




General solution of the chemical master equation and modality of marginal distributions for hierarchic first-order reaction networks

Matthias Reis¹ · Justus A. Kromer² ·
Edda Klipp¹ 

Received: 7 April 2017 / Revised: 10 October 2017 / Published online: 20 January 2018
© The Author(s) 2018. This article is an open access publication

Abstract Multimodality is a phenomenon which complicates the analysis of statistical data based exclusively on mean and variance. Here, we present criteria for multimodality in hierarchic first-order reaction networks, consisting of catalytic and splitting reactions. Those networks are characterized by independent and dependent subnetworks. First, we prove the general solvability of the Chemical Master Equation (CME) for this type of reaction network and thereby extend the class of solvable CME's. Our general solution is analytical in the sense that it allows for a detailed analysis of its statistical properties. Given Poisson/deterministic initial conditions, we then prove the independent species to be Poisson/binomially distributed, while the dependent species exhibit generalized Poisson/Khatri Type B distributions. Generalized Poisson/Khatri Type B distributions are multimodal for an appropriate choice of parameters. We illustrate our criteria for multimodality by several basic models, as well as the well-known two-stage transcription–translation network and Bateman's model from nuclear physics. For both examples, multimodality was previously not reported.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00285-018-1205-2>) contains supplementary material, which is available to authorized users.

✉ Edda Klipp
edda.klipp@hu-berlin.de
Matthias Reis
matthias.reis@hu-berlin.de
Justus A. Kromer
justus.kromer@tu-dresden.de

¹ Institute of Biology, Theoretical Biophysics, Humboldt-Universität zu Berlin, 10117 Berlin, Germany

² Center for Advancing Electronics Dresden, Technische Universität Dresden, 01062 Dresden, Germany

Keywords Hierarchically linear ODE system · Hierarchic reaction network · Chemical master equation · Conditional multimodality · Compound process · Bursting process · Generalized poisson distribution · Generalized binomial distribution · Nuclear decay chain · Transcription–translation model

Mathematics Subject Classification 92C45 · 60J27 · 05A15

1 Introduction

The development of single-molecule methods such as Fluorescence *in situ* Hybridization (FISH) has resulted in considerable advances in fundamental research in biology (Trcek et al. 2011, 2012). Those methods provide researchers with discrete, single-molecule data on (bio-)chemical reaction networks. The measured molecule numbers are often small, requiring a modeling approach that accounts for the discreteness of the molecule numbers (Klipp et al. 2009).

Networks of stochastically reacting molecule species are often described by the Chemical Master Equation (CME) (Gardiner 2009; Kampen 2011). The CME is a difference–differential equation. By using Gillespie’s Stochastic Simulation Algorithm (Gillespie 1977), an exact realization of the CME’s underlying Markov jump process is sampled and an indirect solution is obtained (Gillespie 1992). The CME may also be studied by approximation methods such as finite state projection (Munsky and Khammash 2006), or exactly using analytical approaches. Stationary solutions to the CME are known for a wide range of chemical reaction networks (Kampen 1976; Anderson et al. 2010). However, time-dependent, analytical solutions only exist for monomolecular reaction networks (Gans 1960; Jahnke and Huisinga 2007), consisting exclusively of conversion $S_i \leftrightarrow S_j$, degradation $S_i \rightarrow \emptyset$ and production $\emptyset \rightarrow S_j$ reactions, where S_i defines a species of molecules. In contrast to numerical solution methods, analytical solutions allow for an abstract study of reaction networks, independently of the concrete value of reaction rate constants and network graphs.

Here, we present an extension of the class of analytically solvable CME’s to a subset of first-order reactions, namely hierarchic first-order networks. In addition to monomolecular reactions, first-order reactions include (auto-)catalytic and splitting reactions, and therefore allow for more real-world applications. As a subset of those networks, we define hierarchic networks (Def. 3), characterized by a division into independent and dependent subnetworks. Common examples of hierarchic networks are the transcription–translation model in molecular biology (Friedman et al. 2006; Shahrezaei and Swain 2008), or nuclear decay chains in physics (Bateman 1910; Pressyanov 2002), both of which we discuss here.

In contrast to monomolecular networks, first-order networks result in marginal distributions of individual species that are not unimodal in general, as we show in Theorems 3 and 4. These distributions are poorly characterized by mean and variance, two essential measures used to describe statistical data in natural sciences. On one hand, characterizing multimodal distributions by mean and variance is unintuitive, since the mean is often a very unlikely outcome. On the other hand, many unimodal distributions, such as the Gaussian distribution, are fully specified by the

first two cumulants, i.e. mean and variance, or just the mean in case of the Poisson distribution. Contrastingly, an infinite number of cumulants can be necessary to fully specify the multimodal distributions derived in this article. We therefore argue that CME approximation methods based on the first two cumulants or moments, such as the moment-closure method (Milner et al. 2012), yield insufficient results when reaction networks lead multimodal molecule distributions. Albeit to a lesser degree, this holds true even when more than the first two moments are used to approximate the CME's solution in approaches such as Constantino et al. (2016). Such distributions have already been reported for networks that include two-state or three-state random variables such as promoter switches studied in the context of gene-regulation (Shahrezaei and Swain 2008; Thomas et al. 2014). In the same context, a maximum-entropy approach has been used to approximate multimodal molecule distributions successfully using the first seven moments (Andreychenko et al. 2017).

In the present paper, we develop an exact theoretical description for hierarchic first-order networks, based on probability generating functions. Hierarchic first-order networks are introduced in Sect. 2, followed by the main results in Sect. 3. In Sect. 3.1, we establish the existence of analytical solutions (Theorem 1) and derive the *joint* probability generating functions (Proposition 2), which are complete characterizations of the underlying statistics. Then, the analytical form of the *marginal* distributions of individual species is derived in Sect. 3.2. We show that the independent part of the network exhibits Poissonian and Binomial marginal distributions, while the dependent part is described by Discrete Compound Poisson (DCP) and Kathri Type B (KTB) marginal distributions. Next, we present criteria for the marginal distributions to be multimodal in Sect. 3.3. Since Poissonian/Binomial distributions are generally unimodal, the independent part exhibits unimodal marginal distributions (Theorem 2). In contrast, DCP/KTB marginal distributions from the dependent part can lead to multimodality under quite general conditions (Theorems 3 and 4). We illustrate these general results by several basic models (Sect. 4.1) and two real-world models (Sect. 4.2). Among these are the transcription–translation model (Example 5) and Bateman's model for nuclear decay chains (Example 6). For the former, multimodality was previously only discussed in an extended three-stage model variant (Shahrezaei and Swain 2008), whereas Theorem 3 proves the protein numbers to be multimodally distributed even in the simple two-species version. To the best of our knowledge, our exact solution to the CME for Bateman's model of nuclear decay chains is a novel result, applying the techniques developed in the preceding sections. We apply Theorem 4 to show multimodality for Bateman's model.

The derivation of probability mass functions from generating functions is reviewed in Appendix A.1.

2 Introduction to hierarchic first-order networks

2.1 Chemical master equation and characteristic ODEs

First of all, we define a network of n species with m chemical reactions by the $n \times m$ stoichiometric matrices $\mathbf{Q} = (Q_{ij})$ and $\mathbf{R} = (R_{ij})$, where $Q_{ij}, R_{ij} \in \mathbb{N}_0$. These matrices enable us to write

$$\mathbf{Q}^T \mathbf{S} \xrightarrow{\mathbf{k}} \mathbf{R}^T \mathbf{S},$$

with $\mathbf{S} = (S_1, \dots, S_n)^T$ being the vector of species and $\mathbf{k} = (k_1, \dots, k_m)^T$ the vector of rate constants.

In the present paper, we consider discrete and stochastic models of biochemical reaction networks. Let $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$ be the probability distribution for the number of molecules \mathbf{x} of each species at time t , given the initial condition \mathbf{x}_0 at t_0 . This distribution is described by the Chemical Master Equation, a difference–differential equation defined by¹

$$\partial_t P(\mathbf{x}, t) = \sum_{i=1}^m \left(\alpha_i(\mathbf{x} - \mathbf{n}_i) P(\mathbf{x} - \mathbf{n}_i, t) - \alpha_i(\mathbf{x}) P(\mathbf{x}, t) \right), \tag{1}$$

where

$$\alpha_i(\mathbf{x}) := k_i \prod_{j=1}^n \frac{x_j!}{(x_j - Q_{ji})!} \text{ and } \mathbf{n}_i := \text{col}_i(\mathbf{R} - \mathbf{Q}). \tag{2}$$

In the last equation col_i denotes the i th column of a matrix. The CME (1) is based on the Markovian assumption and is derived in detail in Gardiner (2009) or Kampen (2011). Applicable for (multivariate) distributions of discrete random variables, we introduce the concept of *generating functions*:

Definition 1 (*Generating function*) A probability generating function (PGF) is a power series²

$$g(\mathbf{s}, t) := \sum_{\mathbf{x}=\mathbf{0}}^{\infty} \mathbf{s}^{\mathbf{x}} P(\mathbf{x}, t) \tag{3}$$

defined for $\mathbf{s} \in \mathbb{C}^n$ within some radius of convergence ≤ 1 in each coordinate respectively.

In the univariate case, a large list of correspondences between PGFs and distributions is provided by Johnson et al. (2005). Next, we represent the Master Eq. (1) as a partial differential equation (PDE) in terms of a generating function:

Lemma 1 *The Master Eq. (1) can equivalently be expressed as the partial differential equation*

$$\partial_t g(\mathbf{s}, t) = \sum_{i=1}^m k_i \left(\prod_{j=1}^n s_j^{R_{ji}} - \prod_{j=1}^n s_j^{Q_{ji}} \right) \left(\prod_{j=1}^n \partial_{s_j}^{Q_{ji}} \right) g(\mathbf{s}, t). \tag{4}$$

¹ We suppress the dependence on the initial states \mathbf{x}_0 in the following.

² We suppress the generating function’s dependence on \mathbf{x}_0 and t_0 in the following.

This result can be found in textbooks such as Gardiner (2009), Sect. 11.6.4, p. 297. We provide a proof in “Appendix A.2” to the convenience of the reader.

We define *first-order networks* as those obeying a first-order PDE for the generating function ($\sum_{j=1}^n Q_{ji} \leq 1$), or equivalently a linear ordinary differential equation (ODE) system for the moments. In terms of chemical reactions, first-order networks consist exclusively of reactions of the type $S_j \rightarrow \sum_l R_{li} S_l$, i.e. monomolecular plus splitting and (auto-)catalytic reactions. First-order PDEs are solvable by the method of characteristics,³ a method to convert the PDE solution problem into ODEs. The basic idea is that the solution of the PDE is given by a set of integral curves $\mathbf{s}(\tau)$ called *characteristic curves*, parametrized by some $\tau \in \mathbb{R}$. We obtain the ODE system, describing these curves $\mathbf{s}(\tau)$, by interpreting (4) as a total derivative, i.e.

$$\frac{dg(\mathbf{s}(\tau), t(\tau))}{d\tau} = \frac{\partial g(\mathbf{s}(\tau), \tau)}{\partial t} \frac{dt}{d\tau} + \sum_{i=1}^n \frac{\partial g(\mathbf{s}(\tau), \tau)}{\partial s_i} \frac{ds_i(\tau)}{d\tau} \stackrel{!}{=} f(\mathbf{s}(\tau), \tau)g(\mathbf{s}(\tau), \tau). \tag{5}$$

Then, the *characteristic system of ODEs*, obtained by comparing the coefficients of (5) and (4), reads

$$\begin{aligned} \frac{dt}{d\tau} &= 1 \\ \frac{ds_j(\tau)}{d\tau} &= - \sum_{i=1}^m k_i Q_{ji} \left(\prod_{l=1}^n (s_l(\tau))^{R_{li}} - s_j(\tau) \right) \quad (j \in \{1, \dots, n\}) \end{aligned} \tag{6}$$

$$\frac{dg(\mathbf{s}(\tau), t(\tau))}{d\tau} = \underbrace{\sum_{i=1}^m \sum_{j=1}^n k_i (1 - Q_{ji}) \left(\prod_{l=1}^n (s_l(\tau))^{R_{li}} - 1 \right)}_{:=f(\mathbf{s}(\tau), \tau)} g(\mathbf{s}(\tau), \tau). \tag{7}$$

In case we not only require $\sum_{j=1}^n Q_{ji} \leq 1$, but also $\sum_{j=1}^n R_{ji} \leq 1$, the equations for $\frac{ds_i(\tau)}{d\tau}$ and $\frac{dg(\mathbf{s}(\tau), t(\tau))}{d\tau}$ are linear and we are restricted to *monomolecular reactions*. In consequence, the characteristic ODE system is solvable and we obtain a general solution of the CME, as shown in Gans (1960).⁴ For the complete definition of the ODE system, we specify the initial conditions

$$\begin{aligned} \mathbf{s}^0 &:= \mathbf{s}(0) \\ g(\mathbf{s}^0, 0) &= d(\mathbf{s}^0). \end{aligned} \tag{8}$$

³ See Evans (2010) for an introduction.

⁴ The same result may be obtained without the use of generating functions (Jahnke and Huisinga 2007).

The generating function $d(\mathbf{s}^0)$ specifies the initial distribution. Here, we study *deterministic* initial conditions $d(\mathbf{s}^0) = (\mathbf{s}^0)^{\mathbf{x}_0}$ and *product Poissonian* initial conditions

$$d(\mathbf{s}^0) = \exp \left[\langle \mathbf{x} \rangle_0 \cdot (\mathbf{s}^0 - \mathbf{1}) \right],$$

where $\langle \mathbf{x} \rangle_0$ is the expected number of molecules at $t = 0$.

2.2 Hierarchically linear ODE systems and reaction networks

Compared to monomolecular reaction networks, a more general subset of first-order networks are hierarchically linear networks, which we characterize in this article.

Definition 2 (*Hierarchically linear ODE system*) Let $\mathbf{x}_i(t)$, $\mathbf{f}_i(\cdot, t)$ and $\mathbf{h}_i(\cdot, t)$ be continuously differentiable, vector- and matrix-valued functions respectively.⁵ An ODE system exhibiting the structure

$$\begin{aligned} \frac{d\mathbf{x}_1}{dt} &= \mathbf{f}_1(\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n) + \mathbf{h}_1(\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n)\mathbf{x}_1 \\ \frac{d\mathbf{x}_2}{dt} &= \mathbf{f}_2(\mathbf{x}_3, \mathbf{x}_4, \dots, \mathbf{x}_n) + \mathbf{h}_2(\mathbf{x}_3, \mathbf{x}_4, \dots, \mathbf{x}_n)\mathbf{x}_2 \\ &\vdots \\ \frac{d\mathbf{x}_n}{dt} &= \mathbf{f}_n + \mathbf{h}_n\mathbf{x}_n \end{aligned} \tag{9}$$

is called a *hierarchically linear ODE system*.

The next proposition gives a reason for our interest in hierarchically linear ODE systems:

Proposition 1 *All hierarchically linear ODE systems are solvable by means of matrix exponentials.*

Proof We inspect a system of two levels only, since an extension to the n -dimensional case follows by mathematical induction. Then, the system for $n = 2$

$$\begin{aligned} \dot{\mathbf{x}}_1 &= \mathbf{f}_1(\mathbf{x}_2) + \mathbf{h}_1(\mathbf{x}_2)\mathbf{x}_1 \\ \dot{\mathbf{x}}_2 &= \mathbf{f}_2 + \mathbf{h}_2\mathbf{x}_2 \end{aligned}$$

has a solution in terms of simple matrix exponentials if $\mathbf{h}_1(\mathbf{x}_2(t_1))$ and $\mathbf{h}_1(\mathbf{x}_2(t_2))$ commute, i.e.

$$\mathbf{h}_1(\mathbf{x}_2(t_1))\mathbf{h}_1(\mathbf{x}_2(t_2)) = \mathbf{h}_1(\mathbf{x}_2(t_2))\mathbf{h}_1(\mathbf{x}_2(t_1)) .$$

⁵ The temporal dependence of the functions is suppressed as usual. Furthermore, $\mathbf{x}_i : \mathbb{R}_0^+ \rightarrow \mathbb{C}^{m_i}$, $\mathbf{f}_i : \mathbb{C}^{m_{i+1}} \times \mathbb{C}^{m_{i+2}} \times \dots \times \mathbb{C}^{m_n} \times \mathbb{R}_0^+ \rightarrow \mathbb{C}^{m_i}$ and $\mathbf{h}_i : \mathbb{C}^{m_{i+1}} \times \mathbb{C}^{m_{i+2}} \times \dots \times \mathbb{C}^{m_n} \times \mathbb{R}_0^+ \rightarrow \mathbb{C}^{m_i} \times \mathbb{C}^{m_i}$, where $n, m_i \in \mathbb{N}$.

