a double arrow,

$$\sum_i (\nu_k)_i S_i \rightleftharpoons \sum_i (\nu'_k)_i S_i, \qquad \nu_k, \nu'_k \in \mathcal{C}. \tag{1.3}$$

Consequently, the complex graph becomes an undirected graph.

Consistently with this new definition for reactions, the stoichiometric matrix S and the incidence matrix B have to be redefined. In particular, only one of the two columns representing an original forward-reverse reaction pair is kept in each of these matrices; by convention, denote as *forward reaction* the one whose corresponding vectors are kept, and as *reverse reaction* the one whose corresponding vectors are removed. Clearly, the choice on which column to keep

and which to remove must be the same for both matrices, in order to preserve the fundamental relation $S = ZB$.

One the other hand, elements of the counting vector $Y = \{ Y(t) \}_{t \geq 0}$ are allowed to take negative values

$$Y = \{ Y(t) \in \mathbb{Z}^l,\ t \geq 0 \}; \tag{1.19}$$

indeed, $Y_k(t)$ counts now the *difference* in the number of times forward and reverse reactions R_k have occurred up to time $t \geq 0$, i.e. captures the *net* effect of reversible reaction R_k on the system.

In conclusion, the expression for the state $X(t)$ of the system at time $t \geq 0$ remains unchanged:

$$X(t) = X(0) + \sum_k \zeta_k Y_k(t) = X(0) + S\,Y(t), \qquad t \geq 0, \tag{1.20}$$

where the sum is now considered over reversible reactions.

In the following of this section, the discussion is based on the formulation of CRNs in terms of reversible reactions presented above and extensively developed in [18].

Remark 1.9. This alternative formulation turns out to be particularly useful to derive geometrical properties from system matrices: the fact that the counting vector $Y \in \mathbb{Z}^n$ is now an element of a vector space makes the discussion much easier, since no additional constraints on non-negativity of some linear combinations have to be considered.

Consider the stoichiometric matrix $S \in \mathbb{Z}^{n \times l}$; both its right and left null spaces have a straightforward interpretation.

- Matrix S can be regarded as a (linear) function

  $$S \colon \mathbb{Z}^l \mapsto \mathbb{Z}^n, \tag{1.21}$$

  which maps a vector $Y(t)$ of reaction counts to the effect it has on the system state. From this perspective, its right null space contains counting vectors $Y(t)$ which define a *circuit* (or more circuits) in the state space $\mathbb{S} = \mathbb{N}_0^n$:

  $$Y(t) \in \ker S \quad \Longrightarrow \quad X(t) = X(0) + S\,Y(t) = X(0), \quad t \geq 0.$$

  A basis for $\ker S$ is a set of reaction vectors corresponding to *independent cycles* in the state space.

- Stoichiometric matrix also captures basic conservation laws of the system: indeed, its left null space includes linear combinations of species which are *conserved* for system dynamics:

  $$k \in \ker S^T \quad \Longrightarrow \quad kX(t) = kX(0) + (kS)\,Y(t) = kX(0), \quad t \geq 0,$$

  i.e. $Q = \sum_i k_i S_i$ is a conserved quantity.

  If row vector $k \in \mathbb{Z}_0^n$ has all non-negative entries, Q is called a *conserved moiety*.

Likewise, the image of matrix S also has a useful characterization [1, ch. 3].

Definition 1.10. Let $S \in \mathbb{Z}^{n\times l}$ be the stoichiometric matrix of a CRN.

- The vector subspace $\mathrm{im}S = \mathrm{span}_k \{ \zeta_k \} \subseteq \mathbb{Z}^n$ is the *stoichiometric subspace* of the network.
- For $x \in \mathbb{Z}^n$, the affine space $(x + \mathrm{im}S) \subseteq \mathbb{Z}^n$ is the *stoichiometric compatibility class* for state x.

It is straightforward to show that, whatever the choice of dynamics, the state $X(t)$ of the system remains within a single stoichiometric compatibility class for all time $t \geq 0$:

$$X(t) - X(0) = S\,Y(t) \in \mathrm{im}S \quad \Longrightarrow \quad X(t) \in (X(0) + \mathrm{im}S) \subseteq \mathbb{Z}^n.$$

Moreover, given an initial condition $X(0) \in \mathbb{N}_0^n$, (i.e. having non-negative entries), the state of the system remains in the so called *non-negative stoichiometric compatibility class*

$$(X(0) + \mathrm{im}S) \cap \mathbb{N}_0^n \ \subseteq \mathbb{Z}^n,$$

provided that the choice of dynamics involves a stoichiometrically admissible kinetics.

Theorem 1.11 (Rank-Nullity theorem)**.** *For a generic matrix $S \in \mathbb{Z}^{n\times l}$, the following relations hold:*

$$l = \dim\ker S + \dim\mathrm{im}S \tag{1.22}$$

$$n = \dim\ker S^T + \dim\mathrm{im}S \tag{1.23}$$

Therefore, if the rank of S (i.e. the dimension of $\mathrm{im}S$) is less than n, the stoichiometric subspace is a proper subspace of the state space $\mathbb{S}$. In this case, $\ker S^T$ is non-null, and there exists a conserved quantity in the system.

Consider now the incidence matrix $B \in \mathbb{Z}^{m\times l}$ of the complex graph.

Remark 1.12. In the reversible setting considered here, the complex graph is an undirected graph, even if the form of the incidence matrix B still suggests a direction for graph edges: indeed, B describes the complex graph including *forward* reactions only. However, elements of the counting vector $Y(t)$, which is right-multiplied by matrix B, can take negative values: this means that graph edges can also be walked the other way.

Recall some geometric properties of (directed) incidence matrices [18].

Property 1.13. *Consider a graph $\mathcal{G}$ with m vertices, and denote by $B \in \mathbb{Z}^{m\times l}$ its directed incidence matrix.*

- *The rank of B is given by*

$$\mathrm{rank}B = \dim\mathrm{im}B = m - k, \tag{1.24}$$

where k is the number of connected components. In particular, if $\mathcal{G}$ is connected, $\mathrm{rank}B = m - 1$.

- *The (right) null space of B contains vectors $y \in \mathbb{Z}^l$ which define a circuit (or more circuits) on $\mathcal{G}$, where graph edges can be walked both ways. A basis for* $\ker B$ *is a set of vectors corresponding to independent cycles on the graph.*

Therefore, the kernel of the incidence matrix B contains vectors $Y(t)$ of reaction counts defining circuits in the (reversible) complex graph [15].

A key relation may be established between right null vectors of matrices S and B.

Property 1.14. *A vector $Y(t)$ which is in the kernel of the incidence matrix B is also in the kernel of the stoichiometric matrix S:*

$$Y(t) \in \ker B \quad \Longrightarrow \quad Y(t) \in \ker S,$$

that is, each counting vector $Y(t)$ corresponding to a circuit in the complex graph also corresponds to a circuit in the state space $\mathbb{S}$.

The statement easily follows from the fundamental relation $S = ZB$:

$$B\,Y(t) = 0 \quad \Longrightarrow \quad S\,Y(t) = 0.$$

Intuitively, each circuit in the complex graph (i.e. each closed sequence of reactions) restores the same counts of complexes, and therefore also restores the same counts of chemical species. However, the converse is not necessarily true.

Definition 1.15. The *deficiency* of a chemical reaction network $\{\,\mathcal{S}, \mathcal{C}, \mathcal{R}\,\}$ with stoichiometric matrix S and incidence matrix B is defined as

$$\delta = \dim\ker S - \dim\ker B. \tag{1.25}$$

Note that, from Property 1.14, it directly follows that the deficiency of a CRN is always non-negative.

The deficiency can be interpreted, from the geometric point of view, as the number of independent cycles (closed sequences of reactions) in the state space which cannot be visualized as independent cycles in the complex graph; these cycles are sometimes called *hidden* cycles [15].

An alternative equivalent definition of deficiency is commonly found in literature (e.g. [18]).

Lemma 1.16. *The deficiency can be expressed as*

$$\delta = \dim\mathrm{im}B - \dim\mathrm{im}S = \mathrm{rank}B - \mathrm{rank}S. \tag{1.26}$$

Proof. Applying the Rank-Nullity Theorem 1.11 to matrices S and B, one obtains

$$\dim\ker S = l - \dim\mathrm{im}S, \qquad \dim\ker B = l - \dim\mathrm{im}B. \tag{1.27}$$

The result then follows directly substituting into the definition. □

This equivalent expression allows to give another geometric interpretation of deficiency from a different point of view. Indeed, regarding the complex stoichiometric matrix $Z \in \mathbb{Z}^{n\times m}$ as a function restricted on the proper subspace $\mathrm{im}B \subset \mathbb{Z}^m$,

$$Z\colon \mathrm{im}B \subset \mathbb{Z}^m \mapsto \mathrm{im}S \in \mathbb{Z}^n,$$

deficiency can be associated with the dimensionality reduction caused by Z when mapping $\mathrm{im}B$ into the stoichiometric subspace $\mathrm{im}S$.

Moreover, recalling the definition for the rank of the incidence matrix (1.24), one obtains

$$\delta = m - k - \mathrm{rank}S, \tag{1.28}$$

where m is the number of complexes and k is the number of connected components in the complex graph (which are termed *linkage classes* in this setting) [1, ch. 3].

Remark 1.17. Zero-deficiency is a key property for a chemical reaction network: observations above show that zero-deficiency is equivalent to

$$\ker Z \cap \mathrm{im}B = 0,$$

or to the mapping $Z\colon \mathrm{im}B \mapsto \mathrm{im}S$ being *injective*.

Therefore, in a zero-deficient CRN, there exists a one-to-one correspondence between the stoichiometric subspace $\mathrm{im}S$ and the subspace $\mathrm{im}B \subset \mathbb{Z}^m$, which determines all combinations of complexes counts that are reachable as a consequence of reactions in $\mathcal{R}$. In addition, all cycles in the state space $\mathbb{S}$ are also cycles in the complex graph (and viceversa), i.e. there are no hidden cycles.

1.3 Limit behaviour under classical scaling

In classical chemistry, the number of molecules of species involved in a reaction is usually very high, even for systems contained in relatively small volumes: for this reason, the state of a reaction system is commonly described in terms of chemical concentrations rather than number of molecules.

Definition 1.18. Let $X_i \in \mathbb{N}_0$ be the number of molecules of species S_i present in the system. The *chemical concentration* $C_i^{N_V} \in \mathbb{R}_{\geq 0}$ of species S_i is given by

$$C_i^{N_V} := \frac{X_i}{N_V}, \qquad N_V = N_A V, \tag{1.29}$$

where N_A is Avogadro's number ($\approx 6.02 \times 10^{23}$) and V is the volume of the reaction mixture.

Similarly, reaction rate constants may also be affected by the volume of the reaction mixture.

- For a binary reaction,

  $$S_1 + S_2 \to *,$$

  the reaction rate should vary inversely with the volume, so that it takes the form

  $$\lambda^{N_V}(x) = \frac{\kappa'}{N_V} x_1 x_2.$$

This assumption captures the intuitive idea that, given a fixed number of molecules for both species, reaction happens at faster rate if reactants are confined is a smaller volume (when the system is well-stirred).

- For a unary reaction
$$S_1 \to *,$$
the reaction rate does not dependend on the volume:
$$\lambda^{N_V}(x) = \lambda(x) = \kappa x_1.$$
In this case, the substrate is made of a single chemical species, which transforms into the product complex independently of the volume of the system.

- For an inflow reaction
$$\emptyset \to *,$$
the reaction rate is assumed to vary proportionally with the volume, so that it takes the form
$$\lambda^{N_V}(x) = N_V \kappa'.$$

In general, when the volume of the reaction mixture is included in mass action kinetics, the intensity function (reaction rate) for reaction R_k is given by

$$\lambda_k^{N_V}(x) = \frac{\kappa'_k}{N_V^{|\nu_k|-1}} \frac{x!}{(x-\nu_k)!}, \qquad k = 1, \dots, l, \tag{1.30}$$

where $|\nu_k| = \sum(\nu_k)_i$ is the number of molecules in the substrate. Assuming the number of molecules of species involved in the reaction to be very high, one can write the approximation
$$\frac{x!}{(x-\nu_k)!} \approx x^{\nu_k},$$
so that the intensity function becomes
$$\lambda_k^{N_V}(x) = \frac{\kappa'_k}{N_V^{|\nu_k|-1}} x^{\nu_k} = N_V \kappa'_k \frac{x^{\nu_k}}{N_V^{|\nu_k|}}, \qquad k = 1, \dots, l.$$

Therefore, mass action kinetics can be easily redefined in terms of chemical concentration of species instead of absolute numbers of molecules.

Definition 1.19. Let $c_i = N_V^{-1} x_i$ be the chemical concentration of species S_i. *Mass action kinetics* for concentrations has intensity function for reaction R_k given by
$$\lambda_k^{N_V}(x) = N_V \lambda'_k(c), \qquad k = 1, \dots, l, \tag{1.31}$$
where
$$\lambda'_k(c) := \kappa'_k \prod_{i=1}^{n} c_i^{(\nu_k)_i} = \kappa'_k c^{\nu_k}. \tag{1.32}$$

With these definitions, the state of the system is now described by a vector function of time
$$C^{N_V} = \{\, C^{N_V}(t) \in \mathbb{R}^n_{\geq 0},\ t \geq 0 \,\}, \tag{1.33}$$

representing chemical concentrations of the different species in $\mathcal{S}$. Note that the volume is assumed to be constant over time, i.e. $N_V = N_A V$ does not depend on time.

The stochastic process C^{N_V} can thus be modeled as a continuous time Markov chain with state space $\mathbb{S} = \mathbb{R}^n_{\geq 0}$:

$$C^{N_V}(t) = C^{N_V}(0) + \sum_k N_V^{-1} \xi_k \left(N_V \int_0^t \lambda'_k(C^{N_V}(s))ds \right) \zeta_k, \qquad t \geq 0. \quad (1.34)$$

As already highlighted, quantity $N_V = N_A V$ is very large even if volume V is small: therefore, if we consider a sequence of equations

$$C^N(t) = C^N(0) + \sum_k N^{-1} \xi_k \left(N \int_0^t \lambda'_k(C^N(s))ds \right) \zeta_k, \qquad t \geq 0, \quad (1.35)$$

indexed by N (which is now unbound to the quantity N_V), the limiting stochastic process C for N growing large,

$$C = \lim_{N \to \infty} C^N,$$

if it exists, should reasonably approximate the stochastic process C^N, $N = N_V$.

The following of this section is devoted to characterize the limiting process for the sequence described in (1.35).

Let $l = |\mathcal{R}| < \infty$ and define the vector function

$$F \colon \mathbb{R}^n_{\geq 0} \mapsto \mathbb{R}^n, \qquad F(c) = \sum_k \lambda'_k(c)\zeta_k, \quad (1.36)$$

so that equation (1.35) becomes

$$C^N(t) = C^N(0) + M^N(t) + \int_0^t F(C^N(s))ds, \qquad t \geq 0, \quad (1.37)$$

where

$$M^N(t) = \sum_k N^{-1} \hat{\xi}_k \left(N \int_0^t \lambda'_k(C^N(s))ds \right) \zeta_k, \qquad t \geq 0,$$

and $\hat{\xi}_k(u) = \xi_k(u) - u$.

Theorem 1.20 (Strong law of large numbers for Poisson process)**.** *Let ξ be a unit Poisson process; then, for each $t > 0$,*

$$\lim_{N \to \infty} \sup_{s \leq t} \left| N^{-1} \xi(Ns) - s \right| = 0, \qquad a.s. \quad (1.38)$$

Under suitable hypotheses, the strong law of large numbers for Poisson processes can be directly applied to process M^N in order to prove the following theorem, which may itself be regarded as a law of large numbers for the sequence of stochastic processes described in (1.35) (for the proof, see [7, ch. 11]).

Theorem 1.21. *Let C^N satisfy equation (1.35) and define the deterministic function $c\colon \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ as the solution to the equation*

$$c(t) = c(0) + \int_0^t F(c(s))ds, \qquad t \geq 0. \tag{1.39}$$

Moreover, assume the local Lipschitz condition on F, i.e. for each compact $\mathcal{K} \subset \mathbb{R}^n_{\geq 0}$, there exists M_K such that

$$|F(x) - F(y)| \leq M_K |x - y|, \qquad x, y \in \mathcal{K}. \tag{1.40}$$

If $C^N(0)$ has finite (positive) limit for $N \to \infty$,

$$\lim_{N\to\infty} C^N(0) = c(0) > 0,$$

then, for each $\epsilon > 0$ and for each $t > 0$,

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{s\leq t} |C^N(s) - c(s)| \geq \epsilon\right) = 0. \tag{1.41}$$

As a consequence, the stochastic model for C^{N_V} described in (1.34) could be reasonably approximated, in the large volume limit, by the deterministic model described in (1.39):

$$C^{N_V}(t) \approx C(t) = c(t), \qquad t \geq 0.$$

The approximation procedure presented above is known in literature as *classical scaling* (e.g. [1, ch. 4] and [13]).

Remark 1.22. Equation (1.39) is expressed in integral form, but can be equivalently written as an ordinary differential equation [2]:

$$\dot{c}(t) = F(c(t)) = \sum_k \kappa_k c(t)^{\nu_k} \zeta_k, \qquad t \geq 0.$$

In this form, the equation is well-know in classical chemistry as *deterministic law of mass action*. Note that we are implicitly assuming that a solution to the equation exists for all $t > 0$; uniqueness is guaranteed by the Lipschitz assumption.

1.4 Multiscale approximations

Chemical reactions systems studied in cellular biology renewed the interest in stochastic models, since the approximation under classical scaling proved to be inadequate to properly describe complex phenomena at cellular level. This is essentially due to three reasons:

- the number of molecules involved, at least for some species, may be sufficiently small that the continuous approximation of the deterministic model does not provide a satisfactoy representation of the actual behaviour of the system, or expressing their abundance as concentrations appears inappropriate;

- some species may be present in much greater abundance than others, so that there is no normalization of molecules numbers that is suitable for all species in the system;
- reaction rate constants frequently vary over several orders of magnitude.

Therefore, in order to capture this variability, we would like to explore alternative approaches to produce interesting scaling limit approximation models.

Let $N_0 \gg 1$ denote the (single) scaling parameter for the model: it is still assumed to be large, but has no longer an interpretation in terms of Avogadro's number or volume of the reaction mixture (i.e. it has no physical meaning).

Definition 1.23. For each species S_i, define the *normalized abundance* as

$$Z_i^{N_0} := N_0^{-\alpha_i} X_i, \qquad i = 1, \dots, n, \tag{1.42}$$

where $\alpha_i \geq 0$ should be selected so that $Z_i^{N_0} = O(1)$. Note that the normalized abundance may be the species number ($\alpha_i = 0$), the species concentration ($\alpha_i = 1$) or something else.

The scaling parameter N_0 is usually chosen to be the order of magnitude of the abundance of the most abundant species in the system, so that

$$0 \leq \alpha_i \leq 1, \qquad \forall i = 1, \dots, n.$$

Definition 1.24. For each reaction R_k, define the *normalized reaction rate constant* as

$$\kappa'_k := \kappa_k N_0^{-\beta_k}, \qquad k = 1, \dots, l, \tag{1.43}$$

where $\beta_k \in \mathbb{R}$ should be selected so that $\kappa'_k = O(1)$.

For example, if we suppose

$$\kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_l,$$

it could be reasonable to select coefficients β_i, $i = 1, \dots, l$ so that

$$\beta_1 \geq \beta_2 \geq \cdots \geq \beta_l,$$

although it may be more natural to impose this order separately for unary, binary or inflows reactions, as it happens for *classical scaling* (Remark 1.26).

Similarly to Definition 1.19, intensity functions characterizing mass action kinetics can be rewritten in terms of normalized abundances [2].

Definition 1.25. Let $z_i = N_0^{-\alpha_i} x_i$ be the normalized abundance of species S_i. *Mass action kinetics* for abundances has intensity function for reaction R_k given by

$$\lambda_k^{N_0}(x) = N_0^{\beta_k + \nu_k \cdot \alpha} \lambda'_k(z), \qquad k = 1, \dots, l, \tag{1.44}$$

where

$$\lambda'_k(z) := \kappa'_k \prod_{i=1}^{n} z_i^{(\nu_k)_i} = \kappa'_k z^{\nu_k}. \tag{1.45}$$

Therefore, the stochastic process Z^{N_0}, representing normalized abundances, can be modeled as a continuous time Markov chain: for each species S_i and for $t \geq 0$,

$$Z_i^{N_0}(t) = Z_i^{N_0}(0) + \sum_k N_0^{-\alpha_i} \xi_k \left(N_0^{\beta_k + \nu_k \cdot \alpha} \int_0^t \lambda_k'(Z^{N_0}(s)) ds \right) (\zeta_k)_i.$$

Since N_0 is assumed to be large, we can again consider a sequence of stochastic models, indexed by N, satisfying

$$Z_i^N(t) = Z_i^N(0) + \sum_k N^{-\alpha_i} \xi_k \left(N^{\beta_k + \nu_k \cdot \alpha} \int_0^t \lambda_k'(Z^N(s)) ds \right) (\zeta_k)_i, \qquad (1.46)$$

for each species S_i and for $t \geq 0$, and attempt to approximate the original model (corresponding to $N = N_0$) with the limiting process for N growing large, if such limit exists:

$$Z^{N_0} \approx Z = \lim_{N \to \infty} Z^N. \qquad (1.47)$$

Remark 1.26. The *classical scaling* presented in Section 1.3 is a particular case of the general scaling procedure described in this section, obtained taking

$$\alpha_i = 1, \qquad i = 1, \ldots, n,$$

and

$$\beta_k = 1 - |\nu_k| = 1 - \sum_i (\nu_k)_i, \qquad k = 1, \ldots, l.$$

Remark 1.27. In Definition 1.23, the requirement $Z_i^{N_0} = O(1)$ appears to be quite vague, and still leaves a certain degree of arbitrariness about the selection. To be more precise α_i should be large enough so that the family of stochastic processes $\{ Z_i^N \}$ is stochastically bounded, i.e. for each $\epsilon > 0$ and $t \geq 0$ there exists $K_{\epsilon,t}$ such that

$$\inf_N \mathbb{P} \left(\sup_{s \leq t} Z_i^N(s) \leq K_{\epsilon,t} \right) \geq 1 - \epsilon, \qquad i = 1, \ldots, n.$$

On the other hand, α_i cannot be so large that $\{ Z_i^N \}_N$ converges to zero for $N \to \infty$: for example, a sufficient condition could be the existance of $\delta_{\epsilon,t}$ such that

$$\inf_N \mathbb{P} \left(\inf_{s \leq t} Z_i^N(s) \geq \delta_{\epsilon,t} \right) \geq 1 - \epsilon, \qquad i = 1, \ldots, n.$$

However, it is not uncommon to encounter situations in which $\alpha_i = 0$ and Z_i^N is frequently zero, so this requirement is in general too restrictive [13, sec. 2].

In characterizing the limiting process for the sequence in (1.46), we would like to consider also a scaling in time, and possibly study the model on different time scales. Consider the change of time variable

$$\tau = N^{-\gamma} t, \qquad t \geq 0, \qquad (1.48)$$

and define the stochastic process $Z^{N,\gamma}$ as the time-scaled version of process Z^N:

$$Z^{N,\gamma}(\tau) := Z^N(t) = Z^N(N^\gamma \tau), \qquad t \geq 0. \qquad (1.49)$$

Consequently, equation (1.46) can be restated in terms of the time-scaled process as

$$Z_i^{N,\gamma}(\tau) = Z_i^N(0) + \sum_k N^{-\alpha_i}\xi_k \left(N^{\gamma+\beta_k+\nu_k\cdot\alpha} \int_0^\tau \lambda_k'(Z^{N,\gamma}(s))ds \right) (\zeta_k)_i,$$

for each species S_i and for $\tau \geq 0$.

Different choices of time-scale parameter γ may lead to interesting approximations for different subsets of species. A large number of examples in this sense can be found in [2] and [13, sec. 6].

1.5 Balance conditions and natural time-scales

Remark 1.27 provides reasonable requirements to obtain nontrivial limiting models: these requirements may be extended to the time-scaled family of stochastic processes $\{ Z_i^{N,\gamma} \}$ and place constraints on possible values for parameters α, β and possibly γ.

Consider the Michaelis-Menten model introduced in Section 1.1:

$$S + E \rightleftharpoons ES \to P + E, \qquad E \rightleftharpoons \emptyset.$$

If we introduce scaling parameters α, β and γ, the equation for process $Z_S^{N,\gamma}$ is

$$\begin{aligned} Z_S^{N,\gamma}(\tau) = Z_S^N(0) &- N^{-\alpha_S}\xi_1 \left(N^{\gamma+\beta_1+\alpha_S+\alpha_E} \int_0^\tau \kappa_1' Z_S^{N,\gamma}(s) Z_E^{N,\gamma}(s) ds \right) \\ &+ N^{-\alpha_S}\xi_2 \left(N^{\gamma+\beta_2+\alpha_{ES}} \int_0^\tau \kappa_2' Z_{ES}^{N,\gamma}(s) ds \right), \qquad \tau \geq 0. \end{aligned}$$

Assume that $Z_i^{N,\gamma} = O(1)$ for $i \neq S$ and $Z_S^N(0) = O(1)$. Then $Z_S^{N,\gamma} = O(1)$ if at least one of these constraints holds:

- the rate of production of species S has the same order of magnitude of its rate of consumption, so that $Z_S^{N,\gamma}$ neither explodes nor goes to zero:

$$\beta_2 + \alpha_{ES} = \beta_1 + \alpha_S + \alpha_E;$$

- the rate at which reactions involving species S happen has order of magnitude less or equal to the normalization parameter for its abundance:

$$\gamma + \beta_k + \nu_k \cdot \alpha \leq \alpha_S, \qquad k = 1, 2.$$

In general, we should require parameters α, β and γ to satisfy the following condition [13, sec. 3].

Condition 1.28. Let $\Gamma_i^+ \subseteq \mathcal{R}$ and $\Gamma_i^- \subseteq \mathcal{R}$ be the sets of reactions that result in an increase and decrease in the abundance of species S_i, respectively:

$$\Gamma_i^+ := \{ k \colon (\zeta_k)_i > 0 \}, \qquad \Gamma_i^- := \{ k \colon (\zeta_k)_i < 0 \}, \qquad i = 1, \ldots, n.$$

The *species balance condition* for species S_i holds if at least one of these requirements is satisfied:

1. *Balance equation*

$$\max_{k\in\Gamma_i^+}(\beta_k+\nu_k\cdot\alpha)=\max_{k\in\Gamma_i^-}(\beta_k+\nu_k\cdot\alpha),\qquad i=1,\dots,n; \tag{1.50}$$

2. *Time-scale constraint*

$$\max_{k\in\Gamma_i^+\cup\Gamma_i^-}(\beta_k+\nu_k\cdot\alpha)+\gamma\le\alpha_i,\qquad i=1,\dots,n. \tag{1.51}$$

Note that the time-scale constraint can be equivalently stated as

$$\gamma\le\alpha_i-\max_{k\in\Gamma_i^+\cup\Gamma_i^-}(\beta_k+\nu_k\cdot\alpha),\qquad i=1,\dots,n. \tag{1.52}$$

In particular, if we consider reaction R_k, with $(\zeta_k)_i\neq 0$, the quantity

$$\gamma=\alpha_i-(\beta_k+\nu_k\cdot\alpha)$$

can be regarded as the *natural time-scale* for the normalized reaction counting

$$N^{-\alpha_i}Y_k^{N,\gamma}(\tau)=N^{-\alpha_i}\xi_k\left(N^{\gamma+\beta_k+\nu_k\cdot\alpha}\int_0^\tau\lambda_k'(Z^{N,\gamma}(s))ds\right). \tag{1.53}$$

A similar idea can be extended at the species level.

Definition 1.29. The *natural time-scale* γ_i for the normalized abundance of species S_i is defined as

$$\gamma_i:=\alpha_i-\max_{k\in\Gamma_i^+\cup\Gamma_i^-}(\beta_k+\nu_k\cdot\alpha). \tag{1.54}$$

For $\gamma=\gamma_i$, none of the normalized reaction countings in (1.53) for species S_i should blow up, and at least one should be nontrivial (i.e. should not uniformly converge to zero). To be more precise, let $\Gamma_{i,0}\subseteq\mathcal{R}$ be the set of reactions, involving species S_i, with fastest rates:

$$\Gamma_{i,0}:=\arg\max_{k\in\Gamma_i^+\cup\Gamma_i^-}(\beta_k+\nu_k\cdot\alpha).$$

Then, if $\alpha_i>0$, we reasonably expect (see Theorem 1.20)

$$\lim_{N\to\infty}Z_i^{N,\gamma_i}(\tau)=\lim_{N\to\infty}\left(Z_i^N(0)+\sum_{k\in\Gamma_{i,0}}\int_0^\tau\lambda_k'(Z^{N,\gamma_i}(s))ds\ (\zeta_k)_i\right),$$

while, if $\alpha_i=0$, we expect

$$\lim_{N\to\infty}Z_i^{N,\gamma_i}(\tau)=\lim_{N\to\infty}\left(Z_i^N(0)+\sum_{k\in\Gamma_{i,0}}\xi_k\left(\int_0^\tau\lambda_k'(Z^{N,\gamma_i}(s))ds\right)(\zeta_k)_i\right).$$

Note that time-scales are here associated with single species; therefore, one reaction may determine different time-scales associated with different species (or collection of species).

In conclusion, considering the whole reaction system, we would like Condition 1.28 to hold for each species S_i, $i=1,\dots,n$. In particular, we could require the time-scale constraint, as stated in (1.52), to hold for each species:

$$\gamma\le\alpha_i-\max_{k\in\Gamma_i^+\cup\Gamma_i^-}(\beta_k+\nu_k\cdot\alpha),\qquad\forall i=1,\dots,n. \tag{1.55}$$

Definition 1.30. The *(first) natural time-scale* γ' for a (normalized) reaction system is defined as

$$\gamma' := \min_{i\in\mathcal{S}} \gamma_i = \min_{i\in\mathcal{S}} \left(\alpha_i - \max_{k\in\Gamma_i^+\cap\Gamma_i^-} (\beta_k + \nu_k \cdot \alpha) \right). \tag{1.56}$$

Again, setting $\gamma = \gamma'$, none of the normalized reaction countings in (1.53) should blow up, for each reaction $R_k \in \mathcal{R}$ and each species $S_i \in \mathcal{S}$, and at least one should be nontrivial.

Let $\Gamma_i^\gamma \subseteq \mathcal{R}$ be the set of reactions, involving species S_i, with natural time-scale γ:

$$\Gamma_i^\gamma := \{\, k\colon \alpha_i = \gamma + (\beta_k + \nu_k \cdot \alpha),\ (\zeta_k)_i \neq 0 \,\}.$$

The following Theorem characterizes the limit for the stochastic process of normalized abundances (for the proof, see [13, sec. 4]).

Theorem 1.31. *Let γ' be the first natural time-scale, as defined in* (1.56)*, and assume that $Z_i^N(0)$ has finite positive limit for $N \to \infty$, for each species S_i:*

$$\lim_{N\to\infty} Z_i^N(0) = Z_i(0) > 0, \qquad \forall i = 1, \dots, n.$$

Let $Z^{\gamma'}$ be the stochastic process described, for each species S_i, by equations

$$Z_i^{\gamma'}(\tau) = Z_i(0) + \sum_{k\in\Gamma_i^{\gamma'}} \int_0^\tau \lambda_k'(Z^{\gamma'}(s))ds\ (\zeta_k)_i, \qquad \text{if} \quad \alpha_i > 0,$$

$$Z_i^{\gamma'}(\tau) = Z_i(0) + \sum_{k\in\Gamma_i^{\gamma'}} \xi_k \left(\int_0^\tau \lambda_k'(Z^{\gamma'}(s))ds \right) (\zeta_k)_i, \qquad \text{if} \quad \alpha_i = 0.$$

Then, the stochastic process of normalized abundances $Z^{N,\gamma'}$ converges in distribution (i.e. weakly) to the stochastic process $Z^{\gamma'}$ on time interval $[0, \tau_\infty)$,

$$Z^{N,\gamma'} \Rightarrow Z^{\gamma'}, \qquad \text{on} \quad [0, \tau_\infty), \tag{1.57}$$

where τ_∞ is itself a random (stopping) time, defined as

$$\tau_\infty := \lim_{c\to\infty} \left(\inf \left\{ t\colon \sup_{s\le t} |Z^{\gamma'}(s)| \ge c \right\} \right). \tag{1.58}$$

Remark 1.32. The specification regarding the random time interval $[0, \tau_\infty)$ reflects the possibility for the reaction system to explode in finite time.

For example, consider the following one-species reaction system,

$$2S_1 \to 3S_1, \qquad S_1 \to \emptyset,$$

whose state at time $t \ge 0$ is described by the equation

$$X_1(t) = X_1(0) + \xi_1 \left(\int_0^t \kappa_1 X_1(s)(X_1(s) - 1)ds \right) - \xi_2 \left(\int_0^t \kappa_2 X_1(s)ds \right).$$

This model is actually a birth-and-death process with quadratic birth rates and linear death rates, expressed respectively by

$$\lambda_1(x) = \kappa_1 x(x-1), \qquad \lambda_2(x) = \kappa_2 x.$$

Therefore, if $X_1(0) > 1$, it has positive probability of exploding in finite time.

1.6 Linear combinations and faster time-scales

Condition 1.28 seems a reasonable requirement to prevent normalized abundances from blowing up as N grows large. However, in general, it does not ensure by itself that

$$Z_i^{N,\gamma} = O(1), \qquad i = 1, \dots, n.$$

Indeed, there may be subsets of species such that the collective rate of production is greater than the collective rate of consumption, and therefore their collective normalized abundance blows up as $N \to \infty$, at a suitable time-scale.

Consider again the Michaelis-Menten model,

$$S + E \rightleftharpoons ES \to P + E, \qquad E \rightleftharpoons \emptyset,$$

and focus on species E and ES. If we assume, for example,

$$\alpha_E = \alpha_{ES} = 0, \qquad 0 < \beta_4 < \beta_5 < \beta_1 = \beta_2 = \beta_3, \qquad \gamma = 0,$$

the equations for processes $Z_E^{N,\gamma}$ and $Z_{ES}^{N,\gamma}$ are

$$\begin{aligned} Z_E^{N,\gamma}(\tau) = X_E^N(t) = X_E(0) &- \xi_1 \left(N^{\beta_1} \int_0^t \kappa_1' Z_S^{N,\gamma}(s) Z_E^{N,\gamma}(s) ds \right) \\ &+ \xi_2 \left(N^{\beta_2} \int_0^t \kappa_2' Z_{ES}^{N,\gamma}(s) ds \right) + \xi_3 \left(N^{\beta_3} \int_0^t \kappa_3' Z_{ES}^{N,\gamma}(s) ds \right) \\ &- \xi_4 \left(N^{\beta_4} \int_0^t \kappa_4' Z_E^{N,\gamma}(s) ds \right) + \xi_5 \left(N^{\beta_5} \int_0^t \kappa_5' ds \right), \end{aligned}$$

$$\begin{aligned} Z_{ES}^{N,\gamma}(\tau) = X_{ES}^N(t) = X_{ES}(0) &+ \xi_1 \left(N^{\beta_1} \int_0^t \kappa_1' Z_S^{N,\gamma}(s) Z_E^{N,\gamma}(s) ds \right) \\ &- \xi_2 \left(N^{\beta_2} \int_0^t \kappa_2' Z_{ES}^{N,\gamma}(s) ds \right) - \xi_3 \left(N^{\beta_3} \int_0^t \kappa_3' Z_{ES}^{N,\gamma}(s) ds \right). \end{aligned}$$

Since $\beta_1 = \beta_2 = \beta_3$, the species balance condition is satisfied (balance equation holds), but if we consider the process $Z_E^{N,\gamma} + Z_{ES}^{N,\gamma}$ we obtain the equation

$$\begin{aligned} Z_E^{N,\gamma}(\tau) + Z_{ES}^{N,\gamma}(\tau) &= X_E^N(t) + X_{ES}^N(t) \\ &= X_E(0) + X_{ES}(0) - \xi_4 \left(N^{\beta_4} \int_0^t \kappa_4' Z_E^{N,\gamma}(s) ds \right) \\ &\quad + \xi_5 \left(N^{\beta_5} \int_0^t \kappa_5' ds \right), \end{aligned}$$

and, given that $\beta_4 < \beta_5$, the normalized abundances blow up anyway, at this time-scale. This example suggests the need to consider linear combinations of species.

