



An Analytical Framework for Studying Small-Number Effects in Catalytic Reaction Networks: A Probability Generating Function Approach to Chemical Master Equations

Masaki Nakagawa^{1*} and Yuichi Togashi^{1,2*}

¹ Department of Mathematical and Life Sciences, Graduate School of Science, Hiroshima University, Higashi-Hiroshima, Japan, ² Research Center for the Mathematics on Chromatin Live Dynamics, Hiroshima University, Higashi-Hiroshima, Japan

OPEN ACCESS

Edited by:

Xiaogang Wu,
Institute for Systems Biology, USA

Reviewed by:

Tianshou Zhou,
Sun Yat-Sen University, China
Zuxi Wang,
Huazhong University of Science and
Technology, China
Taichi Haruna,
Kobe University, Japan

*Correspondence:

Masaki Nakagawa
masak1@hiroshima-u.ac.jp;
Yuichi Togashi
togashi@hiroshima-u.ac.jp

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Physiology

Received: 16 December 2015

Accepted: 23 February 2016

Published: 24 March 2016

Citation:

Nakagawa M and Togashi Y (2016) An Analytical Framework for Studying Small-Number Effects in Catalytic Reaction Networks: A Probability Generating Function Approach to Chemical Master Equations. *Front. Physiol.* 7:89. doi: 10.3389/fphys.2016.00089

Cell activities primarily depend on chemical reactions, especially those mediated by enzymes, and this has led to these activities being modeled as catalytic reaction networks. Although deterministic ordinary differential equations of concentrations (rate equations) have been widely used for modeling purposes in the field of systems biology, it has been pointed out that these catalytic reaction networks may behave in a way that is qualitatively different from such deterministic representation when the number of molecules for certain chemical species in the system is small. Apart from this, representing these phenomena by simple binary (on/off) systems that omit the quantities would also not be feasible. As recent experiments have revealed the existence of rare chemical species in cells, the importance of being able to model potential small-number phenomena is being recognized. However, most preceding studies were based on numerical simulations, and theoretical frameworks to analyze these phenomena have not been sufficiently developed. Motivated by the small-number issue, this work aimed to develop an analytical framework for the chemical master equation describing the distributional behavior of catalytic reaction networks. For simplicity, we considered networks consisting of two-body catalytic reactions. We used the probability generating function method to obtain the steady-state solutions of the chemical master equation without specifying the parameters. We obtained the time evolution equations of the first- and second-order moments of concentrations, and the steady-state analytical solution of the chemical master equation under certain conditions. These results led to the rank conservation law, the connecting state to the winner-takes-all state, and analysis of 2-molecules M -species systems. A possible interpretation of the theoretical conclusion for actual biochemical pathways is also discussed.

Keywords: small-number issues, catalytic reaction networks, chemical master equations, probability generating functions, systems biology, mathematical biology, population dynamics

1. INTRODUCTION

Biochemical systems consist of a variety of chemicals, including proteins, nucleic acids, and also small metabolites. Enzymatic reactions, which play an important role in catalyzing many biological reactions, are particularly important to maintain the structure and activity of these systems. Hence, biochemical systems are often modeled as catalytic reaction networks.

These networks are typically analyzed by using deterministic ordinary differential equations with respect to concentrations of chemical species, so-called reaction rate equations [or partial differential equations (reaction-diffusion equations) for spatially distributed, non-uniform cases]; i.e., the concentrations are represented by continuous variables. However, because each chemical actually consists of molecules, the concentration of each species should be a discrete variable. The effects of such discreteness as well as finite-size fluctuations in stochastic reactions become non-negligible if the number of molecules in the system is small. In theory, situations such as these can result in phenomena that cannot be described by rate equations (as well as those equations with additional noise; Togashi and Kaneko, 2001, 2007; Awazu and Kaneko, 2007, 2009).