In this case, the solution can be written⁶ as

$$\begin{aligned} \mathbf{x}_1(t) = & \exp \left\{ \int_0^t \mathbf{h}_1(\mathbf{x}_2(t')) dt' \right\} \mathbf{x}_1^0 \\ & + \exp \left\{ \int_0^t \mathbf{h}_1(\mathbf{x}_2(t')) dt' \right\} \int_0^t \exp \left\{ - \int_0^{t'} \mathbf{h}_1(\mathbf{x}_2(t'')) dt'' \right\} \mathbf{f}_1(\mathbf{x}_2(t')) dt' \end{aligned} \tag{10}$$

$$\mathbf{x}_2(t) = \exp\{\mathbf{h}_2 t\} \mathbf{x}_2^0 + \exp\{\mathbf{h}_2 t\} \int_0^t \exp\{-\mathbf{h}_2 t'\} \mathbf{f}_2 dt' . \tag{11}$$

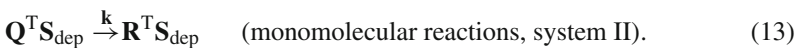
If the matrix-valued function \mathbf{h}_1 does not commute for two different instants of time, the matrix exponent is expressed in terms of Magnus series (see Magnus 1954). Even in that case $[\mathbf{h}_1(\mathbf{x}_2(t_1)), \mathbf{h}_1(\mathbf{x}_2(t_2))] \neq \mathbf{0}$, it is still possible to write down an analytical expression in terms of matrix exponentials. \square

A hierarchically linear ODE system translates to the study of chemical reaction networks modeled by the CME as follows:

Definition 3 (*Hierarchic first-order reaction network*) A first-order reaction network, described by a hierarchically linear characteristic ODE system is called hierarchic first-order reaction network.

2.3 Two-level hierarchic networks

In contrast to the general form of the preceding definition, we focus on *two-level* hierarchic networks, composed of two subsystems. That is, we have n_{ind} independent species $\mathbf{S}_{\text{ind}} = (S_1, \dots, S_{n_{\text{ind}}}, 0, \dots, 0)^T$ and n_{dep} dependent species $\mathbf{S}_{\text{dep}} = (0, \dots, 0, S_{n_{\text{ind}}+1}, \dots, S_n)^T$, where $n_{\text{ind}} + n_{\text{dep}} = n$. We express a two-level hierarchic network in terms of chemical reactions as



The first-order reactions' products are split into two parts: \mathbf{S}_{ind} appears only once per reaction ($\sum_{j=1}^{n_{\text{ind}}} R_{ji} \leq 1$ for $i \in \{1, \dots, m\}$), while an arbitrary number of molecules \mathbf{S}_{dep} is allowed. The monomolecular reactions are defined by $\sum_{j=n_{\text{ind}}+1}^n R_{ji} \leq 1$ for $i \in \{1, \dots, m\}$.

Note that by ignoring \mathbf{S}_{dep} , we might study $\mathbf{Q}^T \mathbf{S}_{\text{ind}} \rightarrow \mathbf{R}^T \mathbf{S}_{\text{ind}}$ *independently* of the monomolecular reactions (13), and obtain another monomolecular network, since $\sum_{j=1}^{n_{\text{ind}}} R_{ji} \leq 1$ for $i \in \{1, \dots, m\}$. We therefore refer to (12) as *independent* part and (13) as *dependent* part of the hierarchic system, since the latter cannot be modeled

⁶ For a review of the solution of a non-homogeneous system see Brannan (2010).

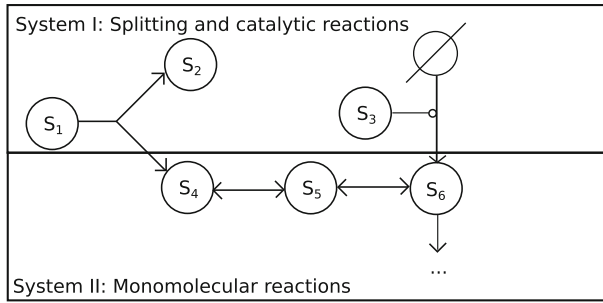


Fig. 1 Scheme depicting a hierarchic first-order network. In system I, all non-monomolecular reactions take place, such as splitting or catalytic reactions, while in system II only monomolecular reactions take place

independently of system I. Figure 1 depicts an example of such a hierarchic first-order two-level system.

Mathematically, (12) and (13) translate to a hierarchically linear characteristic ODE system. That is, the general form (6) is constrained to

$$\frac{ds_j}{dt} = - \sum_{i=1}^m Q_{ji} k_i \left\{ \left[\prod_{l=n_{\text{ind}}+1}^n s_l^{R_{li}} \right] \left(\sum_{l=1}^{n_{\text{ind}}} R_{li} s_l + \delta_{\sum_{l=1}^{n_{\text{ind}}} R_{li}, 0} \right) - Q_{ji} s_j \right\}$$

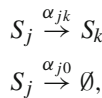
$(j \in \{1, \dots, n_{\text{ind}}\}),$ (14)

$$\frac{ds_j}{dt} = - \sum_{i=1}^m Q_{ji} k_i \left(\sum_{l=n_{\text{ind}}+1}^n R_{li} s_l + \delta_{\sum_{l=n_{\text{ind}}+1}^n R_{li}, 0} - Q_{ji} s_j \right)$$

$(j \in \{n_{\text{ind}} + 1, \dots, n\}).$ (15)

Note that the algebraic hierarchy is inverse to the reaction network hierarchy. The dependent part of the reaction network (system II) is described by an autonomous ODE system, while the independent part (system I) is non-autonomous due to the term $\prod_{l=n_{\text{ind}}+1}^n s_l^{R_{li}}$.

The monomolecular system II can be expressed more compactly as



where $j, k \in \{n_{\text{ind}} + 1, \dots, n\}$ and $\mathbf{A} := (\alpha_{ij})$ is the matrix holding the conversion rates. Using that matrix, we rewrite (15) as

$$\frac{d\mathbf{s}_{\text{dep}}}{dt} = -\mathbf{A}^T (\mathbf{s}_{\text{dep}} - \mathbf{1}).$$
(16)

The inversion of the hierarchic and algebraic structure, mentioned in the previous paragraph, also becomes clear by the traditional rate equations for the concentration vector \mathbf{C}_{dep} of system II

$$\frac{d\mathbf{C}_{\text{dep}}}{dt} = \mathbf{A}\mathbf{C}_{\text{dep}} + \text{influx from system I.} \tag{17}$$

It was shown by Jahnke and Huisinga (2007) using a statistical argument that the solution to the monomolecular CME is a product Poisson distribution, given Poissonian initial conditions, and a multinomial distribution, given deterministic initial conditions. Thereby, the parameters of the distributions are given by the traditional rate Eq. (17). This parametrization is also evident by the characteristic ODEs (16) because these can be expressed using the transposed Jacobian matrix of (17).

However, it also becomes clear that the exact solution to the CME for a general first-order network, is not fully parametrized by the traditional rate equations: The ODE system (13) is structurally different from the traditional rate equations for system I due to the time-dependent coefficients $\prod_{l=n_{\text{ind}}+1}^n s_l^{R_{li}}$. In contrast to the characteristic Eq. (14), the traditional rate equation system is generally autonomous if the rate constants k_i are time-independent as we assume.

In order to highlight the relationship between traditional rate equations and characteristic ODEs for system II, we use the convention

$$\begin{aligned} \mathbf{Q}^T \mathbf{S}_{\text{ind}} &\xrightarrow{\mathbf{k}} \mathbf{R}^T \mathbf{S}_{\text{ind}} + \mathbf{R}^T \mathbf{S}_{\text{dep}}, \\ S_j &\xrightarrow{\alpha_{jk}} S_k, \\ S_j &\xrightarrow{\alpha_{j0}} \emptyset, \\ \emptyset &\xrightarrow{\mathbf{b}} \mathbf{S}, \end{aligned} \tag{18}$$

instead of (12) and (13) to fully express the two-level hierarchic reaction network. Note that we summarized influx reactions to both system I and II in (18), which yields the characteristic equations

$$\begin{aligned} \frac{ds_j}{dt} &= - \sum_{i=1}^m Q_{ji} k_i \left\{ \left[\prod_{l=n_{\text{ind}}+1}^n s_l^{R_{li}} \right] \left(\sum_{l=1}^{n_{\text{ind}}} R_{li} s_l + \delta_{\sum_{l=1}^{n_{\text{ind}}} R_{li}, 0} \right) - Q_{ji} s_j \right\} \\ &\quad (j \in \{1, \dots, n_{\text{ind}}\}) \\ \frac{d\mathbf{s}_{\text{dep}}}{dt} &= -\mathbf{A}^T (\mathbf{s}_{\text{dep}} - \mathbf{1}) \\ \frac{dg(\mathbf{s}, t)}{dt} &= \mathbf{b} \cdot (\mathbf{s} - \mathbf{1}) g(\mathbf{s}, t) \end{aligned}$$

instead of (14) and (15).

3 Main results

3.1 Solvability of hierarchic first-order networks

First, we show that the CME is solvable for any hierarchically linear reaction network.

Theorem 1 (Existence of analytical solutions for hierarchic first-order networks) *The Master equation corresponding to a hierarchic first-order reaction network is analytically solvable.*

Proof Applying Proposition 1, we find that the characteristic system (6) is solvable as a linear equation system as $\mathbf{s} = e^{\mathbf{C}}\mathbf{s}^0 + \mathbf{v}$, where \mathbf{C} is a square matrix and \mathbf{v} a column vector. For the initial condition of (7), we need \mathbf{s}^0 . Since $e^{\mathbf{C}}$ is generally invertible, the characteristic system $\mathbf{s} = e^{\mathbf{C}}\mathbf{s}^0 + \mathbf{v}$ can be solved for \mathbf{s}^0 . This in turn yields the generating function and the distribution, defined as the coefficients of the former. \square

Remark 1 The term ‘‘analytical solution’’, which is often used in a more intuitive sense, requires some explanation. One may argue that the Master equation is a system of linear ODEs and can always be formally solved by a matrix exponential as such. This remains true even for bimolecular networks such as the Lotka–Volterra model (Lotka 1925). However, a formal solution of the Master equation does not tell us anything about the structure of the distribution, if the matrix exponential cannot be calculated explicitly. Contrary to this non-analytical case, the generating function stated in the next proposition earns the word ‘‘analytical’’, even if it also contains a matrix exponential. As shown in Sect. 3.2, the matrix exponential from the next proposition enables us to analyze properties of the distribution and may be used to find the stationary distribution, to compute moments and cumulants, etc.

Proposition 2 *Let $\mathbf{b} = \mathbf{0}$, i.e. no reactions $\emptyset \xrightarrow{\mathbf{b}} \mathbf{S}$. Given that the molecules are initially product Poisson distributed, i.e. $g(\mathbf{s}, 0) = \exp[\langle \mathbf{x} \rangle_0 \cdot (\mathbf{s} - \mathbf{1})]$, the generating function is given by*

$$g(\mathbf{s}, t) = \exp \left\{ \langle \mathbf{x}_1 \rangle_0 \cdot \left[e^{-\int_0^t \mathbf{J}(\mathbf{s}_{dep}, t', t) dt'} \mathbf{s}_{ind} - \int_0^t e^{-\int_0^{t'} \mathbf{J}(\mathbf{s}_{dep}, t'', t) dt''} \mathbf{f}(\mathbf{s}_{dep}, t', t) dt' - \mathbf{1} \right] + \exp(\mathbf{A}t) \langle \mathbf{x}_2 \rangle_0 \cdot (\mathbf{s}_{dep} - \mathbf{1}) \right\}. \tag{19}$$

In the last expression, the Jacobian \mathbf{J} is defined by

$$J_{pq}(\mathbf{s}_{dep}, t', t) := \sum_{i=1}^m Q_{pi} k_i (1 - c_i(\mathbf{s}_{dep}, t', t) R_{qi}), \quad p, q \in \{1, \dots, n_{ind}\} \tag{20}$$

⁷ Without this requirement, we obtain an additional factor $e^{\int_0^t \mathbf{b} \cdot (\mathbf{s}_{ind}(t''') - \mathbf{1}) dt'''}$, where $\mathbf{s}_{ind}(t''')$ is defined in (55). For more compact expressions, this factor was left out without loss of generality.

with⁸ $[\mathbf{J}(\mathbf{s}_{dep}, t_1, t), \mathbf{J}(\mathbf{s}_{dep}, t_2, t)] = \mathbf{0}$ and \mathbf{f} is given by

$$f_j(\mathbf{s}_{dep}, t', t) := - \sum_{i=1}^m Q_{ji} k_i c_i(\mathbf{s}_{dep}, t', t), \quad j \in \{1, \dots, n_{ind}\} \tag{21}$$

where

$$c_i(\mathbf{s}_{dep}, t', t) := \prod_{l=n_{ind}+1}^n \left(\exp \{ - \mathbf{A}(t' - t) \} \epsilon_l \cdot (\mathbf{s}_{dep} - \mathbf{1}) + 1 \right)^{R_{li}}. \tag{22}$$

For deterministic initial conditions, i.e. $P(\mathbf{x}, 0) = \delta_{\mathbf{x}, \mathbf{x}^0}$, where $\mathbf{x}^0 = (x_1^0, \dots, x_n^0)^T$, we obtain

$$g(\mathbf{s}, t) = \left\{ \prod_{i=1}^n \left(\left[e^{-\int_0^t \mathbf{J}(\mathbf{s}_{dep}, t', t) dt'} \mathbf{s}_{ind} - \int_0^t e^{-\int_0^{t'} \mathbf{J}(\mathbf{s}_{dep}, t'', t) dt''} \mathbf{f}(\mathbf{s}_{dep}, t', t) dt' \right] \epsilon_i^{(1)} + \exp(\mathbf{A}t) \epsilon_i^{(2)} \cdot (\mathbf{s}_{dep} - \mathbf{1}) + 1 \right)^{x_i^0} \right\}, \tag{23}$$

with $\epsilon_i^{(1)} := \text{col}_i \mathbf{I}_{n_{ind} \times n_{ind}}$ and $\epsilon_i^{(2)} := \text{col}_i \mathbf{I}_{n_{dep} \times n_{dep}}$.

We postpone the proof to ‘‘Appendix A.3’’ and interpret this result first. Remember that the deterministic or product Poisson distributions, assumed as an initial distributions in the preceding proposition, imply that all random variables are uncorrelated for a fixed time t . This can be shown by using the cumulant generating function $\kappa(\xi) = \log(g(e^\xi))$, that is given for a product Poisson distribution by

$$\kappa(\xi) = \log(e^{\langle x \rangle_1 (e^{\xi_1} - 1) + \dots + \langle x \rangle_n (e^{\xi_n} - 1)}) = \langle x \rangle_1 (e^{\xi_1} - 1) + \dots + \langle x \rangle_n (e^{\xi_n} - 1).$$

Here, ‘‘uncorrelated’’ means that the covariances and higher order mixed cumulants, defined as the coefficients of $\kappa(\xi)$, are zero:

$$\partial_{\xi_1, \dots, \xi_n} \left(\langle x \rangle_1 (e^{\xi_1} - 1) + \dots + \langle x \rangle_n (e^{\xi_n} - 1) \right) \Big|_{\xi=\mathbf{0}} = 0.$$

For monomolecular reaction networks, the variables remain uncorrelated after some time $t > 0$, since a monomolecular system given Poissonian initial conditions stays Poissonian, as shown in Jahnke and Huisinga (2007), Proposition 2. Contrastingly, for first-order processes, covariances between the variables appear for $t > 0$, because the exponent of Eq. (19) is in general not a first-order polynomial. In other words, monomolecular systems stay uncorrelated for uncorrelated initial conditions, while first-order systems do not.

⁸ Without this requirement, Magnus series are needed to express the solution. For more compact expressions we restrict ourselves to $[\mathbf{J}(\mathbf{s}_{dep}, t_1, t), \mathbf{J}(\mathbf{s}_{dep}, t_2, t)] = \mathbf{0}$ without loss of generality.

We understand the correlations in terms of chemical reactions by the difference between the first-order splitting reaction $S_j \rightarrow S_k + S_l$ and the two monomolecular reactions $S_j \rightarrow S_k$ and $S_j \rightarrow S_l$: The molecules S_k and S_l appear simultaneously in the splitting reaction, while the two monomolecular conversions are statistically independent events.

3.2 Analytical form of marginal distributions

As we show in the following, the hierarchic network structure yields marginal distributions that are best interpreted as generalized distributions. We follow Johnson et al. (2005) for the next definition.

Definition 4 (*Generalizing distribution*) The distribution \mathcal{P}_1 is *generalized* by the *generalizing distribution* \mathcal{P}_2 :

$$\mathcal{P} \sim \mathcal{P}_1 \vee \mathcal{P}_2 \Leftrightarrow g(s) = g_1(g_2(s)),$$

where $g_1(s)$ and $g_2(s)$ are the corresponding generating functions.