At first, assume for simplicity that $\alpha_i = \bar{\alpha}$ for all species, and consider the normalized abundance of a linear combination of species $\theta \cdot S$, defined as

$$Z_\theta^{N,\gamma} := \theta \cdot Z^{N,\gamma}, \qquad \theta \in \mathbb{R}^n_{\geq 0}. \tag{1.59}$$

The stochastic process $Z_\theta^{N,\gamma}$ satisfies equation

$$Z_\theta^{N,\gamma}(\tau) = \theta^T Z^{N,\gamma}(\tau) = \theta^T Z^N(0) + \sum_k N^{-\bar{\alpha}} Y_k^{N,\gamma}(\tau)\ \theta^T \zeta_k, \tag{1.60}$$

for $\tau \geq 0$, where

$$Y_k^{N,\gamma}(\tau) = \xi_k \left(N^{\gamma+\beta_k+\nu_k\cdot\alpha} \int_0^\tau \lambda_k'(Z^{N,\gamma}(s))ds \right),$$

for $k = 1, \ldots, l$ are reaction countings.

If normalization parameters α_i, $i = 1, \ldots, n$ are not all equal, the linear combination of species should be normalized with respect to the largest normalization parameter among those of species involved in the combination,

$$\alpha_\theta := \max_{i:\ \theta_i > 0} \alpha_i, \tag{1.61}$$

so that the normalized abundance of a linear combination of species $\theta \cdot S$ is

$$Z_\theta^{N,\gamma} := N^{-\alpha_\theta}\ \theta \cdot \Lambda_N Z^{N,\gamma} = N^{-\alpha_\theta}\ \theta \cdot X^{N,\gamma}, \qquad \theta \in \mathbb{R}^n_{\geq 0}, \tag{1.62}$$

where Λ_N is the diagonal matrix with entries

$$(\Lambda_N)_{ii} = N^{\alpha_i}, \qquad i = 1, \ldots, n.$$

Therefore, the stochastic process $Z_\theta^{N,\gamma}$ satisfies equation

$$\begin{aligned} Z_\theta^{N,\gamma}(\tau) &= N^{-\alpha_\theta} \theta^T \Lambda_N Z^{N,\gamma}(\tau) \\ &= N^{-\alpha_\theta} \theta^T \Lambda_N Z^N(0) + \sum_k N^{-\alpha_\theta} Y_k^{N,\gamma}(\tau)\ \theta^T \zeta_k. \end{aligned} \tag{1.63}$$

Requirements introduced in Remark 1.27 may be applied to linear combinations of species to rule out degenerate limiting models.

As before, assume that $Z_i^{N,\gamma} = O(1)$ for species not entering the linear combination, and $Z_\theta^N(0) = O(1)$. Then, $Z_\theta^{N,\gamma} = O(1)$ if at least one of these constraints holds:

- for species involved in the combination, the collective rate of production has the same order of magnitude of the collective rate of consumption, so that $Z_\theta^{N,\gamma}$ neither explodes nor goes to zero;
- the rate at which reactions with a net effect on the combination happen has order of magnitude less or equal to the common normalization parameter for abundances α_θ.

Species balance condition can thus be easily extended to linear combination of species [13, sec. 3].

Condition 1.33. Let $\Gamma_\theta^+ \subseteq \mathcal{R}$ and $\Gamma_\theta^- \subseteq \mathcal{R}$ be the sets of reactions that result in an increase and decrease in the abundance of the linear combination of species $\theta \cdot S$, respectively:

$$\Gamma_\theta^+ := \{\, k\colon \theta^T \zeta_k > 0 \,\}, \qquad \Gamma_\theta^- := \{\, k\colon \theta^T \zeta_k < 0 \,\}.$$

The *balance condition* for linear combinations of species holds if at least one of these requirements is satisfied for each $\theta \in \mathbb{R}^n_{\geq 0}$:

1. *Balance equation*

$$\max_{k\in\Gamma_\theta^+}(\beta_k+\nu_k\cdot\alpha)=\max_{k\in\Gamma_\theta^-}(\beta_k+\nu_k\cdot\alpha),\qquad \theta\in\mathbb{R}^n_{\geq 0};\tag{1.64}$$

2. *Time-scale constraint*

$$\max_{k\in\Gamma_\theta^+\cup\Gamma_\theta^-}(\beta_k+\nu_k\cdot\alpha)+\gamma\leq\alpha_\theta,\qquad \theta\in\mathbb{R}^n_{\geq 0}.\tag{1.65}$$

Note that, as before, the time-scale constraint can be equivalently stated as

$$\gamma\leq\alpha_\theta-\max_{k\in\Gamma_\theta^+\cup\Gamma_\theta^-}(\beta_k+\nu_k\cdot\alpha),\qquad \theta\in\mathbb{R}^n_{\geq 0}.\tag{1.66}$$

Remark 1.34. Considering $\theta_i>0$ for a single species S_i, Condition 1.33 is equivalent to Condition 1.28, and therefore the former is more general and includes the latter. In particular, time-scale constraint (1.66) is completely equivalent to (1.52), while balance equation (1.64) is a strictly stronger requirement than (1.50).

Definition 1.35. The *natural time-scale* γ_θ for the normalized abundance of linear combination of species $\theta\cdot S$ is defined as

$$\gamma_\theta:=\alpha_\theta-\max_{k\in\Gamma_\theta^+\cup\Gamma_\theta^-}(\beta_k+\nu_k\cdot\alpha).\tag{1.67}$$

For $\gamma=\gamma_\theta$, none of the normalized reaction countings

$$N^{-\alpha_\theta}Y_k^{N,\gamma}(\tau),\qquad k\in\Gamma_\theta^+\cup\Gamma_\theta^-,$$

should blow up, and at least one should be nontrivial (i.e. should not uniformly converge to zero).

Note that, by Definition 1.29, the natural time-scale for the normalized abundance of a linear combination of species is larger than the natural time-scale for the normalized abundances of the single species involved in the combination:

$$\gamma_\theta\geq\min_{i:\,\theta_i>0}\gamma_i.\tag{1.68}$$

Moreover, as observed in Remark 1.34, time-scale constraints (1.66) and (1.52) are equivalent, and therefore, by Definition 1.30,

$$\gamma'=\min_{i\in\mathcal{S}}\gamma_i=\min_{\theta\in\mathbb{R}^n_{\geq 0}}\gamma_\theta,\tag{1.69}$$

where γ' is the (first) natural time-scale. Then, setting $\gamma=\gamma'$, Condition 1.33 holds for each $\theta\in\mathbb{R}^n_{\geq 0}$.

On the other hand, if Condition 1.33 holds for some time-scale γ strictly larger than the first natural time-scale γ', then necessarily balance equation (1.64) holds at least for all $\theta\in\mathbb{R}^n_{\geq 0}$ such that $\gamma_\theta=\gamma'$, because of (1.69).

In general, let $\hat{\gamma}$ be the largest time-scale for which Condition 1.33 holds:

$$\hat{\gamma}:=\sup\{\,\gamma\,:\ \text{Condition 1.33 holds}\,\}.\tag{1.70}$$

This parameter may take different values, depending on whether, and for which values of $\theta\in\mathbb{R}^n_{\geq 0}$, balance equation (1.64) holds:

- if $\hat{\gamma} = \gamma'$, balance equation never holds, or does not hold for some θ such that $\gamma_\theta = \gamma'$;
- if $\hat{\gamma} = \infty$, balance equation holds for each θ;
- if $\gamma' < \hat{\gamma} < \infty$, balance equation holds (at least) for all θ such that $\gamma_\theta < \hat{\gamma}$ and does not hold for some θ such that $\gamma_\theta = \hat{\gamma}$.

As a consequence, if $\hat{\gamma}$ is strictly larger than γ', there exists at least one $\theta \in \mathbb{R}^n_{\geq 0}$ whose natural time-scale γ_θ is larger than the natural time-scale γ', and therefore the reaction system has more than one natural time scale.

Definition 1.36. The *second natural time-scale* γ'' for a (normalized) reaction system is defined as

$$\gamma'' := \inf \{ \gamma_\theta \,:\, \gamma_\theta > \gamma' \}. \tag{1.71}$$

Remark 1.37. The *infimum* in the definition above is actually a *minimum*: indeed, the natural time-scale parameter γ_θ can take only a finite number of values, since the set of reactions $\mathcal{R}$ is finite. Therefore,

$$\gamma' < \gamma'' \leq \hat{\gamma},$$

and γ'' is actually a *second* time-scale.

Example 1.38. Consider the reaction network

$$\emptyset \to A \rightleftharpoons B, \tag{1.72}$$

and assume that

$$\beta_1 = \beta_2 > \beta_3.$$

Suppose both species are *balanced*, i.e. balance equation (1.50) holds for both species:

$$\text{A:} \quad \max(\beta_1, \beta_3 + \alpha_B) = \beta_2 + \alpha_A, \qquad \text{B:} \quad \beta_2 + \alpha_A = \beta_3 + \alpha_B.$$

For example, assume

$$\beta_1 = \beta_2 = \beta > 0, \qquad \beta_3 = 0, \qquad \alpha_A = 0, \qquad \alpha_B = \beta,$$

so that balance equation is satisfied for both A and B. With this choice of scaling parameters, the stochastic processes of normalized abundances satisfy equations

$$\begin{aligned} Z_A^{N,\gamma}(\tau) = Z_A^N(0) + \xi_1(N^{\gamma+\beta}\kappa_1\tau) - \xi_2\left(N^{\gamma+\beta}\int_0^\tau \kappa_2 Z_A^{N,\gamma}(s)ds\right) \\ + \xi_3\left(N^{\gamma+\beta}\int_0^\tau \kappa_3 Z_B^{N,\gamma}(s)ds\right), \end{aligned}$$

$$\begin{aligned} Z_B^{N,\gamma}(\tau) = Z_B^N(0) + N^{-\beta}\xi_2\left(N^{\gamma+\beta}\int_0^\tau \kappa_2 Z_A^{N,\gamma}(s)ds\right) \\ - N^{-\beta}\xi_3\left(N^{\gamma+\beta}\int_0^\tau \kappa_3 Z_B^{N,\gamma}(s)ds\right). \end{aligned}$$

Note that the natural time-scale for normalized abundance of species A is $\gamma_A = -\beta$, while that of species B is $\gamma_B = 0$, and so the first natural time scale for the normalized reaction system is

$$\gamma' = \min(\gamma_A, \gamma_B) = -\beta.$$

Considering linear combinations, for $\theta = (1, 1)$, the normalized abundance of $\theta \cdot S$ is given by

$$Z_\theta^{N,\gamma} := N^{-\beta} Z_A^{N,\gamma} + Z_B^{N,\gamma},$$

and stochastic process $Z_\theta^{N,\gamma}$ satisfies equation

$$Z_\theta^{N,\gamma}(\tau) = N^{-\beta} Z_A^N(0) + Z_B^N(0) + N^{-\beta}\xi_1(N^{\gamma+\beta}\kappa_1\tau). \tag{1.73}$$

However, for this choice of θ,

$$\Gamma_\theta^+ = \{\, 1 \,\}, \qquad \Gamma_\theta^- = \emptyset,$$

and therefore balance equation (1.64) fails. As a consequence, we should require time-scale contraint (1.66) to hold,

$$\gamma \le \alpha_\theta - \beta = 0, \qquad \text{with} \quad \alpha_\theta = \max(\alpha_A, \alpha_B) = \beta.$$

Note that the natural time-scale for the normalized abundance of linear combination $\theta \cdot S$ is $\gamma_\theta = 0$, which is also the second natural time-scale

$$\gamma'' = \inf\{\, \gamma_{\bar\theta} \colon \gamma_{\bar\theta} > \gamma' \,\} = \gamma_\theta = 0,$$

since $\gamma_{\bar\theta} = -\beta = \gamma'$ for all $\bar\theta \neq \theta$.

At the first natural time-scale $\gamma' = -\beta$, the limit process $Z^{\gamma'}$ is described by equations

$$\begin{aligned} Z_A^{\gamma'}(\tau) &= Z_A(0) + \xi_1(\kappa_1\tau) - \xi_2\left(\int_0^\tau \kappa_2 Z_A^{\gamma'}(s)ds\right) + \xi_3\left(\int_0^\tau \kappa_3 Z_B^{\gamma'}(s)ds\right) \\ &= Z_A(0) + \xi_1(\kappa_1\tau) - \xi_2\left(\int_0^\tau \kappa_2 Z_A^{\gamma'}(s)ds\right) + \xi_3\left(\kappa_3 Z_B(0)\tau\right), \end{aligned}$$

$$Z_B^{\gamma'}(\tau) = Z_B(0), \qquad \tau \ge 0.$$

At the second natural time-scale $\gamma'' = 0$, which is the natural time-scale for species B, we could divide the equation for $Z_A^{N,\gamma''}$ by N^β and take the limit for N growing large,

$$\begin{aligned} 0 &= \lim_{N\to\infty} N^{-\beta} Z_A^{N,\gamma''}(\tau) \\ &= \lim_{N\to\infty} N^{-\beta} Z_A^N(0) + N^{-\beta}\xi_1(N^\beta\kappa_1\tau) \\ &\quad - N^{-\beta}\xi_2\left(N^\beta \int_0^\tau \kappa_2 Z_A^{N,\gamma''}(s)ds\right) \\ &\quad + N^{-\beta}\xi_3\left(N^\beta \int_0^\tau \kappa_3 Z_B^{N,\gamma''}(s)ds\right) \\ &= \lim_{N\to\infty} \left(\kappa_1\tau - \int_0^\tau \kappa_2 Z_A^{N,\gamma''}(s)ds + \int_0^\tau \kappa_3 Z_B^{N,\gamma''}(s)ds\right), \end{aligned}$$

so that

$$\lim_{N\to\infty}\left(\int_0^\tau \kappa_2 Z_A^{N,\gamma''}(s)ds - \int_0^\tau \kappa_3 Z_B^{N,\gamma''}(s)ds\right) = \kappa_1\tau, \qquad \tau \geq 0.$$

Therefore, the stochastic process $Z_B^{N,\gamma''}$ converges in distribution to the limit process $Z_B^{\gamma''}$ given by

$$Z_B^{\gamma''}(\tau) = Z_B(0) + \kappa_1\tau, \qquad \tau \geq 0, \tag{1.74}$$

and so does the stochastic process $Z_\theta^{N,\gamma''}$, as follows directly from (1.73),

$$Z_\theta^{N,\gamma''} \Longrightarrow Z_\theta^{\gamma''} = Z_B^{\gamma''}. \tag{1.75}$$

Remark 1.39. In the previous Example 1.38, the largest time-scale for which balance condition for linear combinations holds is

$$\hat{\gamma} = \gamma_\theta = \gamma'',$$

and for time-scales γ strictly larger than $\hat{\gamma}$, normalized abundance for species B grows unbounded:

$$Z_B^{N,\gamma}(\tau) \to \infty, \qquad \forall\tau \geq 0, \quad \gamma > \hat{\gamma}.$$

This remark shows that Condition 1.28, and balance equation (1.50) in particular, is actually not sufficient to prevent normalized abundances from blowing up as N grows large, and time-scale constraint in Condition 1.33 becomes essential here.

1.7 Limit behaviour at second natural time-scale

Consider a reaction system with more than one natural time-scale, i.e. a system for which a second natural time-scale γ'' exists (e.g. see Example 1.38).

Let $\Gamma_\theta^\gamma \in \mathcal{R}$ be the set of reactions, resulting in a net effect on the linear combination $\theta \cdot S$, with natural time-scale γ with respect to θ:

$$\Gamma_\theta^\gamma := \{\, k\colon \alpha_\theta = \gamma + (\beta_k + \nu_k \cdot \alpha),\ \theta^T\zeta_k \neq 0 \,\}.$$

Note that, by Definition 1.35, $\Gamma_\theta^{\gamma_\theta}$ is the set of reactions, resulting in a net effect on the linear combination $\theta \cdot S$, with the fastest rates.

As a result, if $\alpha_\theta > 0$, it would be reasonable to expect

$$\lim_{N\to\infty} Z_\theta^{N,\gamma_\theta}(\tau) = \lim_{N\to\infty}\left(Z_\theta^N(0) + \sum_{k\in\Gamma_\theta^{\gamma_\theta}}\int_0^\tau \lambda_k'(Z^{N,\gamma_\theta}(s))ds\ \theta^T\zeta_k\right),$$

while, if $\alpha_\theta = 0$,

$$\lim_{N\to\infty} Z_\theta^{N,\gamma_\theta}(\tau) = \lim_{N\to\infty}\left(Z_\theta^N(0) + \sum_{k\in\Gamma_\theta^{\gamma_\theta}}\xi_k\left(\int_0^\tau \lambda_k'(Z^{N,\gamma_\theta}(s))ds\right)\theta^T\zeta_k\right).$$

This intuition is certainly true if $\gamma_\theta = \gamma'$, as it follows directly from Theorem 1.31.

Proposition 1.40. *Let $Z_\theta^{\gamma'}$ be the stochastic process defined, for the linear combination of species $\theta \cdot S$, as*

$$Z_\theta^{\gamma'} := \theta^T D^{\alpha_\theta} Z^{\gamma'}, \tag{1.76}$$

where D^{α_θ} is the diagonal matrix with diagonal entry i equal to 1 if species S_i has normalization parameter α_θ.

$$(D^{\alpha_\theta})_{ii} = \begin{cases} 1 & \text{if } \alpha_i = \alpha_\theta \\ 0 & \text{if } \alpha_i \neq \alpha_\theta \end{cases}, \qquad i = 1, \ldots, n.$$

Then, stochastic process $Z_\theta^{N,\gamma'}$ of normalized abundance for the linear combination $\theta \cdot S$ converges in distribution to stochastic process $Z_\theta^{\gamma'}$ on time interval $[0, \tau_\infty)$:

$$Z_\theta^{N,\gamma'} \Rightarrow Z_\theta^{\gamma'}, \qquad \text{on} \quad [0, \tau_\infty). \tag{1.77}$$

Since the reaction system is assumed to have more than one natural time-scale, by Definition 1.36, there exists a linear combination of species $\theta \cdot S$, $\theta \in \mathbb{R}^n_{\geq 0}$, whose natural time-scale γ_θ is equal to γ''.

For this particular choice of θ, it follows from Definition 1.35 that all reactions with natural time-scale γ' with respect to θ have a null effect on the normalized abundance of the linear combination, which is equivalent to state that $\Gamma_\theta^{\gamma'} = \emptyset$. In this case, at the first natural time-scale γ' , the process $Z_\theta^{N,\gamma'}$ converges to the constant process

$$Z_\theta^{\gamma'}(\tau) = Z_\theta(0), \qquad \tau \geq 0.$$

However, at the second natural time-scale γ'', the natural one for the linear combination $\theta \cdot S$, we would expect a result similar to Proposition 1.40 to hold, so that process $Z_\theta^{N,\gamma''}$ converges to a non-degenerate model even at this faster time-scale:

$$Z_\theta^{N,\gamma''} \Rightarrow Z_\theta^{\gamma''}, \qquad \text{on} \quad [0, \tau_\infty). \tag{1.78}$$

In particular, it should be reasonable for $Z_\theta^{N,\gamma''}$ to converge to the stochastic process $Z_\theta^{\gamma''}$ defined as

$$Z_\theta^{\gamma''} := \theta^T D^{\alpha_\theta} Z^{\gamma''}. \tag{1.79}$$

Example 1.41. Consider again the reaction system described in Example 1.38:

$$\emptyset \to A \rightleftharpoons B.$$

For linear combination $\theta = (1, 1)$, the normalized abundance of $\theta \cdot S$ satisfies equation

$$Z_\theta^{N,\gamma}(\tau) = Z_\theta^N(0) + N^{-\beta} \xi_1 (N^{\gamma+\beta} \kappa_1 \tau), \qquad \tau \geq 0.$$

At the first natural time-scale $\gamma = \gamma' = -\beta$, the stochastic process $Z_\theta^{N,\gamma'}$ converges to the constant process

$$Z_\theta^{\gamma'}(\tau) = Z_\theta(0) = Z_B(0), \qquad \tau \geq 0,$$

while at the second natural time-scale $\gamma = \gamma'' = 0$, the stochastic process $Z_\theta^{N,\gamma''}$ converges in distribution to the limit process $Z_\theta^{\gamma''}$ given by

$$Z_\theta^{\gamma''}(\tau) = Z_\theta(0) + \kappa_1 \tau, \qquad \tau \geq 0.$$

Note that the limit process $Z_\theta^{\gamma''}$ is equal to the limit process $Z_B^{\gamma''}$ for species B at its natural time-scale: indeed, species B has normalization parameter $\alpha_B = \beta = \alpha_\theta$, and multiplying process $Z^{\gamma''}$ by vector $\theta^T D^{\alpha_\theta}$ boils down to considering only its component corresponding to species B.

Unfortunately, even if γ'' has a natural interpretation as the second time-scale for the reaction system, the sensible result stated in (1.78) does not hold in general. Indeed, convergence of normalized abundance of linear combinations to a non-degenerate process cannot be guaranteed a priori at the second natural time-scale, unless additional conditions are satisfied.

Moreover, note that the definition of $Z_\theta^{\gamma''}$ given in (1.79) does not even make sense, since limiting process $Z^{\gamma''}$ is actually not well-defined.

Remark 1.42. Convergence of process $Z_\theta^{N,\gamma''}$ to a non-degenerate limiting process may be easily verified in particular cases. Specifically, it follows directly from Theorem 1.31 if intensity functions

$$\lambda'_k(z), \qquad k \in \Gamma_\theta^{\gamma''},$$

take particular forms:

- $\lambda'_k(z)$ does not depend on z, i.e. is a constant,

$$\lambda'_k(Z^{N,\gamma''}) = \kappa_k;$$

- $\lambda'_k(z)$ depends on z only through z_θ,

$$\lambda'_k(Z^{N,\gamma''}) = \lambda'_k(Z_\theta^{N,\gamma''}).$$

In Example 1.38, reaction R_1 is the only reaction in $\Gamma_\theta^{\gamma''}$, and has intensity function

$$\lambda'_1(Z^{N,\gamma''}) = \kappa_1.$$

Therefore, convergence of process $Z_\theta^{N,\gamma''}$ is a consequence of Theorem 1.31.

To generalize the idea of convergence at the second natural time-scale, we need to define specific subspaces of $\mathbb{R}^n$.

Definition 1.43. Let $\mathbb{L}'$ be the linear subspace of $\mathbb{R}^n$ spanned by species with natural time-scale γ_i equal to the first natural time-scale γ':

$$\mathbb{L}' := \operatorname{span}\{\, e_i \,:\, \gamma_i = \gamma' \,\} \subseteq \mathbb{R}^n, \tag{1.80}$$

where $\{\, e_i \,\}_{i=1}^n$ is the canonical basis of $\mathbb{R}^n$, and let Π' be the projection operator onto $\mathbb{L}'$.

Moreover, let $\mathbb{L}''$ be the linear subspace of $\mathbb{R}^n$ spanned by linear combinations of species $\theta \cdot S$ such that, considering only species with natural time-scale γ_i equal to the first natural time-scale γ', the net effect on the linear combination is null for each reaction in $\mathcal{R}$:

$$\mathbb{L}'' := \text{span}\,\{\,\theta \in \mathbb{R}^n_{\geq 0} \,:\, \sum_i \theta_i \delta_i \,(\zeta_k)_i = 0,\ \forall k \in \mathcal{R}\,\} \subseteq \mathbb{R}^n, \tag{1.81}$$

where $\delta_i = 1$ if $\gamma_i = \gamma'$, and let Π'' be the projection operator onto $\mathbb{L}''$.

Intuitively, $\mathbb{L}''$ is the subspace of $\mathbb{R}^n$ on which reactions happen only at the second natural time-scale, i.e. the *slow subspace*, while at the first natural time-scale the projection onto $\mathbb{L}''$ remains in its initial state. On the other hand, $\mathbb{L}'$ may be considered as the *fast subspace*, where reactions happen at the first natural time-scale.

Note that:

- $\mathbb{L}''$ contains the subspace spanned by species with natural time-scale $\gamma_i > \gamma'$, but may be larger;
- since $\mathbb{L}' \cap \mathbb{L}''$ has not necessarily null dimension, Π' and Π'' are, in general, not orthogonal;
- $\mathbb{L}'$ and $\mathbb{L}''$ together cover $\mathbb{R}^n$ entirely, and therefore

$$(I - \Pi'')(x) \in \mathbb{L}', \qquad \forall x \in \mathbb{R}^n.$$

Let $\hat{Z}^{N,\gamma}$ be the projection of $Z^{N,\gamma}$ onto the *slow subspace* $\mathbb{L}''$:

$$\hat{Z}^{N,\gamma} := \Pi''(Z^{N,\gamma}) \ \in \mathbb{L}''. \tag{1.82}$$

In accordance with the intuitive result stated in (1.78), we should expect $\hat{Z}^{N,\gamma''}$ to converge to a non-degenerate model $\hat{Z}^{\gamma''}$ at the second natural time-scale:

$$\hat{Z}^{N,\gamma''} \Longrightarrow \hat{Z}^{\gamma''}, \qquad \text{on} \quad [0, \tau_\infty). \tag{1.83}$$

However, as already mentioned before, convergence is not guaranteed a priori.

Example 1.44. Consider the reaction network

$$\emptyset \to S_1, \qquad \emptyset \to S_2, \qquad S_1 + S_2 \to \emptyset,$$
$$\emptyset \to S_3, \qquad S_1 + S_3 \to \emptyset,$$

and assume that

$$\beta_1 = \beta_2 = \beta_3 = \beta, \qquad \beta_4 = \beta_5 = 0,$$

while $\alpha_i = 0$ for all species.

With this choice of parameters, balance equation (1.50) holds for each species, and the stochastic processes of species numbers satisfy equations

$$X_1^{N,\gamma}(\tau) = X_1(0) + \xi_1(N^{\gamma+\beta}\kappa_1\tau) - \xi_3\left(N^{\gamma+\beta}\int_0^\tau \kappa_3 X_1^{N,\gamma}(s)X_2^{N,\gamma}(s)ds\right)$$
$$- \xi_5\left(N^{\gamma}\int_0^\tau \kappa_5 X_1^{N,\gamma}(s)X_3^{N,\gamma}(s)ds\right),$$

$$X_2^{N,\gamma}(\tau) = X_2(0) + \xi_2(N^{\gamma+\beta}\kappa_2\tau) - \xi_3\left(N^{\gamma+\beta}\int_0^\tau \kappa_3 X_1^{N,\gamma}(s)X_2^{N,\gamma}(s)ds\right),$$
$$X_3^{N,\gamma}(\tau) = X_3(0) + \xi_4(N^{\gamma}\kappa_4\tau) - \xi_5\left(N^{\gamma}\int_0^\tau \kappa_5 X_1^{N,\gamma}(s)X_3^{N,\gamma}(s)ds\right).$$

The natural time-scale for species S_1 and S_2 is $\gamma = -\beta$, while natural time-scale for species S_3 is $\gamma = 0$:

$$\gamma_1 = \gamma_2 = -\beta, \qquad \gamma_3 = 0.$$

Therefore, the first and second natural time-scale for the reaction system are

$$\gamma' = \min_i \gamma_i = -\beta, \qquad \gamma'' = 0.$$

At the first natural time-scale $\gamma' = -\beta$, the limit process $X^{\gamma'}$ is described by equations

$$X_1^{\gamma'}(\tau) = X_1(0) + \xi_1(\kappa_1\tau) - \xi_3\left(\int_0^\tau \kappa_3 X_1^{\gamma'}(s)X_2^{\gamma'}(s)ds\right),$$
$$X_2^{\gamma'}(\tau) = X_2(0) + \xi_2(\kappa_2\tau) - \xi_3\left(\int_0^\tau \kappa_3 X_1^{\gamma'}(s)X_2^{\gamma'}(s)ds\right),$$
$$X_3^{\gamma'}(\tau) = X_3(0), \qquad \tau \geq 0.$$

At the second natural time-scale $\gamma'' = 0$, which is the natural time-scale for species S_3,

$$\lim_{N\to\infty} X_3^{N,\gamma''}(\tau) = \lim_{N\to\infty} X_3(0) + \xi_4(\kappa_4\tau) - \xi_5\left(\int_0^\tau \kappa_5 X_1^{N,\gamma''}(s)X_3^{N,\gamma''}(s)ds\right).$$

However, if $\kappa_1 > \kappa_2$, it can be proved that

$$X_1^{N,\gamma''}(\tau) \to \infty, \qquad \forall\tau \geq 0,$$

and therefore $X_3^{N,\gamma''}(\tau)$ does not converge to a non-degenerate process.

Note that, in this example, subspaces $\mathbb{L}'$ and $\mathbb{L}''$ are given by

$$\mathbb{L}' = \text{span}\{e_1, e_2\}, \qquad \mathbb{L}'' = \text{span}\{e_3\}.$$

The key issue preventing results similar to Theorem 1.31 to hold for the projected process $\hat{Z}^{N,\gamma''}$ is that the equations describing this process do not depend only on the process itself.

Let $W^{N,\gamma}$ be the projection of $Z^{N,\gamma}$ onto the subspace orthogonal to $\mathbb{L}''$,

$$W^{N,\gamma} := (I - \Pi'')(Z^{N,\gamma}) \ \in (\mathbb{L}'')^\perp \subseteq \mathbb{L}', \tag{1.84}$$

so that $Z^{N,\gamma}$ can be decomposed into the sum of two othogonal processes

$$Z^{N,\gamma} = \hat{Z}^{N,\gamma} + W^{N,\gamma}, \qquad \hat{Z}^{N,\gamma} \in \mathbb{L}'', \quad W^{N,\gamma} \in (\mathbb{L}'')^\perp. \tag{1.85}$$

Then, in general, intensity functions λ_k' depend on both processes,

$$\lambda_k'(Z^{N,\gamma}) = \lambda_k'(\hat{Z}^{N,\gamma} + W^{N,\gamma}), \qquad k = 1, \ldots, l,$$

even if we consider only those reactions actually affecting the dynamics of the projection.

As a consequence, the equation describing $\hat{Z}^{N,\gamma''}$ is not self-closed, and a description of $W^{N,\gamma''}$ should be provided in order to properly determine the existence and the form of the limit process.

Remark 1.45. In Example 1.44, the equation for $X_3^{N,\gamma''}$ also depends on $X_1^{N,\gamma''}$:

$$X_3^{N,\gamma''}(\tau) = X_3(0) + \xi_4(\kappa_4\tau) - \xi_5\left(\int_0^\tau \kappa_5 X_1^{N,\gamma}(s)X_3^{N,\gamma}(s)ds\right).$$

As a matter of fact, the reason preventing $X_3^{N,\gamma''}$ to converge is the explosion of $X_1^{N,\gamma''}$ as N grows large. Indeed, even though balance equations are satisfied for the *fast subnetwork*, given by

$$(S_1, S_2) = \Pi'(S),$$

this subnetwork is not *stable*, if $\kappa_1 \neq \kappa_2$.

To guarantee convergence on the slow subspace $\mathbb{L}''$ at the second natural time-scale, additional conditions should be satisfied to ensure that the system on the fast subspace $\mathbb{L}'$ is *stable*, so that the influence of fast components on the slow subspace can be averaged.

In fact, at second natural time-scale, the system on the fast subspace fluctuates very rapidly, and does not converge in a functional sense. However, its behaviour may be partially captured by its so called *occupation measure*.

Definition 1.46. The *occupation measure* $V^{N,\gamma}$ for the projection of stochastic process $Z^{N,\gamma}$ on the fast subspace $\mathbb{L}'$ is the random measure on $\mathbb{L}' \times \mathbb{R}_{\geq 0}$, defined as

$$V^{N,\gamma}(C \times [0,\tau]) := \int_0^\tau 1_C(W^{N,\gamma}(s))\, ds, \qquad C \subseteq \mathbb{L}',\ \tau \geq 0, \tag{1.86}$$

where 1_C is the indicator function for subset $C \subseteq \mathbb{L}'$.

Remark 1.47. An intuitive idea of what the occupation measure represents may be useful to have a better understanding of the dynamics of the system at the second time-scale.

Consider the decomposition of process $Z^{N,\gamma}$ as defined in (1.85):

$$Z^{N,\gamma} = \hat{Z}^{N,\gamma} + W^{N,\gamma}.$$

At the first natural time-scale $\gamma = \gamma'$, the limit process $Z^{\gamma'}$ is well-defined by Theorem 1.31, and the limit process $W^{\gamma'}$ is trivially obtained from projection operation:

$$W^{\gamma'}(\tau) = (I - \Pi'')(Z^{\gamma'}(\tau)), \qquad \tau \geq 0.$$

When time is accelerated to the second time-scale $\gamma = \gamma''$, trajectories of process $W^{N,\gamma}$ are progressively compressed in smaller and smaller time intervals. What happens in the limit essentially depends on the stability of the system on its fast subspace:

- if the system is *unstable*, i.e. explodes in finite time or grows unbounded as $t \to \infty$, probability mass eventually escapes each compact $K \subseteq \mathbb{L}'$, for any time $\tau \geq 0$:

$$1_K(W^{N,\gamma''}(\tau)) \to 0, \qquad \text{a.s.}, \quad K \subseteq \mathbb{L}',\ \tau \geq 0;$$

- if the system is *stable*, the value of the indicator function 1_C, i.e. the probability of being in C, may be appropriately approximated via the stationary distribution π^N of process $W^{N,\gamma'}$, assuming it exists and is unique,

$$1_C(W^{N,\gamma''}(\tau)) \approx \pi^N(C), \qquad C \subseteq \mathbb{L}',\ \tau \geq 0,$$

recalling, in some sense, and ergodic property for process $W^{N,\gamma''}$.

Note that this observations are purely intuitive, and many fundamental details have been disregarded.

Under suitable conditions on convergence of $V^{N,\gamma}$ for $\gamma = \gamma''$ and on its properties with respect to the projection onto the slow subspace, it can be shown that $\hat{Z}^{N,\gamma''}$ converges in distribution to a stochastic process $\hat{Z}^{\gamma''}$ (for complete details and proofs, see [13, sec. 4]).

To simplify, we could say that, if

$$V^{N,\gamma''} \Rightarrow V, \qquad \text{on} \quad \mathbb{L}' \times \mathbb{R}_{\geq 0}, \tag{1.87}$$

then, disregarding minor details,

$$\hat{Z}^{N,\gamma''} = \Pi''(Z^{N,\gamma''}) \ \Rightarrow \ \hat{Z}^{\gamma''}, \qquad \text{on} \quad [0, \tau_\infty), \tag{1.88}$$

where

$$\begin{aligned}\hat{Z}^{\gamma''}(\tau) :=& \Pi'' Z(0) + \sum_{k:\ \rho_k > 0} \int_0^\tau \int_{\mathbb{L}'} \lambda_k'(\hat{Z}^{\gamma''}(s) + w) V(dw \times ds) D^{\rho_k} \zeta_k \\ &+ \sum_{k:\ \rho_k = 0} \xi_k \left(\int_0^\tau \int_{\mathbb{L}'} \lambda_k'(\hat{Z}^{\gamma''}(s) + w) V(dw \times ds) \right) D^{\rho_k} \zeta_k,\end{aligned}$$

and $\rho_k = \gamma'' + \beta_k + \nu_k \cdot \alpha$.

Remark 1.48. The statement above is actually not completely precise: indeed, we are assuming that $V^{N,\gamma''}$ converges to a measure V, and then, given V, the limit $\hat{Z}^{\gamma''}$ exists and is uniquely determined. However, in general, $V^{N,\gamma''}$ depends on $\hat{Z}^{N,\gamma''}$, and therefore convergences of $V^{N,\gamma''}$ and $\hat{Z}^{N,\gamma''}$ should be considered together, i.e. we should prove that the pair converges to the corresponding limit process:

$$\{\, V^{N,\gamma''},\, \hat{Z}^{N,\gamma''} \,\} \ \Rightarrow \ \{\, V,\, \hat{Z}^{\gamma''} \,\}. \tag{1.89}$$

Assuming convergence in (1.89) to hold, it can be shown that, in the limit for N growing large, the occupation measure V takes the form expected from Remark 1.47:

$$V(dw \times ds) = \pi(\hat{Z}^{\gamma''}(s);\, dw)\, ds, \tag{1.90}$$

where $\pi(\hat{Z}^{\gamma''}(s);\cdot)$ is the (conditional) stationary distribution, if it exists and is unique, for process $W^{\gamma'}$ (at the first natural time-scale) given $\hat{Z}^{\gamma''}(s)$.