In contrast, gene regulations are often modeled as a combination of binary (on/off) switches, typically represented by Boolean network models (Kauffman, 1969). However, this approach does not consider the quantities of chemicals such as DNA, mRNA, and proteins. Even for the seemingly digital expression of genes, a stoichiometry involving DNA cannot be ignored, as seen in X-chromosome inactivation (2 to 1) and the trisomy syndrome (2 to 3). Thus, the use of a binary (on/off) representation would also be inappropriate. We therefore need to consider the number of molecules.

Recent experiments have shown the existence in the cell of proteins that only consist of a few molecules each (Taniguchi et al., 2010), and these potential biological phenomena (sometimes referred to as the *small-number issue*) have been gradually recognized by biologists. General theories that would enable predictions in small number situations would be helpful to every biological scientist seeking to understand biochemical processes on any level. Of course, the distributional behavior of such discrete stochastic systems is described by the chemical master equation. However, it is generally difficult to obtain its solution; hence, most efforts have been devoted to the development of approximation and simulation methods and their application (Gillespie, 1976, 1977; Munsky and Khammash, 2006; Lee et al., 2009; Kim and Lee, 2012).

The effects of small numbers in particular systems, e.g., small autocatalytic systems have been mathematically analyzed (Ohkubo et al., 2008; Biancalani et al., 2012, 2014; Houchmandzadeh and Vallade, 2015; Saito and Kaneko, 2015). In addition, a parameter representing the degree of discreteness in the number of molecules has been introduced (Haruna, 2010, 2015). In this work, we pursue analytical frameworks for studying the effects of small numbers in catalytic reaction networks by following an approach that is as general as possible. Our aim is to obtain the steady-state analytical solution for the chemical master equation without specifying parameters, rather

than developing numerical schemes to solve it as was previously done (Kim and Lee, 2012). Furthermore, we try to theoretically describe the long-term behavior of the system by only using information about relationships between elements, thereby implying that we aim to produce results that could be applied to studies in any field.

Our framework provides good operability because our formulas have a specific and satisfyingly simple form, and enables us to obtain the steady state for a wide class of catalytic reaction networks because our framework never specifies any parameters for the networks. We use a probability generating function approach. The probability generating function approach to stochastic chemical kinetics itself has been proposed a long time ago (e.g., Krieger and Gans, 1960; McQuarrie, 1967), after which it spread to biological stochastic kinetics studies (e.g., stochastic gene expression Thattai and van Oudenaarden, 2001; Shahrezaei and Swain, 2008); thus, physicists as well as chemists and mathematical biologists are familiar with the approach. Our main contributions in this paper relate to the efficient usage of the probability generating function (as in Gadgil et al., 2005 regarding first-order reaction networks). Therefore, our approach for obtaining the steady state is understandably easier than that followed in a previous study (Anderson et al., 2010) while, as a consequence, our results are consistent with that study (see Theorem 4.1 and 4.2 in Anderson et al., 2010). Furthermore, since our method uses a procedure based on analytical calculations, it can be easily converted to a computer algorithm.

The present paper is organized as follows. In Section 2.1, we define the catalytic reaction network considered in this paper. The chemical master equation (CME) is provided in Section 2.2. We introduce the probability generating function (PGF) and derive the generating function equation (GFE) in Section 2.3. In subsequent Sections (2.3.1–2.3.3), we show that the GFE introduces the time evolution equation of the first-order and second-order moments of concentrations (we refer to the first-order moment time evolution equation as the pre-rate equation, PRE), and the second-order moment expression of time-averaged concentrations (SME). Section 2.4 is devoted to obtaining the steady-state solutions of the GFE. To simplify the GFE, we neglect the non-catalytic reactions considered as perturbation for the catalytic reaction network if the system is “entirely ergodic.” In Section 2.4.2.3 as the main result, we obtain the probability generating function without winner-takes-all states (PGFwoWTAS) including the solutions of the corresponding rate equation. In Section 3, we describe applications of these results: the rank conservation law, the connecting state to the winner-takes-all state, analysis of 2-molecules M -species systems, and non-autocatalyzation of autocatalytic reaction networks. The prospects of our theory in terms of the small-number issue are briefly discussed in Section 4.