Generalized distributions are related to mixing distributions by Gurland’s theorem (Gurland 1957), if certain requirements are met. Next, we define two classes of generalized distributions that arise as marginal distributions for hierarchic networks.

Definition 5 (*Discrete compound Poisson distribution (DCP)*) A discrete compound Poisson distribution \mathcal{P}_{DCP} is a univariate distribution (Zhang et al. 2014), generalizing the Poisson distribution with another distribution \mathcal{P}_2 :

$$\mathcal{P}_{\text{DCP}} \sim \mathcal{P}_{\text{Poisson}} \vee \mathcal{P}_2.$$

The generating function is defined accordingly:

$$g_{\text{DCP}}(s) := g_{\text{Poisson}}(g_2(s)) = e^{\lambda(g_2(s)-1)} = e^{\sum_{i=1}^{\infty} \lambda \alpha_i (s^i - 1)},$$

where $\lambda > 0$ and $\alpha_i \in [0, 1]$ with $\sum_{i=1}^{\infty} \alpha_i = 1$ are *parameters*. We introduce the notation

$$\text{DCP}^N \Leftrightarrow \text{deg}_s \sum_{i=1}^{\infty} \lambda \alpha_i (s^i - 1) = N. \tag{24}$$

The next type of distribution was previously defined by Khatri and Patel (1961) and will be important for our study of reaction networks given deterministic initial conditions:

Definition 6 (*Khatri’s Type B distribution*) Khatri’s Type B distribution \mathcal{P}_{KTB} is a univariate distribution, generalizing the deterministic distribution with another distribution \mathcal{P}_2 :

$$\mathcal{P}_{\text{KTB}} \sim \mathcal{P}_{\text{Deterministic}} \vee \mathcal{P}_2.$$

The generating function is defined accordingly:

$$g_{\text{KTB}}(s) := (g_2(s))^v = \left(\sum_{i=0}^{\infty} \alpha_i s^i \right)^v,$$

$\alpha_i \in [0, 1]$ are *parameters* and $v \in \mathbb{N}_0$ is the number of trials. Furthermore, $g_2(s)$ must be a generating function, i.e. $\sum_{i=1}^{\infty} \alpha_i = 1$. We introduce the notation

$$\text{KTB}^N : \Leftrightarrow \text{deg}_s \sum_{i=0}^{\infty} \alpha_i s^i = N. \tag{25}$$

These distributions specify the marginal distributions of the dependent part (system II) of hierarchic first-order reaction networks. We will show in Sect. 3.3, that the DCP^N and KTB^N distributions are conditionally multimodal for $N > 1$.

Proposition 3 (Marginal distributions given $\partial_{s_X} \lambda_i(s_X) = 0$) *Let $\lambda_i(s_X)$ be the i th eigenvalue of $\mathbf{J}(s_X, t', t)$. Furthermore, let $\partial_{s_X} \lambda_i(s_X) = 0$, $\mathbf{f} = \mathbf{0}$ and $\mathbf{b} = \mathbf{0}$. Given Poissonian initial conditions, the marginal distribution of any species X from the dependent part of the network is DCP^N , where $N < \infty$. Given deterministic initial conditions, this distribution is KTB^N .*

To prove this statement, we need an auxiliary result:

Lemma 2 *Let $\mathbf{H}(x)$ be an $n \times n$ matrix, whose entries depend polynomially on $x \in \mathbb{C}$. Let $\lambda_i(x)$ be the i th eigenvalue of $\mathbf{H}(x)$. Then, $\partial_x \lambda_i(x) = 0 \Rightarrow \text{deg}_x (e^{\mathbf{H}(x)})_{ij} < \infty$.*

Proof By the Cayley–Hamilton theorem, $\mathbf{H}(x)$ fulfills its own characteristic polynomial $\Delta(\lambda)$, i.e. $\Delta(\mathbf{H}(x)) = \mathbf{0}$. Furthermore, any polynomial $p(\lambda)$ might be expressed as $p(\lambda) = q(\lambda)\Delta(\lambda) + r(\lambda)$, where q is found by long division p/q with remainder r of degree $\leq n - 1$. Since $\Delta(\mathbf{H}(x)) = \mathbf{0}$, we might express⁹ $e^{\mathbf{H}(x)}$ as

$$\underbrace{e^{\mathbf{H}(x)}}_{p(\mathbf{H}(x))} = \underbrace{\sum_{k=0}^{n-1} \alpha_k(x) \mathbf{H}^k(x)}_{r(\mathbf{H}(x))} \tag{26}$$

for some coefficients $\alpha_k(x)$.

Furthermore, the eigenvalues $\lambda_i(x)$ fulfill

$$e^{\lambda_i(x)} = \sum_{k=0}^{\infty} \frac{\lambda_i^k(x)}{k!} = \underbrace{\sum_{k=0}^{n-1} \alpha_k(x) \lambda_i^k(x)}_{:= (\mathbf{L}\alpha)_i} \quad (i \in \{1, \dots, n\}). \tag{27}$$

⁹ The matrix exponential and the following expression, obtained by the Cayley–Hamilton theorem, is reviewed in Rowell (2004).

Here, the second relation follows again from long division by the characteristic polynomial of $\mathbf{H}(x)$. For $\partial_x \lambda_i(x) = 0$, the coefficients of the characteristic polynomial do not depend on x and thus $\partial_x \alpha_k(x) = 0$.

Therefore, the expression (26) introduces only a finite-order dependence on x . \square

Using this Lemma, we prove Proposition 3:

Proof We obtain the marginal distribution of species X by setting all $s_i = 1$ except s_X in Eqs. (19) and (23) respectively. First of all, the integrals in Eq. (19) do not change the order of the polynomial in the exponent with respect to s_X , that is $\deg_{s_X} (e^{-\mathbf{J}(s_X, t, t')})_{ij} = \deg_{s_X} (e^{-\int_0^t \mathbf{J}(s_X, t', t) dt'})_{ij}$. Therefore, we have a DCP^N distribution, where N is given by (24). We find¹⁰

$$N = \deg_{s_X} \left(\mathbf{s}_{\text{ind}}^0(s_X) \cdot \langle \mathbf{x}_1 \rangle_0 \right) = \deg_{s_X} \left(e^{-\mathbf{J}(s_X)} \mathbf{1} \cdot \langle \mathbf{x}_1 \rangle_0 \right) = \deg_{s_X} \sum_{i,j} \left(e^{-\mathbf{J}(s_X)} \right)_{ij} \tag{28}$$

$$\leq \max_{i,j \in \{1, \dots, n_{\text{ind}}\}} \deg_{s_X} \left(e^{-\mathbf{J}(s_X)} \right)_{ij}. \tag{29}$$

Because $\mathbf{J}(s_X)$ depends polynomially on s_X , we apply Lemma 2 to obtain

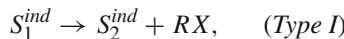
$$N \leq \deg_{s_X} \left(e^{-\mathbf{J}(s_X)} \right)_{ij} < \infty.$$

In case $[\mathbf{J}(s_X, t_1, t), \mathbf{J}(s_X, t_2, t)] \neq \mathbf{0}$, the same argument can be made for the Magnus series. By replacing $\langle \mathbf{x}_1 \rangle_0$ in Eq. (28) with $\varepsilon_i^{(1)} := \text{col}_i \mathbf{I}_{n_{\text{ind}} \times n_{\text{ind}}}$ for deterministic initial conditions, we have the result KTB^N, where $N < \infty$. \square

For simple reaction networks such as those from the next proposition, we show the order of the marginal distributions to be larger than one. We first investigate networks whose independent part is mass-conservative, i.e. $\sum_{l=1}^{n_{\text{ind}}} R_{li} \neq 0 \Rightarrow \mathbf{f} = \mathbf{0}$ and $\sum_{l=1}^{n_{\text{ind}}} Q_{li} \neq 0 \Rightarrow \mathbf{b} = \mathbf{0}$.

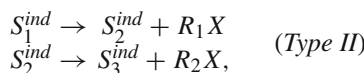
Proposition 4 *Let $\partial_{s_X} \lambda_i(s_X) = 0$, $\mathbf{f} = \mathbf{0}$, $\mathbf{b} = \mathbf{0}$ and $[\mathbf{J}(s_X, t_1, t), \mathbf{J}(s_X, t_2, t)] = \mathbf{0}$. The marginal distribution of any dependent species X is DCP^N (KTB^N for deterministic initial conditions), with $N > 1$ for the following minimal reaction networks:*

1.



where $R > 1$ and $n_{\text{ind}} = 2$.

2.



where $R_1 \geq 1$, $R_2 \geq 1$ and $n_{\text{ind}} = 3$.

¹⁰ We suppress the time-dependence of \mathbf{J} in the following.

Proof We expand the matrix powers in

$$\text{deg}_{s_X} \sum_{p,q} (e^{-\mathbf{J}(s_X)})_{pq} \stackrel{(26)}{=} \text{deg}_{s_X} \sum_{p,q} \left(\sum_{k=0}^{n_{\text{ind}}-1} (-\mathbf{J}(s_X))^k \alpha_k \right)_{pq}, \tag{30}$$

starting with \mathbf{J}^1 :

$$J_{pq}(s_X) \stackrel{(20)}{=} \sum_{i=1}^m k_i Q_{pi} (1 - c_i(s_X) R_{qi}), \quad p, q \in \{1, \dots, n_{\text{ind}}\}$$

For reaction Type I, it suffices to consider the first order \mathbf{J}^1 of (30). Since there are no reactions within the dependent part $\mathbf{A} = \mathbf{0}$, we then have

$$c_i(s_X, t', t) \stackrel{(22)}{=} (1 \cdot (s_X - 1) + 1)^{R_i} = s_X^{R_i} \tag{31}$$

In consequence of (31), we get $\text{deg}_{s_X} \sum_{p,q} (e^{\mathbf{J}(s_X)})_{pq} > 1$ whenever $R_i > 1$. This translates to a splitting reaction of Type I. To see $\partial_{s_X} \lambda_i(s_X) = 0$, consider

$$\begin{aligned} \frac{ds_1}{dt} &= - \underbrace{k_{S_1 \rightarrow S_2+X}}_{:=1} (s_X^R s_2 - s_1) \\ \frac{ds_2}{dt} &= 0 \end{aligned}$$

and

$$\det(\mathbf{J} - \lambda \mathbf{1}) = \det\left(\left(\partial_{s_j} \frac{ds_i}{dt}\right)_{ij} - \lambda \mathbf{1}\right) = \det \begin{pmatrix} 1 - \lambda & -s_X^R \\ 0 & -\lambda \end{pmatrix} = \lambda^2 - \lambda.$$

For reaction Type II, $n_{\text{ind}} = 3$ and we need to calculate the second order \mathbf{J}^2 in (30). We have¹¹

$$\begin{aligned} J_{pq}^2(s_X) &\propto \sum_{r=1}^{n_{\text{ind}}} \left(\sum_{i=1}^m k_i Q_{pi} (1 - s_X^{R_i} R_{ri}) \right) \left(\sum_{j=1}^m k_j Q_{rj} (1 - s_X^{R_j} R_{qj}) \right) \\ &\propto \sum_{r=1}^{n_{\text{ind}}} \sum_{i,j=1}^m k_i k_j Q_{pi} R_{ri} Q_{rj} R_{qj} s_X^{R_i+R_j}. \end{aligned}$$

Note that only $i \neq j$ terms are non-zero due to $\sum_{l=1}^{n_{\text{ind}}} R_{li} \leq 1$. This implies that the products from the independent part must appear in different reactions. In terms of chemical reaction networks, we have

¹¹ Here, “ \propto ” means that lower order terms and constant factors in s_X are dropped.

$$\begin{aligned} Q_{p1}S_p^{\text{ind}} &\rightarrow R_{r1}S_r^{\text{ind}} + R_{\xi 1}X \\ Q_{r2}S_r^{\text{ind}} &\rightarrow R_{q2}S_q^{\text{ind}} + R_{\xi 2}X. \end{aligned}$$

We arbitrarily chose $p = 1, r = 2, q = 3$ to obtain the result. To see $\partial_{s_X} \lambda_i(s_X) = 0$, consider

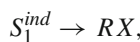
$$\begin{aligned} \frac{ds_1}{dt} &= - \underbrace{k_{S_1 \rightarrow S_2 + R_1 X}}_{:=1} (s_X^{R_1} s_2 - s_1) \\ \frac{ds_2}{dt} &= - \underbrace{k_{S_2 \rightarrow S_3 + R_2 X}}_{:=1} (s_X^{R_2} s_3 - s_2) \\ \frac{ds_3}{dt} &= 0 \end{aligned}$$

and

$$\det \mathbf{J} = \det(\partial_{s_j} \frac{ds_i}{dt})_{ij} = \det \begin{pmatrix} 1 - \lambda & -s_X^{R_1} & 0 \\ 0 & 1 - \lambda & -s_X^{R_2} \\ 0 & 0 & -\lambda \end{pmatrix} = -\lambda (\lambda^2 - 2\lambda + 1).$$

□

Proposition 5 *Let $\mathbf{J}(t', t)$ be independent of s_X , $\mathbf{b} = \mathbf{0}$ and $\partial_{s_X} \mathbf{f}(s_X, t', t) \neq \mathbf{0}$, then the marginal distribution of any species X from the dependent part is DCP^N (KTB^N for deterministic initial conditions), with $N > 1$ for the following minimal reaction network:*



where $R > 1$ and $n_{\text{ind}} = 1$.

Proof We express the reaction $S_1^{\text{ind}} \rightarrow RX$ by

$$f_1(s_X) \stackrel{(21)}{=} -k_1 \left(1 \cdot (s_X - 1) + 1 \right)^R = -k_1 s_X^R$$

and plug the same expression into (19) to obtain $\log g(s_X, t) \propto s_X^R$. In consequence, we get a DCP^N distribution with $N > 1$ whenever $R > 1$. The analogous statement for deterministic initial conditions follows by plugging $f_1(s_X)$ into (23). □

The next proposition holds for the transcription–translation model, as examined in detail in Example 5.

Proposition 6 (Marginal distributions given $\partial_{s_X} \lambda_i(s_X) \neq 0$) *Let $\lambda_i(s_X)$ be the i th eigenvalue of $\mathbf{J}(s_X, t', t)$ and let $\partial_{s_X} \lambda(s_X) \neq 0$. Given Poissonian initial conditions, the marginal distribution of any species X from the dependent part of the network is DCP^∞ (KTB[∞] for deterministic initial conditions).*

Proof In case all eigenvalues are distinct, we solve the system (27) by inverting the matrix¹² $\mathbf{L}(s_X)$, so

$$e^{\mathbf{J}(s_X)} = \sum_{k=0}^{n_{\text{ind}}-1} \sum_{i=0}^{n_{\text{ind}}-1} (\mathbf{L}^{-1}(s_X))_{ki} e^{\lambda_i(s_X)} \mathbf{J}^k(s_X). \tag{32}$$

Since $(\mathbf{L}^{-1})_{ki}$ is a rational function of the eigenvalues, we have

$$\max_{i,j \in \{1, \dots, n_{\text{ind}}\}} \deg_{s_X} (e^{\mathbf{J}(s_X)})_{ij} = \infty. \tag{33}$$

In case there is an eigenvalue λ_j of multiplicity μ , the matrix \mathbf{L} is not invertible. Since

$$\left. \frac{d^{\mu-1} \Delta(\lambda)}{d\lambda^{\mu-1}} \right|_{\lambda_j} = \left. \frac{d^{\mu-1}}{d\lambda^{\mu-1}} (\lambda - \lambda_j)^\mu \prod_{i=1}^{n_{\text{ind}}-\mu} (\lambda - \lambda_i) \right|_{\lambda_j} = 0,$$

we derive (27) for λ and obtain $\mu - 1$ additional equations:

$$\frac{d^i e^{\lambda_j(s_X)}}{d\lambda^i} = \frac{d^i}{d\lambda^i} \sum_{k=0}^{n_{\text{ind}}-1} \alpha_k(s_X) \lambda^k \quad (i \in \{1, \dots, \mu - 1\}).$$

We solve this system together with (27) for $\alpha_k(s_X)$ by inverting a matrix. The entries of this matrix are rational functions of λ , so the terms $e^{\lambda_i(s_X)}$ in (32) do not cancel, i.e. Equation (33) holds. In the last step, we plug Eq. (32) into (19) for Poissonian and into (23) for deterministic initial conditions to obtain the generating function for the marginal distribution of X . Since the exponent of (19) is of infinite degree, we obtain a DCP^∞ class distribution. For deterministic initial conditions, we obtain a KTB^∞ class distribution. □

3.3 Modality of marginal distributions

In this section, we investigate under which conditions the generalized distributions from Proposition 2 have infinitely many modes. We need several definitions for this end:

Definition 7 (*Unimodality*) A distribution p_n is said to be *unimodal* with mode $a \in \mathbb{N}_0$ if

$$p_n \geq p_{n-1}, \forall n \leq a \quad \text{and} \quad p_{n+1} \leq p_n, \forall n \geq a.$$

If this property does not depend on the parameters of the distribution, the latter is said to be *unconditionally unimodal*. A unimodal distribution that results in a unimodal

¹² The time-dependence of \mathbf{L} and \mathbf{J} is suppressed in the following.

distribution upon convolution with another unimodal distribution is called *strongly unimodal*.¹³

The classes DCP^1 and KTB^1 are unconditionally unimodal, since the only members of these sets are the Poissonian and Binomial distribution respectively. For the independent species (system I) of a hierarchic first-order reaction network, we obtain unimodal marginal distributions:

Theorem 2 (Unconditional unimodality of marginal distributions of independent species (system I)) *Given deterministic or Poissonian initial conditions, the species from the independent part of the hierarchic network exhibit unconditionally unimodal marginal distributions at all times $t > 0$.*

Proof Upon setting all $s_i = 1$, except one s_X from the independent part, we obtain a Poissonian distribution for Eq. (19). For deterministic initial conditions, Eq. (23) yields a convolution of binomial and Poisson distributions. Since the binomial and Poisson distribution are strongly unimodal, the resulting distribution is unimodal as well, as shown by Keilson and Gerber (1971). \square

For the second part of a hierarchic first-order network, we need the following definition:

Definition 8 (*Conditional multimodality*) A distribution that is not unconditionally unimodal is called *conditionally multimodal*.