This specific form for the occupation measure allows to redefine intensity functions so that they depend only on the state $\hat{z} \in \mathbb{L}''$ of the projected process. In particular, for reaction R_k, the intensity function may be redefined as

$$\hat{\lambda}'_k(\hat{z}) := \int_{\mathbb{L}'} \lambda'_k(\hat{z}+w)\,\pi(\hat{z};\,dw), \qquad \hat{z} \in \mathbb{L}''. \tag{1.91}$$

Therefore, the limit process $\hat{Z}^{\gamma''}$ satifies equation

$$\begin{aligned}\hat{Z}^{\gamma''}(\tau) :=& \Pi'' Z(0) + \sum_{k:\ \rho_k>0} \int_0^\tau \hat{\lambda}'_k(\hat{Z}^{\gamma''}(s))ds D^{\rho_k}\zeta_k \\ &+ \sum_{k:\ \rho_k=0} \xi_k \left(\int_0^\tau \hat{\lambda}'_k(\hat{Z}^{\gamma''}(s))ds \right) D^{\rho_k}\zeta_k,\end{aligned}$$

where $\rho_k = \gamma'' + \beta_k + \nu_k \cdot \alpha$.

Further details on stochastic averaging methods and a rigorous discussion on the result stated above can be found in [13, sec. 5].

Example 1.49. Consider once more the reaction system described in Example 1.38:

$$\emptyset \to A \rightleftharpoons B.$$

For this system, subspaces $\mathbb{L}'$ and $\mathbb{L}''$ are given by

$$\mathbb{L}' = \operatorname{span}\{\, e_A \,\}, \qquad \mathbb{L}'' = \operatorname{span}\{\, e_B \,\},$$

so that projection operator Π' corresponds to consider species A, while projection operator Π'' corresponds to consider species B.

At the first natural time-scale $\gamma' = -\beta$, the process $Z_A^{N,\gamma'}$ is described by equation

$$Z_A^{N,\gamma'}(\tau) = Z_A(0) + \xi_1(\kappa_1 t) - \xi_2\left(\int_0^\tau \kappa_2 Z_A^{N,\gamma'}(s)ds\right) + \xi_3\left(\int_0^\tau \kappa_3 Z_B^{N,\gamma'}(s)ds\right).$$

Given a fixed value for the normalized abundance of species B, $Z_B^{N,\gamma'}(s) = z_B$, it can be rewritten as

$$Z_A^{N,\gamma'}(\tau) = Z_A(0) + \xi_1(\kappa_1\tau) - \xi_2\left(\int_0^\tau \kappa_2 Z_A^{N,\gamma'}(s)ds\right) + \xi_3(\kappa_3 z_B \tau).$$

This model is actually a birth-and-death process with constant birth rates and linear death rates, expressed respectively by

$$(\lambda'_1 + \lambda'_3)(w) = \kappa_1 + \kappa_3 z_B, \qquad \lambda'_2(w) = \kappa_2 w;$$

specifically, it is an $M/M/\infty$ queue model, whose stationary distribution is the Poisson distribution,

$$\pi(z_B;\cdot) \sim \text{Poisson}\left(\frac{\kappa_1 + \kappa_3 z_B}{\kappa_2}\right).$$

Therefore, the limit occupation measure for process $Z_A^{N,\gamma}$, as N grows large, is

$$V(dw \times ds) = \pi(z_B;\, dw)\, ds,$$

and, the intensity function for reaction R_2,

$$\lambda_2'(z) = \kappa_2 z_A, \qquad z_A \in \mathbb{L}' = \mathbb{N}_{\geq 0},$$

may be redefined in terms of z_B as

$$\begin{aligned}\hat{\lambda}_2'(z_B) &:= \int_{\mathbb{L}'} \lambda_2'(\hat{z}_B + w)\, \pi(z_B; dw) \\ &= \int_{\mathbb{N}_{\geq 0}} \kappa_2 w\, \pi(z_B; dw) = \kappa_1 + \kappa_3 z_B.\end{aligned}$$

As a consequence, at the second time-scale $\gamma'' = 0$, the process $Z_B^{N,\gamma''}$ is described by equation

$$\begin{aligned}Z_B^{N,\gamma''}(\tau) = Z_B^N(0) &+ N^{-\beta}\xi_2 \left(N^\beta \int_0^\tau \left(\kappa_1 + \kappa_3 Z_B^{N,\gamma''}(s) \right) ds \right) \\ &- N^{-\beta}\xi_3 \left(N^\beta \int_0^\tau \kappa_3 Z_B^{N,\gamma''}(s) ds \right),\end{aligned}$$

which converges to

$$\begin{aligned}Z_B^{\gamma''}(\tau) &= Z_B(0) + \int_0^\tau \left(\kappa_1 + \kappa_3 Z_B^{N,\gamma''}(s) \right) ds - \int_0^\tau \kappa_3 Z_B^{N,\gamma''}(s) ds, \\ &= Z_B(0) + \kappa_1 \tau, \qquad \tau \geq 0,\end{aligned}$$

as already stated in (1.74).

Chapter 2

Weak Convergence in Topological Spaces

2.1 Weak convergence in metric spaces

Sections 2.1, 2.2 and 2.3 are entirely inspired by the classic book of Billingsley, *Convergence of Probability Measures* [4]. Refer to this work for additional details and proofs of results stated in these sections.

Let $(\mathcal{X}, \rho)$ be a metric space and let $\mathcal{B}(\mathcal{X})$ be the Borel σ-field on $\mathcal{X}$, i.e. the σ-field generated by open sets (or by closed sets).

Definition 2.1. A *probability measure* μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is a set function

$$\mu \colon \mathcal{B}(\mathcal{X}) \mapsto [0, 1],$$

which satisfies

$$\mu(\emptyset) = 0, \qquad \mu(\mathcal{X}) = 1,$$

and is σ-additive (or countably additive), i.e. for all countable collections $\{ C_n \}_{n \in \mathbb{N}} \subseteq \mathcal{B}(\mathcal{X})$ of pairwise disjoint sets ($C_i \cap C_j = \emptyset$, if $i \neq j$),

$$\mu \left(\bigcup_{i \in \mathbb{N}} C_i \right) = \sum_{i \in \mathbb{N}} \mu(C_i).$$

Definition 2.2. A probability measure μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is said to be *tight* if, for each $\epsilon > 0$, there exists a compact set $K_\epsilon \subseteq \mathcal{X}$ such that

$$\mu(K_\epsilon) > 1 - \epsilon. \tag{2.1}$$

A metric space $(\mathcal{X}, \rho)$ which is separable and complete is called a *Polish space*.

Theorem 2.3. *If $(\mathcal{X}, \rho)$ is a Polish space, then each probability measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is tight.*

Probability measures on $\mathcal{X}$ are defined for all elements in the Borel σ-field $\mathcal{B}(\mathcal{X})$. However, to separate a probability measure μ from all the other probability measures on $\mathcal{X}$, it is sufficient, in general, to consider values of μ (regarded as a function) on a subset of its domain.

Definition 2.4. A subset $\mathcal{A} \subseteq \mathcal{B}(\mathcal{X})$ is called a *separating class* if two probability measures μ and ν that agree on $\mathcal{A}$ necessarily agree on the whole $\mathcal{B}(\mathcal{X})$, that is

$$\mu(A) = \nu(A), \quad \forall A \in \mathcal{A} \quad \text{implies} \quad \mu(A) = \nu(A), \quad \forall A \in \mathcal{B}(\mathcal{X}).$$

In particular, a π-system on $\mathcal{X}$, i.e. a collection of subsets of $\mathcal{X}$ closed under finite intersections, which generates the Borel σ-field $\mathcal{B}(\mathcal{X})$ is a separating class.

Definition 2.5. A sequence $\{\, \mu_n \,\}$ of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is said to *converge weakly* to a probability measure μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ if

$$\int_{\mathcal{X}} f(x)\mu_n(dx) \to \int_{\mathcal{X}} f(x)\mu(dx) \qquad \text{in} \quad \mathbb{R}, \tag{2.2}$$

for every bounded, continuous function $f\colon \mathcal{X} \mapsto \mathbb{R}$. Weak convergence is usually written as

$$\mu_n \Rightarrow \mu. \tag{2.3}$$

It can be proved that two probability measures μ_1 and μ_2 on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ coincide if

$$\int_{\mathcal{X}} f(x)\mu_1(dx) = \int_{\mathcal{X}} f(x)\mu_2(dx),$$

for every bounded, continuous function $f\colon \mathcal{X} \mapsto \mathbb{R}$. Therefore, if a sequence $\{\, \mu_n \,\}_{n\in\mathbb{N}}$ converges to a limit μ, this limit is unique.

The following theorem provides useful conditions equivalent to weak convergence, and each of them could serve as a definition.

Theorem 2.6 (Portmanteau Theorem)**.** *Let $\{\, \mu_n \,\}$ be a sequence of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. The following conditions are equivalent:*

- $\mu_n \Rightarrow \mu$*;*
- *for every bounded, uniformly continuous function $f\colon \mathcal{X} \mapsto \mathbb{R}$,*

$$\int_{\mathcal{X}} f(x)\mu_n(dx) \to \int_{\mathcal{X}} f(x)\mu(dx);$$

- *for every closed set $C \subseteq \mathcal{X}$,*

$$\limsup_n \mu_n(C) \leq \mu(C);$$

- *for every open set $A \subseteq \mathcal{X}$,*

$$\liminf_n \mu_n(A) \geq \mu(C);$$

- *for every μ-continuity set S, i.e. for every set $S \subseteq \mathcal{X}$ whose boundary ∂S satisfies $\mu(\partial S) = 0$,*

$$\mu_n(S) \to \mu(S).$$

The last condition is equivalent to weak convergence if convergence of values $\mu_n(S)$ to $\mu(S)$ holds for every μ-continuity set S. However, to prove weak convergence, it is actually sufficient to check convergence of values $\mu_n(S)$ for a subsets of μ-continuity sets.

Definition 2.7. A subset $\mathcal{A} \subseteq \mathcal{B}(\mathcal{X})$ is called a *convergence-determining class* if, for every sequence of probability measures $\{\,\mu_n\,\}$ and every probability measure μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, convergence of values $\mu_n(S)$ to $\mu(S)$ for every μ-continuity set $S \in \mathcal{A}$ implies weak convergence of μ_n to μ, that is

$$\mu_n(S) \to \mu(S), \quad \forall S \in \mathcal{A},\ S\,\mu\text{-continuity set} \qquad \text{implies} \qquad \mu_n \Rightarrow \mu.$$

Note that a convergence-determining class is obviously a separating class, since the limit measure is unique. However, the converse is not necessarily true.

A further condition equivalent to weak convergence involves subsequences.

Theorem 2.8. *A necessary and sufficient condition for $\mu_n \Rightarrow \mu$ is that each subsequence $\{\,\mu_m\,\} \subseteq \{\,\mu_n\,\}$ contains a further subsequence $\{\,\mu_{m(k)}\,\}$ which converges weakly to μ.*

Let h be a function from $(\mathcal{X}, \rho)$ to another metric space $(\mathcal{Y}, \delta)$,

$$h\colon (\mathcal{X}, \rho) \mapsto (\mathcal{Y}, \delta).$$

If h is measurable, that is, for every set A in $\mathcal{B}(\mathcal{Y})$ (the Borel σ-field on $\mathcal{Y}$),

$$h^{-1}(A) := \{\, x \in \mathcal{X}\colon h(x) \in A \,\} \in \mathcal{B}(\mathcal{X}), \tag{2.4}$$

then each probability measure μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ induces through h a probability measure ν on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, usually called *pushforward measure*, defined as

$$\nu(A) := (\mu \circ h^{-1})(A), \qquad \forall A \in \mathcal{B}(\mathcal{Y}). \tag{2.5}$$

We would like to have conditions under which

$$\mu_n \Rightarrow \mu \qquad \text{implies} \qquad \nu_n \Rightarrow \nu.$$

A simple and sufficient condition is that h is a continuous function. Indeed, for every bounded, continuous function $f\colon \mathcal{Y} \mapsto \mathbb{R}$, the function $f \circ h\colon \mathcal{X} \mapsto \mathbb{R}$ is bounded and continuous, and if we assume $\mu_n \Rightarrow \mu$, then

$$\int_{\mathcal{Y}} f(y)\nu_n(dy) = \int_{\mathcal{X}} f(h(x))\mu_n(dx) \to \int_{\mathcal{X}} f(h(x))\mu(dx) = \int_{\mathcal{Y}} f(y)\nu(dy).$$

However, continuity assumption is not necessary and can be weakened.

Theorem 2.9 (Mapping theorem)**.** *Let $h\colon \mathcal{X} \mapsto \mathcal{Y}$ be a measurable function, and let $D_h \subseteq \mathcal{X}$ be the set of its discontinuity points. If $\mu_n \Rightarrow \mu$ and $\mu(D_h) = 0$, then $\nu_n \Rightarrow \nu$.*

Definition 2.10. A family M of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is said to be *relatively (sequentially) compact* if, for every sequence $\{\,\mu_n\,\}$ of element of M, there exist a subsequence $\{\,\mu_{n(k)}\,\}$ and a probability measure μ such that $\mu_{n(k)} \Rightarrow_k \mu$.

Note that μ is defined on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ but is not necessarily an element of M.

Remark 2.11. Relative compactness is a key property to be verified for sequences of probability measures. Indeed, if a sequence $\{ \mu_n \}$ is relatively compact, each subsequence $\{ \mu_m \} \subseteq \{ \mu_n \}$ contains a further subsequence $\{ \mu_{m(k)} \}$ which converges weakly to some probability measure (say ν); if all these limit measures ν coincide with a measure μ, it follows by Theorem 2.8 that the entire sequence $\{ \mu_n \}$ converges weakly to μ

Definition 2.12. A family M of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is said to be *uniformly tight* if, for each $\epsilon > 0$, there exists a compact set $K_\epsilon \subseteq \mathcal{X}$ such that

$$\mu(K_\epsilon) > 1 - \epsilon, \qquad \text{for all} \quad \mu \in M. \tag{2.6}$$

Theorem 2.13 (Prohorov theorem)**.** *Let M be a family of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.*

- ***Direct Prohorov theorem**. If M is uniformly tight, then it is relatively (sequentially) compact.*
- ***Converse Prohorov theorem**. Assume that $(\mathcal{X}, \rho)$ is a Polish space. If M is relatively (sequentially) compact, then it is uniformly tight.*

Note that the converse theorem contains Theorem 2.3, since an M consisting of a single measure is obviously relatively compact. On the other hand, the direct theorem is essential for applications: it is usually easier to prove uniform tightness than relative compactness.

The theory of weak convergence can be equivalently restated as the theory of convergence *in distribution* (or convergence *in law*).

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ be a metric space, equipped with its Borel σ-field.

Definition 2.14. A measurable function X from $(\Omega, \mathcal{F}, P)$ to $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$,

$$X \colon (\Omega, \mathcal{F}, P) \mapsto (\mathcal{X}, \mathcal{B}(\mathcal{X})),$$

is called a *random element* of $\mathcal{X}$, or an *$\mathcal{X}$-valued random element.*

The terminology *random element* is used in general, but X may take specific names depending on space $\mathcal{X}$ (e.g. random variable, random vector, random function, stochastic process).

Definition 2.15. The *distribution* of a random element X is the pushforward measure P_X induced through X on space $\mathcal{X}$ by probability measure P:

$$P_X(A) := P(X^{-1}(A)) = P(\omega \in \Omega \colon X(\omega) \in A) = P(X \in A), \qquad A \in \mathcal{B}(\mathcal{X}).$$

The distribution of X is also called the *law* of X and denoted by $\mathcal{L}(X)$.

Note that measure P_X is defined on the Borel σ-field $\mathcal{B}(\mathcal{X})$ of metric space $\mathcal{X}$ and contains the essential information about random element X.

Definition 2.16. A sequence $\{ X_n \}$ of random elements of $\mathcal{X}$ is said to *converge in distribution* to a random element X of $\mathcal{X}$ if

$$P_{X_n} \Rightarrow P_X, \quad \text{or, equivalently,} \quad \mathcal{L}(X_n) \Rightarrow \mathcal{L}(X), \tag{2.7}$$

that is

$$\int_{\mathcal{X}} f(x) P_{X_n}(dx) \to \int_{\mathcal{X}} f(x) P_X(dx) \qquad \text{in} \quad \mathbb{R}, \tag{2.8}$$

for every bounded, continuous function $f\colon \mathcal{X} \mapsto \mathbb{R}$, and is written as $X_n \Rightarrow X$.

Equation (2.8) may be restated, in terms of the usual notation adopted in a probabilistic setting, as

$$E\,[f(X_n)] \to E\,[f(X)] \qquad \text{in} \quad \mathbb{R},$$

for every bounded, continuous function $f\colon \mathcal{X} \mapsto \mathbb{R}$

Remark 2.17. A key point is that, for convergence in distribution to hold, random elements $X_1, X_2, \ldots, X$ need not be defined on the same probability space $(\Omega, \mathcal{F}, P)$, that is, their domains may all be distinct. Actually, these spaces are not even considered, as their structure enters the argument only by the probability measures they induce on space $\mathcal{X}$ through functions $X_1, X_2, \ldots, X$.

As a consequence, all definition, properties and theorems stated for probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ may be paraphrased in terms of random elements of $\mathcal{X}$, without any difference.

2.2 Space $C[0,1]$ and uniform metric

Let $C[0,1] = C([0,1]\colon \mathbb{R})$ be space of (uniformly) continuous functions on interval $[0,1]$ with values in $\mathbb{R}$.

The space $C[0,1]$ is naturally equipped with the *uniform norm*,

$$\|x\| := \sup_{t \in [0,1]} |x(t)|, \qquad x \in C[0,1], \tag{2.9}$$

which induces the *uniform metric*

$$\rho(x,y) := \|x - y\| = \sup_{t \in [0,1]} |x(t) - y(t)|, \qquad x, y \in C[0,1]. \tag{2.10}$$

It can be proved that the space $C[0,1]$ with the uniform metric ρ is separable and complete, and so is a Polish space. Therefore, by Theorem 2.3, each probability measure on the Borel σ-field $\mathcal{B}(C[0,1])$ is tight.

A key feature of space $(C[0,1], \rho)$ is that uniform convergence implies pointwise convergence,

$$\rho(x_n, x) \to 0 \qquad \text{implies} \qquad x_n(t) \to x(t), \quad \forall t \in [0,1], \tag{2.11}$$

but the converse is false, as shown by the following example.

Example 2.18. Consider the sequence of functions $\{ x_n \}$, where x_n is the function defined as

$$x_n(t) = \begin{cases} nt & \text{if} \quad 0 \le t < n^{-1} \\ 2 - nt & \text{if} \quad n^{-1} \le t < 2n^{-1} \\ 0 & \text{if} \quad 2n^{-1} \le t \le 1 \end{cases}.$$

The sequence $\{ x_n \}$ converges pointwise to 0, i.e. $x_n(t) \to 0$, for each $t \in [0,1]$, but $\rho(x_n, 0) = 1$ for each n, and therefore $z_n \not\to 0$.

For each $t \in [0,1]$, the *natural projection* of x onto $\mathbb{R}$ is the functional π_t which evaluates function x at t:

$$\pi_t \colon C[0,1] \mapsto \mathbb{R}, \qquad \pi_t(x) = x(t), \quad x \in C[0,1]. \tag{2.12}$$

Similarly, for each finite set $\{ t_1, \ldots, t_k \} \subset [0,1]$, the natural projection of x onto $\mathbb{R}^k$ is the functional $\pi_{t_1,\ldots,t_k}$ which evaluates function x at $t_1, \ldots, t_k$:

$$\pi_{t_1,\ldots,t_k} \colon C[0,1] \mapsto \mathbb{R}^k, \qquad \pi_{t_1,\ldots,t_k}(x) = (x(t_1), \ldots, x(t_k)), \quad x \in C[0,1].$$

Since x is a continuous function, natural projection $\pi_t(x) = x(t)$ is well-defined, for each $x \in C[0,1]$ and for each $t \in [0,1]$. Moreover, with the uniform metric ρ, natural projections are continuous functions.

Because of the continuity of natural projections, mapping theorem applies.

Proposition 2.19. *Let $\{ \mu_n \}$ be a sequence of probability measures on $C[0,1]$ which converges to a probability measure μ. The corresponding sequence of pushforward measures $\{ \mu_n \circ (\pi_{t_1,\ldots,t_k})^{-1} \}$ on $\mathbb{R}^k$ converges to measure $\mu \circ (\pi_{t_1,\ldots,t_k})^{-1}$, for every choice of $t_1, \ldots, t_k \in [0,1]$:*

$$\mu_n \Rightarrow \mu \qquad \textit{implies} \qquad \mu_n \circ (\pi_{t_1,\ldots,t_k})^{-1} \;\Rightarrow\; \mu \circ (\pi_{t_1,\ldots,t_k})^{-1}. \tag{2.13}$$

Probability measure $\mu \circ (\pi_{t_1,\ldots,t_k})^{-1}$ on $\mathbb{R}^k$ is usually referred to as the *finite-dimensional distribution* of μ.

Definition 2.20. Let $\pi_{t_1,\ldots,t_k}$ be the natural projection onto $\mathbb{R}^k$. A subset $F \subseteq C[0,1]$ is a *finite-dimensional set* if it is the pre-image of a subset $H \in \mathcal{B}(\mathbb{R}^k)$ of $\mathbb{R}^k$,

$$\exists\, H \in \mathcal{B}(\mathbb{R}^k) \quad \text{such that} \quad F = (\pi_{t_1,\ldots,t_k})^{-1}(H). \tag{2.14}$$

Let $\mathcal{B}(C[0,1])_f$ be the class of finite-dimensional sets.

Note that finite-dimensional sets are elements of $\mathcal{B}(C[0,1])$, since natural projections are continuous functions.

It can be easily shown that class $\mathcal{B}(C[0,1])_f$ is a π-system on $C[0,1]$ which generates the Borel σ-field $\mathcal{B}(C[0,1])$, and so is a separating class.

Proposition 2.21. *Let μ and ν be probability measures on $C[0,1]$. If finite-dimensional distributions of μ and ν agree for every choice of $t_1, \ldots, t_k$,*

$$(\mu \circ (\pi_{t_1,\ldots,t_k})^{-1})(H) = (\nu \circ (\pi_{t_1,\ldots,t_k})^{-1})(H), \qquad \forall\, H \in \mathcal{B}(\mathbb{R}^k),$$

then μ coincides with ν.

Unfortunately, $\mathcal{B}(C[0,1])_f$ is not a convergence determining class, as shown by the following example.

Example 2.22. Let $\{ z_n \}$ be the sequence of functions introduced in Example 2.18 and let $\mu_n = \delta_{z_n}$, $\mu = \delta_0$ be the Dirac measures centered on functions z_n and 0, respectively. Since we showed that $z_n \not\to 0$, then obviously $\mu_n \not\Rightarrow \mu$.

However, for every choice of $t_1, \ldots, t_k$, there exists n_0 such that

$$\pi_{t_1,\ldots,t_k}(z_n) = \pi_{t_1,\ldots,t_k}(0) = (0,\ldots,0), \qquad \forall\, n \geq n_0,$$

and therefore

$$(\mu_n \circ (\pi_{t_1,\ldots,t_k})^{-1})(H) \to (\mu \circ (\pi_{t_1,\ldots,t_k})^{-1})(H), \qquad \forall\, H \in \mathcal{B}(\mathbb{R}^k).$$

As a result,

$$\mu_n(A) \to \mu(A), \qquad \forall\, A \in \mathcal{B}(C[0,1])_f,$$

(including those that are not μ-continuity sets), even if $\mu_n \not\Rightarrow \mu$, and so $\mathcal{B}(C[0,1])_f$ is not a convergence-determining class.

This example clearly shows that in $C[0,1]$ results about weak convergence go far beyond finite-dimensional theory.

Remark 2.23. To summarize, in $C[0,1]$, weak convergence implies weak convergence of finite-dimensional distributions

$$\mu_n \Rightarrow \mu \qquad \text{implies} \qquad \mu_n \circ (\pi_{t_1,\ldots,t_k})^{-1} \;\Rightarrow\; \mu \circ (\pi_{t_1,\ldots,t_k})^{-1}, \tag{2.15}$$

but the converse is *not* true,

$$\mu_n \circ (\pi_{t_1,\ldots,t_k})^{-1} \;\Rightarrow\; \mu \circ (\pi_{t_1,\ldots,t_k})^{-1} \qquad \cancel{\text{implies}} \qquad \mu_n \Rightarrow \mu. \tag{2.16}$$

Although $\mathcal{B}(C[0,1])_f$ is not a convergence-determining class, the fact that it is a separating class turns out to be useful to show that implication (2.16) actually holds under the assumption of relative compactness of sequence $\{ \mu_n \}$ (see Remark 2.11).

Theorem 2.24. *Let $\{ \mu_n \}$ be a relatively compact sequence of probability measures on $C[0,1]$. If finite-dimensional distributions of $\{ \mu_n \}$ converge to those of a measure μ, then $\mu_n \Rightarrow \mu$:*

$$\mu_n \circ (\pi_{t_1,\ldots,t_k})^{-1} \;\Rightarrow\; \mu \circ (\pi_{t_1,\ldots,t_k})^{-1} \qquad \textit{implies} \qquad \mu_n \Rightarrow \mu. \tag{2.17}$$

Proof. By relative compactness, each subsequence $\{ \mu_m \} \subseteq \{ \mu_n \}$ contains a further subsequence $\{ \mu_{m(k)} \}$ which converges to some probability measure ν, and, by the mapping theorem,

$$\mu_{m(k)} \circ (\pi_{t_1,\ldots,t_k})^{-1} \;\Rightarrow\; \nu \circ (\pi_{t_1,\ldots,t_k})^{-1}.$$

Given that if a sequence converges to a limit, all its subsequences converge to the same limit, since $\{ \mu_{m(k)} \circ (\pi_{t_1,\ldots,t_k})^{-1} \} \subseteq \{ \mu_n \circ (\pi_{t_1,\ldots,t_k})^{-1} \}$, then necessarily

$$\mu \circ (\pi_{t_1,\ldots,t_k})^{-1} = \nu \circ (\pi_{t_1,\ldots,t_k})^{-1}.$$

Therefore, finite-dimensional distributions of μ and ν are identical, and since the class $\mathcal{B}(C[0,1])_f$ is a separating class, it is enough to conclude that $\mu = \nu$ (see Proposition 2.21). As a result, each subsequence contains a further subsequence converging weakly to the same probability measure μ, and it follows by Theorem 2.8 that the entire sequence $\{\,\mu_n\,\}$ converges weakly to μ. □

Theorem 2.25. *Let $\{\,\mu_n\,\}$ be a relatively compact sequence of probability measures on $C[0,1]$. If finite-dimensional distributions of $\{\,\mu_n\,\}$ converge to some probability measure on $\mathbb{R}^k$, for every choice of $t_1, \ldots, t_k \in [0,1]$,*

$$\mu_n \circ (\pi_{t_1,\ldots,t_k})^{-1} \;\Rightarrow\; \mu_{t_1,\ldots,t_k},$$

then there exists a unique probability measure μ having $\{\,\mu_{t_1,\ldots,t_k}\,\}$ as finite-dimensional distributions.

Proof. By relative compactness, some subsequence $\{\,\mu_m\,\} \subseteq \{\,\mu_n\,\}$ converges weakly to some limit μ, and, following an argument similar to proof of Theorem 2.24, we can conclude that there exists a probability measure μ having $\{\,\mu_{t_1,\ldots,t_k}\,\}$ as finite-dimensional distributions. Uniqueness follows from Proposition 2.21 ($\mathcal{B}(C[0,1])_f$ is a separating class). □

Note that, in this last theorem, we do not assume a priori that $\mu_{t_1,\ldots,t_k}$ are finite-dimensional distributions of a certain probability measure μ on $C[0,1]$, and relative compactness is used to show that such a measure exists.

Remark 2.26. The arguments developed to prove Theorem 2.24 (and Theorem 2.25) rely on three elements:

1. a concept of relative compactness;
2. a converse proposition, here established via the mapping theorem (Proposition 2.19);
3. a uniqueness proposition, here following by the fact that $\mathcal{B}(C[0,1])_f$ is a separating class.

This framework is more general, and can be applied to a large class of metric spaces (see Section 2.3).

In order to effectively use Theorem 2.24 to prove weak convergence in $C[0,1]$, it is necessary to investigate the concept of compactness in this space.

Definition 2.27. Let x be an arbitrary function (not necessarily continuous) on interval $[0,1]$ with values in $\mathbb{R}$. The *modulus of continuity* w_x of function x is defined as

$$w_x(\delta) = w(x,\delta) = \sup_{|s-t|\le\delta} |x(s) - x(t)|, \qquad \delta \in (0,1]. \tag{2.18}$$

In words, $w_x(\delta)$ is the largest distance between function values $x(s)$ and $x(t)$ taken at points s and t whose distance does not exceed δ.

It is easy to show that a necessary and sufficient condition for x to be in $C[0,1]$ is

$$\lim_{\delta \to 0^+} w_x(\delta) = 0. \tag{2.19}$$

Relative compactness in $C[0,1]$ is completely characterized by the following theorem.

Theorem 2.28 (Arzelà-Ascoli)**.** *A set K of elements of $C[0,1]$ is relatively compact if and only if two conditions hold:*

1. *uniform boundedness,*

$$\sup_{x \in K} |x(0)| < \infty; \tag{2.20}$$

2. *uniform equicontinuity,*

$$\lim_{\delta \to 0^+} \sup_{x \in K} w_x(\delta) = 0. \tag{2.21}$$

Example 2.29. Let $K = \{ z_n \}$ consist of the sequence of functions introduced in Example 2.18. K is not relatively compact: in particular, it is uniformly bounded, because

$$\sup_n |z_n(0)| = \sup_n 0 = 0 < \infty,$$

but not uniformly continuous, since

$$\sup_n w(z_n, \delta) = 1, \qquad n \geq \delta^{-1}.$$

In fact, uniform convergence does not hold, even if $\{ z_n \}$ converges pointwise to 0-function.

The characterization of relative compactness of a sequence of probability measures on $C[0,1]$ is obtained by simply translating the Arzelà-Ascoli characterization.

Theorem 2.30. *A sequence $\{ \mu_n \}$ of probability measures on $C[0,1]$ is relatively compact if and only if two conditions hold:*

1. *for each $\eta > 0$, there exist L and n_0 such that,*

$$\mu_n \left(x \in C[0,1] \colon |x(0)| \geq L \right) \leq \eta, \qquad n \geq n_0; \tag{2.22}$$

2. *for each $\epsilon > 0$ and $\eta > 0$, there exist $\delta \in (0,1]$ and n_0 such that,*

$$\mu_n \left(x \in C[0,1] \colon w_x(\delta) \geq \epsilon \right) \leq \eta, \qquad n \geq n_0. \tag{2.23}$$

Remark 2.31. The space $C[0,1]$ is a Polish space, and both direct and converse Prohorov theorems (Theorem 2.13) apply. Therefore the concepts of relative compactness and uniform tightness for families of probability measures coincide on $C[0,1]$, and conditions stated in Theorem 2.30 for relative compactness are also conditions for uniform tightness.

The space $C[0,1]$ is the natural space for the description of stochastic processes with continuous trajectories (e.g. Brownian processes).

Let X be a random element (function) of $C[0,1]$, and, for a fixed t, let $X(t) = \pi_t \circ X$ be the random element (variable) on $\mathbb{R}$ defined as the composition of X with the natural projection π_t.

Theorem 2.32. *Let $\{ X_n \}$ be a sequence of random elements of $C[0,1]$. If finite-dimensional distributions of $\{ X_n \}$ converge to those of a random variable element X, i.e. for each choice of $t_1, \ldots, t_k$,*

$$(X_n(t_1), \ldots, X_n(t_k)) \Rightarrow_n (X(t_1), \ldots, X(t_k)), \tag{2.24}$$

and, for each $\epsilon > 0$,

$$\lim_{\delta \to 0^+} \limsup_n P\left(w(X_n, \delta) \geq \epsilon\right) = 0, \tag{2.25}$$

then $X_n \Rightarrow X$.

2.3 Space $D[0,1]$ and Skorohod metric

The space $C[0,1]$ is unsuitable for the description of stochastic processes that contain jumps (e.g. Poisson processes). Therefore, it is useful to enlarge this space to include also some kinds of discontinuous functions.

Let $D[0,1] = D([0,1]\colon \mathbb{R})$ be the space of functions on interval $[0,1]$ with values in $\mathbb{R}$ that are right-continuous and have left limits, that is

1. for $t \in [0,1)$, $x(t+) = \lim_{s \downarrow t^+} x(s)$ exists and $x(t+) = x(t)$;
2. for $t \in (0,1]$, $x(t-) = \lim_{s \uparrow t^-} x(s)$ exists.

Elements of space $D[0,1]$ are usually called *cadlag* functions (from French "continue à droite, limite à gauche") and space $D[0,1]$ is referred to as *Skorohod space*.

Note that the Skorohod space essentially contains functions x with discontinuities of the first kind, where the function value at each discontinuity point t is assumed to be $x(t) = x(t+)$ for convenience. Obviously, continuous functions are cadlag functions, and therefore $C[0,1]$ is a subset of $D[0,1]$.

Remark 2.33. As an extension of $C[0,1]$, the Skorohod space may be equipped with the uniform metric ρ defined in (2.10). However, this metric proves to be unsatisfactory for the purpose of studying weak convergence on this enlarged space, for a number of technical and practical reasons: among them, $D[0,1]$ with the uniform metric is not separable, a fact that may be disadvantageous in probability theory.

Let x and y be cadlag functions, and consider their domain $[0,1]$ to be a time interval. Intuitively, in the uniform metric, x and y are near one another if the graph of x can be superposed onto the graph of y by a uniformly small perturbation of functions values, while evaluation points (time instants) are kept fixed.

In presence of jumps, for x and y to be near one another, it is necessary that points of discontinuity of the two functions coincide: this requirement may be weakened, and uniformly small perturbations of evalutation points may be allowed. From a physical point of view, this amounts to recognise that it is not possible to measure time exactly, and therefore x and y should be considered close to each other also if their graphs can be superposed by a uniformly small deformation of times scale.

This idea, originally introduced by Skorohod [16], leads to the construction of a topology, known as *Skorohod J_1-topology*, on space $D[0,1]$.

Let Λ be the class of strictly increasing and continuous mappings of $[0,1]$ onto itself, that is $\lambda\colon [0,1] \mapsto [0,1]$ is a function in Λ if it is strictly increasing, continuous, $\lambda(0) = 0$ and $\lambda(1) = 1$.

Definition 2.34. A sequence $\{ x_n \}$ of cadlag functions is called *J_1-convergent* to function x if there exists a sequence $\{ \lambda_n \}$ of mappings in Λ such that

$$\sup_{t\in[0,1]} |\lambda_n(t) - t| \to 0 \qquad \text{and} \qquad \sup_{t\in[0,1]} |x_n(\lambda_n(t)) - x(t)| \to 0. \tag{2.26}$$

Note that conditions in (2.26) may be rewritten with respect to the uniform metric ρ as

$$\rho(\lambda_n, I) \to 0 \qquad \text{and} \qquad \rho(x_n \circ \lambda_n, x) \to 0, \tag{2.27}$$

where I is the identity map in Λ, $I(t) = t$.

The class of J_1-convergent sequences defines a (sequential) topology on $D[0,1]$, called J_1-topology, which can be easily metricized. The *Skorohod metric* d on $D[0,1]$ is defined as

$$d(x,y) := \inf_{\lambda\in\Lambda} \max \{ \|\lambda - I\|, \|x \circ \lambda - y\| \}, \qquad x, y \in D[0,1]. \tag{2.28}$$

Remark 2.35. The uniform distance $\rho(x,y)$ between functions x and y may be equivalently defined as the infimum of those $\epsilon > 0$ such that

$$\sup_{t\in[0,1]} |x(t) - y(t)| < \epsilon;$$

their graphs can be superposed by perturbations of function values which do not exceed ϵ.

On the other hand, the Skorohod distance $d(x,y)$ between x and y may be equivalently defined as the infimum of those $\epsilon > 0$ for which there exists $\lambda \in \Lambda$ such that

$$\sup_{t\in[0,1]} |\lambda(t) - t| < \epsilon \qquad \text{and} \qquad \sup_{t\in[0,1]} |x(\lambda(t)) - y(t)| < \epsilon;$$

their graphs can be superposed by perturbations of function values and deformation of time scales which both do not exceed ϵ.

It follows directly from the definition that convergence in ρ implies convergence in d,

$$\rho(x_n, x) \to 0 \qquad \text{implies} \qquad d(x_n, x) \to 0, \tag{2.29}$$

(take $\lambda = I$), but the converse is not true, as shown by the following example.