2. METHODS AND RESULTS

2.1. Catalytic Reaction Networks

Consider an abstract catalytic reaction network consisting of M chemical species and N molecules in a well-stirred reactor

of volume V , as in Awazu and Kaneko (2007). Each species is labeled by each of the integers between 1 and M . The total number of molecules N is always conserved in reaction processes. This chemical reaction system involves both catalytic reactions and non-catalytic reactions, to prevent catalytic reactions from stopping:

(i) Catalytic reactions (two-body catalysis):



where the species $i, j, k \in [1, M]$ represent a substrate, a catalyst, and a product. The reaction rate constant is represented by $R_{ijk} > 0$. If this catalytic reaction does not exist, we specify $R_{ijk} = 0$. Therefore, the catalytic reaction networks are determined by R_{ijk} . In this paper, we impose the following conditions for the catalytic reaction networks;

- (a) $R_{iik} = 0$; Substrate \neq Catalyst.
- (b) $R_{iji} = 0$; Substrate \neq Product.
- (c) $R_{ikk} = 0$; Autocatalytic reactions are not included.
- (d) $\# \{k : R_{ijk} > 0 (\forall i, j)\} = 1$;
One product against a substrate and a catalyst.

(ii) Non-catalytic reactions (one-body reactions):



This reaction exists for all combinations between each species, but a product $j \in [1, M]$ is uniformly-randomly selected from all species 1 to M . The rate constant is set $\varepsilon > 0$ in common, where ε is very small, i.e., $\varepsilon \ll \min\{R_{ijk} > 0\}$.

The state of this catalytic reaction network is specified by the combination of M natural numbers $\mathbf{n} = (n_1, n_2, \dots, n_M)$, where $n_i \in [0, N]$ is the number of molecules of the i th-species. For later convenience, we introduce the following notations: the state space of the catalytic reaction network $W_{M,N}$ (abbr. W) is represented by

$$W := \{\mathbf{n} \in [0, N]^M \mid n_1 + \dots + n_M = N\}, \tag{3}$$

which consists of $\binom{M}{N} := \frac{(N+M-1)!}{(M-1)!N!}$ points; a collection consisting of (one species) winner-takes-all states $I_{M,N}$ (abbr. I) is represented by

$$I := \{\mathbf{n} \in W_{M,N} \mid \exists i \in [1, M] \text{ s.t. } n_i = N\}, \tag{4}$$

which consists of M points. Of course, the winners-take-all states of more than one species can be considered, but we focus on the winner-takes-all states of one species by supposing the system satisfies a certain condition, *entire ergodicity* (see Section 2.4.2.4).

In the present paper, we are interested in the N -dependence of the concentration of each species $x_i = n_i/V$. We basically consider a situation in which the total density of molecules $\rho = N/V$ is conserved, even if the total number of molecules N is changed.

2.2. The Chemical Master Equation

The rate constant R_{ijk} in the catalytic reaction is defined as the number of reactions per unit volume, unit concentration, and unit time. Therefore, the number of reactions per unit time in the catalytic reaction $i \xrightarrow{j} k$, such that the concentrations are x_i and x_j , is

$$R_{ijk}x_i x_j V = R_{ijk} \frac{n_i n_j}{V} = R_{ijk} \frac{n_i n_j}{N} \rho, \tag{5}$$

where $\rho = N/V$ is the total density of molecules in the vessel. On the other hand, the number of reactions per unit time in the non-catalytic reaction $i \rightarrow j$ with probability $1/M$ is