As an example for the preceding definition, we have Neyman's Type A distribution (Neyman 1939), defined by the generating function $\exp(\lambda(e^{\phi(s-1)} - 1))$, where $\lambda > 0$ and $\phi > 0$. Because of the double exponential generation function, it belongs to the class of DCP^∞ distributions that we prove to be conditionally multimodal in the next proposition.

Proposition 7 (Three classes of conditionally multimodal distributions) *The classes of distributions DCP^∞ , KTB^∞ , DCP^2 , KTB^2 and DCP^3 , KTB^3 are conditionally multimodal.*

Proof Since conditional multimodality is defined to be the contrary of unconditional unimodality, a counterexample is enough to prove conditional multimodality. Respective counterexamples are shown in Fig. 2. Figure 2a shows Neyman's Type A distribution, defined by the DCP^∞ generating function $g(s) = \exp(\lambda(e^{\phi(s-1)} - 1))$, can be multimodal.¹⁴ The same is done for the Hermite distribution by the generating function $g(s) = e^{a_1(s-1)+a_2(s^2-1)}$. As a DCP^2 distribution, it is conditionally multimodal, as demonstrated in Fig. 2a. A counterexample for DCP^3 is given by

¹³ Defined by Keilson and Gerber (1971).

¹⁴ Note that the coefficients of $e^{\phi(s-1)}$ are defined as the parameters of DCP^∞ , not ϕ .

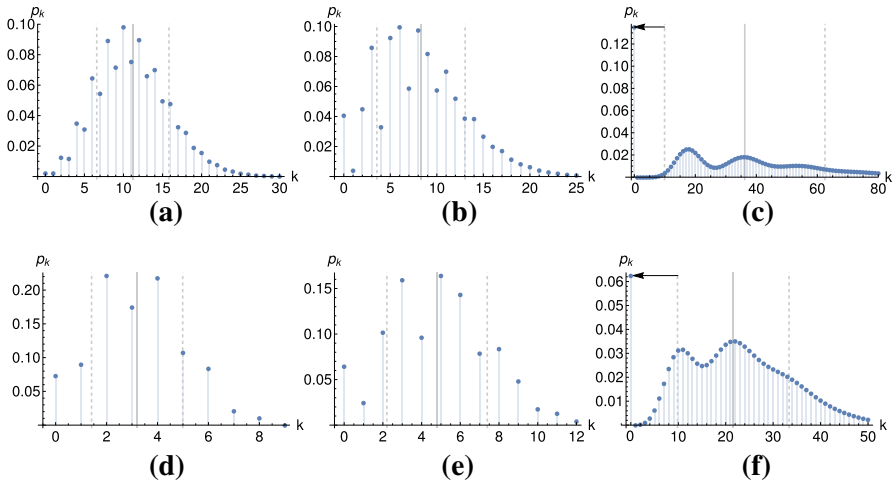


Fig. 2 Some conditionally multimodal distributions. The thick gray lines depict the expectation value, while the dashed lines stand for the standard deviation. The arrow marks the probability for a zero outcome. **a** Hermite (DCP²) distribution, $a_1 = 1, a_2 = 5.1$. **b** Triple stuttering-Poisson (DCP³) distribution, $a_1 = 0.1, a_2 = 1.1, a_3 = 2$. **c** Neyman's type A (DCP[∞]) distribution, $\lambda = 2, \phi = 18.1$. **d** Binomial–Binomial (KTB²) distribution, $p_1 = 0.5, p_2 = 0.8$. **e** Binomial–Binomial (KTB³) distribution, $p_1 = 0.5, p_2 = 0.8$. **f** Binomial–Binomial (KTB^{infinity}) distribution, $p_1 = 0.5, p_2 = 0.8$

$g(s) = e^{a_1(s-1)+a_2(s^2-1)+a_3(s^3-1)}$ from Fig. 2b. Figure 2d–f shows counterexamples for the generalized binomial distributions KTB², KTB³ and KTB[∞]:

$$\begin{aligned}
 g_{\text{KTB}^2}(s) &= (1 - p_1 + p_1(1 - p_2 + p_2s)^2)^4 && (\text{Binomial}(4, p_1) \vee \text{Binomial}(2, p_2)) \\
 g_{\text{KTB}^3}(s) &= (1 - p_1 + p_1(1 - p_2 + p_2s)^3)^4 && (\text{Binomial}(4, p_1) \vee \text{Binomial}(3, p_2)) \\
 g_{\text{KTB}^\infty}(s) &= (1 - p_1 + p_1e^{\lambda(s-1)})^4 && (\text{Binomial}(4, p_1) \vee \text{Poisson}(\lambda)).
 \end{aligned}$$

□

Since we cannot provide an infinite number of counterexamples, one for each DCP^N or KTB^N class, where $N > 1$, we extend the result of the last proposition as a conjecture.

Conjecture 1 *The classes DCP^N and KTB^N are conditionally multimodal for $N > 3$.*

The intuition behind this conjecture is that the DCP^N and KTB^N classes are modeling events that can occur in bursts. In terms of Definition 4, the generalizing distribution \mathcal{P}_2 is modeling the (stochastic) number of events per burst or burst size, whereas the generalized distribution \mathcal{P}_1 models the (stochastic) number of bursts, yielding the total number of events as $\mathcal{P} \sim \mathcal{P}_1 \vee \mathcal{P}_2$. In consequence, the distances between the modes seen in Fig. 2 approximately correspond to multiples of the burst sizes. For the Binomial–Binomial distribution from Fig. 2d the burst size is 2, for Fig. 2e it is 3, and for the Binomial–Poisson distribution from Fig. 2f it is $\lambda = 10.8$. The same applies for the generalized Poisson distributions. For Neyman's Type A (Poisson–Poisson)

distribution shown in Fig. 2c, the modes are clearly visible at multiples of 18.1. Thus, the bursting behavior is linked to conditional multimodality. In consequence, what we are conjecturing is that this link stays the same as we increase the number of possible bursts beyond 3.

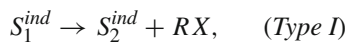
In the next Theorems, we summarize the cases in which conditional multimodality appears for the species of the dependent system II.

Theorem 3 (Sufficient criterion for conditional multimodality, given deterministic or Poissonian initial conditions and $\partial_{s_X} \lambda(s_X) \neq 0$) *Let $\partial_{s_X} \lambda(s_X) \neq 0$. Then the dependent species X in a hierarchic first-order network obeys a conditionally multimodal marginal distribution at all times $t > 0$.*

Proof $\partial_{s_X} \lambda(s_X) \neq 0$ implies DCP^∞ and KTB^∞ class distributions (Proposition 6) that are conditionally multimodal by Proposition 7. □

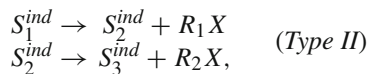
Theorem 4 (Minimal reaction networks exposing conditional multimodality, given deterministic or Poissonian initial conditions and $\partial_{s_X} \lambda(s_X) = 0$) *Let $\partial_{s_X} \lambda(s_X) = 0$, $\mathbf{b} = \mathbf{0}$ and assume Conjecture 1 is true. Then, given Poissonian or deterministic initial conditions, the following minimal hierarchic first-order networks exhibit a conditionally multimodal marginal distribution at all times $t > 0$ for the dependent species X :*

1. For mass-conservative independent networks, i.e. $\mathbf{f} = \mathbf{0}$:
 - (a)



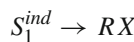
where $R > 1$ and $n_{ind} = 2$.

- (b)



where $R_1 \geq 1, R_2 \geq 1$ and $n_{ind} = 3$.

2. For open independent networks and $\sum_{l=1}^{n_{ind}} R_{li} = 0$ for all reactions, i.e. $\partial_{s_X} \mathbf{J} = \mathbf{0}$ and $\partial_{s_X} \mathbf{f} \neq \mathbf{0}$, the minimal network is:



where $R > 1$ and $n_{ind} = 1$.

Proof By Proposition 4 we know that the distribution class is $\text{DCP}^{N_{\text{Prop4}}}$ whenever $\mathbf{f} = \mathbf{0}$, where the degree $N_{\text{Prop4}} > 1$ for both types of minimal reaction networks. Proposition 5 applies if $\partial_{s_X} \mathbf{f} \neq \mathbf{0}$, and yields the minimal reaction network $S_1^{ind} \rightarrow RX$ for which $N_{\text{Prop5}} > 1$.

These distribution classes are conditionally multimodal if N_{Prop4} and N_{Prop5} are larger than one according to Proposition 7 and Conjecture 1. \square

In general conditional multimodality arises, if the degree of the exponent of (19) is larger than one upon setting all $s_i = 1$ except s_X .

4 Examples

In the following two sections, we illustrate how our results can be used to predict conditional multimodality by several examples. The calculations can be verified using the supplemented Wolfram Mathematica files.

4.1 Basic models

Example 1 (Catalysis) The model of the simple catalytic reaction $X \xrightarrow{k_{cat}} X + Y$ is a trivial example of a hierarchic system, since system II is governed by a time-independent characteristic ODE. By Eqs. (6–7) we have

$$\frac{ds_X}{dt} = -k_{cat}(s_X s_Y - s_X) \tag{34}$$

$$\frac{ds_Y}{dt} = 0 \tag{35}$$

$$\frac{dg}{dt} = 0 \Leftrightarrow g = g(s_X^0, s_Y^0, 0). \tag{36}$$

The solution of (35) simply reads $s_Y(t) = s_Y^0$. We reorder the terms in (34)

$$\frac{ds_X}{dt} = s_X(-k_{cat}(s_Y^0 - 1)),$$

and solve the same equation as

$$s_X(t) = e^{-k_{cat}t(s_Y^0-1)}s_X^0.$$

The generating function, given by (8) for Poissonian initial conditions, reads

$$g_{Pois}(s_X^0, s_Y^0, 0) = \exp\left(\langle x \rangle_0 (s_X^0 - 1) + \langle y \rangle_0 (s_Y^0 - 1)\right).$$

By replacing

$$s_X^0 = e^{k_{cat}t(s_Y^0-1)}s_X,$$

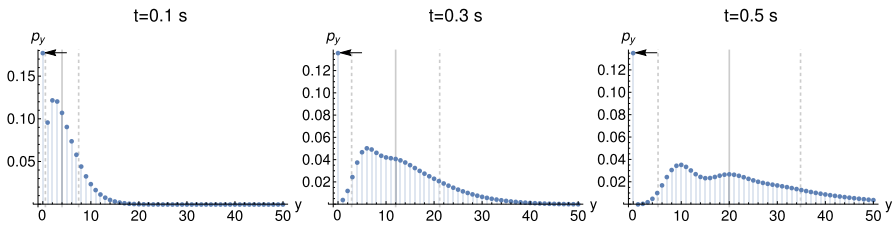


Fig. 3 Marginal distributions for catalysis product $p_y := P(y, t|y_0 = 0, 0)$, given the parameters $k_{\text{cat}} = 20 \text{ s}^{-1}$, deterministic ($y_0 = 0$) and Poissonian initial conditions for $\langle x \rangle_0 = 2$. The thick gray lines mark the expectation value, and the dashed lines the range of one standard deviation from the mean. The arrow marks the probability for a zero outcome

and $s_Y = s_Y^0$ we obtain the generating function

$$g_{\text{GenPois}}(s_X, s_Y, t) = \exp \left(\langle x \rangle_0 (e^{k_{\text{cat}}t(s_Y-1)} s_X - 1) + \langle y \rangle_0 (s_Y - 1) \right). \quad (37)$$

The marginal distribution of Y can be represented as a convolution of Neyman’s type A distribution and a Poisson distribution:

$$\mathcal{P} \sim \underbrace{\left[\text{Poisson}(\langle x \rangle_0) \bigvee \text{Poisson}(k_{\text{cat}}t) \right]}_{\text{Neyman Type A}} * \text{Poisson}(\langle y \rangle_0). \quad (38)$$

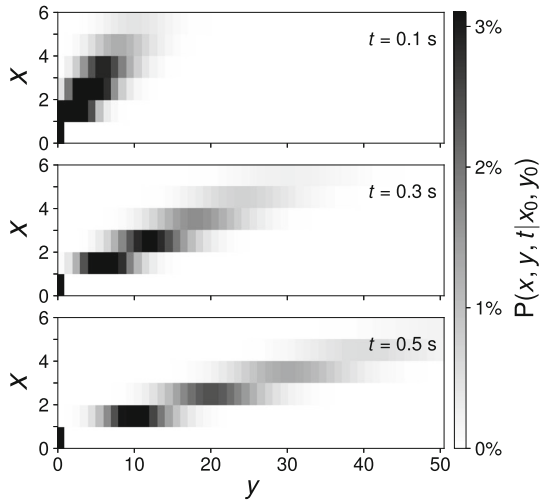
In case we set $y_0 = 0$ as a deterministic initial condition, the distribution is identical to Neyman’s Type A distribution. In Fig. 3, we depict a numerical evaluation of this multimodal distribution. Conditional multimodality is expected by Theorem 3, since the Jacobian exhibits eigenvalues dependent on s_Y . Note that the most likely outcome is zero, is far from the mean (solid gray line in Fig. 3), even outside the standard deviation (dashed gray line). This clearly demonstrates the misleading character of mean and standard deviation for multimodal distributions. The high probability for a zero outcome, depicted by the black arrow in Fig. 3, can be explained by the high likelihood of an “extinction scenario”: If there are initially no molecules of species X , the number of molecules of type Y must always be zero. This phenomenon can be found in many hierarchic networks, as we show in the following figures.

Using deterministic initial conditions, we obtain a contrasting picture, since the resulting marginal distributions are Poissonian:

$$\begin{aligned} g_{\text{det}}(s_X, s_Y, t) &= (s_X)^{x_0} (s_Y)^{y_0} = (e^{k_{\text{cat}}t(s_Y-1)} s_X)^{x_0} (s_Y)^{y_0} \\ \Rightarrow g_{\text{det}}(1, s_Y, t) &= \underbrace{(e^{k_{\text{cat}}t(s_Y-1)})^{x_0}}_{:=g_{\text{cat}}(s_Y)}. \end{aligned} \quad (39)$$

In consequence, no multimodality is observed. However, since $\mathcal{P}_{\text{cat}} \in \text{KTB}^\infty$, Proposition 7 predicts conditionally multimodality. This is not a contradiction because conditional multimodality is a notion defined by the order of the generalizing distribution $g_2(s)$ (see Def. 6), not by the model parameters (here k_{cat}). For a concrete

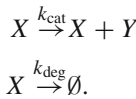
Fig. 4 Plot of the joint distribution for the educt X and product Y for the same parameters and time points as Fig. 3



model, only subsets of the parameter space, defined by the set of coefficients of $g_2(s)$, are reachable by appropriate choices of the model parameters (here k_{cat}). In our model this subset does not contain to multimodal distributions. The joint distribution is plotted in Fig. 4.

To obtain a multimodal distribution for deterministic initial conditions, we add a degradation reaction, as shown in the next example.