Example 2.36. Consider the sequence of functions $\{ x_n \}$, where x_n is defined as

$$x_n(t) = 1_{[0,\tau+n^{-1}]}(t), \qquad \tau \in [0,1), \quad t \in [0,1].$$

The sequence $\{ x_n \}$ converges in the Skorohod metric to function

$$x(t) = 1_{[0,\tau]}(t), \qquad t \in [0,1] :$$

indeed, if we define continuous time deformation λ_n so that it is piecewise linear and $\lambda_n(\tau + n^{-1}) = \tau$, then

$$\sup_{t\in[0,1]} |\lambda_n(t) - t| = n^{-1} \qquad \text{and} \qquad \sup_{t\in[0,1]} |x_n(\lambda_n(t)) - x(t)| = 0.$$

However, $\{ x_n \}$ does not converge to x in the uniform metric, since

$$\rho(x_n, x) = \sup_{t\in[0,1]} |x_n(t) - x(t)| = 1, \qquad \forall n.$$

Moreover, note that pointwise convergence $x_n(t) \to x(t)$ fails for $t = \tau$:

$$x_n(\tau) = 1, \quad \forall n, \qquad x(\tau) = 0.$$

As shown in this example, convergence in Skorohod metric does not imply uniform convergence, neither pointwise convergence. However, it does with some additional restrictions and requirements:

- convergence in Skorohod metric implies pointwise convergence for continuity points, that is, if t is a continuity point of x, then

$$d(x_n, x) \to 0 \qquad \text{implies} \qquad x_n(t) \to x(t);$$

- convergence in Skorohod metric implies uniform convergence if the limit function is continuous, that is, if $x \in C[0,1]$, then

$$d(x_n, x) \to 0 \qquad \text{implies} \qquad \rho(x_n, x) \to 0.$$

Therefore, the Skorohod metric on $D[0,1]$ restricted to $C[0,1]$ is equivalent to the uniform metric, meaning that they induce the same topology on $C[0,1]$.

It can be proved that the space $D[0,1]$ with the Skorohod metric d is separable but not complete. However, it is possible to define on $D[0,1]$ another metric $d*$ which is equivalent to d (both induce Skorohod J_1-topology) but under which $D[0,1]$ is also complete. Therefore, $D[0,1]$ with metric $d*$ is a Polish space.

Note that separability is a topological property, i.e is a property of the J_1-topology, and does not depend on d, while completeness is a property of metric spaces, and so strictly depends on the chosen metric d or $d*$.

Remark 2.37. The intuition leading to metric $d*$ is slighlty less obvious than the one proposed to define d. However, if some properties of metric spaces are needed, the Skorohod space $D[0,1]$ is usually equipped with $d*$ to take advantage from completeness, which facilitates characterization of compact sets. On the other hand, when only topological properties are concerned, both metrics can be used, since they induce the same topology.

In order to investigate compactness in $D[0,1]$, it is useful to define a quantity that plays in $D[0,1]$ the same role played in $C[0,1]$ by the modulus of continuity.

Let $\{ T_i \}$ be a finite partition of interval $[0,1]$, such that

$$T_i = [t_{i-1}, t_i], \quad i = 1, \ldots, k, \qquad 0 = t_0 < t_1 < \cdots < t_k = 1;$$

a partition $\{ T_i \}$ is called δ-*sparse* if

$$|t_i - t_{i-1}| > \delta, \qquad i = 1, \ldots, k. \tag{2.30}$$

For $x \in D[0,1]$ and $T \subset [0,1]$, define

$$w_x(T) := w(x,T) = \sup_{s,t \in T} |x(s) - x(t)|, \tag{2.31}$$

as the largest distance between function values $x(s)$ and $x(t)$ at points s, t in T.

Proposition 2.38. *For each $x \in D[0,1]$ and each $\epsilon > 0$, there exists a partition $\{ T_i \}$ such that*

$$w_x(T_i) = \sup_{s,t \in T_i} |x(s) - x(t)| < \epsilon, \qquad i = 1, \ldots, k \tag{2.32}$$

It follows from Proposition 2.38 that:

- for each $\epsilon > 0$, there can be at most finitely many points t at which the jump $|x(t) - x(t-)|$ exceeds ϵ, and therefore, x has at most a countable number of discontinuities;
- x is bounded,
$$\|x\| = \sup_{t \in [0,1]} |x(t)| < \infty. \tag{2.33}$$

Definition 2.39. Let x be an arbitrary function on interval $[0,1]$ with values in $\mathbb{R}$. The *cadlag modulus* w'_x of function x is defined as

$$w'_x(\delta) = w'(x,\delta) = \inf_{\{ T_i \}} \max_i w_x(T_i), \qquad \delta \in (0,1), \tag{2.34}$$

where the infimum is taken over all δ-sparse partitions $\{ T_i \}$.

In words, $w'_x(\delta)$ is the infimum over δ-sparse partitions of the maximum of the largest distances between function values $x(s)$ and $x(t)$ taken at points s and t belonging to the same partition interval.

As a consequence of Proposition 2.38, a necessary and sufficient condition for x to be in $D[0,1]$ is

$$\lim_{\delta \to 0^+} w'_x(\delta) = 0. \tag{2.35}$$

For an arbitrary function $x \colon [0,1] \mapsto \mathbb{R}$, it is possible to derive relations between its modulus of continuity w_x and its cadlag modulus w'_x.

- Since $C[0,1] \subseteq D[0,1]$, then obviously $w_x(\delta) \to 0$ implies $w'_x(\delta) \to 0$: indeed, it can be verified that
$$w'_x(\delta) \leq w_x(2\delta), \qquad \delta \in (0, 1/2).$$

- A similar inequality in the opposite direction cannot exist, since $w_x(\delta) \not\to 0$ if x has discontinuities; however, it holds that

$$w_x(\delta) \leq 2w'_x(\delta) + j(x),$$

where $j(x)$ is the maximum absolute jump in x,

$$j(x) := \sup_{t\in[0,1]} |x(t) - x(t-)|.$$

Note that the supremum is a maximum because only a finite number of jumps can exceed a given positive threshold (see Proposition 2.38).

- If x is continuous, i.e. $x \in C[0,1]$,

$$w_x(\delta) \leq 2w'_x(\delta);$$

the modulus of continuity and the cadlag modulus are essentially equivalent for continuous functions.

Relative compactness in $D[0,1]$ is characterized by an analogue of Arzelà-Ascoli theorem, essentially obtained by substituting the modulus of continuity with the cadlag modulus.

Theorem 2.40. *A set K of elements of $D[0,1]$ is relatively compact if and only if two conditions hold:*

1. *uniform boundedness,*
$$\sup_{x\in K} \|x\| < \infty; \tag{2.36}$$
2. *an analogue of uniform equicontinuity,*
$$\lim_{\delta\to 0^+} \sup_{x\in K} w'_x(\delta) = 0. \tag{2.37}$$

To prove weak convergence in $C[0,1]$ (Theorem 2.24) we rely on the three elements listed in Remark 2.26. Given the many similarities highlighted above between $C[0,1]$ and $D[0,1]$, it would be natural to try to adapt those arguments to $D[0,1]$.

Since $D[0,1]$ with metric $d*$ is a Polish space, relative compactness and uniform tightness for families of probability measures coincide, as it happens for $C[0,1]$ (see Remark 2.31), and there is no difficulty on that point. However, things becomes slighlty more complicated when it comes to finite-dimensional distributions.

Natural projections, as defined in (2.12), play in $D[0,1]$ the same role they play in $C[0,1]$. The only (fundamental) difference involves continuity, as a consequence of the fact that convergence in d does not imply pointwise convergence. In particular, the following properties hold:

- each natural projection π_t, $t \in [0,1]$, is a measurable function;
- natural projections π_0 and π_1 are continuous functions;

- for $t \in (0,1)$, natural projection π_t is continuous at x if and only if x is continuous at t;
- the class of finite dimensional sets $\mathcal{B}(D[0,1])_f$ is a separating class, and so finite-dimensional distributions are enough to separate probability measures on $D[0,1]$.

Let $\{\,\mu_n\,\}$ be a sequence of probability measures on $D[0,1]$ which converges to a probability measure μ. Since natural projections are not everywhere continuous on $D[0,1]$, mapping theorem cannot be applied, and therefore, in general,

$$\mu_n \Rightarrow \mu \qquad \cancel{\text{implies}} \qquad \mu_n \circ (\pi_{t_1,\dots,t_k})^{-1} \;\Rightarrow\; \mu \circ (\pi_{t_1,\dots,t_k})^{-1}. \tag{2.38}$$

Moreover, as it happens in $C[0,1]$, $\mathcal{B}(D[0,1])_f$ is not a convergence determining class, and so

$$\mu_n \circ (\pi_{t_1,\dots,t_k})^{-1} \;\Rightarrow\; \mu \circ (\pi_{t_1,\dots,t_k})^{-1} \qquad \cancel{\text{implies}} \qquad \mu_n \Rightarrow \mu; \tag{2.39}$$

indeed, Example 2.22 applies also to $D[0,1]$.

However, it is possible to establish in $D[0,1]$ a positive relation between weak convergence and convergence of finite-dimensional distribution that is similar to the one established for $C[0,1]$ thanks to the mapping theorem.

For a probability measure μ on $D[0,1]$, let $\mathcal{T}_\mu$ be the set of $t \in [0,1]$ for which the natural projection π_t is continuous almost surely, i.e. is discontinuous on a set of null μ-measure:

$$\mathcal{T}_\mu := \{\, t \in [0,1] \colon \pi_t \text{ continuous } \mu\text{-a.s.} \,\}. \tag{2.40}$$

It can be easily proved that:

- since π_0 and π_1 are everywhere continuous, points 0 and 1 are in $\mathcal{T}_\mu$;
- for $t \in (0,1)$, since π_t is continuous at x if and only if x in continuous at t, then $t \in \mathcal{T}_\mu$ if and only if
$$\mu(J_t) = 0, \qquad J_t := \{\, x \in D[0,1] \colon x(t) \neq x(t-) \,\}; \tag{2.41}$$
- since $x \in D[0,1]$ has at most a countable number of jumps, $\mu(J_t) > 0$ for at most a countable number of points t, and the complement of $\mathcal{T}_\mu$ in $[0,1]$ is at most countable.

Therefore, if $t_1, \dots, t_k \in \mathcal{T}_\mu$, the natural projection $\pi_{t_1,\dots,t_k}$ is continuous μ-a.s., and mapping theorem applies.

Proposition 2.41. *Let $\{\,\mu_n\,\}$ be a sequence of probability measures on $D[0,1]$ converging to a probability measure μ. Then, for every choice of $t_1, \dots, t_k \in \mathcal{T}_\mu$,*

$$\mu_n \Rightarrow \mu \qquad \textit{implies} \qquad \mu_n \circ (\pi_{t_1,\dots,t_k})^{-1} \;\Rightarrow\; \mu \circ (\pi_{t_1,\dots,t_k})^{-1}. \tag{2.42}$$

As a consequence, an analogue of Theorem 2.24 holds in $D[0,1]$.

Theorem 2.42. *Let $\{\,\mu_n\,\}$ be a relatively compact sequence of probability measures on $D[0,1]$. If finite-dimensional distributions of $\{\,\mu_n\,\}$ converge to those of a measure μ, when restricted to $\mathcal{T}_\mu$, then $\mu_n \Rightarrow \mu$:*

$$\mu_n \circ (\pi_{t_1,\dots,t_k})^{-1} \;\Rightarrow\; \mu \circ (\pi_{t_1,\dots,t_k})^{-1} \qquad \textit{implies} \qquad \mu_n \Rightarrow \mu. \tag{2.43}$$

To conclude, the characterization of relative compactness of a sequence of probability measures on $D[0,1]$ is obtained by translating the characterization of relative compactness in $D[0,1]$ (Theorem 2.40).

Theorem 2.43. *A sequence $\{\mu_n\}$ of probability measures on $D[0,1]$ is relatively compact if and only if two conditions hold:*

1. *for each $\eta > 0$, there exist L and n_0 such that,*

$$\mu_n\left(x \in D[0,1] \colon \|x\| \geq L\right) \leq \eta, \qquad n \geq n_0; \tag{2.44}$$

2. *for each $\epsilon > 0$ and $\eta > 0$, there exist $\delta \in (0,1]$ and n_0 such that,*

$$\mu_n\left(x \in D[0,1] \colon w'_x(\delta) \geq \epsilon\right) \leq \eta, \qquad n \geq n_0. \tag{2.45}$$

2.4 Skorohod Representation Theorem

Let $(\mathcal{X}, \rho)$ be a metric space and let X and Y be $\mathcal{X}$-valued random elements (see Definition 2.14). There exist different concepts of "equality" between X and Y.

1. X and Y are equal as functions from a probability space $(\Omega, \mathcal{F}, P)$ to $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$,

$$X(\omega) = Y(\omega), \qquad \forall\, \omega \in \Omega. \tag{2.46}$$

2. X and Y coincide as functions from $(\Omega, \mathcal{F}, P)$ to $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, except for a set $N \subseteq \Omega$ having P-measure zero,

$$X(\omega) = Y(\omega), \qquad \forall\, \omega \in \Omega \setminus N, \quad P(N) = 0; \tag{2.47}$$

   this situation is usually called *almost sure equality* and is written as

$$X = Y \qquad P\text{-a.s.}. \tag{2.48}$$

3. The distributions of X and Y on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ coincide as probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ (see Definition 2.15),

$$P_X(A) = P_Y(A), \qquad \forall\, A \in \mathcal{B}(\mathcal{X}) \tag{2.49}$$

   this situation is usually called *equality in distribution* and is written as

$$X \sim Y \qquad \text{or} \qquad \mathcal{L}(X) = \mathcal{L}(Y). \tag{2.50}$$

There is a fundamental difference between the first two concepts and the third one: for equality and almost sure equality to hold, X and Y must be defined on the same probability space, while this is not necessary for equality in distribution, since only the resulting probability measures on space $\mathcal{X}$ are considered (see Remark 2.17).

Let $\{X_n\}$ be a sequence of random elements of $\mathcal{X}$ and assume they are defined on the same probability space $(\Omega, \mathcal{F}, P)$; a different (and stronger) concept of convergence for sequence $\{X_n\}$ can be defined.

Definition 2.44. A sequence $\{ X_n \}$ of random elements of $\mathcal{X}$ defined on the same probability space $(\Omega, \mathcal{F}, P)$ is said to *converge almost surely* to a random element X of $\mathcal{X}$, defined on $(\Omega, \mathcal{F}, P)$, if

$$\rho(X_n(\omega), X(\omega)) \to 0, \qquad \forall\, w \in \Omega \setminus N, \quad P(N) = 0, \tag{2.51}$$

and is written as

$$X_n \to X \qquad P\text{-a.s.}. \tag{2.52}$$

Almost sure convergence clearly implies convergence in distribution, and is a sort of pointwise convergence for random elements $\{ X_n \}$ regarded as functions on a common probability space $(\Omega, \mathcal{F}, P)$. Indeed, if $\{ X_n \}$ converges to X almost surely, then $X_n(\omega)$ converges to some element $X(\omega)$ in the metric space $(\mathcal{X}, \rho)$, while if only convergence in distribution holds, $X_n(\omega)$ need not converge anywhere in $(\mathcal{X}, \rho)$.

On Polish spaces, it is possible to define an almost surely convergent representation for sequences, known as *(almost sure) Skorohod representation* [16].

Definition 2.45. Let $(\mathcal{X}, \rho)$ be a Polish space. A sequence $\{ X_n \}$ of random elements of $\mathcal{X}$ is said to admit an *(almost sure) Skorohod representation* if there exist a sequence of $\mathcal{X}$-valued random elements $\{ Y_n \}$ and an $\mathcal{X}$-valued random element Y, defined on the unit interval $([0,1], \mathcal{B}([0,1]))$ equipped with the Lebesgue measure ℓ, such that

$$X_n \sim Y_n, \quad \forall n, \tag{2.53}$$

$$Y_n \to Y \qquad \ell\text{-a.s.}. \tag{2.54}$$

Clearly, a sequence $\{ X_n \}$ which admits a Skorohod representation is convergent in distribution to a random element X, whose distribution coincides with that of Y:

$$X_n \Rightarrow X \sim Y. \tag{2.55}$$

The converse is also true, as stated in the following result (for the proof, see the original article by Skorohod [16] or [4, sec. 6]).

Theorem 2.46 (Skorohod Representation Theorem)**.** *Let $(\mathcal{X}, \rho)$ be a Polish space and let $\{ X_n \}$ be a sequence of random elements of $\mathcal{X}$ which converges in distribution to a random element X of $\mathcal{X}$,*

$$X_n \Rightarrow X.$$

Then, $\{ X_n \}$ admits a Skorohod representation with $Y \sim X$.

Therefore, Theorem 2.46 establishes an equivalence between convergence in distribution and existence of a Skorohod representation for sequences of random elements.

Remark 2.47. Almost sure convergence in (2.54) can be actually extended, without loss of generality, to each $\omega \in [0,1]$:

$$\rho(Y_n(\omega), Y(\omega)) \to 0, \qquad \forall\, \omega \in [0,1]. \tag{2.56}$$

Indeed, adjusting values of random elements $\{ Y_n \}$ on a subset $N \subseteq [0,1]$ such that $\ell(N) = 0$ does not affect equality in distribution (2.53) nor convergence in distribution (2.55).

A generalization of Theorem 2.46 was proved in [5].

Theorem 2.48 (Blackwell-Dubbins-Fernique)**.** *Let $(\mathcal{X}, \rho)$ be a Polish space and let μ be a probability measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Then, there exists a measurable function Y_μ defined on the unit interval $([0,1], \mathcal{B}([0,1]), \ell)$ with values in $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that*

$$\ell \circ Y_\mu^{-1} = \mu. \tag{2.57}$$

Moreover, if a sequence $\{ \mu_n \}$ of probability measure on $\mathcal{X}$ converges in distribution to probability measure μ, then

$$Y_{\mu_n} \to Y_\mu \qquad \ell\text{-a.s.}. \tag{2.58}$$

Therefore, if $(\mathcal{X}, \rho)$ is a Polish space, there exists a *Skorohod parametrization* of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

In general, it can be proved that, for an arbitraty metric space $(\mathcal{X}, \rho)$, there exists a Skorohod parametrization of the subset of *tight* probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ (recall that on Polish spaces, every probability measure is tight) [6].

The Skorohod representation is frequently used to trivialize proofs in the theory of convergence in distribution on Polish spaces. However, Skorohod himself applied this representation in a uncommon and interesting way [16].

Let $D[0,1]$ be the Skorohod space, and let $\mathbb{Q}$ be a countable dense subset of $[0,1]$ containing 1. Functions in $D[0,1]$ are completely determined by their values on dense subsets of the domain, and therefore there exists a *one-to-one* mapping

$$x \in D[0,1] \mapsto \{ x(q) \}_{q \in \mathbb{Q}} \in \mathbb{R}^{\mathbb{Q}}. \tag{2.59}$$

The space $\mathbb{R}^{\mathbb{Q}}$ is Polish space and the class of finite-dimensional sets is a convergence-determining class (unlike in $C[0,1]$ and $D[0,1]$).

Let $\{ X_n \}$ be a sequence of random elements of $D[0,1]$. If finite-dimensional distributions of $\{ X_n \}$ converge for each choice of $q_1, \ldots, q_k \in \mathbb{Q}$, then it is possible to define a random element X of $\mathbb{R}^{\mathbb{Q}}$ such that

$$\{ X_n(q) \}_{q \in \mathbb{Q}} \Rightarrow X \qquad \text{on} \quad \mathbb{R}^{\mathbb{Q}}. \tag{2.60}$$

By the Skorohod Representation Theorem, there exist random elements $\{ Y_n \}$, Y of $\mathbb{R}^{\mathbb{Q}}$, defined on the Lebesgue interval, such that

$$X \sim Y, \qquad \{ X_n(q) \}_{q \in \mathbb{Q}} \sim Y_n, \quad \forall n, \tag{2.61}$$

$$Y_n(q, \omega) \to Y(q, \omega), \qquad q \in \mathbb{Q}, \quad \omega \in [0,1]. \tag{2.62}$$

Moreover, it is possible to define, for each random element Y_n, Y of $\mathbb{R}^{\mathbb{Q}}$, a corresponding random element Z_n, Z of $D[0,1]$ as

$$Z_n(t, \omega) := \lim_{q \downarrow t^+} Y_n(q, \omega), \quad t \in [0,1), \qquad Z_n(1, \omega) := Y_n(1, \omega), \tag{2.63}$$

for almost each $\omega \in [0,1]$.

As a result, the sequence $\{ Z_n \}$ is a representation of the original sequence $\{ X_n \}$ which preserves convergence on dense subset $\mathbb{Q}$ and which is independent of any topology defined on $D[0,1]$.

This representation is relevant in at least two situations.

1. Assuming uniform tightness (and thus relative compactness, via direct Prohorov Theorem) of the sequence $\{ \mathcal{L}(X_n) \}$ with respect to some topology on $D[0,1]$ (not necessarily J_1), every subsequence $\{ Z_m \} \subseteq \{ Z_n \}$ contains a further subsequence $\{ Z_{m(k)} \}$ which is convergent almost surely,

$$Z_{m(k)} \to_k Z \qquad \ell\text{-a.s.},$$

and this is enough to conclude convergence in distribution,

$$X_n \Rightarrow Z \qquad \text{in} \quad D[0,1].$$

2. Assuming convergence of finite-dimensional distributions of $\{ X_n \}$ on a countable dense subset $\mathbb{Q} \subseteq [0,1]$ and relative compactness of the sequence $\{ \mathcal{L}(X_n) \}$, then uniform tightness is implied, independently of whether the topology defined on $D[0,1]$ makes it a Polish space or not.

2.5 Sequential topological spaces and space $\mathcal{P}$

Definitions and results stated in this section are entirely taken from [10, sec. 4]; see that article for additional references.

Let $\mathcal{X}$ be a generic space.

Definition 2.49. The space $\mathcal{X}$ is said of *type* $\mathcal{L}$ if

- among all sequences of elements of $\mathcal{X}$, a class of *convergent* sequences is distinguished;
- for each convergent sequence $\{ x_n \}$, there exists a unique point x in $\mathcal{X}$, called the *limit*, and convergence is denoted as $x_n \to x$.

Convergent sequences defined above must satisfy two conditions:

1. for every $x \in \mathcal{X}$, the constant sequence $\{ x, x, \dots \}$ is convergent to x;
2. if $x_n \to x$, each subsequence $\{ x_m \} \subseteq \{ x_n \}$ is convergent, and converges to the same limit x.

The identification of convergent sequences, together with their limits, allows to define closed sets, and therefore a topology on $\mathcal{X}$.

Definition 2.50. A set $C \subseteq \mathcal{X}$ is *closed* if, for every convergent sequence $\{ x_n \} \subseteq C$, its limit x is in C.

Closed sets induce a topology $\mathcal{O}(\to)$ on $\mathcal{X}$, called *sequential topology*, and the space $(\mathcal{X}, \mathcal{O}(\to))$ is called *sequential (topological) space*.

Note that the two very simple conditions stated in Definition 2.49 are enough to define a topology on $\mathcal{X}$.

The topology $\mathcal{O}(\to)$ defines in turn a new class of convergent sequences (those converging in the topology), which is, in general, larger than the initial class of convergent sequences. To distinguish the two classes, sequences converging in the induced topology $\mathcal{O}(\to)$ are usually called convergent *a posteriori*, while sequences identified in Definition 2.49 are called convergent *a priori*.

A characterization of sequences convergent a posteriori is the following.

Theorem 2.51 (Kantorowich-Kisynski)**.** *A sequence* $\{ x_n \}$ *of elements of* $\mathcal{X}$ *is convergent a posteriori if and only if every subsequence* $\{ x_m \} \subseteq \{ x_n \}$ *contains a further subsequence* $\{ x_{m(k)} \}$ *which is convergent a priori.*

Therefore, the initial class of convergent sequences (convergent a priori) may be enlarged to include also sequences convergent a posteriori.

Definition 2.52. The space $\mathcal{X}$ is said of *type* $\mathcal{L}*$ if it is of type $\mathcal{L}$ and convergent sequences satisfy an additional condition:

3. if every subsequence $\{ x_m \}$ of $\{ x_n \}$ contains a further subsequence $\{ x_{m(k)} \}$ which is converges to x, then the entire sequence $\{ x_n \}$ converges to x.

Convergence in $\mathcal{L}*$ is denoted as $x_n \xrightarrow{*} x$.

Note that, because of Theorem 2.51, in spaces of type $\mathcal{L}*$ convergence a posteriori coincides with convergence a priori. Moreover, convergence "$\xrightarrow{*}$" is already the usual convergence of sequences in the induced sequential topological space $(\mathcal{X}, \mathcal{O}(\to)) = (\mathcal{X}, \mathcal{O}(\xrightarrow{*}))$.

Example 2.53. There are at least two well-known examples in which a notion of convergence a priori is weakened to a notion of convergence a posteriori.

1. Let $(\Omega, \mathcal{F}, P)$ be a probability space and consider real-valued random variables: if convergence *almost surely* is identified with convergence a priori, then convergence a posteriori is the convergence *in probability.*
2. Let $\mathcal{X} = \mathbb{R}$ and consider a sequence $\{ \epsilon_n > 0 \}$ such that $\epsilon_n \downarrow 0$. Assume that convergence a priori $x_n \to x$ means

$$|x_n - x| < \epsilon_n, \qquad n \geq 0,$$

   i.e. $\{ x_n \}$ converges to x at given rate $\{ \epsilon_n \}$. Then convergence a posteriori is the usual convergence of real numbers.

The additional condition stated in Definition 2.52 is slighlty more complicated to be checked than those stated in Definition 2.49. However, it is not actually needed for the verification of relative sequential compactness.

Proposition 2.54. *A set* $K \subseteq \mathcal{X}$ *is relatively sequentially compact with respect to a priori convergence* "$\to$" *if and only if it is relatively sequentially compact with respect to a posteriori convergence* "$\xrightarrow{*}$"*.*

Let $(\mathcal{X}, \rho)$ be a metric space and denote by $\mathcal{P}(\mathcal{X})$ be the space of tight probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$:

$$\mathcal{P}(\mathcal{X}) := \{ \mu \text{ prob. measure on } (\mathcal{X}, \mathcal{B}(\mathcal{X})) \ : \ \mu \text{ is tight} \}. \tag{2.64}$$

Note that if $\mathcal{X}$ is a Polish space, then, by Theorem 2.3, $\mathcal{P}(\mathcal{X})$ contains all probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

The notion of weak convergence of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, introduced in Definition 2.5, is actually a notion of convergence on the space $\mathcal{P}(\mathcal{X})$. It can be easily checked that weak convergence satisfies conditions in Definition 2.49:

- for each convergent sequence $\{ \mu_n \}$, there exists a unique limit $\mu \in \mathcal{P}(\mathcal{X})$;

1. for every $\mu \in \mathcal{X}$, the constant sequence $\{ \mu, \mu, \dots \}$ is convergent to μ;
2. if $\mu_n \to \mu$, each subsequence $\{ \mu_m \} \subseteq \{ \mu_n \}$ is convergent, and converges to the same limit μ.

Moreover, weak convergence also satisfies the additional condition in Definition 2.52, as a consequence of Theorem 2.8:

3. if every subsequence $\{ \mu_m \}$ of $\{ \mu_n \}$ contains a further subsequence $\{ \mu_{m(k)} \}$ which is converges to μ, then the entire sequence $\{ \mu_n \}$ converges to μ.

Therefore, $\mathcal{P}(\mathcal{X})$ is a space of type $\mathcal{L}*$ and weak convergence naturally induces on $\mathcal{P}(\mathcal{X})$ a sequential topology $\mathcal{O}(\Rightarrow)$.

Remark 2.55. The notion of relative (sequential) compactness for a family of probability measures, given in Definition 2.10, actually corresponds to the notion of relative compactness in the sequential topology $\mathcal{O}(\Rightarrow)$ on $\mathcal{P}(\mathcal{X})$. Moreover, Prohorov theorems are results on relative compactness in the sequential topology induced by weak convergence.

2.6 Weak convergence in non-metric spaces

Let $(\mathcal{X}, \tau)$ be a topological, non metrizable space and let $\mathcal{B}(\mathcal{X})$ be the Borel σ-field on $\mathcal{X}$.

In a non-metric space $\mathcal{X}$, the notion of weak convergence, introduced in Definition 2.5, is still valid, since continuity of functions on $\mathcal{X}$ is a topological property. However, it brings many disturbing problems, even when restricted to the space $\mathcal{P}(\mathcal{X})$ of tight probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

Consider the infinite dimensional Hilbert space $(H, \langle , \rangle)$ equipped with the weak topology $\tau = \sigma(H, H)$. In [8] (reported in [10] and [12]), the author suggests an example of a sequence $\{ X_n \}$ of random elements of H which is convergent in distribution to $X = 0$, but has no uniformly tight subsequences. The following observations summarize the notable features of this example.

- There are weakly convergent sequences of probability measures on $(H, \mathcal{B}(H))$ for which no subsequence is uniformly tight, and therefore the direct Prohorov theorem, although still valid on (H, τ), looses its role as fundamental tool for investigating weak convergence.
- Relative (sequential) compactness on $\mathcal{P}(H)$ could be completely characterized, but it appears very difficult to check for it without uniform tightness.
- It can be proved that every uniformly tight sequence $\{ X_n \}$ contains a subsequence $\{ X_{n(k)} \}$ for which there exists a Skorohod representation (Definition 2.45), and therefore $\{ X_{n(k)} \}$ is convergent in distribution [12].
- There are sequences convergent in distribution for which no subsequence has a Skorohod representation.

The third observation highlights a close relation between the direct Prohorov theorem and the Skorohod representation for subsequences, which in turn implies convergence in distribution. Unfortunately, a Skorohod representation is not available for every sequence which converges in distribution, as the last observation states.

However, in some other non-metric spaces, convergence in distribution and existence of a Skorohod representation are essentially equivalent [12].

Proposition 2.56. *In distribution spaces $\mathcal{S}'$ and $\mathcal{D}'$, a sequence $\{ X_n \}$ of random elements converges in distribution to a random element X if and only if every subsequence $\{ X_m \} \subseteq \{ X_n \}$ contains a further subsequence $\{ X_{m(k)} \}$ which admits a Skorohod representation (with $Y \sim X$).*

Note that this Proposition is essentially a reformulation of Theorem 2.8, where convergence in distribution for subsequences is replaced with the existence of a Skorohod representation. It is therefore natural to raise questions on how stronger the existence of a Skorohod representation is with respect to convergence in distribution, and whether results similar Proposition 2.56 may be extended to construct a consistent theory.

The theory of convergence of probability measures presented hereunder has been devised by Jakubowski in [10] and [12] (see those articles for additional details and proofs), and holds in topological spaces satisfying a specific condition.

Condition 2.57. Let $(\mathcal{X}, \tau)$ be a topological space. There exists a countable family

$$\mathcal{F} := \{ f_n \colon \mathcal{X} \mapsto [-1, 1],\, n \in \mathbb{N} \},$$

of τ-continuous functions which separate points of $\mathcal{X}$, i.e. for each $x, y \in \mathcal{X}$,

$$x \neq y \quad \text{implies} \quad \exists\, f_n \in \mathcal{F} \quad \text{s.t.} \quad f_n(x) \neq f_n(y).$$

This condition is not too restrictive, but is sufficient for $(\mathcal{X}, \tau)$ to have some important properties. The most remarkable feature is the following.

Proposition 2.58. *Let $(\mathcal{X}, \tau)$ be a topological space satifying Condition 2.57 and let μ be a tight probability measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, i.e. $\mu \in \mathcal{P}(\mathcal{X})$. Then there exists a measurable function Y_μ define on the unit interval $([0,1], \mathcal{B}([0,1]), \ell)$ with values in $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that*

$$\ell \circ Y_\mu^{-1} = \mu. \tag{2.65}$$

In words, each tight probability measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ has a Skorohod parametrization.

Next, we need a result on the existence of a Skorohod representation on such topological spaces, extending Theorem 2.46 (for the proof, see [12]).

Theorem 2.59 (Jakubowski)**.** *Let $(\mathcal{X}, \tau)$ be a topological space satifying Condition 2.57 and let $\{ X_n \}$ be a uniformly tight sequence of random elements of $\mathcal{X}$. Then, there exist:*

- *a subsequence $\{ X_{n(k)} \}$ of $\{ X_n \}$,*

- *a sequence of $\mathcal{X}$-valued random elements $\{ Y_k \}$ and an $\mathcal{X}$-valued random element Y, defined on the unit interval $([0,1], \mathcal{B}([0,1]))$ equipped with the Lebesgue measure ℓ,*

such that

$$X_{n(k)} \sim Y_k, \quad \forall k, \tag{2.66}$$

$$Y_k \to Y \qquad \ell\text{-a.s.}. \tag{2.67}$$

Moreover, for each $\epsilon > 0$ there exists a compact $K_\epsilon \subseteq \mathcal{X}$ such that

$$\mathbb{P}\left(\{ \omega \in [0,1] \colon Y_k(\omega) \in K_\epsilon, \, \forall\, k \}\right) > 1 - \epsilon. \tag{2.68}$$

The almost sure convergence in (2.67), with the additional condition stated in (2.68), is called almost sure convergence *in compacts*, and the Skorohod representation with convergence strengthened to almost sure convergence in compacts is called *strong Skorohod representation.*

Therefore, adopting this new terminology, Theorem 2.59 may be restated as follows.

Theorem 2.60. *Let $(\mathcal{X}, \tau)$ be a topological space satifying Condition 2.57 and let $\{ X_n \}$ be a uniformly tight sequence of random elements of $\mathcal{X}$. Then there exists a subsequence $\{ X_{n(k)} \}$ of $\{ X_n \}$ which admits the strong Skorohod representation.*

This theorem is essentially a direct Prohorov theorem, where convergence in distribution is replaced with the existence of a strong Skorohod representation.

Remark 2.61. In the assumptions of Theorem 2.59, if additionally $X_n \Rightarrow X$, then there exists a (strong) Skorohod representation for subsequences: every subsequence $\{ X_m \} \subseteq \{ X_n \}$ contains a further subsequence $\{ X_{n(k)} \}$ which admits a (strong) Skorohod representation with $Y \sim X$.

Given the central role that the (strong) Skorohod representation seem to play for the theory of weak convergence in non-metric spaces, Jakubowski proposes in [10] a new definition of convergence for random elements taking values on a topological space, which is stronger than usual convergence in distribution.

Definition 2.62. Let $(\mathcal{X}, \tau)$ be a topological space satifying Condition 2.57. A sequence $\{ X_n \}$ of random elements of $\mathcal{X}$ is said to converge in the sense of Jakubowski to a random element X of $\mathcal{X}$, and is written as

$$X_n \overset{*}{\Rightarrow} X, \tag{2.69}$$

if every subsequence $\{ X_m \} \subseteq \{ X_n \}$ contains a further subsequence $\{ X_{m(k)} \}$ which admits a strong Skorohod representation with $Y \sim X$.

This novel notion of convergence for random element of $\mathcal{X}$ is in fact a notion of convergence in the space $\mathcal{P}(\mathcal{X})$ of tight probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

It is easy to check that convergence is the sense of Jakubowski satisfies conditions in Definition 2.49 (Proposition 2.58 is essential here), as well as the additional condition in Definition 2.52. Therefore, $\mathcal{P}(\mathcal{X})$ equipped with convergence in Definition 2.62 is of type $\mathcal{L}*$ and this notion of convergence induces on $\mathcal{P}(\mathcal{X})$ a sequential topology $\mathcal{O}(\overset{*}{\Rightarrow})$, which is sometimes referred to as the sequential topology *induced by the strong Skorohod representation.*

Remark 2.63. In light of Theorem 2.60, one may be tempted to define the new notion of convergence by means of the strong Skorohod representation for full sequences: a sequence $\{ X_n \}$ converges in the sense of Jakubowski to X if it admits a strong Skorohod representation with $Y \sim X$.

However, this notion of convergence in $\mathcal{P}(\mathcal{X})$ does not satisfy the additional condition in Definition 2.52. Therefore, $\mathcal{P}(\mathcal{X})$ equipped with this convergence is only of type $\mathcal{L}$, that is, convergence a priori does not coincide with convergence a posteriori. Instead, convergence in Definition 2.62 is obtained weakening this notion of convergence (by means of Theorem 2.51) to include sequences convergent a posteriori, so that it assumes a topological meaning.