$$\varepsilon x_i V \times \frac{1}{M} = \frac{\varepsilon}{M} n_i. \tag{6}$$

Then, the time-evolution of the probability $P(\mathbf{n}, t)$, with which the system is in the state \mathbf{n} at time t , obeys the chemical master equation (CME):

$$\begin{aligned} \frac{dP(\mathbf{n}, t)}{dt} &= \frac{\rho}{N} \sum_{i,j,k} R_{ijk} (E_i^{+1} E_k^{-1} - 1) n_i n_j P(\mathbf{n}, t) \\ &+ \frac{\varepsilon}{M} \sum_{i,j} (E_i^{+1} E_j^{-1} - 1) n_i P(\mathbf{n}, t), \end{aligned} \tag{7}$$

where $E_i^{\pm m}$ are step operators, i.e.,

$$E_i^{\pm m} f(n_1, \dots, n_i, \dots, n_M) := f(n_1, \dots, n_i \pm m, \dots, n_M). \tag{8}$$

Of course, the state space on which the probability $P(\cdot, t)$ is supported, is the previously described $W_{M,N}$.

2.3. Probability Generating Function Method

The probability generating function (PGF) is useful to analyze the CME:

$$\phi(\mathbf{z}, t) := \sum_{n_1, \dots, n_M} P(\mathbf{n}, t) z_1^{n_1} \dots z_M^{n_M}. \tag{9}$$

Note that the following expressions are translated to differential forms of the PGF:

$$n_i P(\mathbf{n}, t) \mapsto z_i \frac{\partial}{\partial z_i} \phi(\mathbf{z}, t), \tag{10a}$$

$$n_i n_j P(\mathbf{n}, t) \mapsto z_i z_j \frac{\partial}{\partial z_i} \frac{\partial}{\partial z_j} \phi(\mathbf{z}, t) \quad (i \neq j), \tag{10b}$$

$$E_i^{+1} E_j^{-1} n_i P(\mathbf{n}, t) \mapsto z_j \frac{\partial}{\partial z_i} \phi(\mathbf{z}, t) \quad (i \neq j), \tag{10c}$$

$$E_i^{+1} E_k^{-1} n_i n_j P(\mathbf{n}, t) \mapsto z_j z_k \frac{\partial}{\partial z_i} \frac{\partial}{\partial z_j} \phi(\mathbf{z}, t) \quad (i \neq j \neq k \neq i). \tag{10d}$$

Therefore, rewriting the CME to the equation governing the PGF enables the generating function equation (GFE) to be obtained:

$$\begin{aligned} \frac{\partial \phi(\mathbf{z}, t)}{\partial t} &= \frac{\rho}{N} \sum_{i,j,k} R_{ijk} (z_k - z_i) z_j \frac{\partial}{\partial z_i} \frac{\partial}{\partial z_j} \phi(\mathbf{z}, t) \\ &+ \frac{\varepsilon}{M} \sum_{i,j} (z_i - z_j) \frac{\partial}{\partial z_j} \phi(\mathbf{z}, t). \end{aligned} \quad (11)$$

The GFE consists of continuous variables, unlike the CME, which consists of discrete variables.

Once we obtain the PGF ϕ as a solution to the GFE, we can derive all statistics of the catalytic reaction network; for example, the ensemble averages (first-order moments) and second-order moments become

$$\langle n_i \rangle(t) = \left. \frac{\partial}{\partial z_i} \phi(\mathbf{z}, t) \right|_{\mathbf{z}=\mathbf{1}}, \quad (12a)$$

$$\langle n_i n_j \rangle(t) = \left. \frac{\partial^2}{\partial z_i \partial z_j} \phi(\mathbf{z}, t) \right|_{\mathbf{z}=\mathbf{1}} \quad (i \neq j), \quad (12b)$$

$$\langle n_l (n_l - 1) \rangle(t) = \left. \frac{\partial^2}{\partial z_l^2} \phi(\mathbf{z}, t) \right|_{\mathbf{z}=\mathbf{1}}, \quad (12c)$$

and the marginal distributions become

$$p_i(n, t) = \left. \frac{1}{n!} \frac{\partial^n}{\partial z_i^n} \phi(\mathbf{z}, t) \right|_{z_i=0; z_j=1(j \neq i)} \quad (0 \leq n \leq N), \quad (13)$$

where $p_i(n, t)$ is defined by

$$p_i(n, t) := \sum_{\substack{n_1, \dots, n_M \\ (\text{except } n_i)}} P(\mathbf{n}, t) \Big|_{n_i=n}. \quad (14)$$