Example 2 (Catalysis with degradation) We see a different picture for deterministic initial conditions once we add a degradation mechanism:



These reactions result in the characteristic system

$$\begin{aligned}
 \frac{ds_X}{dt} &= -k_{cat}(s_X s_Y - s_X) - k_{deg}(1 - s_X) \\
 \frac{ds_Y}{dt} &= 0,
 \end{aligned}$$

solved by

$$s_X(t) = \frac{(s_X(k_{cat}(1 - s_Y) + k_{deg}) - k_{deg})e^{t(k_{cat}(1 - s_Y) + k_{deg})} + k_{deg}}{k_{cat}(1 - s_Y) - k_{deg}}.$$

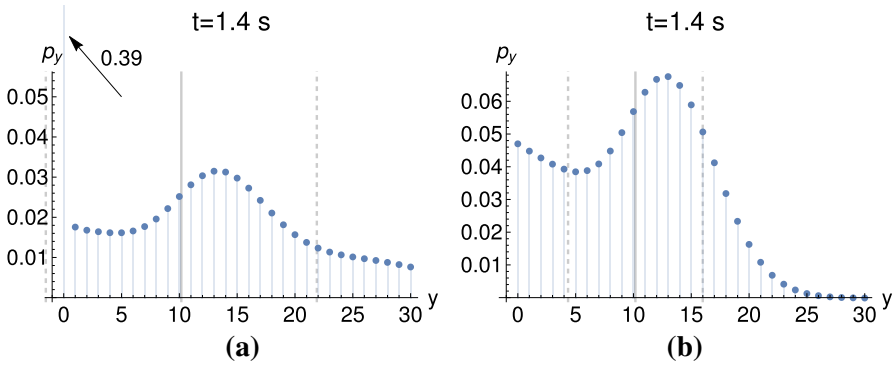


Fig. 5 Marginal distributions $p_y := P(y, t | y_0 = 0, 0)$ at $t = 1.4$ s. Parameters: $k_{\text{cat}} = 10.1 \text{ s}^{-1}$, $k_{\text{deg}} = 0.5 \text{ s}^{-1}$. The thick gray lines depict the expectation value, while the dashed lines stand for the standard deviation. The arrow marks the probability for a zero outcome. **a** Poissonian initial conditions: $\langle x \rangle_0 = 1, \langle y \rangle_0 = 0.001$. **b** Deterministic initial conditions: $x_0 = 1, y_0 = 0$

Deterministic initial conditions result in the following generating function:

$$g(s_X, s_Y, t) = \left(\frac{e^{-t(k_{\text{cat}}(1-s_Y)+k_{\text{deg}})} (-s_X(k_{\text{cat}}(1-s_Y) + k_{\text{deg}}) + k_{\text{deg}} (-e^{t(k_{\text{cat}}(1-s_Y)+k_{\text{deg}})} + k_{\text{deg}}))}{k_{\text{cat}}(s_Y - 1) - k_{\text{deg}}} \right)^{x_0} (s_Y)^{y_0}.$$

Figure 5 shows the comparison between Poissonian and deterministic initial conditions. For Poissonian initial conditions, the scenario of starting without any molecules of species X is possible. This “extinction scenario” does not lead to the production of any Y molecules at any time and is seen by the large peak at $y = 0$ in Fig. 5a. For deterministic initial conditions, we set $x_0 = 1$ and Fig. 5b therefore does not exhibit a peak of the same size at $y = 0$. However, after some time, the species X may well become extinct and therefore the mode at $y = 0$ still appears, albeit with a lower size compared to Fig. 5a.

Example 3 (Simple splitting) The simple splitting reaction $X \xrightarrow{k_1} Y + Z$, obeys the characteristic system

$$\begin{aligned} \frac{ds_X}{dt} &= -k_1(s_Y s_Z - s_X) \\ \frac{ds_Y}{dt} &= 0 \\ \frac{ds_Z}{dt} &= 0. \end{aligned}$$

It is solved by

$$\begin{aligned} s_X(t) &= s_X^0 e^{k_1 t} + s_Y s_Z (1 - e^{k_1 t}) \\ \Leftrightarrow s_X^0 &= e^{-k_1 t} s_X(t) - s_Y s_Z (e^{-k_1 t} - 1). \end{aligned}$$

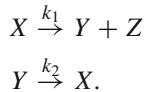
Poissonian initial conditions yield the generating function

$$g(s_X, s_Y, s_Z, t) = \exp\left\{\langle x_1 \rangle_0 (e^{-k_1 t} s_X(t) - s_Y s_Z (e^{-k_1 t} - 1) - 1) + \langle x_2 \rangle_0 (s_Y - 1) + \langle x_3 \rangle_0 (s_Z - 1)\right\}. \tag{40}$$

Therefore, the marginal distributions are Poissonian and the joint distribution is a multivariate Poisson distribution, studied in detail by Kawamura (1979). The product Poisson distribution differs from the multivariate Poisson distribution by the fact that the variables Y and Z correlate, yet the marginal distributions stay unimodal for both.

Without the explicit solution (40), we prove unconditional unimodality by noting that either Y or Z may be assigned to the independent part of the network. In consequence, Theorem 2 can be applied to both Y and Z .

Example 4 (Additional conversion) Things get more interesting as we add a conversion reaction to the same network:



The characteristic system is given by

$$\begin{aligned} \frac{ds_X}{dt} &= -k_1(s_Y s_Z - s_X) \\ \frac{ds_Y}{dt} &= -k_2(s_X - s_Y) \\ \frac{ds_Z}{dt} &= 0, \end{aligned}$$

and the Jacobian by

$$\mathbf{J}(s_Z) = \begin{pmatrix} k_1 & -k_1 s_Z \\ -k_2 & k_2 \end{pmatrix}.$$

The characteristic polynomial

$$\det(\lambda \mathbf{I} - \begin{pmatrix} k_1 & -k_1 s_Z \\ -k_2 & k_2 \end{pmatrix}) = -k_1 \lambda - k_2 \lambda - k_1 k_2 s_Z + k_1 k_2 + \lambda^2$$

depends on s_Z , which implies $\partial_{s_Z} \lambda(s_Z) \neq 0$. Proposition 6 therefore applies and the resulting distributions are conditionally multimodal. Multiple modes, obtained by assuming Poisson distributed initial numbers, are visible in the marginal distribution Fig. 6 as well as the joint distribution Fig. 7.

By adding a degradation reaction $X \xrightarrow{k_3} \emptyset$ to this model and setting $k_2 \gg k_1$, we obtain a model that behaves like Example 2. In consequence, multimodal distributions

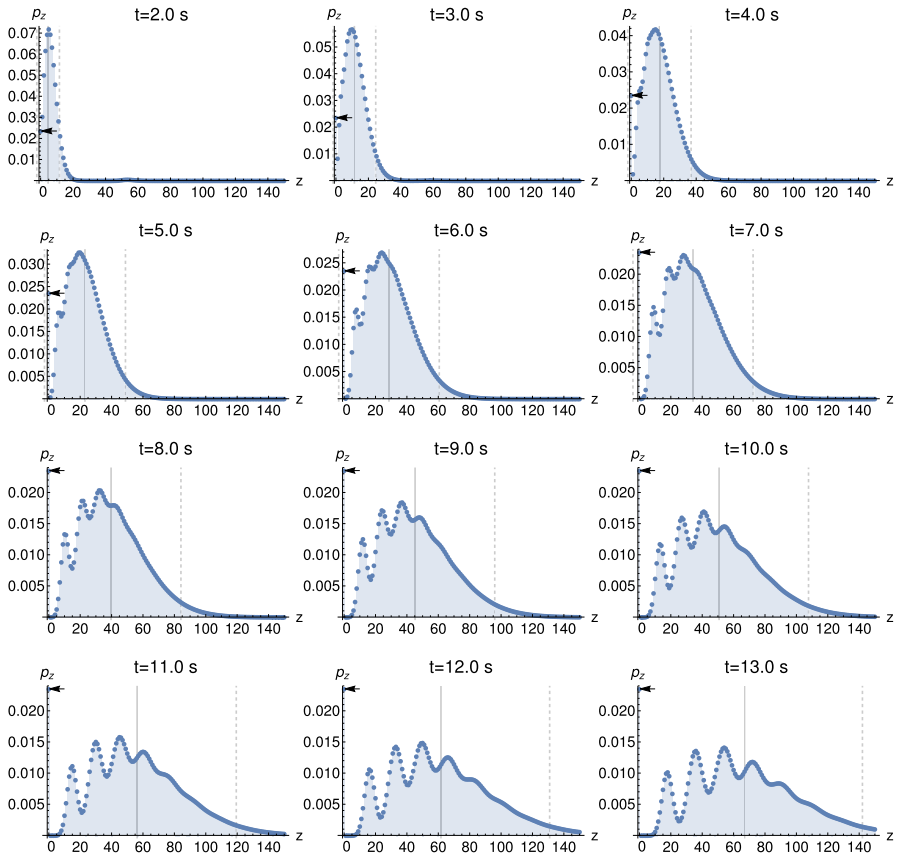
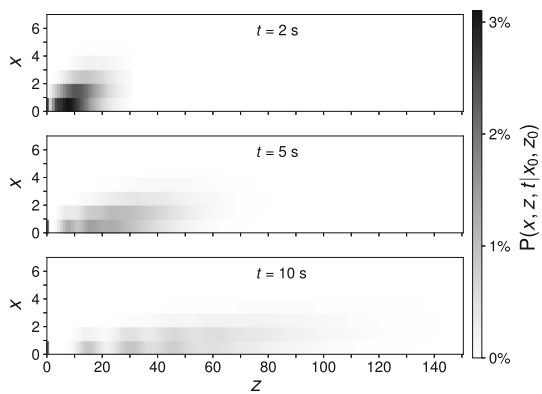


Fig. 6 Marginal distribution $p_z := P(z, t|z_0 = 0, 0)$, given Poissonian initial conditions $\langle x \rangle_0 = 1.875$, $\langle y \rangle_0 = 1.875$, as well as the deterministic initial condition $\xi_Z = 0$. Parameters are $k_1 = 7.0 \text{ s}^{-1}$, $k_2 = 1.875 \text{ s}^{-1}$. The thick black line represents the mean while the dashed lines depict the standard deviation. The arrow marks the probability for a zero outcome

Fig. 7 Plot of the joint distribution for the educt X and product Z for the same parameters as Fig. 6



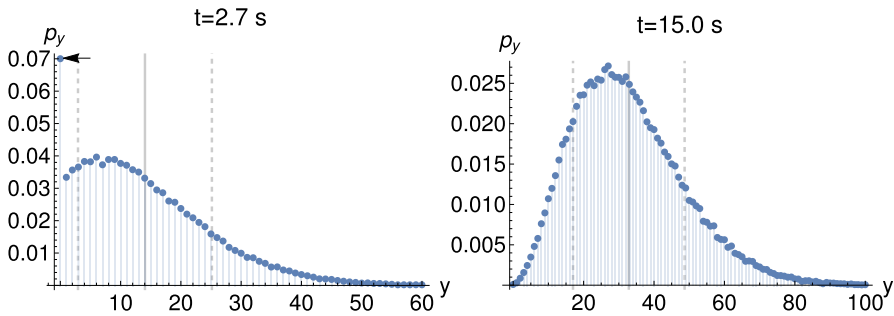
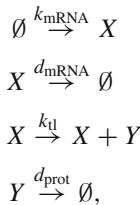


Fig. 8 Marginal distribution of protein numbers $p_y := P(y, t|y_0 = 0, 0)$ created from 10^5 simulated trajectories. Parameters: $t = 2.7$ s and $t = 15.0$ s, $k_{deg}^{mRNA} = 1.2 \text{ s}^{-1}$, $k_{syn}^{mRNA} = 1.2 \text{ s}^{-1}$, $k_{deg}^{prot} = 0.3 \text{ s}^{-1}$, $k_{transl} = 10.0 \text{ s}^{-1}$, deterministic initial conditions with $x_0 = y_0 = 0$. The thick gray lines depict the expectation value, while the dashed lines stand for the standard deviation. The arrow marks the probability for a zero outcome in the left hand side plot

appear for deterministic and Poissonian initial conditions for $k_1 = k_{cat}$ and $k_3 = k_{deg}$ (data not shown).

4.2 Real world models

Example 5 (Transcription–translation model) The two-stage model, which excludes any effects of promoter activity, consists of the reactions



where X are mRNA molecules and Y proteins. We use the implementation of Gillespie’s algorithm by Maarleveld et al. (2013) to simulate this model.

The appearance of multimodal distributions is due to the reaction $X \rightarrow X + Y$, studied in Example 2. As mentioned there, the Jacobian’s eigenvalues depend on s_Y , conditional multimodality is expected. Figures 8 and 9 show that the protein distribution can be multimodal for deterministic and Poissonian initial conditions respectively. Previously, the two-stage transcription–translation model was not shown to be conditionally multimodal (Shahrezaei and Swain 2008). Given small degradation rates, the apparent difference between the shape of Neyman’s Type A distribution (see Figs. 2c and 3), solving the CME for $X \rightarrow X + Y$, and the simulation results is rather small. Furthermore, the previously described “extinction scenario” is visible by the large peak at $y = 0$ in Figs. 8 and 9. This highlights the usefulness of basic models for the study of more complex models.

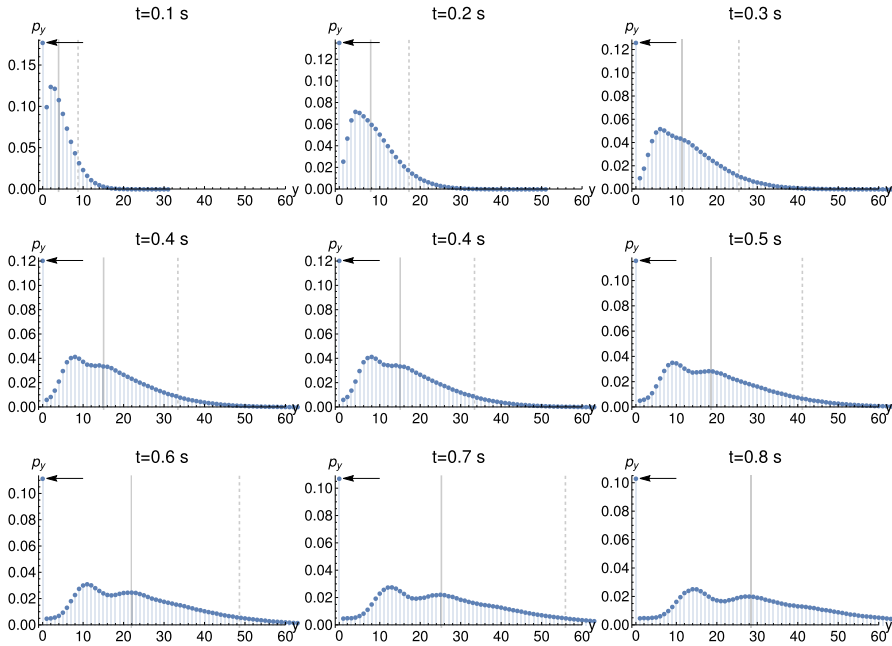
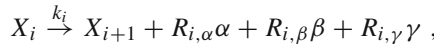


Fig. 9 Marginal distribution of proteins $p_y := P(y, t|y_0 = 0, 0)$, simulated using Poissonian initial conditions for the mRNA, $\langle x \rangle_0 = 2$. Parameters: $k_{\text{deg}}^{\text{mRNA}} = 0.2 \text{ s}^{-1}$, $k_{\text{syn}}^{\text{mRNA}} = 0.4 \text{ s}^{-1}$, $k_{\text{deg}}^{\text{prot}} = 0.3 \text{ s}^{-1}$, $k_{\text{transl}} = 20.0 \text{ s}^{-1}$. The number of simulated trajectories was 10^6 . The thick gray lines depict the expectation value, while the dashed lines stand for the standard deviation. The arrow marks the probability for a zero outcome

The right panel in Fig. 8 shows that the probability of still having zero proteins after 15 s is much smaller than at 2.6 s. Therefore, the corresponding mode disappears. However, the distribution class is still conditionally multimodal as predicted by Theorem 3. Remember that conditional multimodality is a notion defined by the order of the generalizing distribution $g_2(s)$ (see Def. 6), not by the model parameters.

Example 6 (Natural decay series including decay particles) In an extension of the classical model by Bateman (1910), we now formulate the reaction network for nuclear decay chains including decay particles α , β and γ . The stoichiometry of the model is



where $R_{i,\alpha}$, $R_{i,\beta}$ and $R_{i,\gamma}$ are the numbers of decay particles involved in each reaction.