Convergence in the sense of Jakubowski may be applied in many cases of practical interest, is quite operational and proves to be more satisfactory than usual convergence in distribution, when working on non-metric spaces.

In particular, assume that $(\mathcal{X}, \tau)$ is a topological space satifying Condition 2.57 and space $\mathcal{P}(\mathcal{X})$ is equipped with the sequential topology $\mathcal{O}(\overset{*}{\Rightarrow})$ induced by the strong Skorohod representation (i.e. induced by convergence in the sense of Jakubowski).

- Theorem 2.60 is essentially the direct Prohorov theorem for families of random elements of $\mathcal{X}$:

  $$\text{uniform tightness} \quad \text{implies} \quad \text{relative compactness in } \mathcal{P}(\mathcal{X}).$$

- Convergence in $\mathcal{P}(\mathcal{X})$ implies uniform tightness, as a consequence of almost sure convergence *in compacts*. The converse Prohorov theorem does not hold in general, but it does in presence of some additional criteria [10, sec. 6].

To conclude, it may be interesting to discuss the relation between this new notion of convergence and usual convergence in distribution.

- The sequential topology $\mathcal{O}(\overset{*}{\Rightarrow})$ induced by convergence in the sense of Jakubowski is finer than the sequential topology $\mathcal{O}(\Rightarrow)$ induced by convergence in distribution. The example described at the beginning of this section shows that in general these two topologies do not coincide, and $\mathcal{O}(\overset{*}{\Rightarrow})$ may be strictly finer than $\mathcal{O}(\Rightarrow)$.

- If $(\mathcal{X}, \tau)$ is a metric space, the notions of convergence in the sense of Jakubowski and convergence in distribution are equivalent, that is, topologies $\mathcal{O}(\overset{*}{\Rightarrow})$ and $\mathcal{O}(\Rightarrow)$ coincide.

2.7 Jakubowski S-topology on $D[0,1]$

Let $D[0,1]$ be the Skorohod space, i.e. the space of functions on interval $[0,1]$ with values in $\mathbb{R}$ that are right-continuous and have left limits.

In [9] and [11], Jakubowski proposes a new sequential topology on $D[0,1]$, which arises from quite natural criteria of compactness but cannot be metricized. However, space $D[0,1]$ equipped with this topology satisfies Condition 2.57,

and therefore it is possible to build a satisfactory theory of convergence for probability measures on this space (see Section 2.6).

In order to better explain the arguments leading to the definition of the above-mentioned sequential topology, Jakubowski starts discussing a simple example.

Let $\mathbb{V}^+ \subseteq D[0,1]$ be the space of *non-negative* and *non-decreasing* cadlag functions and let $\mathbb{Q}$ be a countable dense subset of $[0,1]$ containing 1. Moreover, let K be a set of elements of $\mathbb{V}^+$ and assume that

$$\sup_{x \in K} v(1) < \infty. \tag{2.70}$$

Then, it can be shown that:

- there exist a sequence $\{ x_n \} \subseteq K$ and an element $\widetilde{x} \in \mathbb{R}^{\mathbb{Q}}$ such that

$$x_n(q) \to \widetilde{x}(q), \qquad q \in \mathbb{Q}; \tag{2.71}$$

- the cadlag function $x \in D[0,1]$, defined as

$$x(t) := \lim_{q \downarrow t^+} \widetilde{x}(q), \quad t \in [0,1), \qquad x(1) = \widetilde{x}(1), \tag{2.72}$$

  belongs to $\mathbb{V}^+$;

- pointwise convergence holds for continuity points of x, that is, for each $t \in [0,1]$ such that x is continuous at t,

$$x_n(t) \to x(t). \tag{2.73}$$

Moreover, each element $x \in \mathbb{V}^+$ determines a finite measure μ on $([0,1], \mathcal{B}([0,1]))$, given by

$$\mu([0,t]) = x(t), \qquad t \in [0,1]. \tag{2.74}$$

Therefore, a notion of convergence on the space $\mathbb{V}^+$ may be defined by means of weak convergence of the corresponding finite measures. It can be proved that this notion of convergence is equivalent to the one in (2.73).

As a consequence, condition (2.70) restricted to $\mathbb{V}^+$ is a criterion of relative compactness with respect to a quite natural sequential topology, induced by weak convergence of the corresponding finite measures.

A similar procedure may be followed for space $D[0,1]$. Let $\mathbb{Q}$ be a countable dense subset of $[0,1]$ containing 1 and let K be a set of elements of $D[0,1]$ such that

$$\sup_{x \in K} \|x\| < \infty. \tag{2.75}$$

Again, it can be shown that

- there exist a sequence $\{ x_n \} \subseteq K$ and an element $\widetilde{x} \in \mathbb{R}^{\mathbb{Q}}$ such that

$$x_n(q) \to \widetilde{x}(q), \qquad q \in \mathbb{Q}. \tag{2.76}$$

However, in this case, constructing a candidate limit $x \in D[0,1]$ is not as easy as in (2.72), and additional conditions beyond (2.75) should be considered, in order to guarantee that each expression

$$x(t) := \lim_{q \downarrow t^+} \widetilde{x}(q), \qquad t \in [0,1), \tag{2.77}$$

is actually well-defined and x belongs to $D[0,1]$. Moreover, even when such construction is possible, it is not clear whether the sequence $\{ x_n \}$ converges to x in some topology on $D[0,1]$.

The new sequential topology proposed by Jakubowski allows to naturally extend the reasoning presented above for space $\mathbb{V}^+$ to the entire space $D[0,1]$.

Definition 2.64. Let $x \in D[0,1]$ be a cadlag function. The *total variation* $V(x)$ of function x is defined as

$$V(x) := \sup_n \{ |x(0)| + \sum_{i=1}^{n} |x(t_i) - x(t_{i-1})| \,:\, 0 = t_0 < t_1 < \cdots < t_n = 1 \}.$$

The subspace of $D[0,1]$ containing cadlag functions with finite total variation is denoted by $\mathbb{V}$:

$$\mathbb{V} := \{ x \in D[0,1] \,:\, V(x) < \infty \} \subseteq D[0,1]. \tag{2.78}$$

Each element $x \in \mathbb{V}$ determines a (finite) signed measure ν on $([0,1], \mathcal{B}([0,1]))$ given by

$$\nu([0,t]) = x(t), \qquad t \in [0,1]. \tag{2.79}$$

Similarly to $\mathbb{V}^+$, weak convergence of the corresponding signed measures defines a notion of convergence on the space $\mathbb{V}$.

Definition 2.65. A sequence $\{ x_n \}$ of elements of $\mathbb{V}$ is said to *converge weakly-$*$* to an element $x \in \mathbb{V}$ if the corresponding signed measures $\{ \nu_n \}$, defined in (2.79), converge weakly to the signed measure ν corresponding to x, that is, if

$$\int_0^1 f(t) \nu_n(dt) \to \int_0^1 f(t) \nu(dt), \tag{2.80}$$

for every continuous function $f \colon [0,1] \mapsto \mathbb{R}$. In this section, weak-$*$ convergence is written as

$$x_n \xrightarrow{w} x. \tag{2.81}$$

This notion of convergence is used in an essential way to define the *Jakubowski S-convergence* on $D[0,1]$.

Definition 2.66. A sequence $\{ x_n \}$ of cadlag functions is called *S-convergent* to a function x if, for each $\epsilon > 0$, there exist functions $\{ \widetilde{x}_{n,\epsilon} \}$ and $\widetilde{x}_\epsilon$ in $\mathbb{V}$ which are ϵ-uniformly close to the corresponding functions $\{ x_n \}$ and x and are weakly-$*$ convergent, that is

$$\| x_n - \widetilde{x}_{n,\epsilon} \| \le \epsilon, \qquad \| x - \widetilde{x}_\epsilon \| \le \epsilon, \tag{2.82}$$

$$\widetilde{x}_{n,\epsilon} \xrightarrow{w} \widetilde{x}_\epsilon. \tag{2.83}$$

In this section, S-convergence is written as

$$x_n \to_S x. \tag{2.84}$$

Remark 2.67. The construction of functions $\{\, \widetilde{x}_{n,\epsilon} \,\}$ and $\widetilde{x}_\epsilon$ is quite standard. For example, consider function x and, for each $\epsilon > 0$, define time instants $\{\, \tau_k^\epsilon(x) \,\}$ recursively as

$$\tau_0^\epsilon(x) := 0, \qquad \tau_k^\epsilon(x) := \inf \{\, t > \tau_{k+1}^\epsilon(x) \,:\, |x(t) - x(\tau_{k-1}^\epsilon(x))| > \epsilon \,\},$$

with the convention that $\inf \emptyset = \infty$. Then, function $\widetilde{x}_\epsilon$ is naturally defined as

$$\widetilde{x}_\epsilon(t) = x(\tau_k^\epsilon(x)), \qquad \tau_k^\epsilon(x) \le t < \tau_{k+1}^\epsilon(x), \quad t \in [0,1].$$

An similar procedure is adopted to construct functions $\{\, \widetilde{x}_{n,\epsilon} \,\}$.

It is possible to prove that the Skorohod space equipped with S-convergence is of type $\mathcal{L}$, that is, satisfies conditions in Definition 2.49. Therefore, we have enough information to characterize closed sets and define a sequential topology on $D[0,1]$ (see Definition 2.50), which is called *Jakubowski S-topology*.

Remark 2.68. Note that the Skorohod space equipped with S-convergence is *not* of type $\mathcal{L}*$, i.e. does not satisfy the additional condition in Definition 2.52. Therefore, S-convergence, which is convergence a priori, may be weakened to $S*$-convergence by means of Theorem 2.51, in order to include sequences convergent a posteriori.

However, this further complication is not always necessary, since in many cases checking properties for S-convergence is enough to have them verified for the induced sequential topology S. In particular, relative compactness with respect to S-convergence coincides with relative compactness in the S-topology (Theorem 2.54).

A typical example of S-convergent sequence is the following.

Example 2.69. Consider the sequence of functions $\{\, x_n \,\}$, where x_n is defined as

$$x_n(t) = 1_{[\tau, \tau + n^{-1}]}(t), \qquad \tau \in [0,1), \quad t \in [0,1].$$

The sequence $\{\, x_n \,\}$ converges in the S-topology to the null function

$$x(t) = 0, \qquad t \in [0,1].$$

Indeed, for each $n \ge 1$, $V(x_n) = 2 < \infty$, and so $x_n \in \mathbb{V}$, which means that we can take $\widetilde{x}_{n,\epsilon} = x_n$, for each $\epsilon > 0$. Moreover, the corresponding signed measure is given by

$$\nu_n(dt) = \delta_\tau(t) - \delta_{\tau + n^{-1}}(t),$$

and therefore

$$\int_0^1 f(t) \nu_n(dt) = f(\tau) - f(\tau + n^{-1}) \;\to\; 0,$$

for every continuous function $f \colon [0,1] \mapsto \mathbb{R}$.

However, $\{\, x_n \,\}$ does not converge to x in the Skorohod metric d, since for each $\delta > 0$ there exists n_0 such that

$$w'(x_n, \delta) = 1, \qquad n \ge n_0,$$

which means that $\{ x_n \}$ is not relatively compact. Clearly, $\{ x_n \}$ neither converges in the uniform metric ρ, even if the candidate limit x is a continuous function.

Finally, note that pointwise convergence $x_n(t) \to x(t)$ fails for $t = \tau$:

$$x_n(\tau) = 1, \quad \forall n, \qquad x(\tau) = 0.$$

As shown in this example, S-convergence does not imply pointwise convergence, exactly as it happens for convergence in the Skorohod metric. However, is does outside a countable set.

Proposition 2.70. *Let $\{ x_n \}$ be a sequence of cadlag functions converging to x in the S-topology,*

$$x_n \to_S x.$$

Then, there exists a countable set $D \subseteq [0,1]$ such that

$$x_n(t) \to x(t), \qquad t \in [0,1] \setminus D. \tag{2.85}$$

In order to investigate compactness with respect to the S-topology, we need first to introduce the notions of upcrossings and oscillations.

Definition 2.71. Let $x \in D[0,1]$ be a cadlag function.

- The number of *upcrossings* $N^{a,b}(x)$ of levels $a < b$ for function x is defined as follows: $N^{a,b}(x) \geq k$ if there exist time instants

  $$0 \leq t_1 < t_2 < \cdots < t_{2k-1} < t_{2k} \leq 1$$

  such that

  $$x(t_{2i-1}) \leq a, \qquad x(t_{2i}) \geq b, \qquad i = 1, \ldots, k.$$

- The number of *oscillations* $N_\eta(x)$ of size $\eta > 0$ for function x is defined as follows: $N_\eta(x) \geq k$ if there exist time instants

  $$0 \leq t_1 < t_2 < \cdots < t_{2k-1} < t_{2k} \leq 1$$

  such that

  $$|x(t_{2i-1}) - x(t_{2i})| \geq \eta, \qquad i = 1, \ldots, k.$$

The most remarkable feature of the S-topology, and the main reason for its use in probability theory, is represented by its very natural criteria of relative compactness (for the proof, see [9, sec. 2]).

Theorem 2.72. *A set K of elements of $D[0,1]$ is relatively (sequentially) compact with respect to the S-topology if and only if it is uniformly bounded*

$$\sup_{x \in K} \|x\| < \infty, \tag{2.86}$$

and at least one of the following equivalent conditions hold:

1. *for each $a < b$,*

   $$\sup_{x \in K} N^{a,b}(x) < \infty; \tag{2.87}$$

2. for each $\eta > 0$,

$$\sup_{x \in K} N_\eta(x) < \infty. \tag{2.88}$$

Note that the alternative conditions in (2.87) and (2.88) are precisely the additional requirements needed to guarantee that the construction of the candidate limit in (2.77) is well-defined. In fact, the S-topology is exactly the topology on $D[0,1]$ in which the sequence $\{ x_n \}$ converges to the candidate limit x obtained with such a procedure.

Proposition 2.73. *Let* $\mathbb{Q}$ *be a dense subset of* $[0,1]$ *containing* 1 *and let* $\{ x_n \}$ *be a relatively compact sequence of cadlag functions, with respect to the* S*-topology. If there exists an element* $x \in D[0,1]$ *such that*

$$x_n(q) \to x(q), \qquad \forall\, q \in \mathbb{Q},$$

then $\{ x_n \}$ *converges to* x *in the* S*-topology.*

It is interesting to point out that natural projections are useful for the identification of the limit even though they are *nowhere* continuous in the S-topology.

Indeed, for each natural projection π_t, $t \in [0,1)$ and each cadlag function $y \in D[0,1]$, consider the sequence $\{ x_n \}$ defined in Example 2.69 with $\tau = t$. It is straightforward to prove that

$$y + x_n \to_S y, \qquad (y + x_n)(t) = y(t) + 1 \not\to y(t),$$

which means that π_t is not continuous at each point $y \in D[0,1]$.

Other fundamental results about S-topology are summarized in the following theorem (proofs and further details can be found in [9, sec. 2]).

Theorem 2.74. *Consider the Skorohod space* $D[0,1]$ *equipped with the sequential topology* S*.*

- *The topological space* $(D[0,1], S)$ *is a Hausdorff space and cannot be metricized.*
- *There exists a countable family of* S*-continuous functions which separates points of* $D[0,1]$*, and therefore Condition 2.57 is satisfied.*
- *The Borel* σ*-field on* $D[0,1]$ *coincides with the* σ*-field generated by natural projections:*

$$\mathcal{B}_S = \mathcal{B}(D[0,1], S) = \sigma \{ \pi_t,\ t \in [0,1] \}.$$

A direct consequence of the form of the Borel σ-field $\mathcal{B}_S$ is that every probability measure on $(D[0,1], \mathcal{B}_S)$ is tight, and the notion of random element of $(D[0,1], \mathcal{B}_S)$ coincides with the notion of stochastic process with trajectories in $D[0,1]$.

The fact that Condition 2.57 holds in $(D[0,1], S)$ allows to consider on this topological space the notion of convergence for random elements on $D[0,1]$ devised by Jakubowski, and described in Section 2.6. In particular, Definition 2.62 of convergence in the sense of Jakubowski is restated here.

Definition 2.75. Let $(D[0,1], S)$ be the Skorohod space of cadlag functions equipped with the S-topology. A sequence $\{ X_n \}$ of random elements of $D[0,1]$ is said to converge in the sense of Jakubowski to a random element X, and is written as

$$X_n \overset{*}{\Rightarrow} X, \tag{2.89}$$

if every subsequence $\{ X_m \} \subseteq \{ X_n \}$ contains a further subsequence $\{ X_{m(k)} \}$ which admits a strong Skorohod representation with $Y \sim X$.

In the space $\mathcal{P}(D[0,1], \mathcal{B}_S)$ of (tight) probability measures on $(D[0,1], \mathcal{B}_S)$, equipped with convergence in the sense of Jakubowski, both the direct and the converse Prohorov theorems hold (see Section 2.6 and Theorem 2.60).

Theorem 2.76. *Let M be a family of stochastic processes with trajectories in $D[0,1]$. Then M is uniformly tight with respect to the S-topology if and only if it is relatively (sequentially) compact with respect to convergence in the sense of Jakubowski.*

As a consequence, we can work with probability measures on $(D[0,1], \mathcal{B}_S)$ much like they were defined on a Polish space, despite the fact that $(D[0,1], S)$ is not even metrizable.

The fact that natural projections, although nowhere continuous, are enough to uniquely identify the limit (Proposition 2.73), translates into the context of probability measures as the unique identification of the limit by means of finite-dimensional distributions. Therefore, an analogue of Theorem 2.42 holds, even if the mapping theorem can *never* be applied in this setting (see [9, sec. 3] for the proof).

Theorem 2.77. *Let $\mathbb{Q}$ be a dense subset of $[0,1]$ containing 1 and let $\{ X_n \}$ be a relatively compact sequence of stochastic processes with trajectories in $D[0,1]$. If finite-dimensional distributions of $\{ X_n \}$ converge to those of a stochastic process $X \in D[0,1]$ when restricted to $\mathbb{Q}$, then $X_n \overset{*}{\Rightarrow} X$: if, for every choice of $q_1, \dots, q_k \in \mathbb{Q}$,*

$$(X_n(q_1), \dots, X_n(q_k)) \Rightarrow_n (X(q_1), \dots, X(q_k)), \tag{2.90}$$

then $X_n \overset{*}{\Rightarrow} X$.

Remark 2.78. The construction of the candidate limit X, starting from its finite-dimensional distributions, is precisely the one described in Section 2.4 for a generic stochastic process on $D[0,1]$, independently of the chosen topology. This observation highlights once more the fundamental role of the Skorohod Representation Theorem in the theory of convergence in distribution.

A slighlty improved result is the following [9, sec. 3].

Theorem 2.79. *Let M be a relatively compact family of stochastic processes with trajectories in $D[0,1]$. Then there exist*

- *a sequence $\{ X_n \} \subseteq M$,*
- *a stochastic process X with trajectories in $D[0,1]$,*

- *a countable subset* $D \subseteq [0,1)$,

such that, for every choice of $t_1, \dots, t_k \in [0,1] \setminus D$,

$$(X_n(t_1), \dots, X_n(t_k)) \Rightarrow_n (X(t_1), \dots, X(t_k)), \tag{2.91}$$

and, in particular, $X_n \overset{*}{\Rightarrow} X$.

Similarly to the case of the Skorohod J_1-topology, the characterization of uniform tightness (or, equivalently, relative compactness) of a family of probability measures on $(D[0,1], \mathcal{B}_S)$ is obtained translating the characterization of relative compactness in $(D[0,1], S)$ (Theorem 2.72).

Theorem 2.80. *A family* $\{ X_\alpha \}$ *of stochastic processes with trajectories in* $D[0,1]$ *is uniformly tight with respect to the* S*-topology if and only if*

- $\{ \|X_\alpha\| \}$ *is a uniformly tight family of* $\mathbb{R}$*-valued random variables,*

and at least one of the following equivalent conditions hold

1. *for each* $a < b$, $\{ N^{a,b}(X_\alpha) \}$ *is a uniformly tight family of* $\mathbb{N}$*-valued random variables;*
2. *for each* $\eta > 0$, $\{ N_\eta(X_\alpha) \}$ *is a uniformly tight family of* $\mathbb{N}$*-valued random variables.*

To conclude this section, we highlight some facts which may hopefully help clarifying the relation between S-topology and J_1-topology on the Skorohod space $D[0,1]$ [11]:

- $D[0,1]$ equipped with the Skorohod metric $d*$, inducing the J_1-topology, is separable and complete;
- $D[0,1]$ equipped with the S-topology is a non-metrizable, Hausdorff topological space;
- J_1-convergence implies S-convergence,

  $$x_n \to_{J_1} x \qquad \text{implies} \qquad x_n \to_S x,$$

  and therefore the S-topology is *weaker* than the J_1-topology;
- addition is not sequentially continuous in the J_1-topology, i.e.

  $$x_n \to_{J_1} x, \quad y_n \to_{J_1} y \qquad \cancel{\text{implies}} \qquad x_n + y_n \to_{J_1} x + y;$$

- addition is sequentially continuous in the S-topology, i.e.

  $$x_n \to_S x, \quad y_n \to_S y \qquad \text{implies} \qquad x_n + y_n \to_S x + y.$$

The last observation is the reason for the typical phenomenon of self-cancelling oscillations in the S-topology (see Example 2.69), which is instead not allowed in the J_1-topology.

Chapter 3

Togashi-Kaneko Model

3.1 Original 4-dim model

In the study of biochemical reactions at cellular level, is not infrequent to encounter models for which an approximation based on classical scaling turns out to be unsatisfactory. In particular, the intrinsic discreteness of the stochastic process describing species counts cannot always be disregarded, even when normalized abundances correspond to species concentrations. A well-known example in cellular biology is the signal transduction process, in which even a very small number of molecules may trigger a biochemical cascade, eventually determining a switch in the state of a cell.

In [17], Togashi and Kaneko propose a chemical reaction system which displays this sort of peculiar behaviour. Its dynamics is characterized by switches between patterns where some species are present in very small or vanishing quantities, while others are abundant. Similarly to signal transduction processes, switches are triggered by the inflow of a single molecule belonging to a species which was previously extinct, and the switching phase involves a sequence of fast reactions.

Such a drastic effect of a single molecule inflow on the system dynamics is highly non-linear (a negligible perturbation is extremely amplified) and cannot be captured by the classical continuous approximation. In fact, switches are intrinsically tied with the underling discrete structure of the stochastic model, and, for this reason, are called *Discreteness Induced Transitions* [3].

The general Togashi-Kaneko model involves species $S_1, \dots, S_n$ and is characterized by two classes of reactions:

- (fast) autocatalytic reactions,

$$R_i^A: \quad S_i + S_{i+1} \to 2S_{i+1}, \qquad i = 1, \dots, n, \tag{3.1}$$

  with the convention that $S_{n+1} = S_1$;

- (slow) inflow and outflow reactions,

$$R_i^I: \quad \emptyset \to S_i, \qquad R_i^O: \quad S_i \to \emptyset, \qquad i = 1, \dots, n. \tag{3.2}$$

Autocatalytic reactions are very common in this type of cellular processes and are actually those responsible for the amplification effect which causes switches.

According to mass action kinetics (Definition 1.6), reaction rates are expressed as:

$$\begin{aligned} \lambda_i^A(x) &= \kappa_i x_i x_{i+1}, \qquad i = 1, \dots, n, \\ \lambda_i^I(x) &= \eta_i, \qquad \lambda_i^O(x) = \delta_i x_i, \qquad i = 1, \dots, n. \end{aligned} \tag{3.3}$$

As highlighted by the form of intensity functions, autocatalytic reaction R_i^A may happen only if both S_i and S_{i+1} are together present in the system, that is, transitions are triggered by an inflow of a species which was previously extinct (thus preventing the autocatalytic cascade to start).

The number of molecules of species S_i at time $t \geq 0$ is described by the following equation:

$$\begin{aligned} X_i(t) = X_i(0) &+ \xi_{i-1}^A \left(\int_0^t \kappa_{i-1} X_{i-1}(s) X_i(s) ds \right) \\ &- \xi_i^A \left(\int_0^t \kappa_i X_i(s) X_{i+1}(s) ds \right) \\ &+ \xi_i^I \left(\int_0^t \eta_i ds \right) - \xi_i^O \left(\int_0^t \delta_i X_i(s) ds \right). \end{aligned} \tag{3.4}$$

Notice that, assuming no autocatalytic cascade involving species S_i is running, either S_i is extinct or both S_{i-1} and S_{i+1} are extinct. This observation suggests some sort of alternation pattern between extinct and abundant species, which is confirmed by numerical experiments when autocatalytic reactions are sufficiently faster than inflows and outflows.

In their article [17], Togashi and Kaneko consider the case for $n = 4$ and approximate the system dynamics according to the classical scaling procedure (see Section 1.3):

- species abundances are expressed in terms of their concentrations,

$$C_i^N = \frac{X_i}{N}, \qquad i = 1, \dots, n; \tag{3.5}$$

- reaction rate constants for autocatalytic reactions are given by

$$\kappa_i = \frac{\kappa_i'}{N}, \qquad i = 1, \dots, n; \tag{3.6}$$

- reaction rate constants for inflow and outflow reactions are expressed as

$$\eta_i = \eta_i' N, \qquad \delta_i = \delta_i', \qquad i = 1, \dots, n. \tag{3.7}$$

For further simplification, the authors assume

$$\kappa_i' = 1, \qquad \eta_i' = \delta_i' = D, \qquad \forall\, i = 1, \dots, n. \tag{3.8}$$

Therefore, under the above assumptions, the concentration of species S_i at time $t \geq 0$ is given by

$$\begin{aligned} C_i^N(t) = C_i^N(0) &+ N^{-1}\xi_{i-1}^A \left(N \int_0^t C_{i-1}^N(s) C_i^N(s) ds \right) \\ &- N^{-1}\xi_i^A \left(N \int_0^t C_i^N(s) C_{i+1}^N(s) ds \right) \\ &+ N^{-1}\xi_i^I \left(N \int_0^t D ds \right) - N^{-1}\xi_i^O \left(N \int_0^t D\, C_i^N(s) ds \right). \end{aligned} \tag{3.9}$$

The limit as N grows large is obtained according to Theorem 1.21. Indeed, assuming that vector $C^N(0)$ has a finite (positive) limit for $N \to \infty$,

$$\lim_{N\to\infty} C^N(0) = c(0) > 0, \tag{3.10}$$

the stochastic solution of (3.9) converges to the deterministic solution of equation

$$\begin{aligned} c_i(t) = c_i(0) &+ \int_0^t c_{i-1}(s) c_i(s) ds - \int_0^t c_i(s) c_{i+1}(s) ds \\ &+ \int_0^t D ds - \int_0^t D\, c_i(s) ds, \qquad t \geq 0. \end{aligned} \tag{3.11}$$

The resulting system of equations can be restated in differential form as

$$\begin{aligned} \dot{c}_i &= c_{i-1}\, c_i - c_i\, c_{i+1} + D - D\, c_i \\ &= c_i(c_{i-1} - c_{i+1}) + D(1 - c_i), \qquad i = 1, \ldots, 4. \end{aligned} \tag{3.12}$$

It is easy to show that the ordinary differential system above has an equilibrium for

$$c_i = c_i^e = 1, \qquad i = 1, \ldots, n, \tag{3.13}$$

and such equilibrium is asymptotically stable, since the Jacobian matrix, evaluated at this point, has eigenvalues

$$\lambda_1 = \lambda_3 = -D, \qquad \lambda_2 = -D - 2i, \qquad \lambda_4 = -D + 2i.$$

Note that the presence of two complex conjugate eigenvalues makes the equilibrium point an attractive focus in the subspace generated by the two corresponding eigenvectors,

$$x^{(2)} = \begin{bmatrix} 1 & i & -1 & -i \end{bmatrix}, \qquad x^{(4)} = \begin{bmatrix} 1 & -i & -1 & i \end{bmatrix},$$

i.e. the subspace on which the sums of species counts for S_1 and S_3 and for S_2 and S_4, respectively, are constant, showing again an alternation pattern between species.

Numerical experiments are carried out setting the scaling parameter to

$$N = 256,$$

and initializing species concentrations at the equilibrium point $c_i^e = 1$, $i = 1, \ldots, 4$; in order to model slow inflow and outflow reactions, their common normalized reaction rate constant is set to

$$D = 1/16.$$

Results are displayed in Figure 3.1:

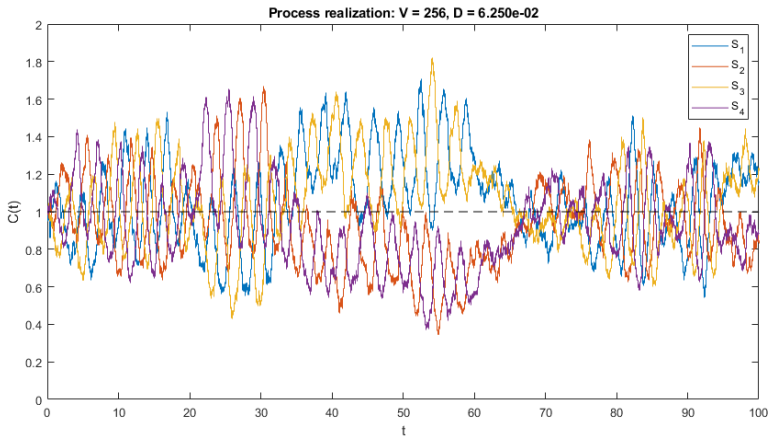

Figure 3.1: Realization of stochastic process C^N of species concentrations for the Togashi-Kaneko model with $n = 4$ species. Parameters are set to $N = 256$ and $D = 1/16$. Trajectories fluctuate around the stable equilibrium.

- species concentrations fluctuate around the deterministic stable equilibrium;
- species S_1 and S_3 (respectively, S_2 and S_4) are paired and oscillate around the attractive focus in the subspace generated by $x^{(2)}$ and $x^{(4)}$, with frequency $f = 1/\pi$.

However, if the diffusion parameter D is progressively decreased, making inflows and outflows more and more rare, the anticipated phenomenon of Discreteness Induced Transitions appears; setting, for example,

$$D = 1/4096,$$

we obtain results as displayed in Figure 3.2 (see also [17, fig. 1]).

In the initial transition phase:

- species S_3 happens to become extint;
- molecules of species S_2 are produced but not consumed (since the autocatalytic reaction consuming molecules of S_2 does not happen, being S_3 extinct);
- at the same time, molecules of S_1 are consumed, until S_1 also becomes extinct;
- the system enters a configuration in which both S_1 and S_3 are extinct, so that all autocatalytic reactions are stopped, concentration of S_2 is high, while concentration of S_4 is low (call this pattern $2H4L$ [3]).

In the long run:

- autocatalytic reactions are stopped most of the time, and are triggered by the inflows of molecules of extinct species;

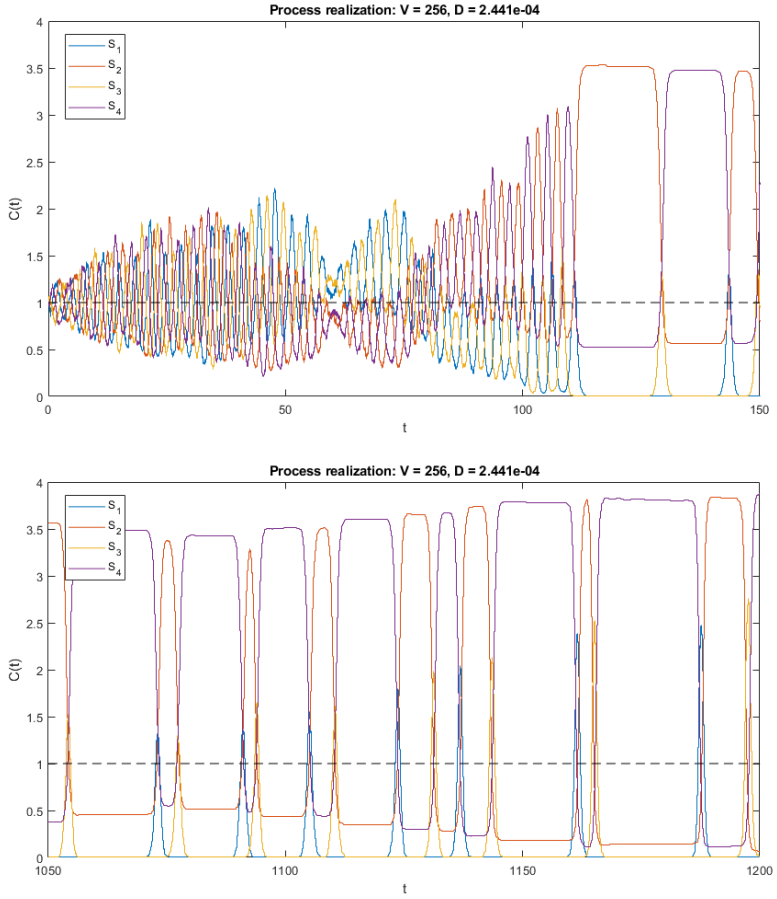

Figure 3.2: Realization of stochastic process C^N of species concentrations for the Togashi-Kaneko model with $n = 4$ species. Parameters are set to $N = 256$ and $D = 1/4096$. Discreteness Induced Transitions appear.

- the system alternates four different configurations, each with a pair of extinct species (S_1 and S_3 or S_2 and S_4) and a pair of species which are present with higher and lower concentrations, respectively;
- switches between configurations are driven by autocatalytic reactions, and a pattern like $2H4L$ is far more often followed by pattern $2L4H$ than by the other opposite configurations $1H3L$ and $1L3H$.

Clearly, results displayed in Figure 3.2 do not contradict the theoretical result in (3.11): for N growing large, the system dynamics converges to the deterministic model for each value of D. However, these numerical experiments suggest that, with a different scaling of reaction rate constants, it is possible to obtain limiting models featuring Discreteness Induced Transitions.

3.2 Simplified 2-dim model

In order to simplify the analysis of Discreteness Induced Transitions and obtain meaningful results, in the following sections we consider the case with $n = 2$.

The reaction network is represented by the following complex graph

$$\begin{aligned} S_1 + S_2 \to 2S_2, \qquad & S_1 + S_2 \to 2S_1, \\ S_1 \rightleftharpoons \emptyset, \qquad & S_2 \rightleftharpoons \emptyset, \end{aligned} \tag{3.14}$$

and reation rates are again given by mass action kinetics

$$\begin{aligned} \lambda_i^A(x) = \kappa_i x_1 x_2, \qquad & i = 1, 2, \\ \lambda_i^I(x) = \eta_i, \qquad \lambda_i^O(x) = \delta_i x_i, \qquad & i = 1, 2. \end{aligned} \tag{3.15}$$

Similarly to the original case with $n = 4$, approximate the system dynamics according to the classical scaling:

$$C_i^N = \frac{X_i}{N}, \qquad \kappa_i = \frac{\kappa_i'}{N}, \qquad \eta_i = \eta_i' N, \qquad \delta_i = \delta_i', \qquad i = 1, 2, \tag{3.16}$$

and further assume that

$$\kappa_i' = 1, \qquad \eta_i' = \delta_i' = D, \qquad \forall\, i = 1, 2. \tag{3.17}$$

As N grows large, the stochastic solution converges to the deterministic solution of equations

$$\dot{c}_i = D(1 - c_i), \qquad i = 1, 2. \tag{3.18}$$

In this case, deterministic equations are disjoint, and there exists an asymptotically stable equilibrium for $c_i^e = 1$, $i = 1, 2$, since the Jacobian matrix is diagonal with diagonal entries equal to $-D$.

Again, numerical experiments are carried out setting the scaling parameter to

$$N = 256,$$

and initializing species concentrations at equilibrium $c_i^e = 1$, $i = 1, 2$. Figure 3.3 shows results obtained setting the diffusion parameter to

$$D = 1/16.$$

Exactly as in the case with $n = 4$, trajectories fluctuate around the deterministic stable equilibrium; no periodic oscillations are observed, since the Jacobian matrix has no complex eigenvalues and the equilibrium is a stable node.