The generating function PDE corresponding to the Master equation is

$$\partial_t g = \sum_{i=1}^{n-1} k_i (s_\alpha^{R_{\alpha,i}} s_\beta^{R_{\beta,i}} s_\gamma^{R_{\gamma,i}} s_{i+1} - s_i) \partial_{s_i} g . \tag{41}$$

As before, we solve the PDE (41) using the method of characteristics. The characteristic ODEs corresponding to Eq. (41) are:

$$\frac{ds_i}{dt} = -k_i (s_\alpha^{R_{\alpha,i}} s_\beta^{R_{\beta,i}} s_\gamma^{R_{\gamma,i}} s_{i+1} - s_i) \quad (i \in 1, \dots, n - 1) \tag{42}$$

$$\frac{ds_n}{dt} = 0 \tag{43}$$

$$\frac{ds_\alpha}{dt} = \frac{ds_\beta}{dt} = \frac{ds_\gamma}{dt} = 0$$

$$\frac{dg(\mathbf{s}(t), t)}{dt} = 0 \Rightarrow g(\mathbf{s}(t), t) = f(\mathbf{s}^0). \tag{44}$$

These equations represent a hierarchically linear ODE system. It can be interpreted as being composed of two subsystems, just as in Theorem 2. System I consists of all isotopes, while system II represents the α , β and γ particles. System II is a trivial example of a monomolecular reaction network, since no reactions take place within the system, we just have to consider the influx of particles from system I. We note that the Jacobian matrix of the characteristic system is identical to the rate equation's transposed, as discussed in Sect. 2.3. The ODE system is homogeneous and we have

$$\begin{aligned} \mathbf{s} &= e^{\mathbf{A}(\mathbf{s}_{\alpha,\beta,\gamma})t} \mathbf{s}^0 \\ \Leftrightarrow \mathbf{s}^0 &= e^{-\mathbf{A}(\mathbf{s}_{\alpha,\beta,\gamma})t} \mathbf{s}, \end{aligned}$$

where $\mathbf{A}(\mathbf{s}_{\alpha,\beta,\gamma}) = (\partial_{s_j} \frac{ds_i}{dt})$ is the Jacobian. We express the generating function for Poissonian initial conditions by this matrix as

$$\begin{aligned} g(\mathbf{s}, t) &= \exp \left\{ \langle \mathbf{x} \rangle_0 \cdot (\mathbf{s}^0 - \mathbf{1}) \right\} \\ &= \exp \left\{ \langle \mathbf{x}_{\alpha,\beta,\gamma} \rangle_0 \cdot (\mathbf{s}_{\alpha,\beta,\gamma} - \mathbf{1}) + \langle \mathbf{x}_{1,\dots,n} \rangle_0 \cdot (\exp(-\mathbf{A}(\mathbf{s}_{\alpha,\beta,\gamma})t) \mathbf{s} - \mathbf{1}) \right\} \\ &= \exp \left\{ \langle \mathbf{x}_{\alpha,\beta,\gamma} \rangle_0 \cdot (\mathbf{s}_{\alpha,\beta,\gamma} - \mathbf{1}) + \exp(-\mathbf{A}(\mathbf{s}_{\alpha,\beta,\gamma})^T t) \langle \mathbf{x}_{1,\dots,n} \rangle_0 \cdot \mathbf{s} - \langle \mathbf{x}_{1,\dots,n} \rangle_0 \cdot \mathbf{1} \right\}. \end{aligned}$$

In the last line, we have used the identity $\mathbf{A}\mathbf{b} \cdot \mathbf{c} = \mathbf{b} \cdot \mathbf{A}^T \mathbf{c} = \mathbf{A}^T \mathbf{c} \cdot \mathbf{b}$, which holds for any square matrix \mathbf{A} and two vectors \mathbf{b}, \mathbf{c} of the same dimension.¹⁵ The insight of the equivalence of the characteristic system and the transposed rate equation system, as discussed in Sect. 2.3, greatly simplifies the following computations. In consequence, the solution of the Master equation is not much more difficult than the solution of the rate equations, as provided by Bateman (1910) originally.

¹⁵ This identity can be seen by writing out the components of the scalar and matrix product:

$$\sum_{i=1}^n (\mathbf{A}\mathbf{b})_i c_i = \sum_{i=1}^n b_i (\mathbf{A}^T \mathbf{c})_i = \sum_{i=1}^n \sum_{j=1}^n A_{ij} b_j c_i = \sum_{i=1}^n b_i \sum_{j=1}^n A_{ij}^T c_j = \sum_{i=1}^n b_i \sum_{j=1}^n A_{ji} c_j.$$

The difference to Bateman’s system is that the negative Jacobian matrix \mathbf{A}^T depends on the constants s_α, s_β and s_γ :

$$-\mathbf{A}^T(\mathbf{s}_{\alpha,\beta,\gamma}) = \begin{pmatrix} -k_1 & 0 & \dots & & 0 \\ k_1 s_\alpha^{R_{\alpha,1}} s_\beta^{R_{\beta,1}} s_\gamma^{R_{\gamma,1}} & & & & \\ 0 & \ddots & & & \\ & & -k_i & & 0 \\ & & k_i s_\alpha^{R_{\alpha,i}} s_\beta^{R_{\beta,i}} s_\gamma^{R_{\gamma,i}} & & \\ & & & \ddots & \\ & & & & -k_{n-1} & 0 \\ 0 & & & & k_{n-1} s_\alpha^{R_{\alpha,n-1}} s_\beta^{R_{\beta,n-1}} s_\gamma^{R_{\gamma,n-1}} & 0 \end{pmatrix}.$$

Since this matrix is triangular, the eigenvalues are the diagonal entries. Because the eigenvalues do not depend on $\mathbf{s}_{\alpha,\beta,\gamma}$, the marginal distribution of the corresponding particles obeys Proposition 3. Furthermore, Theorem 4 predicts that conditional multimodality depending on the choice of $R_{\alpha,i}, R_{\beta,i}$ and $R_{\gamma,i}$. Note that Theorem 4 gives us this information without the exact solution of the system presented here.

We now proceed along the lines of Pressyanov (2002) and solve the rate equation system, transposed to the characteristic system, to obtain $\exp(-\mathbf{A}^T(\mathbf{s}_{\alpha,\beta,\gamma})t)\langle \mathbf{x}_1, \dots, \mathbf{n} \rangle_0$. The computations can be found in ‘‘Appendix A.4’’ and yield

$$\begin{aligned} & \left(\exp(-\mathbf{A}^T(\mathbf{s}_{\alpha,\beta,\gamma})t)\langle \mathbf{x}_1, \dots, \mathbf{n} \rangle_0 \right)_i \\ &= \langle x_i \rangle_0 e^{-k_i t} + \sum_{m=1}^{i-1} \langle x_m \rangle_0 \left(\prod_{l=m}^{i-1} k_l s_\alpha^{R_{\alpha,l}} s_\beta^{R_{\beta,l}} s_\gamma^{R_{\gamma,l}} \right) \sum_{l=m}^i \frac{e^{-k_l t}}{\prod_{j=m, j \neq l}^i (k_j - k_l)}. \end{aligned}$$

In consequence, we use the matrix exponential to express the generating function as

$$\begin{aligned} g(\mathbf{s}, t) &= \exp \left\{ \langle \mathbf{x}_{\alpha,\beta,\gamma} \rangle_0 \cdot (\mathbf{s}_{\alpha,\beta,\gamma} - \mathbf{1}) \right. \\ &+ \sum_{i=1}^n \left[\left(\langle x_i \rangle_0 e^{-k_i t} \right. \right. \\ &+ \left. \left. \sum_{m=1}^{i-1} \langle x_m \rangle_0 \left(\prod_{l=m}^{i-1} k_l s_\alpha^{R_{\alpha,l}} s_\beta^{R_{\beta,l}} s_\gamma^{R_{\gamma,l}} \right) \sum_{l=m}^i \frac{e^{-k_l t}}{\prod_{j=m, j \neq l}^i (k_j - k_l)} \right) (s_i - \langle x_i \rangle_0) \right] \left. \right\}. \end{aligned}$$

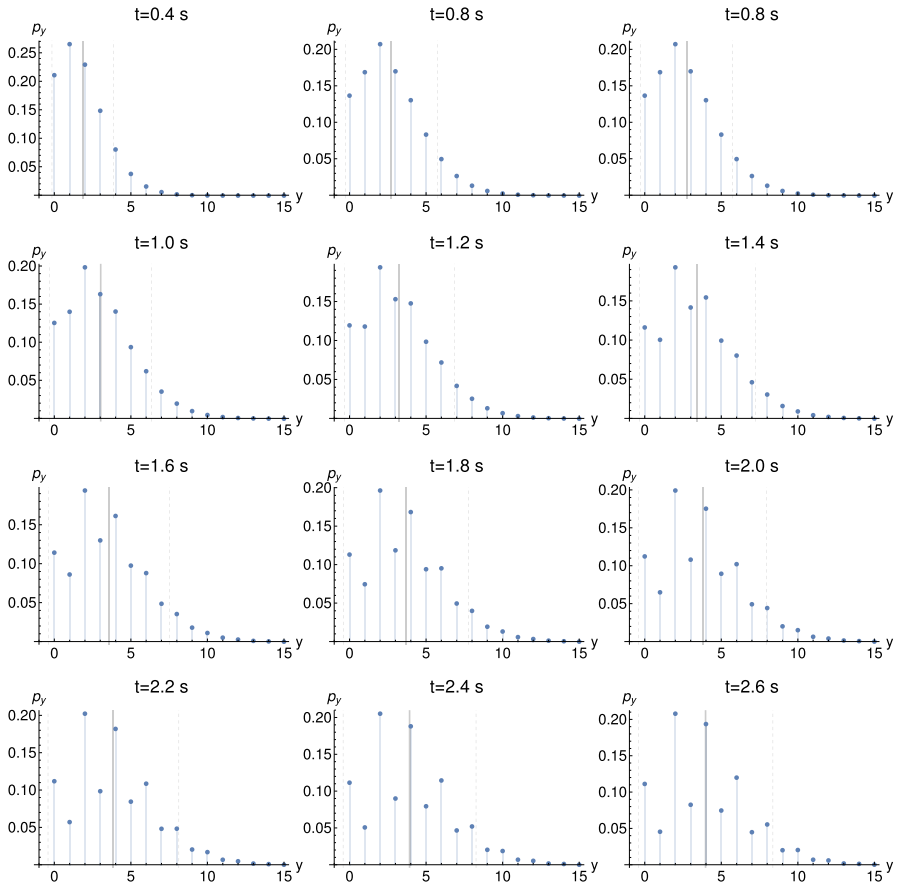


Fig. 10 Marginal distribution $p_y := P(y, t|y_0 = 0, 0)$ of decay particles Y . Parameters: $\langle x_1 \rangle_0 = 2$, $\langle x_2 \rangle_0 = 0.1$, $\langle x_3 \rangle_0 = 0.001$, $k_1 = 3.1 \text{ s}^{-1}$, $k_2 = 1 \text{ s}^{-1}$. The thick black line represents the mean while the dashed lines depict the standard deviation

Interestingly, the further the isotopes are downstream of the decay chain, the more they contribute to the higher order cumulants¹⁶ of the α , β and γ particles.¹⁷ An insight like that is not easily gained by doing stochastic simulations, because the simulation results alone would not directly point to this fact. We also see that the distribution is not simply characterized by mean and covariance.

Next, we consider multimodality of the marginal distributions. Since the eigenvalues of $\mathbf{A}^T(\mathbf{s}_{\alpha, \beta, \gamma})$ do not depend on $\mathbf{s}_{\alpha, \beta, \gamma}$, Theorem 4 applies and predicts conditional multimodality for several minimal networks. First, we consider networks corresponding to case 1. We evaluate the solution for the decay chain

¹⁶ The cumulants are defined as the coefficients of the cumulant generating function $\kappa(\xi) := \log(g(e^\xi))$ at $\xi = 0$.

¹⁷ This can be seen by looking at the number of factors $\prod_{l=m}^{i-1}$.



and plot the marginal distributions for Y in Fig. 10. The plots, depicting the temporal evolution of the marginal distribution, clearly demonstrate multimodality. Here, (45–46) represent the minimal reaction network Type I from Theorem 4. A minimal reaction network of Type II is obtained by reducing the network to



Finally, by assigning X_2 to the dependent part of the network, we obtain an even smaller network as $X_1 \xrightarrow{k_1} (X_2) + 2Y$. Now the independent part of the network is open. If we ignore X_2 , this represents case 2 in Theorem 4.

Note that X_2 can either be assigned to the independent or the dependent part of the reaction network. However, in either case Theorem 4 applies and predicts conditional multimodality.

5 Conclusions

In this article, we analyzed the CME for hierarchic first-order reaction networks, based on a general solution method. We showed in Theorem 1 that hierarchic reaction networks are generally treatable in an analytic manner. We derived the analytical solution for the joint probability generating function for Poisson and deterministic initial conditions. Next we analyzed the multimodality of resulting marginal distributions of individual species. The analysis revealed that the marginal distributions of species from the independent part of the network are unimodal (Theorem 2), while the dependent part yields conditionally multimodal distributions (Theorems 3 and 4). Furthermore, we presented several basic models of hierarchic reaction networks, which we consider insightful for the understanding of larger reaction networks. We illustrated this point by showing the similarity between the catalysis basic model and the transcription–translation model. As a proof of principle, we showed that even more complex models such as the nuclear decay chain presented here, are amenable to an exact analytical treatment.

Underlining the prevalence of multimodality, we saw that even trivial networks like the one from Example 1, consisting of the reaction $X \rightarrow X + Y$, are conditionally multimodal given Poissonian initial conditions. We also showed that deterministic initial conditions give rise to multimodal distributions in Examples 2 and 5. For the two-stage transcription–translation model, multimodality was previously not reported (Shahrezaei and Swain 2008).

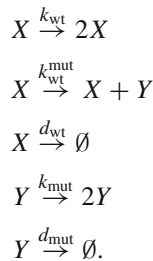
Since the main Theorems 3 and 4 only require knowledge about the dependence of the eigenvalues of the Jacobian matrix on the variable of interest, it is sufficient to compute the characteristic polynomial of the Jacobian matrix. However, the characteristic polynomial of any square matrix is algorithmically computable and so our Theorems are algorithmically applicable in practice. We furthermore demonstrated

how insight about a network may be gained by simply dividing the network into dependent and independent parts (see Example 3), without doing any calculations or simulations. Therefore, we believe that our abstract treatment of hierarchic reaction networks may inspire other researchers to analyze and design networks with respect to their multimodality properties.

6 Further developments

As a future project, we suggest the development of a numerical software to determine the parameter regions of multimodality of our basic models. In that study, the collection of basic models maybe extended, since the ones presented here are far from the only ones whose generating function may be written in a compact form. It might also be interesting to study n -fold hierarchies instead of the two-level hierarchies presented here.

Furthermore, we note that the Luria–Delbrück model (Zheng 1999), describing the division of wild-type cells X into mutants Y may be investigated in a similar way we proposed. One of the many variants of the model is defined by



Even though the reaction equations show the hierarchic character of the model, they do not fit our definition of hierarchic first-order networks. This fact can be seen by the characteristic ODEs:

$$\frac{ds_X}{dt} = -k_{wt}(s_X^2 - s_X) - k_{wt}^{mut}(s_X s_Y - s_X) - d_{wt}(1 - s_X) \tag{48}$$

$$\frac{ds_Y}{dt} = -k_{mut}(s_Y^2 - s_Y) - d_{mut}(1 - s_Y)$$

$$\frac{dg}{dt} = 0. \tag{49}$$

The system is not hierarchically linear due to the s_X^2 and s_Y^2 terms, which is why we excluded autocatalytic reactions $X \rightarrow 2X$ from our study. However, (48) and (49) are examples of Riccati differential equations, and may be linearized as such by a coordinate transformation. In this way, an exact solution of the system can be obtained,¹⁸ even though it involves large expressions involving hypergeometric functions.

¹⁸ Write the author for a sketch of the calculations.

Acknowledgements The authors thank Jakob Löber for fruitful discussions. Matthias Reis was funded by the Leibniz Graduate School of Molecular Biophysics.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

A.1 How to obtain distributions from generating functions

Before finally coming to the proof of Proposition 2, we discuss how the probability mass function (PMF), defined as the coefficients of the generating function, can be found. In the first-order case, this is not as easy as in the monomolecular case, where the product Poisson and multinomial distributions are obtained (Jahnke and Huisinga 2007). Even the PMF of the relatively simple Neyman Type A distribution given by $g(s) = e^{\lambda(\exp(\phi(s-1))-1)}$, almost identical to Eq. (38) from Example 1, is defined only in terms of a multinomial expansion (Zhang et al. 2014). In general, the distribution corresponding to the generating function of Eq. (19) is not expressible in terms of elementary functions, except for special cases.