On the contrary, Figure 3.4 displays results obtained for

$$D = 1/4096.$$

The dynamics is here extremely simplified:

- species S_1 becomes extinct, preventing autocatalytic reactions to happen;
- inflows of molecules of S_1 trigger autocatalytic cascades, which occasionally lead to a switch in the roles of the two species, i.e. S_2 is extinct and S_1 has high concentration;

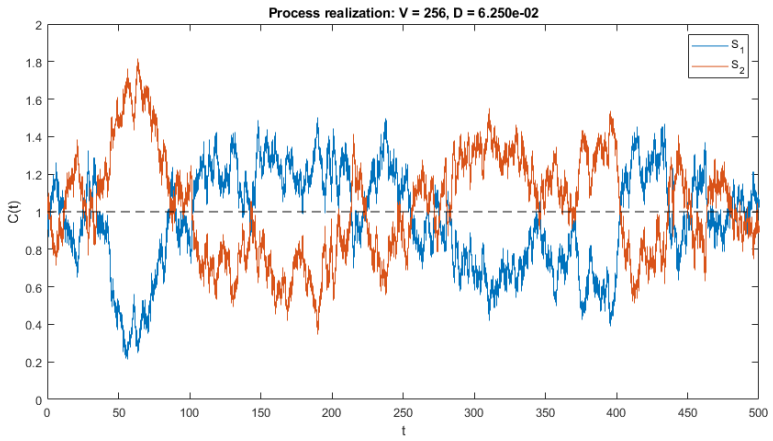

Figure 3.3: Realization of stochastic process C^N of species concentrations for the Togashi-Kaneko model with $n = 2$ species. Parameters are set to $N = 256$ and $D = 1/16$. Trajectories fluctuate around the stable equilibrium, since $ND = 16$.

- autocatalytic cascades frequently fail to produce a switch, and result instead in peculiar peaks, reaching various levels between 0 and the value taken by the sum process;
- the sum of the two species, denoted as

$$W := C_1 + C_2, \tag{3.19}$$

  is not affected by autocatalytic reactions, and fluctuates around its equilibrium point $w^e = c_1^e + c_2^e = 2$.

In order to have a better understanding of the behaviour of the 2-dim Togashi-Kaneko reaction system, it is interesting to look at its stationary distribution. Assume that network (3.14) has reaction rate constants given by

$$\kappa_1 = \kappa_2 = \kappa, \qquad \delta_1 = \delta_2 = \delta. \tag{3.20}$$

It is proved in [3] that there exists a unique stationary distribution π which factorizes as

$$\pi(x) = \mu(x_1|n)\,\nu(n), \tag{3.21}$$

where $n = x_1 + x_2$ is the total number of molecules in the system. In particular:

- ν is a Poisson distribution,

$$\nu(n) = \frac{\gamma^n}{n!}\exp(-\gamma), \qquad n \geq 0, \tag{3.22}$$

  with parameter

$$\gamma = \frac{\eta_1 + \eta_2}{\delta},$$

  and is the stationary distribution for the stochastic process describing the sum of the two species counts;

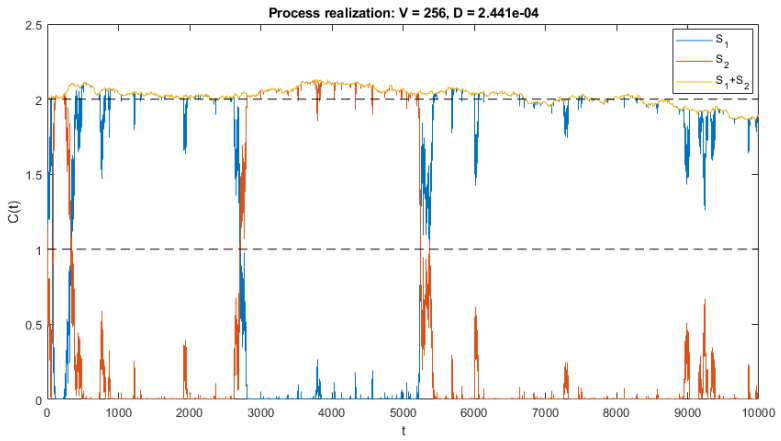

Figure 3.4: Realization of stochastic process C^N of species concentrations for the Togashi-Kaneko model with $n = 2$ species. Parameters are set to $N = 256$ and $D = 1/4096$. Discreteness Induced Transitions appear, since $ND = 1/16$.

- $\mu(\cdot|n)$ is a beta-binomial distribution,

$$\mu(i|n) = \binom{n}{i} \frac{B(i+\alpha, n-i+\beta)}{B(\alpha,\beta)}, \qquad i = 0, \dots, n, \tag{3.23}$$

  where B is the beta function, with parameters

$$\alpha = \frac{\delta\eta_1}{\kappa(\eta_1+\eta_2)}, \qquad \beta = \frac{\delta\eta_2}{\kappa(\eta_1+\eta_2)},$$

  and represents the number of moleculs of S_1, compared to S_2, given their sum n.

Note that, under the further assumption

$$\eta_1 = \eta_2 = \eta,$$

the parameters α and β of the beta-binomial distribution coincide and no more depend on η:

$$\alpha = \beta = \frac{\delta}{2\kappa}. \tag{3.24}$$

This form for the stationary distribution can be checked by direct verification; further details and explicit calculations can be found in [3].

The beta-binomial distribution may assume different shapes depending on the parameter values (see Figure 3.5):

- if $\alpha, \beta > 1$, the distribution is unimodal, and symmetric if $\alpha = \beta$;
- if $\alpha, \beta < 1$, the distribution is bimodal, with probability mass concentrated at boundaries;

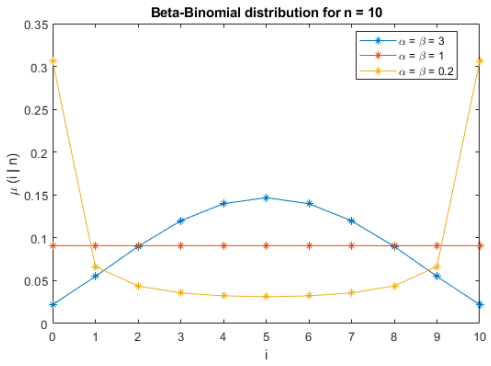

Figure 3.5: Beta-Binomial distribution for $n = 10$.

- if $\alpha = \beta = 1$, the distribution reduces to the discrete uniform on $\{0, \ldots, n\}$.

The explicit form for the stationary distribution for reaction system (3.14) turns out to be extremely useful if we consider again the approximation under classical scaling, described in (3.16), with reaction rate constants given by (3.17).

In particular, parameters α and β of the beta-binomial distribution coincide and are equal to

$$\alpha = \beta = \frac{ND}{2}; \tag{3.25}$$

as a consequence, for large N, the behaviour of the system dynamics in the long run is determined by the quantity ND:

- for $ND >> 1$, trajectories fluctuate around the deterministic equilibrium, since the beta-binomial density is unimodal and symmetric, with a sharp peak around $n/2$ (see Figure 3.3, where $ND = 16$);
- for $ND << 1$, Discreteness Induced Transitions appears, since the beta-binomial distribution is bimodal, with probability mass concentrated in 0 and n (see Figure 3.4, where $ND = 1/16$).

It can be actually shown that, for $ND \to 0$, the probability mass of the beta-binomial distribution is confined at the boundaries, for each n:

$$\mu(0, n) + \mu(n, n) \to 1, \qquad ND \to 0. \tag{3.26}$$

On the other hand, the parameter γ of the Poisson distribution ν for the sum of the two species counts is given by

$$\gamma = 2N; \tag{3.27}$$

for large N, the normalized sum $W = C_1 + C_2$ concentrates around its equilibrium point $w^e = 2$, since

$$\mathbb{E}[W] = \frac{\gamma}{N} = 2, \qquad \mathrm{Var}(W) = \frac{\gamma}{N^2} = \frac{2}{N} \to 0. \tag{3.28}$$

The following sections are devoted to further investigate properties of the 2-dim Togashi-Kaneko model in the limit for N growing large, preserving, at the same time, such peculiar Discreteness Induced Transitions. To this end, classical scaling procedure is set aside, and reaction rates constants are scaled so that

$$\alpha = \beta = \frac{\delta}{2\kappa} \to 0, \qquad \text{as} \quad N \to \infty. \tag{3.29}$$

3.3 Multiscale approximations

The chemical reaction system proposed by Togashi and Kaneko, in the case with $n = 2$, is represented by the complex graph

$$\begin{gathered} S_1 + S_2 \to 2S_2, \qquad S_1 + S_2 \to 2S_1, \\ S_1 \rightleftharpoons \emptyset, \qquad S_2 \rightleftharpoons \emptyset. \end{gathered} \tag{3.14}$$

with reactions rates given by

$$\begin{gathered} \lambda_i^A(x) = \kappa_i x_1 x_2, \qquad i = 1, 2, \\ \lambda_i^I(x) = \eta_i, \qquad \lambda_i^O(x) = \delta_i x_i, \qquad i = 1, 2. \end{gathered} \tag{3.15}$$

The stochastic process X describing species counts is modeled as the continuous time Markov chain solving equations, for $t \geq 0$,

$$\begin{aligned} X_1(t) = X_1(0) &- \xi_1^A \left(\int_0^t \kappa_1 X_1(s) X_2(s) ds \right) \\ &+ \xi_2^A \left(\int_0^t \kappa_2 X_1(s) X_2(s) ds \right) \\ &+ \xi_1^I(\eta_1 t) - \xi_1^O \left(\int_0^t \delta_1 X_1(s) ds \right), \end{aligned} \tag{3.30}$$

$$\begin{aligned} X_2(t) = X_2(0) &+ \xi_1^A \left(\int_0^t \kappa_1 X_1(s) X_2(s) ds \right) \\ &- \xi_2^A \left(\int_0^t \kappa_2 X_1(s) X_2(s) ds \right) \\ &+ \xi_2^I(\eta_2 t) - \xi_2^O \left(\int_0^t \delta_2 X_2(s) ds \right). \end{aligned} \tag{3.31}$$

As discussed in the Section 3.2, the classical scaling procedure has to be set aside in order to preserve Discreteness Induced Transitions. However, some of the assumptions considered there should be kept not to loose fundamental properties of the model:

- species abundances are expressed in terms of their concentrations,

$$C_i^N = \frac{X_i}{N}, \qquad i = 1, \ldots, n; \tag{3.32}$$

- reaction rates constants coincide for the two species,

$$\kappa_1 = \kappa_2 = \kappa, \qquad \eta_1 = \eta_2 = \eta, \qquad \delta_1 = \delta_2 = \delta; \tag{3.33}$$

- the relationship between reaction rate constants for inflows and outflows remains fixed,
$$\frac{\eta}{\delta} = N, \tag{3.34}$$
since this relationship determines the behaviour of the process $W = C_1 + C_2$.

To keep coherence with the previous discussion, we may arbitrarily assume that the reaction rate constant for autocatalytic reactions scales as

$$\kappa = \frac{\kappa'}{N}; \tag{3.35}$$

reaction rate constants for inflow and outflow reactions are scaled according to generic scaling parameters β_I and β_O,

$$\eta = \eta' N^{\beta_I}, \qquad \delta = \delta' N^{\beta_O}, \tag{3.36}$$

with the constraint that $\beta_I = \beta_O + 1 = \beta$.

Therefore, the stochastic process C^N of normalized abundances is described by the following equations:

$$\begin{aligned} C_1^N(t) = C_1^N(0) &- N^{-1}\xi_1^A\left(N\int_0^t \kappa' C_1^N(s)C_2^N(s)ds\right) \\ &+ N^{-1}\xi_2^A\left(N\int_0^t \kappa' C_1^N(s)C_2^N(s)ds\right) \\ &+ N^{-1}\xi_1^I(N^\beta\eta' t) - N^{-1}\xi_1^O\left(N^\beta\int_0^t \delta' C_1^N(s)ds\right), \end{aligned} \tag{3.37}$$

$$\begin{aligned} C_2^N(t) = C_2^N(0) &+ N^{-1}\xi_1^A\left(N\int_0^t \kappa' C_1^N(s)C_2^N(s)ds\right) \\ &- N^{-1}\xi_2^A\left(N\int_0^t \kappa' C_1^N(s)C_2^N(s)ds\right) \\ &+ N^{-1}\xi_2^I(N^\beta\eta' t) - N^{-1}\xi_2^O\left(N^\beta\int_0^t \delta' C_2^N(s)ds\right), \end{aligned} \tag{3.38}$$

For the continuation of the discussion, assume that

$$\lim_{N\to\infty} C^N(0) = c(0) > 0. \tag{3.39}$$

For $\beta = 1$, we resort to classical scaling, and the stochastic process C^N converges to the solution c of the deterministic system of equations,

$$c_i(t) = c_i(0) + \eta t - \int_0^t \delta c_i(s)ds, \qquad i = 1, 2; \tag{3.40}$$

the two equations are disjoint, and their solution is given by

$$c_i(t) = \left(c_i(0) - \frac{\eta'}{\delta'}\right)\exp\left(-\delta' t\right) + \frac{\eta'}{\delta'}, \qquad i = 1, 2. \tag{3.41}$$

Note that, if $\eta' = \delta'$ as assumed in the previous sections, the asymptotically stable equilibrium is $c_i^e = 1$, $i = 1, 2$.

However, as observed at the end of Section 3.2, in order to preserve Discreteness Induced Transitions for N growing large, we need the parameter(s) $\alpha = \beta$ of the beta-binomial distribution in (3.23) to converge to zero as N grows, i.e. we should require

$$\frac{\delta}{2\kappa} = \frac{\delta'}{2\kappa'} N^{\beta} \to 0, \qquad \text{as} \quad N \to \infty, \tag{3.42}$$

and therefore we must take

$$\beta = -\alpha, \qquad \alpha > 0. \tag{3.43}$$

Moreover, we may also consider a scaling in time, expressed as

$$\tau = N^{-\gamma} t, \qquad t \geq 0. \tag{3.44}$$

As a consequence, the time-scaled stochastic process $C^{N,\gamma}$ is described by the equations

$$\begin{aligned} C_1^{N,\gamma}(\tau) = C_1^N(0) &- N^{-1}\xi_1^A \left(N^{1+\gamma} \int_0^\tau \kappa' C_1^{N,\gamma}(s) C_2^{N,\gamma}(s) ds \right) \\ &+ N^{-1}\xi_2^A \left(N^{1+\gamma} \int_0^\tau \kappa' C_1^{N,\gamma}(s) C_2^{N,\gamma}(s) ds \right) \\ &+ N^{-1}\xi_1^I (N^{\gamma-\alpha}\eta'\tau) - N^{-1}\xi_1^O \left(N^{\gamma-\alpha} \int_0^\tau \delta' C_1^{N,\gamma}(s) ds \right), \end{aligned} \tag{3.45}$$

$$\begin{aligned} C_2^{N,\gamma}(\tau) = C_2^N(0) &+ N^{-1}\xi_1^A \left(N^{1+\gamma} \int_0^\tau \kappa' C_1^{N,\gamma}(s) C_2^{N,\gamma}(s) ds \right) \\ &- N^{-1}\xi_2^A \left(N^{1+\gamma} \int_0^\tau \kappa' C_1^{N,\gamma}(s) C_2^{N,\gamma}(s) ds \right) \\ &+ N^{-1}\xi_2^I (N^{\gamma-\alpha}\eta'\tau) - N^{-1}\xi_2^O \left(N^{\gamma-\alpha} \int_0^\tau \delta' C_2^{N,\gamma}(s) ds \right), \end{aligned} \tag{3.46}$$

It is easy to check that the balance equation (1.50) is verified for both species: considering the fastest reactions (i.e the autocatalytic reactions), the rate of production has the same order of magnitude of the rate of consumption. Moreover, both species have the same natural time-scale

$$\gamma_i = 1 - \max\{\, 1, -\alpha \,\} = 0, \qquad i = 1, 2,$$

and therefore the first natural time-scale for the (normalized) reaction system is

$$\gamma' = \min_i \gamma_i = 0. \tag{3.47}$$

For $\gamma = \gamma'$, the processes of normalized reaction countings for autocatalytic reactions converge to their deterministic equivalent,

$$N^{-1}\xi_1^A \left(N \int_0^\tau \kappa' C_1^{N,\gamma'}(s) C_2^{N,\gamma'}(s) ds \right) \;\to\; \int_0^\tau \kappa' C_1^{N,\gamma'}(s) C_2^{N,\gamma'}(s) ds,$$

while reaction countings for inflows and outflows converge to zero,

$$\xi_i^I (N^{-\alpha}\eta'\tau) \;\to\; 0, \qquad \xi_i^O \left(N^{-\alpha} \int_0^\tau \delta' C_i^{N,\gamma'}(s) ds \right) \;\to\; 0, \qquad i = 1, 2.$$

As a consequence, the stochastic process $C^{N,\gamma'}$ of species concentrations converges to the solution c of the deterministic system of equations (Theorem 1.31):

$$\begin{aligned} c_i(\tau) &= c_i(0) - \int_0^\tau \kappa' c_1(s)c_2(s)ds + \int_0^\tau \kappa' c_1(s)c_2(s)ds \\ &= c_i(0), \qquad i = 1, 2. \end{aligned} \tag{3.48}$$

In other words, the limiting process at the first natural time-scale is the constant process, that is, nothing happens at this time-scale.

It is interesting to note that the first natural time-scale is defined so that none of the normalized reaction countings should blow up and at least one should be non trivial, as it actually happens in this case (Definition 1.30). However, despite such characterization, the fact that normalized countings for autocatalytic reactions are somewhat perfectly balanced results in a constant dynamics for the limiting model.

Considering linear combinations of the two species (see Section 1.6), the only combination $\theta \cdot S$ which raises some interest is obtained for $\theta = (1, 1)$, corresponding to the sum of their counts. The normalized abundance (concentration) of the sum process is given by,

$$W^{N,\gamma} := C_1^{N,\gamma} + C_2^{N,\gamma}, \tag{3.49}$$

and the stochastic process $W^{N,\gamma}$ satisfies equation

$$\begin{aligned} W^{N,\gamma}(\tau) = W^N(0) &+ N^{-1}\xi_1^I(N^{\gamma-\alpha}\eta'\tau) + N^{-1}\xi_2^I(N^{\gamma-\alpha}\eta'\tau) \\ &- N^{-1}\xi_1^O\left(N^{\gamma-\alpha}\int_0^\tau \delta' C_1^{N,\gamma}(s)ds\right) \\ &- N^{-1}\xi_2^O\left(N^{\gamma-\alpha}\int_0^\tau \delta' C_2^{N,\gamma}(s)ds\right). \end{aligned} \tag{3.50}$$

Note that the balance equation (1.64) is also verified for this linear combination, and thus for each linear combination $\theta \in \mathbb{R}^n_{\geq 0}$.

The natural time-scale for the concentration of the sum is

$$\gamma_\theta = 1 - \max\{-\alpha\} = 1 + \alpha,$$

which is also the second natural time-scale,

$$\gamma'' = \min\{\gamma_{\bar\theta} \colon \ \gamma_{\bar\theta} > \gamma'\} = \gamma_\theta = 1 + \alpha, \tag{3.51}$$

since $\gamma_{\bar\theta} = \gamma' = 0$ for all $\bar\theta \neq \theta$.

As discussed in Section 1.7, convergence of (3.50) to a non-degenerate model is not guaranteed a priori at the second natural time-scale, at least in the general case. Indeed, some additional requirements concerning stability of the reaction system on its fast subspace should be satisfied, in order to average the influence of fast components on the slow subspace.

However, in this particular example, it is possible to bypass this issue relying on a peculiar property of this model, namely *lumpability*.

Consider a continuous time Markov chain $X = \{X(t) \colon t \geq 0\}$ on state space $\mathbb{S}$ with transition rates $Q(x, y)$, $x, y \in \mathbb{S}$. Let $\{\mathcal{P}_i \colon i \in \mathcal{I}\}$ be a partition of the

state space $\mathbb{S}$ and let $\mathcal{L}$ be the function mapping each element $x \in \mathbb{S}$ into the index of the partition it belongs to:

$$\mathcal{L}\colon \mathbb{S} \mapsto \mathcal{I}, \qquad \mathcal{L}(x) = i \quad \text{if and only if} \quad x \in \mathcal{P}_i.$$

Definition 3.1. A continous time Markov chain $\{ X(t)\colon t \geq 0 \}$ on state space $\mathbb{S}$ is called *(strongly) lumpable* with respect to a partition $\{ \mathcal{P}_i \}$ if the *lumped* process $\bar{X}$, defined as

$$\bar{X}(t) = \mathcal{L}(X(t)), \qquad t \geq 0,$$

is itself a continuous time Markov chain, for any choice of the initial distribution.

A necessary condition for lumpability is that, for every $i, j \in \mathcal{I}$ and every $x, y \in \mathcal{P}_i$,

$$\sum_{z \in \mathcal{P}_j} Q(x, z) = \sum_{z \in \mathcal{P}_j} Q(y, z) := \bar{Q}(i, j); \tag{3.52}$$

the quantities $\bar{Q}(i, j)$ are the transition rates for the lumped process $\bar{X}$.

With these concepts in mind, consider the stochastic process X of species counts, introduced at the beginning of this section, which is a continuous time Markov chain on the state space $\mathbb{S} = \mathbb{N}_0^2$, and let $\{ \mathcal{P}_n\colon n \in \mathbb{N} \}$ be the partition of $\mathbb{S}$ defined as

$$\mathcal{P}_n := \{ x \in \mathbb{S}\colon x_1 + x_2 = n \}. \tag{3.53}$$

It is proved in [3] that process X is lumpable with respect to partition $\{ \mathcal{P}_n \}$, under the assumptions in (3.33).

The lumped process $\bar{X}$ describes the total number of molecules in the system, since, by definition of $\{ \mathcal{P}_n \}$,

$$\bar{X}(t) = n \quad \text{if and only if} \quad X_1(t) + X_2(t) = n.$$

In particular, its transition rates are obtained according to (3.52):

- the rate at which the total number of species is increased by one unit is equal to the sum of inflow reaction rates of the two species:

$$\bar{Q}(n, n+1) = \lambda^I(n) := \eta_1 + \eta_2 = 2\eta, \qquad n \geq 0;$$

- the rate at which the total number of species is decreased by one unit is equal to the sum of outflow reaction rates of the two species:

$$\bar{Q}(n, n-1) = \lambda^O(n) := \delta_1 x_1 + \delta_2 x_2 = \delta n, \qquad n \geq 1.$$

Note that the transition rate (i.e. intensity function) for inflows is constant, while the transition rate for outflows depends on state x only through $n = x_1 + x_2$, highlighting the essential property which makes lumped processes Markov chains.

Remark 3.2. A consequence of the definition of partition $\{ \mathcal{P}_n \}$ is that autocatalytic reactions can be completely disregarded when considering the lumped process. Indeed, such reactions do not affect the total number of species in the system, and therefore the set of states visited during an autocatalytic sequence is contained in a single element of the partition.

Lumpability is not affected by the scaling procedure, since reaction rates for inflows (respectively, outflows) of the two species are multiplied by the same constant, which includes scalings of state space, reation rate constants and time. In particular, after rescaling, reaction rates can be rewritten as

$$\lambda^I(w) = N^{\gamma-\alpha}2\eta', \qquad \lambda^O(w) = N^{\gamma-\alpha}\delta' w.$$

Therefore, the stochastic process $W^{N,\gamma}$ describing the concentration of the sum of the two species is itself a continuous time Markov chain satisfying equation

$$\begin{aligned} W^{N,\gamma}(\tau) = W^N(0) &+ N^{-1}\xi^I(N^{\gamma-\alpha}2\eta'\tau) \\ &- N^{-1}\xi^O\left(N^{\gamma-\alpha}\int_0^\tau \delta' W^{N,\gamma}(s)ds\right). \end{aligned} \tag{3.54}$$

where ξ^I and ξ^O are independent unit Poisson processes.

It is interesting to observe how expression (3.54) seems to derive directly from (3.50) as a trivial consequence of a well-know property of Poisson processes.

Proposition 3.3 (Superposition property)**.** *Let $Y_1, \ldots, Y_l$ be Poisson processes with intensity functions $\lambda_1, \ldots, \lambda_l$, respectively, that is*

$$Y_k(t) := \xi_k\left(\int_0^t \lambda_k(s)\,ds\right), \qquad t \geq 0,$$

where $\xi_1, \ldots, \xi_l$ are independent unit Poisson processes. Then, the process Y defined as

$$Y(t) = \sum_k Y_k(t), \qquad t \geq 0,$$

is a Poisson process with intensity function

$$\lambda(t) = \sum_k \lambda_k(t), \qquad t \geq 0.$$

In light of this property, it should be reasonable to express reaction countings for inflows as a single counting process

$$\begin{aligned} Y^I(\tau) &= Y_1^I(\tau) + Y_2^I(\tau) \\ &= \xi_1^I(N^{\gamma-\alpha}\eta'\tau) + \xi_2^I(N^{\gamma-\alpha}\eta'\tau) \\ &= \xi^I(N^{\gamma-\alpha}2\eta'\tau), \end{aligned} \tag{3.55}$$

while reaction countings for outflows may be rewritten as

$$\begin{aligned} Y^O(\tau) &= Y_1^O(\tau) + Y_2^O(\tau) \\ &= \xi_1^O\left(N^{\gamma-\alpha}\int_0^\tau \delta' C_1^{N,\gamma}(s)ds\right) \\ &\quad + \xi_2^O\left(N^{\gamma-\alpha}\int_0^\tau \delta' C_2^{N,\gamma}(s)ds\right) \\ &= \xi^O\left(N^{\gamma-\alpha}\int_0^\tau \delta'\left(C_1^{N,\gamma}(s) + C_2^{N,\gamma}(s)\right)ds\right) \\ &= \xi^O\left(N^{\gamma-\alpha}\int_0^\tau \delta' W^{N,\gamma}(s)ds\right). \end{aligned} \tag{3.56}$$

This approach seems correct in the case of inflows, since reaction rates do not depend on the state, i.e. are constant. On the contrary, reaction rates for outflow reactions depend on the concentrations of the two species, which in turn are not independent from each other. Therefore, it is not clear a priori if the superposition property could be applied in this context.

Regardless of whether (3.54) derives from lumpability or superposition, at the second natural time-scale $\gamma'' = 1 + \alpha$, it becomes

$$\begin{aligned} W^{N,\gamma''}(\tau) = W^N(0) &+ N^{-1}\xi^I(N\,2\eta'\tau) \\ &- N^{-1}\xi^O\left(N\int_0^\tau \delta' W^{N,\gamma''}(s)ds\right). \end{aligned} \tag{3.57}$$

By Theorem 1.21, under the assumption stated in (3.39), that is

$$\lim_{N\to\infty} W^N(0) = w(0) = c_1(0) + c_2(0) > 0, \tag{3.58}$$

the process $W^{N,\gamma''}$ converges to the solution w of the deterministic equation,

$$w(\tau) = w(0) + 2\eta'\tau - \int_0^\tau \delta' w(s)ds, \qquad \tau \geq 0, \tag{3.59}$$

which is given by

$$w(\tau) = \left(w(0) - \frac{2\eta'}{\delta'}\right)\exp\left(-\delta'\tau\right) + \frac{2\eta'}{\delta'}, \qquad \tau \geq 0. \tag{3.60}$$

Note that, for $\eta' = \delta'$, the asymptotically stable equilibrium is $w^e = 2$.

Numerical experiments are carried out in order to confirm empirically results discussed above. In particular, parameters are set to the following values:

$$\kappa' = \eta' = \delta' = 1, \qquad \alpha = 1, \qquad c_1(0) = c_2(0) = 2.$$

Figures 3.6 and 3.7 display sample trajectories for growing values of the scaling parameter

$$N = 16,\ 64,\ 256,\ 1024,$$

at the first and second natural time-scales, respectively.

As anticipated, at the first natural time-scale, the process of species concentrations converges to the constant process

$$c_i(\tau) = c_i(0) = 2, \qquad i = 1, 2.$$

To be precise, inflows and outflows are more and more rare, so that the sum process is essentially constant, while autocatalytic reactions have a non-negligible impact on process trajectories, but their effects are balanced in such a way that species concentrations fluctuate nearer and nearer to their initial values.

On the contrary, at the second natural time-scale, the process describing the sum of the two species converges to the deterministic process

$$w(\tau) = 2 + (w(0) - 2)\,e^{-\tau} = 2 + 2e^{-\tau},$$

and the process of species concentrations fluctuates rapidly, occasionally displaying the peculiar Discreteness Induced Transitions. In particular, inflows and

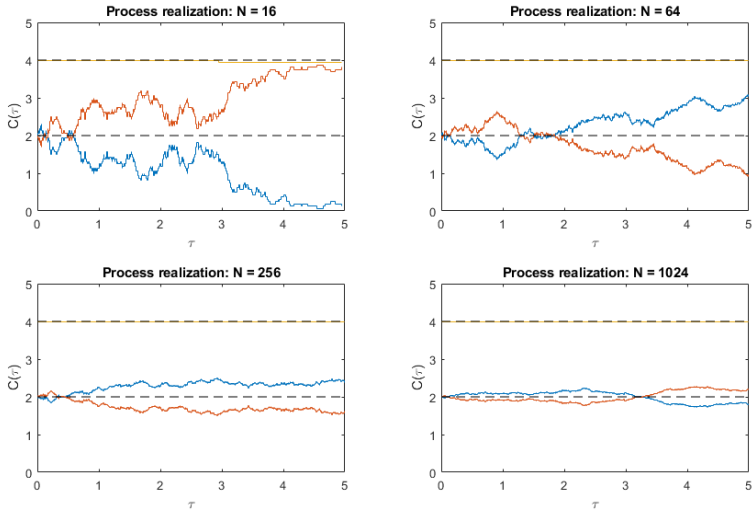

Figure 3.6: Realizations of stochastic process $C^{N,\gamma'}$ of species concentrations at the first natural time-scale $\gamma' = 0$, for growing values of the scaling parameter N.

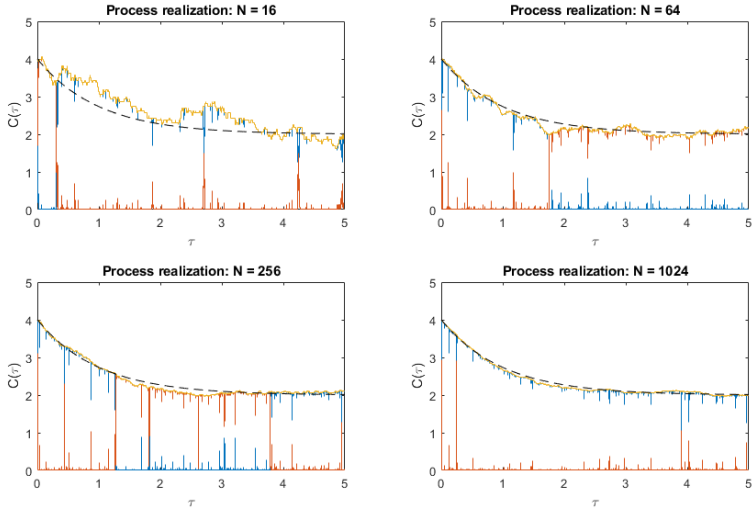

Figure 3.7: Realizations of stochastic process $C^{N,\gamma''}$ of species concentrations at the second natural time-scale $\gamma'' = 1 + \alpha = 2$, for growing values of the scaling parameter N.

outflows determine the dynamics of the sum process, while autocatalytic reactions are responsible for peaks and switches, which apparently happen instantaneously.

It is interesting to focus our attention on autocatalytic cascades:

- at the first natural time-scale $\gamma' = 0$, time goes by too slowly to detect complete transitions, and only small fluctuations near to the initial value are observed;
- at the second natural time-scale $\gamma'' = 1 + \alpha$, time passes too fast, and autocatalytic cascades apparently occur in an infinitesimal time span.

Therefore, it seems reasonable to wonder if there exists a time scale $\bar{\gamma}$ in between γ' and γ'' at which autocatalytic cascades happen in finite but not vanishing time, while inflows and outflows are still too slow to be observed.

Consider the intermediate time-scale $\bar{\gamma} = 1$; the stochastic process $C^{N,\bar{\gamma}}$ is described by the equations

$$\begin{aligned} C_1^{N,\bar{\gamma}}(\tau) = C_1^N(0) &- N^{-1}\xi_1^A\left(N^2\int_0^\tau \kappa' C_1^{N,\bar{\gamma}}(s)C_2^{N,\bar{\gamma}}(s)ds\right) \\ &+ N^{-1}\xi_2^A\left(N^2\int_0^\tau \kappa' C_1^{N,\bar{\gamma}}(s)C_2^{N,\bar{\gamma}}(s)ds\right) \\ &+ N^{-1}\xi_1^I(N^{1-\alpha}\eta'\tau) - N^{-1}\xi_1^O\left(N^{1-\alpha}\int_0^\tau \delta' C_1^{N,\bar{\gamma}}(s)ds\right), \end{aligned} \tag{3.61}$$

$$\begin{aligned} C_2^{N,\bar{\gamma}}(\tau) = C_2^N(0) &+ N^{-1}\xi_1^A\left(N^2\int_0^\tau \kappa' C_1^{N,\bar{\gamma}}(s)C_2^{N,\bar{\gamma}}(s)ds\right) \\ &- N^{-1}\xi_2^A\left(N^2\int_0^\tau \kappa' C_1^{N,\bar{\gamma}}(s)C_2^{N,\bar{\gamma}}(s)ds\right) \\ &+ N^{-1}\xi_2^I(N^{1-\alpha}\eta'\tau) - N^{-1}\xi_2^O\left(N^{1-\alpha}\int_0^\tau \delta' C_2^{N,\bar{\gamma}}(s)ds\right), \end{aligned} \tag{3.62}$$

At this time-scale, normalized reaction countings for inflows and outflows converge to zero, i.e. are too slow to be detected:

$$N^{-1}\xi_i^I(N^{1-\alpha}\eta'\tau) \;\to\; 0, \qquad N^{-1}\xi_i^O\left(N^{1-\alpha}\int_0^\tau \delta' C_i^{N,\gamma'}(s)ds\right) \;\to\; 0.$$

In order to characterize the limit for normalized countings of autocatalytic reactions, we need the following result on functional convergence for Poisson processes (see [1, app. A] for the proof, and [7, ch. 7] for generalizations).

Theorem 3.4 (Functional Central Limit Theorem for Poisson process)**.** *Let ξ be a unit Poisson process and let W^N be the stochastic process defined as*

$$W^N(t) = \frac{1}{\sqrt{N}}\left(\xi(Nt) - Nt\right), \qquad t \geq 0.$$

Then, W^N converges weakly to a standard Brownian motion W,

$$W^N \;\Rightarrow\; W.$$

In light of this Theorem, define process W_i^N as

$$W_i^N(u^N(\tau)) := N^{-1}\xi_i^A\left(N^2u^N(\tau)\right) - Nu^N(\tau), \qquad i = 1, 2, \tag{3.63}$$

where

$$u^N(\tau) = \int_0^\tau \kappa' C_1^{N,\bar{\gamma}}(s)C_2^{N,\bar{\gamma}}(s)ds,$$

so that normalized countings for autocatalytic reactions may be rewritten as

$$N^{-1}\xi_i^A\left(N^2u^N(\tau)\right) = W_i^N(u^N(\tau)) + Nu^N(\tau), \qquad i = 1, 2. \tag{3.64}$$

Therefore, the stochastic process $C_i^{N,\bar{\gamma}}$ is given by

$$\begin{aligned} C_i^{N,\bar{\gamma}}(\tau) &= C_i^N(0) - W_i^N(u^N(\tau)) - Nu^N(\tau) \\ &\quad + W_{i+1}^N(u^N(\tau)) + Nu^N(\tau) \\ &= C_i^N(0) - W_i^N(u^N(\tau)) + W_{i+1}^N(u^N(\tau)). \end{aligned} \tag{3.65}$$

By the functional central limit theorem stated above, process $C^{N,\bar{\gamma}}$ converges in distribution to the stochastic process $C^{\bar{\gamma}}$, solution of the following system of equations:

$$C_i^{\bar{\gamma}}(\tau) = C_i(0) - W_i(u(\tau)) + W_{i+1}(u(\tau)), \qquad i = 1, 2, \tag{3.66}$$

where W_1 and W_2 are standard Brownian motions and

$$u(\tau) = \int_0^\tau \kappa' C_1^{\bar{\gamma}}(s)C_2^{\bar{\gamma}}(s)ds.$$

Using properties of Brownian motions, the system in (3.66) may be rewritten as

$$C_1^{\bar{\gamma}}(\tau) = C_1(0) + W\left(\int_0^\tau 2\kappa' C_1^{\bar{\gamma}}(s)C_2^{\bar{\gamma}}(s)ds\right), \tag{3.67}$$

$$C_2^{\bar{\gamma}}(\tau) = C_2(0) - W\left(\int_0^\tau 2\kappa' C_1^{\bar{\gamma}}(s)C_2^{\bar{\gamma}}(s)ds\right), \tag{3.68}$$

where W is a standard Brownian motion.