Generally, the coefficients of a generating function $g(\mathbf{s}, t)$ are given by the multivariate Cauchy integral formula

$$P(\mathbf{x}, t) = \frac{1}{(2\pi i)^n} \oint_{\mathcal{C}_1 \times \dots \times \mathcal{C}_n} \frac{g(\mathbf{s}, t)}{\mathbf{s}^{\mathbf{x}+1}} ds. \quad (50)$$

This identity is due to the definition of the generating function as a power series. Here, \mathcal{C}_i are discs encircling the origin on the complex (hyper)plane. See Pemantle and Wilson (2013), Appendix A, for the technicalities of the definition of multivariate Cauchy integrals.

Asymptotic joint distributions $|\mathbf{x}| \rightarrow \infty$ can be obtained by applying a saddle point integration approach to the Cauchy integral (50). Pemantle and Wilson (2013) present these techniques, developed over the course of the last 20 years, in an excellent textbook. The resulting asymptotic distributions are often in very good numerical agreement with the exact ones, even for low $|\mathbf{x}|$, and are a convenient tool to analyze the distributions.

If only a numerical evaluation of the distributions is needed, the Discrete Fourier Transform (DFT) is obtained by defining $s := e^{-2\pi i \omega}$ with $\omega \in [0, 2\pi]$ and the application of an inverse Fast-Fourier Transform (FFT) leads to a fast computation of the distributions. Recently, Xu and Minin (2015) have made efforts to increase the efficiency of these computations. Numerical integration of (50) is of course also a possibility.

Lastly, it shall be mentioned that there are a number of algorithms based on recurrences (Zheng 1999; Zhang et al. 2014), applicable for one dimensional distributions. The most important of these is the *Lévy-Adelson-Panjer recursion* for the DCP^N, given

by

$$p_{k+1} = \frac{\lambda}{k+1} \sum_{j=0}^k (k+1-j)\alpha_{k+1-j} p_j. \tag{51}$$

In practice, the sum $\sum_{j=0}^k$ often converges after a constant number of terms (i.e. the convergence is linear,¹⁹) so the computational cost of evaluating n probabilities is $O(n)$. Khatri and Patel (1961) provide recurrences for KTB^N distributions, needed for the deterministic initial condition setting.

The generating function can also be used directly for parameter estimation, without calculating the distribution first. Marcheselli et al. (2008) propose a method for parameter estimation that is based exclusively of generating functions.

A.2 Proof of Lemma 1

Proof The proof is adapted from Gardiner (2009), Sect. 11.6.4, p. 297. We first multiply with $\mathbf{s}^{\mathbf{x}}$ and sum over all \mathbf{x} to obtain

$$\begin{aligned} \partial_t g(\mathbf{s}, t) &= \sum_{\mathbf{x}=0}^{\infty} \sum_{i=1}^m \mathbf{s}^{\mathbf{x}} (\alpha_i(\mathbf{x} - \mathbf{n}_i) P(\mathbf{x} - \mathbf{n}_i, t) - \alpha_i(\mathbf{x}) P(\mathbf{x}, t)) \\ &= \sum_{i=1}^m \left(\sum_{\mathbf{x}'=0}^{\infty} \mathbf{s}^{\mathbf{x}'+\mathbf{n}_i} \alpha_i(\mathbf{x}') P(\mathbf{x}', t) - \sum_{\mathbf{x}=0}^{\infty} \mathbf{s}^{\mathbf{x}} \alpha_i(\mathbf{x}) P(\mathbf{x}, t) \right). \end{aligned} \tag{52}$$

Using the definition of the stochastic propensities $\alpha_i(\mathbf{x}) = k_i \prod_{j=1}^n \frac{x_j!}{(x_j - Q_{ji})!}$, we find the identities

$$\begin{aligned} \mathbf{s}^{\mathbf{x}} \alpha_i(\mathbf{x}) &= k_i \prod_{j=1}^n \frac{s_j^{x_j} x_j!}{(x_j - Q_{ji})!} = k_i \prod_{j=1}^n s_j^{x_j} (x_j(x_j - 1)(x_j - 2) \cdots (x_j - Q_{ji} + 1)) \\ &= k_i \prod_{j=1}^n s_j^{Q_{ji}} \left(\partial_{s_j}^{Q_{ji}} s_j^{x_j} \right) \end{aligned}$$

and

$$\mathbf{s}^{\mathbf{x}+\mathbf{n}_i} \alpha_i(\mathbf{x}) = k_i \prod_{j=1}^n \frac{s_j^{x_j+n_{ji}} x_j!}{(x_j - Q_{ji})!} = k_i \prod_{j=1}^n s_j^{R_{ji}} \left(\partial_{s_j}^{Q_{ji}} s_j^{x_j} \right).$$

Plugging these into Eq. (52), the desired Eq. (4) is obtained. □

¹⁹ Deuffhard and Hohmann (2003), p. 85, Definition 7.

A.3 Proof of solution formula for twofold hierarchy

Proof We start by writing out the PDE for a system of two parts, corresponding to the variables $\mathbf{s}_{\text{ind}} := (s_1, \dots, s_{n_{\text{ind}}})^T$ and $\mathbf{s}_{\text{dep}} := (s_{n_{\text{ind}}+1}, \dots, s_n)^T$, as depicted in an exemplary manner in Fig. 1.

The PDE for the generating function, derived from Lemma 1, then reads

$$\begin{aligned} \frac{\partial g(\mathbf{s}, t)}{\partial t} &= \sum_{i=1}^m k_i \left(c_i(\mathbf{s}_{\text{dep}}) \left(\sum_{l=1}^{n_{\text{ind}}} R_{li} s_l + \delta_{\sum_{l=1}^{n_{\text{ind}}} R_{li}, 0} \right) - \sum_{l=1}^{n_{\text{ind}}} Q_{li} s_l \right) \left(\sum_{l=1}^{n_{\text{ind}}} Q_{li} \partial_{s_l} \right) g \\ &+ \sum_{i=n_{\text{ind}}+1}^N \sum_{j=n_{\text{ind}}+1}^N \alpha_{ij} (s_j - s_i) \partial_{s_j} g(\mathbf{s}, t) \\ &+ \sum_{i=n_{\text{ind}}+1}^N \alpha_{i0} (1 - s_i) \partial_{s_i} g(\mathbf{s}, t) + \mathbf{b} \cdot (\mathbf{s} - \mathbf{1}) g(\mathbf{s}, t). \end{aligned}$$

$c_i(\mathbf{s}_{\text{dep}}(t', t)) := \prod_{l=n_{\text{ind}}+1}^n s_l^{R_{li}}(t', t)$ is defined to be the function representing the produced molecules that do not belong to part I. The characteristic ODEs are obtained in the same manner as for monomolecular systems and read for part I as:

$$\frac{ds_j}{dt} = - \sum_{i=1}^m Q_{ji} k_i \left(c_i(\mathbf{s}_{\text{dep}}) \left(\sum_{l=1}^{n_{\text{ind}}} R_{li} s_l + \delta_{\sum_{l=1}^{n_{\text{ind}}} R_{li}, 0} \right) - Q_{ji} s_j \right) \quad (k_i \in \mathbb{R}^+).$$

Part II is monomolecular and the characteristic equations are:

$$\frac{d\mathbf{s}_{\text{dep}}}{dt} = -\mathbf{A}^T(\mathbf{s}_{\text{dep}} - \mathbf{1}).$$

The total derivative of $g(\mathbf{s}(t), t)$ reads

$$\frac{dg(\mathbf{s}(t), t)}{dt} = \mathbf{b} \cdot (\mathbf{s} - \mathbf{1}) g(\mathbf{s}, t). \tag{53}$$

The Jacobian matrix for part I is given by

$$\begin{aligned} J_{jk} &= -\partial_{s_k} \sum_{i=1}^m Q_{ji} k_i \left(c_i(\mathbf{s}_{\text{dep}}) \left(\sum_{l=1}^{n_{\text{ind}}} R_{li} s_l(t) + \delta_{\sum_{l=1}^{n_{\text{ind}}} R_{li}, 0} \right) - Q_{ji} s_j(t) \right) \quad (k \in 1, \dots, n_{\text{ind}}) \\ &= - \sum_{i=1}^m Q_{ji} k_i \left(c_i(\mathbf{s}_{\text{dep}}) \sum_{l=1}^{n_{\text{ind}}} R_{li} \delta_{kl} - 1 \right) \\ &= \sum_{i=1}^m Q_{ji} k_i (1 - c_i(\mathbf{s}_{\text{dep}}) R_{ki}). \end{aligned}$$

If the Jacobian commutes for two different times, i.e. $\mathbf{J}(t_1)\mathbf{J}(t_2) = \mathbf{J}(t_2)\mathbf{J}(t_1)$, the homogeneous part of this system is solved by $\exp\left\{\int_0^t \mathbf{J}(\mathbf{s}_{\text{dep}}(t'))dt'\right\}$. If it does not commute, $\exp\left\{\int_0^t \mathbf{J}(\mathbf{s}_{\text{dep}}(t'))dt'\right\}$ is the first term of the Magnus expansion

$$e^{\Omega(\mathbf{J}(t))} = \exp\left\{\int_0^t \mathbf{J}(\mathbf{s}_{\text{dep}}(t'))dt' + \int_0^t \int_0^{t_1} [\mathbf{J}(\mathbf{s}_{\text{dep}}(t_1)), \mathbf{J}(\mathbf{s}_{\text{dep}}(t_2))]dt_2dt_1 + \dots\right\}.$$

The nonhomogeneous part is

$$f_j(t) := -\sum_{i=1}^m Q_{ji}k_i c_i(\mathbf{s}_{\text{dep}}(t)) \tag{54}$$

for the l th component of system I, so the complete solution reads

$$\begin{aligned} \mathbf{s}_{\text{ind}}(t) = & \exp\left\{\int_0^t \mathbf{J}(\mathbf{s}_{\text{dep}}(t'))dt'\right\} \mathbf{s}_{\text{ind}}^0 + \\ & \exp\left\{\int_0^t \mathbf{J}(\mathbf{s}_{\text{dep}}(t'))dt'\right\} \int_0^t \exp\left\{-\int_0^{t'} \mathbf{J}(\mathbf{s}_{\text{dep}}(t''))dt''\right\} \mathbf{f}(\mathbf{s}_{\text{dep}}(t'))dt' \end{aligned} \tag{55}$$

$$\mathbf{s}_{\text{dep}}(t) = \exp(-\mathbf{A}^T t) (\mathbf{s}_{\text{dep}}^0 - \mathbf{1}) + \mathbf{1}. \tag{56}$$

From Eq. (53), we know that

$$g(\mathbf{s}(t), t) = d(\mathbf{s}^0) \exp\left(\int_0^t \mathbf{b} \cdot (\mathbf{s}(t') - \mathbf{1})dt'\right)$$

where $d(\mathbf{s}^0)$ is an arbitrary function to be determined by the initial conditions. The integral over $c_i(\mathbf{s}_{\text{dep}}(t', t)) = \prod_{l=n_{\text{ind}}+1}^n (s_l(t', t))^{R_{li}}$ appears both in the homogeneous and in the inhomogeneous part of the equations for $\mathbf{s}_{\text{ind}}(t', t)$:

$$\begin{aligned} \int_0^t \prod_{l=n_{\text{ind}}+1}^n s_l^{R_{li}}(t', t) dt' &= \int_0^t \prod_{l=n_{\text{ind}}+1}^n \left(\exp\left\{-\mathbf{A}^T t'\right\} (\mathbf{s}_{\text{dep}}^0 - \mathbf{1}) \cdot \epsilon_l + 1\right)^{R_{li}} dt' \\ &= \int_0^t \prod_{l=n_{\text{ind}}+1}^n \left(\exp\left\{-\mathbf{A}^T t'\right\} \exp(\mathbf{A}^T t) (\mathbf{s}_{\text{dep}} - \mathbf{1}) \cdot \epsilon_l + 1\right)^{R_{li}} dt' \\ &= \int_0^t \prod_{l=n_{\text{ind}}+1}^n \left(\exp\left\{-\mathbf{A}(t' - t)\right\} \epsilon_l \cdot (\mathbf{s}_{\text{dep}} - \mathbf{1}) + 1\right)^{R_{li}} dt'. \end{aligned} \tag{57}$$

In the last line, we made use of the identity $\mathbf{A}\mathbf{b} \cdot \mathbf{c} = \mathbf{b} \cdot \mathbf{A}^T \mathbf{c} = \mathbf{A}^T \mathbf{c} \cdot \mathbf{b}$. We note that the integral can be expressed completely in terms of the fundamental matrix of the second, monomolecular reaction network. Thus, the property that the statistics of

monomolecular networks can be fully parametrized by the solution of its traditional reaction rate equations translates to hierarchic first-order networks.

Using Poissonian initial conditions, we find $d(\mathbf{s}^0)$:

$$g(\mathbf{s}^0, t = 0) = \exp(\langle \mathbf{x} \rangle_0 \cdot (\mathbf{s}^0 - \mathbf{1})) = d(\mathbf{s}^0). \tag{58}$$

We solve Eq. (55) for $\mathbf{s}_{\text{ind}}^0$:

$$\mathbf{s}_{\text{ind}}^0 = \exp \left\{ - \int_0^t \mathbf{J}(\mathbf{s}_{\text{dep}}(t')) dt' \right\} \mathbf{s}_{\text{ind}} - \int_0^t \exp \left\{ - \int_0^{t'} \mathbf{J}(\mathbf{s}_{\text{dep}}(t'')) dt'' \right\} \mathbf{f}(\mathbf{s}_{\text{dep}}(t')) dt'.$$

The last integral is over a double exponential function and a simple exponential, hidden in

$$f_j(\mathbf{s}_{\text{dep}}, t', t) := - \sum_{i=1}^m Q_{ji} k_i \prod_{l=\text{ind}+1}^n \left(\exp \left\{ - \mathbf{A}(t' - t) \right\} \epsilon_l \cdot (\mathbf{s}_{\text{dep}} - \mathbf{1}) + 1 \right)^{R_{li}}$$

and is not easily expressed in terms of elementary functions. A series expansion can be found, but is not very simple, which is why we keep the integral as it is. For $\mathbf{s}_{\text{dep}}^0$, we obtain the same result as in the monomolecular case, namely

$$\mathbf{s}_{\text{dep}}^0 = \exp(\mathbf{A}^T t) (\mathbf{s}_{\text{dep}} - \mathbf{1}) + \mathbf{1}.$$

The complete generating function is

$$\begin{aligned} g(\mathbf{s}, t) &= \exp \left\{ \langle \mathbf{x}_1 \rangle_0 \cdot \underbrace{\left(e^{-\int_0^t \mathbf{J}(\mathbf{s}_{\text{dep}}, t', t) dt'} \mathbf{s}_{\text{ind}} - \int_0^t e^{-\int_0^{t'} \mathbf{J}(\mathbf{s}_{\text{dep}}, t'', t) dt''} \mathbf{f}(\mathbf{s}_{\text{dep}}, t', t) dt' - \mathbf{1} \right)}_{:=\mathbf{s}_{\text{ind}}(t)} \right. \\ &\quad \left. + \langle \mathbf{x}_2 \rangle_0 \cdot \exp(\mathbf{A}^T t) (\mathbf{s}_{\text{dep}} - \mathbf{1}) + \int_0^t \mathbf{b} \cdot (\mathbf{s}_{\text{ind}}(t''') - \mathbf{1}) dt''' \right\} \\ &= \exp \left\{ \left(e^{-\int_0^t \mathbf{J}^T(\mathbf{s}_{\text{dep}}, t', t) dt'} \langle \mathbf{x}_1 \rangle_0 \cdot \mathbf{s}_{\text{ind}} \right. \right. \\ &\quad \left. \left. - \int_0^t e^{-\int_0^{t'} \mathbf{J}^T(\mathbf{s}_{\text{dep}}, t'', t) dt''} \langle \mathbf{x}_1 \rangle_0 \cdot \mathbf{f}(\mathbf{s}_{\text{dep}}, t', t) dt' - \langle \mathbf{x}_1 \rangle_0 \right) \right. \\ &\quad \left. + \exp(\mathbf{A} t) \langle \mathbf{x}_2 \rangle_0 \cdot (\mathbf{s}_{\text{dep}} - \mathbf{1}) + \int_0^t \mathbf{b} \cdot (\mathbf{s}_{\text{ind}}(t''') - \mathbf{1}) dt''' \right\}. \tag{59} \end{aligned}$$

If the initial condition is deterministic, that is $P(\mathbf{x}, 0) = \delta_{\mathbf{x}, \mathbf{x}^0}$, where $\mathbf{x}^0 = (x_1^0, \dots, x_n^0)^T$, we have

$$g(\mathbf{s}(0), 0) = \sum_{\mathbf{x}=\mathbf{0}}^{\infty} \delta_{\mathbf{x}, \mathbf{x}^0}(\mathbf{s}_0) \mathbf{x}^0 = \prod_{i=1}^n (s_i^0)^{x_i^0} = d(\mathbf{s}^0). \tag{60}$$