Numerical simulations are carried out assuming the same experimental setting discussed above (Figures 3.6 and 3.7), and results are displayed in Figure 3.8.

At this intermediate time-scale, inflow and outflow reactions are more and more rare, and the sum is nearly constant, while autocatalytic cascades can be approximated by a Brownian motion. Note that, since inflows and outflows are too slow to be detected at this time-scale, once the autocatalytic cascade ends (i.e one of the species is extinct), the process trajectory remains constant.

3.4 Characterization of the limit process

At the second natural time-scale $\gamma'' = 1 + \alpha$, the process $W^{N,\gamma''}$ describing the concentration of the sum of the two species satisfies equation

$$\begin{aligned} W^{N,\gamma''}(\tau) &= W^N(0) + N^{-1}\xi^I(N\,2\eta'\tau) \\ &\quad - N^{-1}\xi^O\left(N\int_0^\tau \delta' W^{N,\gamma''}(s)ds\right), \end{aligned} \tag{3.57}$$

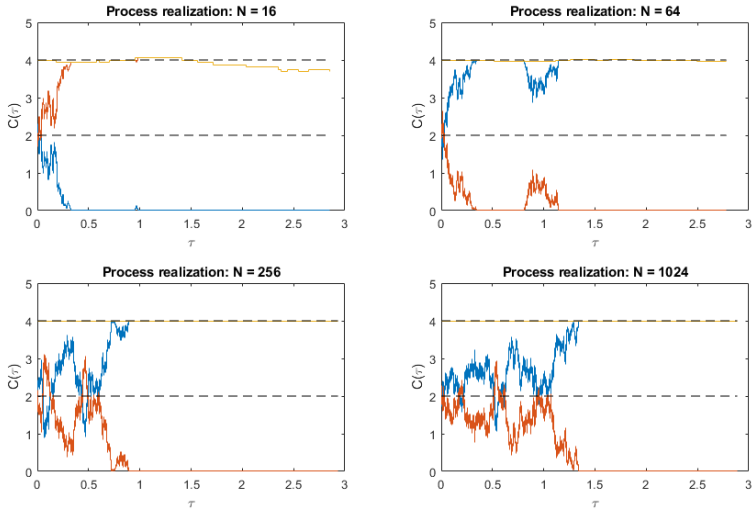

Figure 3.8: Realizations of stochastic process $C^{N,\bar{\gamma}}$ of species concentrations at the intermediate time-scale $\bar{\gamma} = 1$, for growing values of the scaling parameter N.

and converges, for N growing large, to the deterministic process

$$w(\tau) = \left(w(0) - \frac{2\eta'}{\delta'}\right) \exp\left(-\delta'\tau\right) + \frac{2\eta'}{\delta'}, \qquad \tau \geq 0, \tag{3.60}$$

as already discussed in the previous section.

On the contrary, the bidimensional process $C^{N,\gamma''}$ describing the concentrations of the single species is characterized by fast fluctuations, driven by autocatalytic reactions, which determine sharp peaks of different heights and, occasionally, Discreteness Induced Transitions (see Figure 3.7).

The purpose of this section is to derive some useful properties for this process and, eventually, identify a candidate limit, for N growing large. To this end, a single component (say, C_1) is considered, being the other completely determined by the former and their sum W.

Figure 3.9 provides a closer look at a sample trajectory of process $C_1^{N,\gamma''}$, describing concentration of species S_1. Some empirical observations may be collected:

- autocatalytic cascades, either leading to peaks or switches, are almost instantaneous;
- the process $C_1^{N,\gamma''}$ is either equal to zero or to the sum process $W^{N,\gamma''}$ for most of the time;
- the sum process $W^{N,\gamma''}$ is approximately constant during autocatalytic cascades.

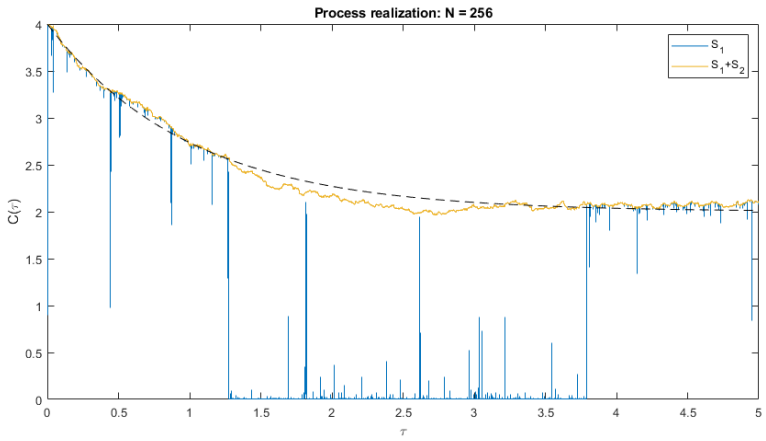

Figure 3.9: Realization of stochastic process $C_1^{N,\gamma''}$ describing concentration of species S_1 at the second natural time-scale $\gamma'' = 1 + \alpha$, for $N = 256$ and other parameters set as for Figures 3.6 and 3.7.

It is then natural to ask whether these properties could be stated in a more rigorous form and actually characterize the limit process. In particular, we may try to answer the following questions:

1. do autocatalytic cascades occur instantaneously?
2. are autocatalytic reactions stopped for almost all the time?
3. is the sum process constant during autocatalytic cascades?

In order to discuss the first two questions, we initially assume that the third one has a positive answer: the fact that such assumption is actually verified, under suitable conditions, is proved later on.

Suppose that, at time $\hat{\tau}-$, species S_1 is extinct and species S_2 has concentration s^-, equal to the concentration W of the sum,

$$C_1^{N,\gamma}(\hat{\tau}-) = 0, \qquad W^{N,\gamma}(\hat{\tau}-) = C_2^{N,\gamma}(\hat{\tau}-) = s^- = \frac{S-1}{N};$$

then, at time $\hat{\tau}$, an inflow of species S_1 occurs, triggering an autocatalytic cascade:

$$C_1^{N,\gamma}(\hat{\tau}) = \frac{1}{N}, \qquad W^{N,\gamma}(\hat{\tau}) = s = \frac{S}{N}. \tag{3.69}$$

Since the sum process is assumed to be constant for the duration of the cascade, the process describing the number of molecules of species S_1 can be modeled as a continuous time birth-and-death Markov chain on the state space

$$\mathbb{S} = \{\, 0, 1, \ldots, S \,\},$$

with absorbing states 0 and $S = sN$. The key properties of this process are gathered in the Appendix to this chapter: in particular, results about the exit distribution and the expected exit time are used in the following of this section.

Let T_E^N be the duration of the autocatalytic cascade, i.e. the time it takes the Markov chain to reach an absorbing state, starting from state $x = 1$:

$$T_E^N(s) := T_E^N(s) = \inf \{\, t \geq 0 \,:\, C_1^{N,\gamma}(\hat{\tau} + t) \in \{\, 0, s \,\} \,\}\,. \tag{3.70}$$

Proposition 3.5. *The random (stopping) time T_E^N defined in* (3.70) *converges in probability to* 0 *at the second natural time-scale γ'',*

$$\lim_N \mathbb{P}\left(T_E^N \geq \epsilon\right) = 0, \qquad \forall \epsilon > 0.$$

Proof. Taking into consideration the scaling of reaction rate constant κ and the scaling in time,

$$\kappa = \frac{\kappa'}{N}, \qquad \tau = N^{-\gamma} t,$$

the expected value of T_E^N, for large N, is approximated as

$$\mathbb{E}\left[T_E^N\right] \approx \frac{\log N}{\kappa' s}\, N^{-\gamma}.$$

Since T_E^N is a non-negative random variable, by Markov's inequality we obtain, for each $\epsilon > 0$,

$$\mathbb{P}\left(T_E^N \geq \epsilon\right) \leq \frac{1}{\epsilon}\, \mathbb{E}\left[T_E^N\right].$$

At the second natural time-scale, $\gamma = 1 + \alpha$, and therefore,

$$\mathbb{P}\left(T_E^N \geq \epsilon\right) \leq \frac{\log N}{\epsilon\, \kappa' s}\, N^{-\gamma} = \frac{\log N}{\epsilon\, \kappa' s}\, N^{-1-\alpha} \;\to\; 0,$$

for N growing large. □

As a consequence, we may give a positive answer to the first question: autocatalytic cascades occur instantaneously at the second time-scale, in the sense that their duration converges to zero in probability. Note that this Proposition holds true not only at the second natural time-scale $\gamma'' = 1 + \alpha$, but at every time-scale strictly faster than the first natural time-scale $\gamma' = 0$.

Remark 3.6. In the proof of Proposition 3.5, the limit procedure is correct assuming that $s = O(1)$. This assumption is not restrictive, since process $W^{N,\gamma''}$ converges to the deterministic process (3.60), and therefore, for each $\tau > 0$ and each $\epsilon > 0$, there exists $m > 0$ such that, for sufficiently large N,

$$\mathbb{P}\left(\inf_{\sigma \leq \tau} W^{N,\gamma''}(\sigma) \geq m\right) \geq 1 - \epsilon. \tag{3.71}$$

Likewise, for each $\tau > 0$ and each $\epsilon > 0$, there exists $M < \infty$ such that, for sufficiently large N,

$$\mathbb{P}\left(\sup_{\sigma \leq \tau} W^{N,\gamma''}(\sigma) \leq M\right) \geq 1 - \epsilon. \tag{3.72}$$

In order to answer the second question, we must consider the entire process $C_1^{N,\gamma}$ on a finite time interval $[0, T]$. An autocatalytic cascade is triggered whenever:

- species S_1 is extinct and an inflow of S_1 occurrs;
- species S_2 is extinct, that is, the concentration of S_1 is equal to the concentration W of the sum, and an inflow of S_2 occurs.

By assumption, the sum process $W^{N,\gamma}$ is constant during autocatalytic cascades, and thus inflows (and outflows) of species happen only when one of the species is extinct. Moreover, reaction rates of inflows are independent from species concentrations.

Therefore, whatever the species which is extinct and whatever the concentration of the non-extinct species, the number $Y^T(\tau)$ of inflows triggering an autocatalytic cascade, up to time τ, is a counting process with the same rate of the inflow reactions:

$$Y^T(\tau) := \xi^T \left(N^{\gamma-\alpha} \eta' \tau \right), \tag{3.73}$$

where ξ^T is a unit Poisson process.

In addition, the duration of each cascade depends only on the value of the sum process $W^{N,\gamma}$ at the time the cascade is triggered, and so

$$T_E^N \left(W^{N,\gamma}(\tau_i) \right), \qquad i = 1, \dots, Y^T(\tau),$$

where $\{ \tau_i \}$ are the jump times of process Y^T, are conditionally independent random variables.

Let $\bar{T}_E^N(\tau)$ be the total duration of the autocatalytic cascades, up to time τ:

$$\bar{T}_E^N(\tau) := \sum_{i=1}^{Y^T(\tau)} T_E^N \left(W^{N,\gamma}(\tau_i) \right). \tag{3.74}$$

Proposition 3.7. *The random variable $\bar{T}_E^N(\tau)$ defined in* (3.74) *converges in probability to* 0 *at the second natural time-scale γ'', for each $\tau \geq 0$,*

$$\lim_N \mathbb{P} \left(\bar{T}_E^N(\tau) \geq \epsilon \right) = 0, \qquad \forall \epsilon > 0.$$

Proof. Conditioning on the counting process $Y^T(\tau)$, we can write the expected value of $\bar{T}_E^N(\tau)$ as

$$\begin{aligned} \mathbb{E}\left[\bar{T}_E^N(\tau) \right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{Y^T(\tau)} T_E^N \left(W^{N,\gamma}(\tau_i) \right) \;\middle|\; Y^T(\tau) \right] \right] \\ &= \mathbb{E}\left[\sum_{i=1}^{Y^T(\tau)} \mathbb{E}\left[T_E^N \left(W^{N,\gamma}(\tau_i) \right) \right] \right]. \end{aligned}$$

As observed in the proof of Proposition 3.5, the expected value of stopping time $T_E^N(s)$ is approximated, for large N, as

$$\mathbb{E}\left[T_E^N(s) \right] \approx \frac{\log N}{\kappa' \, s} \, N^{-\gamma}.$$

By Remark 3.6, we may assume that, for sufficiently large N,

$$W^{N,\gamma}(\tau_i) \geq m > 0, \qquad i = 1, \dots, Y^T(\tau),$$

and therefore obtain

$$\mathbb{E}\left[T_E^N\left(W^{N,\gamma}(\tau_i)\right)\right] \leq \frac{\log N}{\kappa' m}\, N^{-\gamma}, \qquad i = 1, \ldots, Y^T(\tau).$$

As a consequence,

$$\mathbb{E}\left[\bar{T}_E^N(\tau)\right] \leq \mathbb{E}\left[Y^T(\tau)\ \frac{\log N}{\kappa' m}\, N^{-\gamma}\right] = \mathbb{E}\left[Y^T(\tau)\right]\ \frac{\log N}{\kappa' m}\, N^{-\gamma},$$

and, by definition (3.73) of counting (Poisson) process $Y^T(\tau)$,

$$\mathbb{E}\left[\bar{T}_E^N(\tau)\right] \leq \frac{\eta'\tau}{\kappa' m}\, \log N\, N^{-\alpha}.$$

Since $\bar{T}_E^N(\tau)$ is a non-negative random variable, by Markov's inequality we obtain, for each $\epsilon > 0$,

$$\mathbb{P}\left(\bar{T}_E^N(\tau) \geq \epsilon\right) \leq \frac{1}{\epsilon}\, \mathbb{E}\left[\bar{T}_E^N(\tau)\right],$$

which implies that, for each $\tau \geq 0$,

$$\mathbb{P}\left(\bar{T}_E^N(\tau) \geq \epsilon\right) \leq \frac{\eta'\tau}{\epsilon\, \kappa' m}\, \log N\, N^{-\alpha}\ \to\ 0,$$

for N growing large. □

This Proposition gives a positive answer to the second question: since the total duration of autocatalytic cascades converges to zero in probability, autocatalytic reactions are stopped for almost all the time. More formally, we may say that the Lebesgue measure of the subset of time interval $[0, T]$ containing time instants at which the process $C_1^{N,\gamma}$ is different from both zero and $W^{N,\gamma}$ converges to 0 in probability:

$$\lim_N \mathbb{P}\left(\ell\left(\left\{\, \tau \in [0,T]\ :\ C_1^{N,\gamma}(\tau) \notin \left\{\, 0, W^{N,\gamma}(\tau)\,\right\}\right\}\right) \geq \epsilon\right) = 0, \tag{3.75}$$

where ℓ is the Lebesgue measure on $[0, T]$.

Remark 3.8. Proposition 3.7 seems to hold independently of the time-scale. However, for time-scales $\gamma < \alpha$, the number $Y^T(\tau)$ of inflows triggering autocatalytic cascades, up to time τ, converges to 0,

$$Y^T(\tau) = \xi^T\left(N^{\gamma-\alpha}\eta'\tau\right)\ \to\ 0, \qquad \gamma < \alpha;$$

consequently, the total duration of such cascades trivially goes to 0, since none of them happens in the limit for large N.

The fact that the sum process $W^{N,\gamma}$ remains constant while the rapid sequence of autocatalytic reactions occur has been used in an essential way for proving results above. Indeed, this assumption is fundamental to model the autocatalytic process as a continuous time birth-and-death Markov chain which is homogeneous in time, that is, whose reaction rates do not depend on time. However, whether such assumption actually makes sense has to be verified rigorously.

Suppose the setting introduced in (3.69) holds, that is, an inflow of species S_1, previously extinct, occurs at time $\hat{\tau}$:

$$C_1^{N,\gamma}(\hat{\tau}) = \frac{1}{N}, \qquad W^{N,\gamma}(\hat{\tau}) = s = \frac{S}{N}. \tag{3.76}$$

Then, the dynamics of the number of molecules of S_1 is described by the above-mentioned Markov chain up until the first inflow or outflow reaction happens, changing the value of the sum process.

Since the countings of inflows and outflows of process $W^{N,\gamma}$ are modeled as Poisson counting processes (3.54), the time until an inflow or outflow occur has exponential distribution. In particular:

- the process Y^I counting inflows is a Poisson process with constant intensity function
$$\lambda^I(\cdot) = N^{\gamma-\alpha} 2\eta',$$
so that the time T_I^N until the first inflow, starting from time $\hat{\tau}$, has exponential distribution with parameter λ^I,
$$T_I^N \sim \mathrm{Exp}\left(N^{\gamma-\alpha} 2\eta'\right); \tag{3.77}$$

- the process Y^O counting outflows is a Poisson process with intensity function
$$\lambda^O(w) = N^{\gamma-\alpha}\delta' w,$$
so that the time T_O^N until the first outflow, starting from time $\hat{\tau}$, has exponential distribution with parameter λ^O,
$$T_O^N \sim \mathrm{Exp}\left(N^{\gamma-\alpha}\delta' W^{N,\gamma}(\hat{\tau})\right) = \mathrm{Exp}\left(N^{\gamma-\alpha}\delta' s\right). \tag{3.78}$$

Therefore, the time T_{IO}^N until the first inflow or outflow occurs, starting from $\hat{\tau}$, has also exponential distribution with parameter $\lambda^I + \lambda^O$:

$$T_{IO}^N = \min(T_I^N, T_O^N) \sim \mathrm{Exp}\left(N^{\gamma-\alpha}(2\eta' + \delta' s)\right). \tag{3.79}$$

As a result, the question on whether the sum process $W^{N,\gamma}$ is constant or not during the autocatalytic cascade may be reformulated asking if, in the limit for large N, the duration of the autocatalytic sequence T_E^N is smaller that the time T_{IO}^N until the first inflow or outflow, that is if

$$\lim_{N\to\infty} \mathbb{P}\left(T_E^N \le T_{IO}^N\right) = 1 \tag{3.80}$$

Proposition 3.9. *Let T_E^N and T_{IO}^N be random variables defined in* (3.70) *and* (3.79)*, respectively. Then,*

$$\mathbb{P}\left(T_E^N \ge T_{IO}^N\right) \le O\left(\log^2 N\, N^{-\alpha}\right), \qquad \textit{for} \quad N \to \infty.$$

Proof. Since T_{IO}^N has exponential distribution,

$$T_{IO}^N \sim \mathrm{Exp}\left(N^{\gamma-\alpha}\rho\right), \qquad \text{with} \quad \rho = 2\eta' + \delta' s,$$

the probability of T_E^N being larger than T_{IO}^N can be rewritten conditioning on the value of T_{IO}^N:

$$\mathbb{P}\left(T_E^N \geq T_{IO}^N\right) = \int_0^\infty \mathbb{P}\left(T_E^N \geq t\right) f^N(t)\, dt,$$

where $f^N(t)$ is the probability density function of T_{IO}^N,

$$f^N(t) = \rho\, N^{\gamma-\alpha} \exp\left(-N^{\gamma-\alpha}\rho\, t\right)$$

By Markov's inequality we obtain,

$$\mathbb{P}\left(T_E^N \geq t\right) \leq \min\left\{1\ ,\ \frac{1}{t}\,\mathbb{E}\left[T_E^N\right]\right\},$$

and therefore

$$\begin{aligned}\mathbb{P}\left(T_E^N \geq T_{IO}^N\right) &\leq \int_0^\infty \min\left\{1\,,\ \frac{1}{t}\,\mathbb{E}\left[T_E^N\right]\right\} f^N(t)\, dt \\ &= \int_0^{\mathbb{E}\left[T_E^N\right]} f^N(t)\, dt + \int_{\mathbb{E}\left[T_E^N\right]}^\infty \frac{1}{t}\,\mathbb{E}\left[T_E^N\right] f^N(t)\, dt.\end{aligned}$$

Recall that the expected value of stopping time T_E^N is approximated, for large N, as

$$\mathbb{E}\left[T_E^N\right] \approx \frac{\log N}{\kappa'\, s}\, N^{-\gamma}.$$

Introduce the change of integration variable

$$z = N^{\gamma-\alpha}\rho\, t,$$

and substitute the expressions for $\mathbb{E}\left[T_E^N\right]$ and $f^N(t)$:

$$\begin{aligned}\mathbb{P}\left(T_E^N \geq T_{IO}^N\right) &\leq \int_0^{E(N)} e^{-z}\, dz + \int_{E(N)}^\infty \frac{1}{z}\, E(N) e^{-z}\, dz \\ &\leq E(N)\left(1 + \int_{E(N)}^\infty \frac{1}{z}\, e^{-z}\, dz\right),\end{aligned}$$

where

$$E(N) = \frac{\rho}{\kappa'\, s}\, \log N\, N^{-\alpha}.$$

The *exponential integral* $E_1 \colon \mathbb{C} \mapsto \mathbb{R}$ is the non-elementary function defined as

$$E_1(z) = \int_z^\infty \frac{1}{z} e^{-z}\, dz;$$

it can be expressed in the form of convergent series as

$$E_1(z) = -\gamma^{EM} - \log z - \sum_{k=1}^\infty \frac{(-z)^k}{k k!},$$

where γ^{EM} is a constant, so that, for $z \to 0$,

$$E_1(z) \approx -\log z.$$

Therefore, using the fact that $E(N) \to 0$ for large N, we obtain

$$\mathbb{P}\left(T_E^N \geq T_{IO}^N\right) \leq E(N)\left(1 - \log E(N)\right).$$

Substituting the expression for $E(N)$,

$$1 - \log E(N) = 1 - \log\left(\frac{\rho}{\kappa' s} \log N\, N^{-\alpha}\right) \approx \alpha \log N,$$

which leads to the result:

$$\mathbb{P}\left(T_E^N \geq T_{IO}^N\right) \leq \frac{\alpha\rho}{\kappa' s} \log^2 N\, N^{-\alpha}.$$

for large N. □

This result guarantees that, when a *single* autocatalytic cascade is triggered, the probability that an inflow or outflow happens before the conclusion of the autocatalytic sequence converges to 0 as N grows large. However, in order to claim that this is true for *every* autocatalytic cascade, we have to consider the entire process on a finite time interval $[0, T]$.

In particular, consider the process Y^T describing the number of inflows triggering an autocatalytic cascade:

$$Y^T(\tau) = \xi^T\left(N^{\gamma-\alpha}\eta'\tau\right); \tag{3.73}$$

for $\gamma \leq \alpha$, the number of such inflows is finite or converging to 0, and therefore the result in Proposition 3.9 is enough to conclude that no inflows or outflows occur during autocatalytic cascades. On the contrary, for faster time-scales, the number of such inflows grows unbounded for large N, and a conclusion is not trivial.

Remark 3.10. The number of inflows triggering an autocatalytic cascade is modeled as a Poisson counting process Y^T with intensity function

$$\lambda^T(c) = N^{\gamma-\alpha}\eta'.$$

However, such expression for λ^T would be correct under the assumption that no inflows happen during autocatalytic cascade, that is exactly what we want to prove hereunder.

On the other hand, this counting process actually overestimates the true Y^T, possibly considering inflows which instead occur during autocatalytic cascades. Therefore, assuming this expression for λ^T does not affect the argument, since in the proof of Proposition 3.12 the counting process $\hat{Y}^T$ is eventually bounded from above.

Before stating a formal result, recall a well-know property of Poisson processes.

Proposition 3.11 (Thinning property)**.** *Let Y be a Poisson process with intensity function λ,*

$$Y(t) = \xi\left(\int_0^t \lambda(s)ds\right), \qquad t \geq 0,$$

where ξ is a unit Poisson process, and assume that each arrival has a given property with probability p, possibly depending on time.

Then, the process $\hat{Y}$ counting arrivals of Y having such property (thinned process) is a Poisson process with intensity function λp,

$$\hat{Y}(t) = \xi\left(\int_0^t \lambda(s)p(s)ds\right), \qquad t \geq 0.$$

Consider again the process Y^T describing the number of inflows triggering an autocatalytic cascade. For each of such inflows, the probability p^N that another inflow or outflow happens before the conclusion of the autocatalytic sequence depends only on the value of the sum process $W^{N,\gamma}$ at the time the cascade is triggered:

$$p^N = p^N\left(W^{N,\gamma}(\tau_i)\right), \qquad i = 1, \ldots, Y^T(\tau),$$

where $\{\,\tau_i\,\}$ are the jump times of process Y^T.

Let $\hat{Y}^T$ be the thinned process, derived from Y^T, which describes the number of autocatalytic cascades that are still running when another inflow or outflow occur:

$$\hat{Y}^T(\tau) := \xi^T\left(N^{\gamma-\alpha}\eta' \int_0^\tau p^N\left(W^{N,\gamma}(s)\right) ds\right). \tag{3.81}$$

Proposition 3.12. *The random variable $\hat{Y}^T(\tau)$ defined in* (3.81) *converges in probability to* 0 *at the second natural time-scale γ'', for each $\tau \geq 0$,*

$$\lim_N \mathbb{P}\left(\hat{Y}^T(\tau) \geq \epsilon\right) = 0, \qquad \forall \epsilon > 0,$$

under the assumption that $\alpha > 1$.

Proof. By Proposition 3.9, the probability $p^N(s)$ is bounded above:

$$p^N(s) = \mathbb{P}\left(T_E^N(s) \geq T_{IO}^N(s)\right) \leq \frac{\alpha\rho}{\kappa' s} \log^2 N\, N^{-\alpha}.$$

By Remark 3.6, we may assume that, for sufficiently large N,

$$W^{N,\gamma}(\tau) \geq m > 0, \qquad \tau \geq 0,$$

and therefore obtain

$$p^N\left(W^{N,\gamma}(\tau)\right) \leq \frac{\alpha\rho}{\kappa' m} \log^2 N\, N^{-\alpha}, \qquad \tau \geq 0.$$

As a consequence, the intensity function of the thinned process may be bounded above by a quantity which does not depend on time:

$$N^{\gamma-\alpha}\eta'\, p^N\left(W^{N,\gamma}(\tau)\right) \leq \eta' \frac{\alpha\rho}{\kappa' m} \log^2 N\, N^{\gamma-2\alpha};$$

then, the expected value of $\hat{Y}^T(\tau)$ can be bounded as well:

$$\mathbb{E}\left[\hat{Y}^T(\tau)\right] \leq \eta'\tau \frac{\alpha\rho}{\kappa' m} \log^2 N\, N^{\gamma-2\alpha}.$$

By Markov's inequality we obtain, for each $\epsilon > 0$,

$$\mathbb{P}\left(\hat{Y}^T(\tau) \geq \epsilon\right) \leq \frac{1}{\epsilon}\,\mathbb{E}\left[\hat{Y}^T(\tau))\right],$$

which implies that, for each $\tau \geq 0$,

$$\mathbb{P}\left(\hat{Y}^T(\tau) \geq \epsilon\right) \leq \eta' \tau \frac{\alpha\rho}{\epsilon\,\kappa'\,m}\,\log^2 N\, N^{\gamma-2\alpha},$$

for large N. At the second time-scale $\gamma = 1+\alpha$, and so

$$\lim_{N\to\infty} \mathbb{P}\left(\hat{Y}^T(\tau) \geq \epsilon\right) = 0,$$

if $\gamma = 1+\alpha < 2\alpha$, that is, $\alpha > 1$. □

The fundamental consequence of Proposition 3.12 is that, in the study of the limit process at the second natural time-scale, we may restrict to consider the space $\mathcal{S}$ of those processes for which autocatalytic cascades and inflow/outflow reactions are separated, in the sense that no inflows or outflows happen while an autocatalytic sequence is running:

$$\mathcal{S} := \{\, C^{N,\gamma''} \in D[0,T] \,:\, \hat{Y}^T(T) = 0 \,\}\,. \tag{3.82}$$

Indeed, in the limit for large N, the probability measure of this subset of the Skorohod space $D[0,T]$ converges to 1, that is, for each $\epsilon > 0$, there exists N_0 such that

$$\mathbb{P}\left(C^{N,\gamma''} \in \mathcal{S}\right) \geq 1-\epsilon, \qquad \forall N \geq N_0. \tag{3.83}$$

The procedure of thinning the counting process Y^T of inflows triggering autocatalytic cascades turns out to be useful to derive other interesting properties of the limit process. In particular, we may focus on the number of Discreteness Induced Transitions and on the number of peaks (i.e. failed transitions) of height at least $h > 0$, that is, peaks which cause a variation in species concentration of at least h before returning to their original values.

Proposition 3.13. *Let Y^T be the counting process defined in* (3.73)*. At the second natural time-scale γ'', in the limit for N growing large,*

- *the number $Y^{DIT}(\tau)$ of Discreteness Induced Transitions, up to time τ, is a Poisson process*

$$Y^{DIT}(\tau) := \xi^T\left(\eta' \int_0^\tau \frac{1}{w(s)}\,ds\right), \qquad \tau \geq 0; \tag{3.84}$$

- *the number $Y^{Peak}(h;\tau)$ of peaks of height* at least *h, up to time τ, is a Poisson process*

$$Y^{Peak}(h;\tau) := \xi^T\left(\eta' \int_0^\tau \frac{1}{h}\left(1-\frac{h}{w(s)}\right)^+ ds\right), \qquad \tau \geq 0. \tag{3.85}$$

Proof. Suppose that an inflow of species S_1, previously extinct, occurs at time $\hat{\tau}$, as in (3.69):

$$C_1^{N,\gamma}(\hat{\tau}) = \frac{1}{N}, \qquad W^{N,\gamma}(\hat{\tau}) = s = \frac{S}{N}.$$

The probability that the resulting autocatalytic sequence leads to a transition is given by

$$p^{DIT} = p^{DIT}(s) = \frac{1}{S} = \frac{1}{sN},$$

and so the process $Y^{N,DIT}$ describing the number of transitions is obtained by thinning process Y^T with probability p^{DIT}:

$$Y^{N,DIT}(\tau) = \xi^T \left(N^{\gamma-\alpha-1}\eta' \int_0^\tau \frac{1}{W^{N,\gamma}(s)}\, ds \right), \qquad \tau \geq 0.$$

At the second time-scale $\gamma = 1+\alpha$ and the sum process $W^{N,\gamma''}$ converges to the deterministic process w defined in (3.60), so that

$$Y^{N,DIT} \;\Rightarrow\; Y^{DIT}, \qquad Y^{DIT}(\tau) = \xi^T \left(\eta' \int_0^\tau \frac{1}{w(s)}\, ds \right), \quad \tau \geq 0.$$

Similarly, the probability that the resulting autocatalytic sequence leads to a peak of height *at least* $h < s$ is given by

$$p^{Peak}(h) = p^{Peak}(h;s) = \frac{1}{hN}\left(1 - \frac{h}{s}\right);$$

clearly, if $h \geq s$ the definition of peak does not make sense, since the process either completes a transition (if $h = s$) or cannot reach an height of h moving inside the interval of values $[0, s]$.

The process $Y^{N,Peak}(h;\cdot)$ describing the number of peaks of height at least h is obtained by thinning process Y^T with probability $p^{Peak}(h)$:

$$Y^{N,Peak}(h;\tau) = \xi^T \left(N^{\gamma-\alpha-1}\eta' \int_0^\tau \frac{1}{h}\left(1 - \frac{h}{W^{N,\gamma}(s)}\right)^+ ds \right), \qquad \tau \geq 0.$$

Again, at the second time-scale $\gamma = 1+\alpha$ and the sum process $W^{N,\gamma''}$ converges to w, so that

$$Y^{Peak}(h;\tau) = \xi^T \left(\eta' \int_0^\tau \frac{1}{h}\left(1 - \frac{h}{w(s)}\right)^+ ds \right), \qquad \tau \geq 0. \qquad \square$$

The following result is analogous to this Proposition, and will be used in the next section.

Corollary 3.14. *The number $Y^V(h;\tau)$ of autocatalytic cascades which cause a variation in species concentration of* at least *h, up to time τ, is a Poisson process*

$$Y^V(h;\tau) := \xi^T \left(\eta' \int_0^\tau \frac{1}{h}\left\{h \leq w(s)\right\} ds \right), \qquad \tau \geq 0. \tag{3.86}$$

Proof. Again, suppose that an inflow of species S_1, previously extinct, occurs at time $\hat{\tau}$:

$$C_1^{N,\gamma}(\hat{\tau}) = \frac{1}{N}, \qquad W^{N,\gamma}(\hat{\tau}) = s = \frac{S}{N}.$$

The probability that the resulting autocatalytic sequence reaches *at least* a height of $h \leq s$ is given by

$$p^V(h) = p^V(h;s) = \frac{1}{hN};$$

as before, if $h > s$ the definition does not make sense, since the process cannot reach an height of h moving inside the interval $[0, s]$.

The process $Y^{N,V}(h;\tau)$ is therefore obtained by thinning process Y^T with probability $p^V(h)$:

$$Y^{N,V}(h;\tau) = \xi^T\left(N^{\gamma-\alpha-1}\eta' \int_0^\tau \frac{1}{h}\left\{h \le W^{N,\gamma}(s)\right\} ds\right), \qquad \tau \ge 0;$$

at the second time-scale $\gamma = 1+\alpha$ and process $W^{N,\gamma''}$ converges to w, so that (3.86) follows. □

A consequence of Proposition 3.13 is that, even if the number of inflows triggering autocatalytic cascades (and thus, the number of such cascades) grows unbounded for large N at the second time-scale γ'', the number of Discreteness Induced Transitions and peaks of finite height $h > 0$ is actually finite and depends on N only through process $W^{N,\gamma''}$, which in turn converges to a non-degenerate process w.

Remark 3.15. It is interesting to observe the way probability $p^{Peak}(h)$ of a peak of height at least h changes with the value of h:

$$p^{Peak}(h) = p^{Peak}(h;s) \sim \frac{s-h}{h}, \qquad h \in [0,s]. \tag{3.87}$$

In particular, this probability is null only for $h = s$, since once the transition is completed it is not possible to return to the original concentrations within the same autocatalytic sequence. On the contrary, such probability is positive for each value of h strictly smaller than s.

Indeed, as displayed in Figure 3.9, peaks may reach any height with positive probability, possibly very near s, without resulting in a transition. However, for value of h larger than $s/2$, the probability of a transition is larger than the probability of a peak, as one may reasonably expect.

To conclude this section, we commit to identifying a candidate limit C for the sequence of processes $\{C^{N,\gamma''}\}$ of species concentrations at the second natural time-scale:

$$\lim_{N\to\infty} C^{N,\gamma''} \text{ "=" } C. \tag{3.88}$$

In the light of results stated above, the limit process should satisfy the following requirements:

- it is a cadlag process, so that we can discuss convergence properties in the Skorohod space (see Section 2.3 for definition and properties):

$$C \in D[0,T];$$

- species concentrations are either equal to zero or to the deterministic process w for almost all the times, because of Proposition 3.7:

$$\ell\left(\{\tau \in [0,T] \,:\, C_1(\tau), C_2(\tau) \notin \{0, w(\tau)\}\}\right) = 0;$$

- the number of switches (Discreteness Induced Transitions) is modeled as a Poisson process Y^{DIT}, described in Proposition 3.13:

$$Y^{DIT}(\tau) := \xi^T\left(\eta' \int_0^\tau \frac{1}{w(s)}\, ds\right), \qquad \tau \ge 0;$$

The first requirement seems to be quite natural, but is actually the critical one, as it solves the issue of defining the limit process for time instant at which autocatalytic cascades happen.

Indeed, suppose that at time $\hat{\tau}$ an autocatalytic cascade is triggered: since, at the second time-scale, the duration of such cascades converges to 0 in the limit for large N, we may informally say that the process visits all the states reached during the autocatalytic sequence "at the same time". It is therefore natural to ask what value should the limit process take at time $\hat{\tau}$.

The fact that the limit process is assumed to be right-continuous easily answers to this question. In particular, assume that

$$C_1(\hat{\tau}-) = 0, \qquad C_2(\hat{\tau}-) = w(\hat{\tau});$$

- if the autocatalytic cascade leads to a transition, the process values at time $\hat{\tau}$ are switched between the two components,

  $$C_1(\hat{\tau}) = w(\hat{\tau}), \qquad C_2(\hat{\tau}) = 0,$$

  so that C is right-continuous at $\hat{\tau}$;

- if the autocatalytic cascade leads to a peak (of any height), the process values at time $\hat{\tau}$ do not change,

  $$C_1(\hat{\tau}) = 0, \qquad C_2(\hat{\tau}) = w(\hat{\tau}),$$

  so that the C is actually continuous at $\hat{\tau}$.

In other words, the value of the limit process at time $\hat{\tau}$ is the value taken *at the end* of the "instantaneous" autocatalytic process.

The last point to address is what value should the limit process take at the starting time $\tau = 0$.

Proposition 3.16. *Suppose that*

$$\lim_{N\to\infty} C_1^N(0) = c_1(0) > 0, \qquad \lim_{N\to\infty} C_2^N(0) = c_2(0) > 0.$$

The limit process C at time $\tau = 0$ should be defined as

$$C_1(0) = \begin{cases} 0 & \text{if} \quad I = 0 \\ w(0) & \text{if} \quad I = 1 \end{cases}, \qquad C_2(0) = w(0) - C_1(0),$$

where I is a Bernoulli random variable,

$$I \sim \text{Bernoulli}\left(p = \frac{c_1(0)}{w(0)}\right), \qquad w(0) = c_1(0) + c_2(0).$$

Proof. Since $c_1(0)$ and $c_2(0)$ are both non-zero, an autocatalytic sequence is immediately triggered (i.e. autocatalytic reactions are not stopped in the initial state). Let T_E^N be the duration of such autocatalytic sequence:

$$T_E^N = \inf\{\, t \geq 0 \colon C_1^{N,\gamma}(t) \in \{\, 0, w(0) \,\}\,\}.$$

The expected value of T_E^N, for large N, is approximated as

$$\mathbb{E}\left[T_E^N\right] \approx \frac{K}{\kappa'} N^{1-\gamma}, \qquad K = O(1).$$

At the second natural time-scale $\gamma = 1 + \alpha$, and therefore, for N growing large,

$$\mathbb{P}\left(T_E^N \geq \epsilon\right) \leq \frac{K}{\epsilon\, \kappa'} N^{1-\gamma} \;\to\; 0,$$

which in turn implies that the process "instantaneously" reaches one of the two absorbing states.

The probabilities that $C_1^{N,\gamma}$ ends up in $W^N(0)$ or 0, respectively, are equal to

$$F^N\left(C_1^N(0), W^N(0)\right) = \frac{C_1^N(0)}{W^N(0)} \quad \to \quad \frac{c_1(0)}{w(0)} := p$$
$$F^N\left(C_1^N(0), 0\right) = 1 - \frac{C_1^N(0)}{W^N(0)} \quad \to \quad 1 - \frac{c_1(0)}{w(0)} = 1 - p.$$

The statement of the Proposition follows by the arguments on right-continuity of the limit process already discussed in the case of a generic autocatalytic cascade. □

As a result, the candidate limit is defined as follows.

Definition 3.17. Let w be the deterministic process defined in (3.60),

$$w(\tau) = \left(w(0) - \frac{2\eta'}{\delta'}\right) \exp\left(-\delta'\tau\right) + \frac{2\eta'}{\delta'}, \qquad \tau \geq 0,$$

and let Y^{DIT} be the counting process defined in (3.84),

$$Y^{DIT}(\tau) = \xi^T\left(\eta' \int_0^\tau \frac{1}{w(s)}\, ds\right), \qquad \tau \geq 0.$$

Moreover, suppose that

$$\lim_{N\to\infty} C_1^N(0) = c_1(0) > 0, \qquad \lim_{N\to\infty} C_2^N(0) = c_2(0) > 0, \tag{3.89}$$

and let I be a Bernoulli random variable,

$$I \sim \text{Bernoulli}\left(p = \frac{c_1(0)}{w(0)}\right), \qquad w(0) = c_1(0) + c_2(0). \tag{3.90}$$

The *candidate* limit process C is defined, for each $\tau \geq 0$, as

$$C_1(\tau) = \begin{cases} 0 & \text{if} \quad Y^{DIT}(\tau) + I \text{ is even} \\ w(\tau) & \text{if} \quad Y^{DIT}(\tau) + I \text{ is odd} \end{cases} \tag{3.91}$$

$$C_2(\tau) = w(\tau) - C_1(\tau). \tag{3.92}$$

3.5 Convergence to the candidate limit process

Definition 3.17 suggests a candidate limit process C for the sequence of processes $\{ C^{N,\gamma''} \}$ of species concentrations at the second natural time-scale,

$$\lim_{N\to\infty} C^{N,\gamma''} \text{ "=" } C. \tag{3.93}$$

Such candidate limit is expressly defined on the Skorohod space $D[0,T]$ of cadlag functions, and therefore we would like to establish weak convergence (i.e. convergence in distribution) of the sequence $\{ C^{N,\gamma''} \}$ to C:

$$C^{N,\gamma''} \Rightarrow C.$$

However, it is not clear at all whether there exists a topology on the Skorohod space with respect to which such convergence in distribution actually holds.

The process $W^{N,\gamma''}$ describing the concentration of the sum of the two species, satisfying (3.54), converges in probability, and therefore in distribution, to the deterministic process w defined in (3.60):

$$W^{N,\gamma''} \;\Rightarrow\; w. \tag{3.94}$$

The topology on $D[0,T]$ implicitly assumed for this class of discontinuous processes is the Skorohod J_1-topology (Definition 2.34). However, since the deterministic limit process w is continuous, that is

$$w \in C[0,T],$$

the above convergence holds true if $D[0,T]$ is equipped with the topology induced by the uniform metric.

On the contrary, the process $C_1^{N,\gamma''}$ describing the concentration of species S_1 does not converge to the candidate limit process C_1, defined in (3.91), nor to any other limit process in the Skorohod J_1-topology. Indeed, the sequence of stochastic processes $\{ C_1^{N,\gamma''} \}$ is not relatively compact with respect to this topology.

Proposition 3.18. *Let $\{ C_1^{N,\gamma''} \}$ be the sequence of stochastic processes defined in (3.45), with $\gamma = \gamma''$. This sequence is not relatively compact with respect to the Skorohod J_1-topology on $D[0,T]$.*

In particular, there exist $\epsilon > 0$ and $\eta > 0$ for which, for every $\delta \in (0,T]$, there exist N_0 such that

$$\mathbb{P}\left(w'\left(C_1^{N,\gamma''}, \delta \right) \geq \epsilon \right) \geq \eta, \qquad \forall N \geq N_0.$$

Proof. Theorem 2.43 provides a necessary condition for a sequence $\{ \mu_n \}$ of probability measures on $D[0,T]$ to be relatively compact: for each $\epsilon > 0$ and $\theta > 0$, there exist $\delta \in (0,T]$ and n_0 such that

$$\mu_n \left(x \in D[0,T] \colon w'(x,\delta) \geq \epsilon \right) \leq \theta, \qquad n \geq n_0.$$

Such condition can be reformulated for a sequence $\{ C_1^{N,\gamma''} \}$ of random elements of $D[0,T]$ (random cadlag functions) as follows: for each $\epsilon > 0$ and $\theta > 0$, there exist $\delta \in (0,T]$ and N_0 such that

$$\mathbb{P}\left(w'\left(C_1^{N,\gamma''}, \delta \right) \geq \epsilon \right) \leq \theta, \qquad N \geq N_0.$$

In order to show that this condition does not hold, one should find $\epsilon > 0$ and $\theta > 0$ for which, for every $\delta \in (0,T]$ and every N_0, there exists $N \geq N_0$ such that

$$\mathbb{P}\left(w'\left(C_1^{N,\gamma''}, \delta \right) \geq \epsilon \right) \geq \theta.$$

Fix a height $\epsilon > 0$ strictly lower than the lowest value taken by deterministic function w:

$$0 < \epsilon < \min_{\tau \in [0,T]} w(\tau);$$

the number of autocatalytic cascades which cause a variation in species concentration of at least ϵ is a Poisson process described in Corollary 3.14:

$$Y^{N,V}(\epsilon; T) = \xi^T \left(\eta' \int_0^T \frac{1}{\epsilon} \left\{ \epsilon \leq W^{N,\gamma}(s) \right\} ds \right), \qquad \tau \geq 0.$$

By Remark 3.6 we may fix $\theta_1 > 0$ and find the corresponding N_1 such that

$$\mathbb{P}\left(\inf_{\tau \leq T} W^{N,\gamma''}(\tau) \geq \epsilon \right) \geq 1 - \theta_1, \qquad N \geq N_1;$$

conditioning on the event above, the probability that there exists at least one autocatalytic cascade causing a variation in concentration of at least ϵ is given by

$$\mathbb{P}\left(Y^{N,V}(\epsilon; T) > 0 \right) = \exp\left(-\frac{\eta'}{\epsilon} T \right).$$

Note that the expression for $Y^{N,V}(\epsilon; T)$ holds provided that the number $\hat{Y}^T(T)$ of autocatalytic cascades that are still running when another inflow of outflow occur is 0; by Proposition 3.12, one can find N_2 such that this condition is satisfied with sufficiently large probability,

$$\mathbb{P}\left(\hat{Y}^T(T) = 0 \right) \geq 1 - \theta_2, \qquad N \geq N_2.$$

Moreover, by Proposition 3.7, the total duration $\hat{T}_E^N(T)$ of autocatalytic cascades converges to 0 in probability: therefore, for every $\delta \in [0,T]$, one can find N_3 such that

$$\mathbb{P}\left(\bar{T}_E^N(T) \leq \delta \right) \geq 1 - \theta_3, \qquad N \geq N_3,$$

for a previously fixed $\theta_3 > 0$. Therefore, if we assume that an autocatalytic cascade causing a variation in concentration of at least ϵ happens, for every δ-sparse partition $\{ T_i \}$ there exist a time interval containing such autocatalytic cascade entirely, which implies that

$$w'\left(C_1^{N,\gamma''}, \delta \right) \geq \epsilon.$$

To summarize, for the height $\epsilon > 0$ fixed at the beginning, we may choose $\hat{\theta} > 0$ and fix

$$\theta = (1 - \hat{\theta}) \exp\left(-\frac{\eta'}{\epsilon} T \right).$$

Then, for every $\delta \in [0, T]$, one can find N_0 such that the events

$$\inf_{\tau \leq T} W^{N,\gamma''}(\tau) \geq \epsilon, \qquad \hat{Y}^T(T) = 0, \qquad \bar{T}_E^N(T) \leq \delta,$$

jointly happen with probability larger than $1 - \hat{\theta}$, for every $N \geq N_0$.

Given such events, the probability that there exists at least one autocatalytic cascade causing a variation in concentration of at least ϵ and occurring in a time span smaller than δ is given by

$$\mathbb{P}\left(Y^{N,V}(\epsilon; T) > 0 \right) = \exp\left(-\frac{\eta'}{\epsilon} T \right) > 0.$$

This implies that

$$\mathbb{P}\left(w'\left(C_1^{N,\gamma''}, \delta \right) \geq \epsilon \right) \geq \theta. \qquad \square$$

The presence of Discreteness Induced Transitions and peaks of finite height is the reason why the sequence $\{ C_1^{N,\gamma''} \}$ is not relatively compact in the Skorohod J_1-topology. In particular:

- Discreteness Induced Transitions prevent relative compactness to hold because they result from a sequence of infinitesimal jumps which coalesce in the limit into a single jump of finite height;
- peaks prevent relative compactness to hold because they reach a finite height and return back within an infinitesimal time span.

In consequence of Proposition 3.18, there exists at least a subsequence of $\{ C_1^{N,\gamma''} \}$ not containing any further subsequence which is convergent in distribution. This in turn implies that the entire sequence cannot converge to any random element of $D[0, T]$ equipped with the Skorohod J_1-topology.

Although convergence with respect to the classical Skorohod topology has to be set aside for lack of relative compactness, the limit process suggested in Definition 3.17 seems a reasonable candidate. A first step towards obtaining convergence in distribution to such candidate limit is to find a topology on the Skorohod space of cadlag functions with respect to which the sequence $\{ C_1^{N,\gamma''} \}$ is relatively compact.

In light of the main criticalities highlighted above, preventing relative compactness with respect to J_1-topology to hold, we are looking for a topology on $D[0, T]$ satisfying the following requirements:

- a cascade of infinitesimal jumps leading to a finite variation within an infinitesimal time span converges to a single jump;
- a peak of finite height occurring within an infinitesimal time span is self-cancelling in the limit;

- the topology is strictly weaker than Skorohod J_1-topology.

The Jakubowski S-topology, described in Section 2.7, perfectly fits this framework, since it is weaker than the J_1-topology and features self-cancelling oscillations (see Example 2.69). As a matter of fact, the sequence $\{ C_1^{N,\gamma''} \}$ is uniformly tight with respect to the S-topology, as proved in Proposition 3.21.

The following results are preliminary to the proof of that Proposition; however, they are stated independently, being of interest on their own.

Lemma 3.19. *Let $\{ X_n \}$ be a sequence of non-negative random variables. Assume that there exists a constant $M < \infty$ such that*

$$\mathbb{E}[X_n] \leq M, \qquad \textit{for all} \quad n \geq 0.$$

Then, the sequence $\{ X_n \}$ is uniformly tight.

Proof. By definition, a sequence of $\mathbb{R}$-valued random variables is uniformly tight if, for each $\epsilon > 0$, there exists a compact set $K_\epsilon \subseteq \mathbb{R}$ such that

$$\mathbb{P}(X_n \in K_\epsilon) \geq 1 - \epsilon, \qquad \text{for all} \quad n \geq 0.$$

Since X_n is a non-negative random variable, by Markov's inequality we obtain

$$\mathbb{P}(X_n \geq L) \leq \frac{1}{L}\mathbb{E}[X_n], \qquad \text{for all} \quad n \geq 0;$$

moreover, by hypothesis,

$$\mathbb{E}[X_n] \leq M, \qquad \text{for all} \quad n \geq 0,$$

and therefore

$$\mathbb{P}(X_n \geq L) \leq \frac{M}{L}, \qquad \text{for all} \quad n \geq 0.$$

As a consequence, for each $\epsilon > 0$, there exists a compact set (i.e. closed and bounded)

$$K_\epsilon := \left[0 \ , \ L = \frac{M}{\epsilon}\right] \subseteq \mathbb{R},$$

such that

$$\mathbb{P}(X_n \in K_\epsilon) = \mathbb{P}(X_n \leq L) \geq 1 - \frac{M}{L} \geq 1 - \epsilon, \qquad \text{for all} \quad n \geq 0. \qquad \square$$

Proposition 3.20. *Let $U^N(a, b; \tau)$ be the process describing the number of upcrossings of levels $a < b$, up to time τ, for process $C_1^{N,\gamma}$ at the second natural time scale γ'' (see Definition 2.71).*

Then, for every $a \geq 0$, $a < b$, there exists a constant $M < \infty$ such that, for sufficiently large N,

$$\mathbb{E}\left[U^N(a, b; T)\right] \leq M.$$

Proof. Consider first the case for $a = 0$, $b > 0$.

The number $U^N(0,b;T)$ of upcrossings of levels $0 < b$, up to time T, is controlled by the number $Y^{N,V}(b;T)$ of autocatalytic cascades causing a variation in concentration of at least b:

$$U^N(0,b;T) \leq Y^{N,V}(b;T).$$

Indeed, some of these cascades are triggered when process $C_1^{N,\gamma''}$ is in state 0: such cascades lead to a single upcrossing of levels $0 < b$, since the process cannot move back to 0 and then reach state b again within the same autocatalytic sequence.

From the proof of Corollary 3.14, we obtain

$$\mathbb{E}\left[Y^{N,V}(b;T)\right] = \eta' \int_0^T \frac{1}{b}\left\{b \leq W^{N,\gamma}(s)\right\} ds \leq \frac{\eta' T}{b},$$

and therefore

$$\mathbb{E}\left[U^N(0,b;T)\right] \leq \mathbb{E}\left[Y^{N,V}(b;T)\right] \leq \frac{\eta' T}{b}.$$

Consider now the general case for $a > 0$, $b > a$. Suppose that, at time $\hat{\tau}$, the process $C_1^{N,\gamma''}$ is in state $a > 0$, with $a < s$, for the first time within the ongoing autocatalytic cascade:

$$C_1^{N,\gamma}(\hat{\tau}) = a = \frac{A}{N}, \qquad W^{N,\gamma}(\hat{\tau}) = s = \frac{S}{N}.$$

As derived in the Appendix, the number $\hat{U}(a,b)$ of upcrossing of levels $a < b$ before the end of the autocatalytic sequence has a geometric distribution,

$$\hat{U}(a,b) = \hat{U}(a,b;s) \sim \text{Geometric}\left(\frac{s}{b} \cdot \frac{b-a}{s-a}\right).$$

Clearly, if $b > s$ this expression does not make sense, since the process cannot reach state b moving inside the interval of values $[0,s]$.

As a consequence, we obtain

$$\mathbb{E}\left[\hat{U}(a,b)\right] = \frac{a}{s} \cdot \left(\frac{s-b}{b-a}\right)^+.$$

In the particular case $b = s$, the expected number of upcrossing is actually equal to the probability of ending up in s before reaching 0, that is

$$\mathbb{E}\left[\hat{U}(a,s)\right] = \frac{a}{s}\left\{a < s\right\}.$$

Let $Y^U(a;\tau)$ be the number of autocatalytic cascades reaching state $a > 0$, up to time τ. The number of upcrossings in each cascade depends only on the value of the sum process at that time, and so

$$\hat{U}\left(a,b;\, W^{N,\gamma''}(\tau_i)\right), \qquad i = 1,\ldots,Y^U(a;\tau),$$

where $\{\tau_i\}$ are the jump times of process Y^U, are conditionally independent random variables.

The total number $U^N(a,b;T)$ of upcrossings of levels $a<b$, up to time T, can be expressed as

$$U^N(a,b;T) = \sum_{i=1}^{Y^U(a;T)} \hat{U}\left(a,b;\, W^{N,\gamma''}(\tau_i)\right).$$

Conditioning on the counting process $Y^U(T)$, we can write the expected value of $U^N(a,b;T)$ as

$$\begin{aligned}\mathbb{E}\left[U^N(a,b;T)\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{Y^U(a;T)} \hat{U}\left(a,b;\, W^{N,\gamma''}(\tau_i)\right) \;\middle|\; Y^U(a;T)\right]\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{Y^U(a;T)} \mathbb{E}\left[\hat{U}\left(a,b;\, W^{N,\gamma''}(\tau_i)\right)\right]\right].\end{aligned}$$

By Remark 3.6, for sufficiently large N, we may assume that

$$0<m\le W^{N,\gamma''}(\tau_i)\le M<\infty, \qquad i=1,\dots,Y^U(a;T),$$

and therefore obtain

$$\mathbb{E}\left[\hat{U}\left(a,b;W^{N,\gamma''}(\tau_i)\right)\right] \le \frac{a\,(M-b)}{m\,(b-a)}, \qquad i=1,\dots,Y^U(a;T).$$

As a consequence,

$$\mathbb{E}\left[U^N(a,b;T)\right] \le \mathbb{E}\left[Y^U(a;T)\,\frac{a\,(M-b)}{m\,(b-a)}\right] = \mathbb{E}\left[Y^U(a;T)\right]\,\frac{a\,(M-b)}{m\,(b-a)}.$$

The number $Y^U(a;T)$ of autocatalytic cascades reaching state a, up to time T, is controlled by the number $Y^{N,V}(c;T)$ of autocatalytic cascades causing a variation in concentration of at least c,

$$Y^U(a;T)\le Y^{N,V}(c;T), \qquad \text{where} \quad c=\min\{\,a,\,b-a\,\}.$$

Indeed, either the process $C_1^{N,\gamma''}$ reaches state a from state 0, with a variation of concentration of at least a, or reaches state a from the state s of the sum process at the time the autocatalytic cascade is triggered. However, in case level b is larger than state s, such autocatalytic sequence can be disregarded, since no upcrossing can happen. Therefore, in this case, $b<s$ and the variation of concentration is at least $b-a$.

From the proof of Corollary 3.14, we obtain

$$\mathbb{E}\left[Y^{N,V}(c;T)\right] = \eta'\int_0^T \frac{1}{c}\left\{c\le W^{N,\gamma}(s)\right\}ds \le \frac{\eta' T}{c},$$

and therefore

$$\mathbb{E}\left[Y^U(a;T)\right] \le \mathbb{E}\left[Y^{N,V}(c;T)\right] \le \frac{\eta' T}{c}.$$

In conclusion, we have that

$$\mathbb{E}\left[U^N(a,b;T)\right] \le \mathbb{E}\left[Y^U(a;T)\right]\,\frac{a\,(M-b)}{m\,(b-a)} \le \frac{\eta' T}{c}\,\frac{a\,(M-b)}{m\,(b-a)}. \qquad \square$$

Proposition 3.21. *Let* $\{ C_1^{N,\gamma''} \}$ *be the sequence of stochastic processes defined in* (3.45)*, with* $\gamma = \gamma''$*. This sequence is uniformly tight with respect to the Jakubowsky S-topology.*

Proof. Theorem 2.80 provides necessary and sufficient conditions for a familiy $\{ C_1^{N,\gamma''} \}$ of random elements of $D[0,T]$ to be uniformly tight in the S-topology:

1. the sequence $\{ \|C_1^{N,\gamma''}\| \}$ is a uniformly tight family of $\mathbb{R}$-valued random variables;
2. for each $a < b$, the sequence $\{ U^N(a,b;T) \}$ is a uniformly tight family of $\mathbb{N}$-valued random variables.

By definition of process $C_1^{N,\gamma''}$, for each $N \geq 0$,

$$0 \leq C_1^{N,\gamma''}(\tau) \leq W^{N,\gamma''}(\tau), \qquad \tau \geq 0,$$

and therefore,

$$\|C_1^{N,\gamma''}\| := \sup_{\tau \leq T} C_1^{N,\gamma''}(\tau) \ \leq \ \sup_{\tau \leq T} W^{N,\gamma''}(\tau) =: \|W^{N,\gamma''}\|.$$

By Remark 3.6, for each $\epsilon > 0$, there exist $M < \infty$ and N_0 such that,

$$\mathbb{P}\left(\sup_{\tau \leq T} W^{N,\gamma''}(\tau) \leq M \right) \geq 1 - \epsilon, \qquad N \geq N_0.$$

As a consequence, for each $\epsilon > 0$, there exist a compact set

$$K_\epsilon := [0, M],$$

and N_0 such that

$$\mathbb{P}\left(\|C_1^{N,\gamma''}\| \in K_\epsilon \right) \geq 1 - \epsilon, \qquad N \geq N_0,$$

which means that the sequence $\{ \|C_1^{N,\gamma''}\| \}$ is uniformly tight.

By Proposition 3.20, for every $a \geq 0$, $a < b$, there exist a constant $M < \infty$ and N_0 such that

$$\mathbb{E}\left[U^N(a,b;T) \right] \leq M, \qquad N \geq N_0.$$

Since $\{ U^N(a,b;T) \}$ is a sequence of non-negative random variables, the above condition is enough to conclude that it is uniformly tight, in consequence of Lemma 3.19. □

Theorem 2.76 contains both the direct and converse Prohorov theorems for the space $\mathcal{P}(D[0,1], \mathcal{B}_S)$ of tight probability measures on $(D[0,1], \mathcal{B}_S)$, equipped with convergence in the sense of Jakubowski. This notion of convergence is introduced in Definition 2.62 and represent the "equivalent" of weak convergence for non-metric topological spaces.

As a consequence, since the sequence $\{ C_1^{N,\gamma''} \}$ is uniformly tight with respect to the S-topology, as proved in Proposition 3.21, then it is relatively (sequentially) compact with respect to convergence in the sense of Jakubowski. By definition,

this means that each subsequence contains a further subsequence which admits a strong Skorohod representation, as defined in Theorem 2.59; in particular, these further subsequences are convergent in distribution in the classical sense.

In order to prove convergence of the sequence $\{ C_1^{N,\gamma''} \}$ to the candidate limit C_1 in the space $\mathcal{P}(D[0,1], \mathcal{B}_S)$,

$$C_1^{N,\gamma''} \stackrel{*}{\Rightarrow} C_1,$$

we still need to prove convergence of finite-dimensional distributions, that is, we should verify that, for every choice of $\tau_1, \ldots, \tau_k \in [0,T]$,

$$\left(C_1^{N,\gamma''}(\tau_1), \ldots, C_1^{N,\gamma''}(\tau_k) \right) \Rightarrow_N (C_1(\tau_1), \ldots, C_1(\tau_k)) .$$

As highlighted in Theorem 2.77, it is actually sufficient to prove convergence of finite-dimensional distribution on a dense subset $\mathbb{Q} \subseteq [0,T]$ containing T: for every choice of $q_1, \ldots, q_k \in \mathbb{Q}$,

$$\left(C_1^{N,\gamma''}(q_1), \ldots, C_1^{N,\gamma''}(q_k) \right) \Rightarrow_N (C_1(q_1), \ldots, C_1(q_k)) .$$

Formal results and empirical observations collected in these sections suggest that convergence of finite-dimensional distributions holds. However, a rigorous proof of this fact has not been devised yet.

To conclude, consider again the process $W^{N,\gamma''}$ describing the concentration of the sum of the two species; as recalled at the beginning of this section, it converges in distribution with respect to the Skorohod J_1-topology to process w,

$$W^{N,\gamma''} \Rightarrow w.$$

Since the S-topology is weaker than the J_1-topology, convergence with respect to the latter implies convergence with respect to the former: as a consequence,

$$W^{N,\gamma''} \stackrel{*}{\Rightarrow} w.$$

Moreover, addition is sequentially continuous in the S-topology, that is

$$x_n \to_S x, \quad y_n \to_S y \qquad \text{implies} \qquad x_n + y_n \to_S x + y.$$

Therefore, if we assume that the sequence $\{ C_1^{N,\gamma''} \}$ converges to the candidate limit C_1, then it directly follows that sequence $\{ C_2^{N,\gamma''} \}$ converges to its corresponding candidate limit C_2:

$$C_2^{N,\gamma''} := W^{N,\gamma''} - C_1^{N,\gamma''} \stackrel{*}{\Rightarrow} w - C_1 =: C_2.$$

Appendix: Continuous time birth-and-death Markov chain with absorbing states

Consider a continuous time birth-and-death Markov chain $\{ X(t) \colon t \geq 0 \}$ on the state space

$$\mathbb{S} = \{ 0, \ldots, S \},$$

with birth and death rates given by, respectively,

$$\lambda(x) = x(S-x), \qquad \mu(x) = x(S-x), \qquad x \in \mathbb{S}.$$

States can be classified into *transient states* $\mathcal{T}$,

$$\mathcal{T} = \{\, 1, \ldots, S-1 \,\},$$

and *recurrent (absorbing) states* $\mathcal{R}$,

$$\mathcal{R} = \{\, 0, S \,\}.$$

The transition rate matrix Q (after potential reordering) could be partitioned as

$$Q = \begin{bmatrix} Q_{T,T} & Q_{T,R} \\ 0 & Q_{R,R} \end{bmatrix} \in \mathbb{R}^{(S+1)\times(S+1)},$$

where $Q_{T,T}$ contains transition rates among transient states:

$$Q(x, x+1) = \lambda(x) = x(S-x), \qquad Q(x, x-1) = \mu(x) = x(S-x),$$
$$Q(x,x) = -\lambda(x) - \mu(x) = -2x(S-x).$$

Note that, in this case, $Q_{R,R} = 0$.

Moreover, let P be the transition matrix of the embedded discrete time Markov chain, denoted by $\{\, Y(t) \colon t \in \mathbb{N} \,\}$:

$$P = \begin{bmatrix} P_{T,T} & P_{T,R} \\ 0 & P_{R,R} \end{bmatrix} \in \mathbb{R}^{(S+1)\times(S+1)},$$

where

$$P(x,y) = \frac{Q(x,y)}{Q(x,x)}, \qquad x \in \mathcal{T}.$$

Exit distribution

Let $F(x,y)$ be the probability that, starting from $x \in \mathcal{T}$, the process visits $y \in \mathcal{R}$ as the first absorbing state, i.e. the process dies in y:

$$F(x,y) = \mathbb{P}_x(y = \arg\min_{z \in \mathcal{R}} \tau_z),$$

where τ_z is the hitting time for state z.

Conditioning on the first step, one can show that

$$F = P_{T,T}F + P_{T,R},$$

and therefore, since matrix $I - P_{T,T}$ is invertible,

$$F = (I - P_{T,T})^{-1} P_{T,R}.$$

By direct computation, one obtains:

$$F(x,0) = 1 - \frac{x}{S}, \qquad F(x,S) = \frac{x}{S}.$$

Expected number of visits to transient state

Let $N(x, y)$ be the expected number of visits to transient state $y \in \mathcal{T}$, starting from $x \in \mathcal{T}$:

$$N(x, y) = \mathbb{E}_x[1_{Y(t)=y}].$$

One can show that

$$N = \sum_{n=0}^{\infty} (P_{T,T})^n,$$

and, since matrix $P_{T,T}$ is substochastic, it can be rewritten as

$$N = (I - P_{T,T})^{-1},$$

By direct computation, one obtains:

$$N(x, y) = 2\min(x, y)\left(1 - \frac{\max(x, y)}{S}\right), \qquad x, y \in \mathcal{T}.$$

Expected time spent in transient state

Let $T(x, y)$ be the expected time spent in transient state $y \in \mathcal{T}$, starting from $x \in \mathcal{T}$:

$$T(x, y) = \mathbb{E}_x[1_{X(t)=y}].$$

One can easily show that

$$T(x, y) = -Q_{T,T}^{-1} = \frac{N(x, y)}{\lambda(y) + \mu(y)},$$

and therefore

$$T(x, y) = \frac{1}{S}\min\left(\frac{x}{y}, \frac{S - x}{S - y}\right).$$

Expected time to reach a absorbing state

Let $T(x)$ be the expected time it takes the process to enter $\mathcal{R}$, starting from $x \in \mathcal{T}$:

$$T(x) = \mathbb{E}_x[\min_{z \in \mathcal{R}} \tau_z].$$

It directly follows that

$$T(x) = \sum_{y \in \mathcal{T}} T(x, y),$$

and therefore

$$T(x) = \frac{S - x}{S} \sum_{y=1}^{x} \frac{1}{S - y} + \frac{x}{S} \sum_{y=1}^{S-x} \frac{1}{S - y} - \frac{1}{S}.$$

Scaling of state space with respect to N

Let us rewrite the elements of the state space as functions of a common parameter N:

$$x = kN, \qquad S = sN, \qquad k \in [0, s].$$

The exit distribution is easily adapted to this new definition,

$$F^N(k,0) = F(kN,0) = 1 - \frac{k}{s}, \qquad F^N(k,s) = F(kN,sN) = \frac{k}{s}.$$

Likewise, the expected time to reach a absorbing state becomes

$$T^N(k) = T(kN) = \left(1 - \frac{k}{s}\right) \sum_{z=1/N}^{k} \frac{1}{N} \frac{1}{s-z} \; + \; \frac{k}{s} \sum_{z=1/N}^{s-k} \frac{1}{N} \frac{1}{s-z} \; - \; \frac{1}{sN}.$$

For large N, the above expression can be approximated as follows:

$$\begin{aligned} T^N(k) = T(kN) &\approx \left(1 - \frac{k}{s}\right) \int_0^k \frac{1}{s-z} \, dz + \frac{k}{s} \int_0^{s-k} \frac{1}{s-z} \, dz \\ &= -\left(1 - \frac{k}{s}\right) \log\left(1 - \frac{k}{s}\right) - \frac{k}{s} \log\left(\frac{k}{s}\right). \end{aligned}$$

Note that the maximum is reached for $k = s/2$ (i.e. for $x = S/2$),

$$T^N(s/2) = T(S/2) \approx \log 2,$$

and, for $x = 1$, as N grows large,

$$k = \frac{x}{N} \; \to \; 0, \qquad T(1) = T^N(1/N) \approx \frac{\log(sN)}{sN} \; \to \; 0.$$

Finally, if reaction rates are multiplied by a constant $\alpha = \alpha(N)$, the exit distribution remains unchanged, while the expected time to reach an absorbing state is given by

$$T^{N,\alpha}(k) = \frac{1}{\alpha(N)} \, T^N(k).$$

Number of upcrossing distribution

Let $U(x,y)$ be the number of upcrossings of levels $x, y \in \mathcal{T}$, with $x < y$, starting from x.

Recall that $U(x,y)$ is defined as follows: $U(x,y) \geq k$ if there exist time instants

$$0 \leq t_1 < t_2 < \cdots < t_{2k-1} < t_{2k}$$

such that

$$Y(t_{2i-1}) \leq x, \qquad Y(t_{2i}) \geq y, \qquad i = 1, \ldots, k.$$

If the process is in state x, the probability of moving up to y and then come back to x before reaching an absorbing state is given by

$$p^U(x,y) = \frac{x}{y} \cdot \frac{S-y}{S-x}.$$

This expression derives from the result on the exit distribution stated above, and uses in an essential way the fact that transition probabilities of the embedded discrete time Markov chain do not depend on the state. Indeed, embedded discrete time Markov chains defined on the subsets of the state space $\mathcal{S}$,

$$\mathbb{S}_1 = \{\, 0, \ldots, y \,\} \subseteq \mathcal{S}, \qquad \mathbb{S}_2 = \{\, x, \ldots, S \,\} \subseteq \mathcal{S},$$

are "equivalent" to the one defined on the entire state space.

As a consequence, the number $U(x, y)$ of upcrossings of levels $x < y$, starting from x, is a geometric random variable with parameter $1 - p^U(x, y)$:

$$U(x, y) \sim \text{Geometric}\left(\frac{S}{y} \cdot \frac{y - x}{S - x}\right).$$

Assume that elements of the state space are rewritten as functions of a common parameter N,

$$x = kN, \qquad y = hN, \qquad S = sN, \qquad k, h \in [0, s];$$

the number of upcrossing is easily adapted to this new definition:

$$U^N(k, h) = U(kN, hN) \sim \text{Geometric}\left(\frac{s}{h} \cdot \frac{h - k}{s - k}\right).$$

Bibliography

[1] D.F. Anderson and T.G. Kurtz. *Stochastic Analysis of Biochemical Systems*. Vol. 1. Mathematical Biosciences Institute Lecture Series. Springer, 2015.

[2] K. Ball et al. "Asymptotic Analysis of Multiscale Approximations to Reaction Networks". In: *The Annals of Applied Probability* 16.4 (2006), pp. 1925–1961.

[3] E. Bibbona, J. Kim, and C. Wiuf. "Stationary Distributions of Systems with Discreteness Induced Transitions". In: *Journal of The Royal Society Interface* 17.168 (2020).

[4] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 1999.

[5] D. Blackwell and L. Dubbins. "An Extension of Skorohod's Almost Sure Representation Theorem". In: *Proceedings of the American Mathematical Society* 89 (1983), pp. 691–692.

[6] V.I. Bogachev and A.V. Kolesnikov. "Open Mappings of Probability Measures and the Skorohod Representation Theorem". In: *Theory of Probability and Its Applications* 46.1 (2002), pp. 20–38.

[7] S.N. Ethier and T.G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 2005.

[8] X. Fernique. "Processus linéaires, processus généralisés". In: *Annales de l'Institut Fourier* 17.1 (1967), pp. 1–92.

[9] A. Jakubowski. "A Non-Skorohod Topology on the Skorohod Space". In: *Electronic Journal of Probability* 2.4 (1997), pp. 1–21.

[10] A. Jakubowski. "From Convergence of Functions to Convergence of Stochastic Processes. On Skorohod's Sequential Approach to Convergence in Distribution". In: (1997).

[11] A. Jakubowski. "New Characterizations of the S Topology on the Skorokhod Space". In: *Electronic Communications in Probability* 23.2 (2018), pp. 1–16.

[12] A. Jakubowski. "The Almost Sure Skorohod Representation for Subsequences in Non-Metric Spaces". In: *Theory of Probability and Its Applications* 42.1 (1997), pp. 167–174.

[13] H. Kang and T.G. Kurtz. "Separation of Time-Scales and Model Reduction for Stochastic Reaction Networks". In: *The Annals of Applied Probability* 23.2 (2013), pp. 529–583.

[14] T.G. Kurtz. “Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes”. In: *Journal of Applied Probability* 7.1 (1970), pp. 49–58.

[15] M. Polettini, A. Wachtel, and M. Esposito. “Dissipation in Noisy Chemical Networks: The Role of Deficiency”. In: *The Journal of Chemical Physics* 143.184103 (2015).

[16] A.V. Skorohod. “Limit Theorems for Stochastic Processes”. In: *Theory of Probability and its Applications* 1.3 (1956), pp. 261–290.

[17] Y. Togashi and K. Kaneko. “Transitions Induced by the Discreteness of Molecules in a Small Autocatalytic System”. In: *Physical Review Letters* 86.2459 (2001).

[18] A. Van Der Schaft, S. Rao, and B. Jayawardhana. “On the Mathematical Structure of Balanced Chemical Reaction Networks Governed by Mass Action Kinetics”. In: *SIAM Journal on Applied Mathematics* 73.2 (2013), pp. 953–973.