This yields the result, using $\epsilon_i := \text{col}_i \mathbf{I}_{n \times n}$ and $\epsilon_i^{(1)} := (\epsilon_{i,1}, \dots, \epsilon_{i,n_{\text{ind}}})^T$, $\epsilon_i^{(2)} := (\epsilon_{i,n_{\text{ind}}+1}, \dots, \epsilon_{i,n})^T$:

$$\begin{aligned}
 d(\mathbf{s}^0) &= \prod_{i=1}^n (s_i^0)^{x_i^0} = \prod_{i=1}^n (\mathbf{s}^0 \cdot \epsilon_i)^{x_i^0} \\
 &= \prod_{i=1}^n \left(\left[e^{-\int_0^t \mathbf{J}(\mathbf{s}_{\text{dep}}, t') dt'} \mathbf{s}_{\text{ind}} - \int_0^t e^{-\int_0^{t'} \mathbf{J}(\mathbf{s}_{\text{dep}}, t'') dt''} \mathbf{f}(\mathbf{s}_{\text{dep}}, t', t) dt' \right] \cdot \epsilon_i^{(1)} \right. \\
 &\quad \left. + \left(e^{\mathbf{A}^T t} (\mathbf{s}_{\text{dep}} - \mathbf{1}) + \mathbf{1} \right) \cdot \epsilon_i^{(2)} \right)^{x_i^0} \tag{61}
 \end{aligned}$$

$$\begin{aligned}
 &= \prod_{i=1}^n \left(e^{-\int_0^t \mathbf{J}^T(\mathbf{s}_{\text{dep}}(t')) dt'} \epsilon_i^{(1)} \cdot \mathbf{s}_{\text{ind}} - \int_0^t e^{-\int_0^{t'} \mathbf{J}^T(\mathbf{s}_{\text{dep}}(t'')) dt''} \epsilon_i^{(1)} \cdot \mathbf{f}(\mathbf{s}_2(t')) dt' \right. \\
 &\quad \left. + \left(\exp(\mathbf{A}t) \epsilon_i^{(2)} \cdot (\mathbf{s}_{\text{dep}} - \mathbf{1}) + \mathbf{1} \right) \right)^{x_i^0}. \tag{62}
 \end{aligned}$$

□

A.4 Solution of characteristic ODE system for decay chains

The ODE system, describing the temporal evolution of the concentration of isotope x_i , is transposed to the characteristic system (43–44) of the PDE (41). More precisely, the system for x_i is identical to the original one from Bateman (1910), except for the factors²⁰ representing the decay particles, $s_\alpha^{R_{\alpha,i}} s_\beta^{R_{\beta,i}} s_\gamma^{R_{\gamma,i}}$:

$$\frac{dx_1}{dt} = -k_1 x_1 \tag{63}$$

$$\frac{dx_2}{dt} = k_1 s_\alpha^{R_{\alpha,1}} s_\beta^{R_{\beta,1}} s_\gamma^{R_{\gamma,1}} x_1 - k_2 x_2$$

⋮

$$\frac{dx_n}{dt} = k_{n-1} s_\alpha^{R_{\alpha,n-1}} s_\beta^{R_{\beta,n-1}} s_\gamma^{R_{\gamma,n-1}} x_{n-1} - k_n x_n. \tag{64}$$

Due to this similarity, we are able to adapt the calculations from Pressyanov (2002) almost one-to-one. We take the Laplace transform²¹ on both sides of Eqs. (63) and (64), thereby introducing the complex variable σ , setting $x_i^0 = 0$ for $i \geq 2$, and obtain

²⁰ In the following, we suppress the dependence of \mathbf{x} on $s_\alpha^{R_{\alpha,i}} s_\beta^{R_{\beta,i}} s_\gamma^{R_{\gamma,i}}$ for better readability.

²¹ See Brannan (2010) for an introduction.

$$\begin{aligned}
 \int_0^\infty e^{-\sigma t} \frac{dx_1}{dt} dt &= \int_0^\infty e^{-\sigma t} dx_1 = e^{-\sigma t} x_1(t) \Big|_0^\infty + \sigma \int_0^\infty e^{-\sigma t} x_1(t) dt \\
 &= -x_1^0 + \sigma \tilde{x}_1(\sigma) \\
 &= -k_1 \tilde{x}_1(\sigma), \tag{65} \\
 \int_0^\infty e^{-\sigma t} dx_2 &= \sigma \tilde{x}_2(\sigma) \\
 &= k_1 s_\alpha^{R_{\alpha,1}} s_\beta^{R_{\beta,1}} s_\gamma^{R_{\gamma,1}} \tilde{x}_1(\sigma) - k_2 \tilde{x}_2(\sigma) \\
 &\vdots \\
 \int_0^\infty e^{-\sigma t} \frac{dx_n}{dt} dt &= \sigma \tilde{x}_n(\sigma) \\
 &= k_{n-1} s_\alpha^{R_{\alpha,n-1}} s_\beta^{R_{\beta,n-1}} s_\gamma^{R_{\gamma,n-1}} \tilde{x}_{n-1}(\sigma) - k_n \tilde{x}_n(\sigma). \tag{66}
 \end{aligned}$$

In Laplace domain, we solve a linear equation system and invert the Laplace transform $\tilde{\mathbf{x}}(\sigma)$ to obtain $\mathbf{x}(t)$. The solution of Eqs. (65–66) is given by

$$\begin{aligned}
 \tilde{x}_1(\sigma) &= \frac{x_1^0}{k_1 + \sigma}, \\
 \tilde{x}_2(\sigma) &= \frac{k_1 s_\alpha^{R_{\alpha,1}} s_\beta^{R_{\beta,1}} s_\gamma^{R_{\gamma,1}}}{k_2 + \sigma} \tilde{x}_1(\sigma), \\
 &\vdots \\
 \tilde{x}_n(\sigma) &= \frac{k_{n-1} s_\alpha^{R_{\alpha,n-1}} s_\beta^{R_{\beta,n-1}} s_\gamma^{R_{\gamma,n-1}}}{k_n + \sigma} \tilde{x}_{n-1}(\sigma). \tag{67}
 \end{aligned}$$

The last Eq. (67) represents the starting point of a linear recurrence, which, if the initial conditions x_2, \dots, x_n are set to zero,²² can be solved by

$$\tilde{x}_n(\sigma) = \frac{\prod_{i=1}^{n-1} k_i s_\alpha^{R_{\alpha,i}} s_\beta^{R_{\beta,i}} s_\gamma^{R_{\gamma,i}}}{\prod_{i=1}^n (k_i + \sigma)} x_1^0.$$

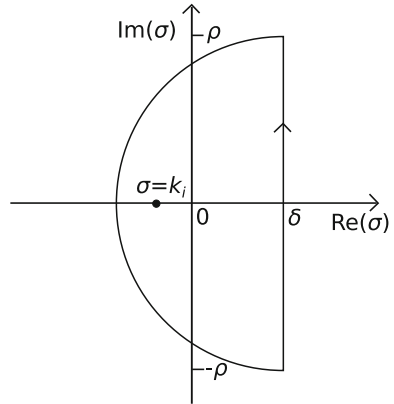
Now, we go back from Laplace domain to the time domain, by performing the Bromwich integral

$$x_n(t) = \frac{1}{2\pi i} \int_{\delta-i\infty}^{\delta+i\infty} e^{\sigma t} \tilde{x}_n(\sigma) d\sigma.$$

Here, δ defines an integration path parallel to the complex line that is situated right of all singularities of $\tilde{x}_n(\sigma)$. We may choose $\delta = 0$, since all poles are on the negative part

²² We will shortly investigate general initial conditions.

Fig. 11 Integration path used for the solution of the Bromwich integral



of the real axis. By closing the integration path through the addition of a semicircle (see Fig. 11), we obtain an integral that can be solved by applying the residue theorem, since the contour is closed²³:

$$\frac{1}{2\pi i} \oint_{\mathcal{C}} e^{\sigma t} \tilde{x}_n(\sigma) d\sigma = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{\sigma t} \tilde{x}_n(\sigma) d\sigma + \lim_{\rho \rightarrow \infty} \frac{1}{2\pi i} \int_{\pi/2}^{3\pi/2} e^{\rho e^{i\phi} t} \frac{\prod_{i=1}^{n-1} k_i S_{\alpha}^{R_{\alpha,i}} S_{\beta}^{R_{\beta,i}} S_{\gamma}^{R_{\gamma,i}}}{\prod_{i=1}^n (k_i + \rho e^{i\phi})} x_1^0 i \rho e^{i\phi} d\phi.$$

Here, the contribution of the circular part to the value of the integral along \mathcal{C} is zero.²⁴ we may calculate $x_n(t)$ by the residue formula

$$x_n(t) = \frac{1}{2\pi i} \oint_{\mathcal{C}} e^{\sigma t} \tilde{x}_n(\sigma) d\sigma = \sum_{j=1}^n \text{Res}_j(e^{\sigma t} \tilde{x}_n(\sigma)),$$

where the sum is over all poles. In our case, we have n simple poles and so the residue theorem yields

$$\begin{aligned} \text{Res}_j(e^{\sigma t} \tilde{x}_n(\sigma)) &= \lim_{\sigma \rightarrow -k_j} ((\sigma + k_j) e^{\sigma t} \tilde{x}_n(\sigma)) \\ &= e^{-k_j t} \frac{\prod_{i=1}^{n-1} k_i S_{\alpha}^{R_{\alpha,i}} S_{\beta}^{R_{\beta,i}} S_{\gamma}^{R_{\gamma,i}}}{\prod_{i=1, i \neq j}^n (k_i - k_j)} x_1^0. \end{aligned}$$

²³ See Boas (2006) for a review of the solution methods for this kind of integral.

²⁴ To see this, apply the integral triangular inequality.

Summing over all residues, we get

$$x_n(t) = x_1^0 \left(\prod_{i=1}^{n-1} k_i s_\alpha^{R_{\alpha,i}} s_\beta^{R_{\beta,i}} s_\gamma^{R_{\gamma,i}} \right) \sum_{j=1}^n \frac{e^{-k_j t}}{\prod_{i=1, i \neq j}^n (k_i - k_j)}.$$

The only thing left to do is to generalize the initial conditions for x_2, \dots, x_{n-1} . If we set $x_m \neq 0$, we obtain for example

$$\begin{aligned} x_i(t) &= x_1^0 \left(\prod_{l=1}^{i-1} k_l s_\alpha^{R_{\alpha,l}} s_\beta^{R_{\beta,l}} s_\gamma^{R_{\gamma,l}} \right) \sum_{l=1}^i \frac{e^{-k_l t}}{\prod_{j=1, j \neq l}^i (k_j - k_l)} \\ &+ x_m^0 \left(\prod_{l=m}^{i-1} k_l s_\alpha^{R_{\alpha,l}} s_\beta^{R_{\beta,l}} s_\gamma^{R_{\gamma,l}} \right) \sum_{l=m}^i \frac{e^{-k_l t}}{\prod_{j=m, j \neq l}^i (k_j - k_l)}. \end{aligned}$$

Applying this idea to all x_i^0 , we conclude the solution of the rate equations with

$$\begin{aligned} x_i(t) &= x_i^0 e^{-k_i t} + \\ &\sum_{m=1}^{i-1} x_m^0 \left(\prod_{l=m}^{i-1} k_l s_\alpha^{R_{\alpha,l}} s_\beta^{R_{\beta,l}} s_\gamma^{R_{\gamma,l}} \right) \sum_{l=m}^i \frac{e^{-k_l t}}{\prod_{j=m, j \neq l}^i (k_j - k_l)}. \end{aligned}$$

References

- Anderson DF, Craciun G, Kurtz TG (2010) Product-form stationary distributions for deficiency zero chemical reaction networks. *Bull Math Biol* 72(8):1947–1970
- Andreychenko A, Bortolussi L, Grima R, Thomas P, Wolf V (2017) Distribution approximations for the chemical master equation: comparison of the method of moments and the system size expansion. In: *Modeling cellular systems*. Springer, Cham, pp 39–66
- Bateman H (1910) The solution of a system of differential equations occurring in the theory of radioactive transformations. *Math Proc Camb Philos Soc* 15(V):423–427
- Boas ML (2006) *Mathematical methods in the physical sciences*. Wiley, Hoboken
- Brannan JR (2010) *Differential equations with boundary value problems: an introduction to modern methods and applications*. Wiley, New York
- Constantino PH, Vlysidis M, Smadbeck P, Kaznessis YN (2016) Modeling stochasticity in biochemical reaction networks. *J Phys D Appl Phys* 49(9):093001
- Deufhard P, Hohmann A (2003) *Numerical analysis in modern scientific computing*. In: *Texts in applied mathematics*, vol 43. Springer, New York
- Evans LC (2010) *Partial differential equations*. American Mathematical Society, Providence
- Friedman N, Cai L, Xie X (2006) Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys Rev Lett* 97(16):1–4
- Gans PJ (1960) Open first-order stochastic processes. *J Chem Phys* 33(3):691
- Gardiner C (2009) *Stochastic methods*. Springer, Berlin
- Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81(25):2340–2361
- Gillespie DT (1992) A rigorous derivation of the chemical master equation. *Physica A* 188(1–3):404–425
- Gurland J (1957) Some interrelations among compound and generalized distributions. *Biometrika* 44(1–2):265–268
- Jahnke T, Huisinga W (2007) Solving the chemical master equation for monomolecular reaction systems analytically. *J Math Biol* 54(1):1–26

- Johnson NL, Kemp AW, Kotz S (2005) Univariate discrete distributions, 3rd edn. Wiley, Hoboken
- Kampen VNG (1976) The equilibrium distribution of a chemical mixture. *Phys Lett A* 59(5):333–334
- Kawamura K (1979) The structure of multivariate Poisson distribution. *Kodai Math J* 2(3):337–345
- Keilson J, Gerber H (1971) Some results for discrete unimodality. *J Am Stat Assoc* 66(334):386
- Khatri CG, Patel IR (1961) Three classes of univariate discrete distributions. *Biometrics* 17(4):567
- Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H, Herwig R (2009) *Systems biology*. Wiley-VCH, Weinheim
- Lotka AJ (1925) *Elements of physical biology*. Dover Publications, New York
- Maarleveld TR, Olivier BG, Bruggeman FJ (2013) StochPy: a comprehensive, user-friendly tool for simulating stochastic biological processes. *PLoS ONE* 8(11):e79345
- Magnus W (1954) On the exponential solution of differential equations for a linear operator. *Commun Pure Appl Math* 7(4):649–673
- Marcheselli M, Baccini A, Barabesi L (2008) Parameter estimation for the discrete stable family. *Commun Stat Theory Methods* 37(6):815–830
- Milner P, Gillespie CS, Wilkinson DJ (2012) Moment closure based parameter inference of stochastic kinetic models. *Stat Comput* 23(2):287–295
- Munsky B, Khammash M (2006) The finite state projection algorithm for the solution of the chemical master equation. *J Chem Phys* 124(4):044104
- Neyman J (1939) On a new class of “contagious” distributions, applicable in entomology and bacteriology. *Ann Math Stat* 10(1):35–57
- Pemantle R, Wilson MC (2013) *Analytic combinatorics in several variables*. Cambridge University Press, Cambridge
- Pressyanov DS (2002) Short solution of the radioactive decay chain equations. *Am J Phys* 70(4):444
- Rowell D (2004) Computing the matrix exponential the Cayley–Hamilton method. In: *Lecture notes on advanced system dynamics and control*. Technical report, Massachusetts Institute of Technology
- Shahrezaei V, Swain PS (2008) Analytical distributions for stochastic gene expression. *Proc Natl Acad Sci USA* 105(45):17256–61
- Thomas P, Popović N, Grima R (2014) Phenotypic switching in gene regulatory networks. *Proc Nat Acad Sci USA* 111(19):6994–6999
- Trcek T, Larson DR, Moldón A, Query CC, Singer RH (2011) Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast. *Cell* 147(7):1484–97
- Trcek T, Chao JA, Larson DR, Park HY, Zenklusen D, Shenoy SM, Singer RH (2012) Single-mRNA counting using fluorescent in situ hybridization in budding yeast. *Nat Protoc* 7(2):408–19
- van Kampen NG (2011) *Stochastic processes in physics and chemistry*. North-Holland Personal Library, Amsterdam, London, New York
- Xu J, Minin VN (2015) Efficient transition probability computation for continuous-time branching processes via compressed sensing. [arXiv:1503.02644](https://arxiv.org/abs/1503.02644)
- Zhang H, Liu Y, Li B (2014) Notes on discrete compound Poisson model with applications to risk theory. *Insur Math Econ* 59:325–336
- Zheng Q (1999) Progress of a half century in the study of the Luria–Delbrück distribution. *Math Biosci* 162(1):1–32