2.3.1. The Pre-rate Equation

Rate equations are differential equations for the concentrations x_i of chemical species $i = 1, \dots, M$ when $N \rightarrow \infty$. We use the GFE to derive a formula, which is reminiscent of the rate equation.

Differentiating both sides of Equation (11) by z_i , substituting $\mathbf{z} = \mathbf{1}$, and writing $x_i = n_i/V$, leads to the following pre-rate equation (PRE);

$$\frac{d}{dt} \langle x_i \rangle = \sum_{j,k} (R_{jki} \langle x_j x_k \rangle - R_{ikj} \langle x_i x_k \rangle) + \varepsilon \left(\frac{\rho}{M} - \langle x_i \rangle \right). \quad (15)$$

Note that, as we are considering a system of which the total number of molecules N is conserved, the concentration conservation law (CCL) must be satisfied;

$$\sum_i \langle x_i \rangle(t) = \rho \quad (\forall t), \quad (16)$$

where the PRE (Equation 15) is consistent with the CCL (Equation 16).

If the independence $\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle$ ($i \neq j$) holds, the PRE (Equation 15) becomes the same expression as the rate equation. On the other hand, the PRE does not explicitly include extensive variables, such as the total number of molecules

N . Therefore, the formula (Equation 15) always holds for an arbitrary molecular number N unless the total molecular density ρ is changed.

2.3.2. Second-Order Moment Expression for Time-Averaged Concentrations

We suppose the following ergodicity to replace ensemble-averages with time-averages; $\bar{n}_i = \langle n_i \rangle_*$ i.e.,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T n_i(t) dt = \sum_{n_1, \dots, n_M} n_i P_*(\mathbf{n}), \quad (17)$$

where $P_*(\mathbf{n})$ is the steady-state solution of the CME corresponding to almost all initial conditions. Considering the steady state in the PRE [that is, taking $t \rightarrow \infty$ on both sides of Equation (15)], one obtains the following second-order moment expression (SME) for time-averaged concentrations:

$$\bar{x}_i = \frac{\rho}{M} + \frac{1}{\varepsilon} \sum_{j,k} (R_{jki} \bar{x}_j \bar{x}_k - R_{ikj} \bar{x}_i \bar{x}_k). \quad (18)$$

The second term on the right hand side of Equation (18) represents the difference from the uniform concentration ρ/M .

If the independence $\bar{x}_i \bar{x}_j = \bar{x}_i \bar{x}_j$ holds, Equation (18) becomes an equation that only includes the concentrations \bar{x}_i . Therefore, in combination with the time-averaged version of the CCL (Equation 16), the concentrations \bar{x}_i can be determined without including unknown quantities. However, the actual concentrations \bar{x}_i depend on their second-order moments $\bar{x}_i \bar{x}_j$.

2.3.3. Time-Evolution of Second-Order Moments

Determining the concentrations \bar{x}_i from the SME (Equation 18) requires us to know their second-order moments $\bar{x}_i \bar{x}_j$.

Differentiating both sides of the GFE (Equation 11) by z_l and z_m ($l \neq m$), substituting $\mathbf{z} = \mathbf{1}$, and writing $x_i = n_i/V$, then, after simplification, the following time-evolution equation of second-order moments (TESM) is obtained:

$$\begin{aligned} \frac{d}{dt} \langle x_l x_m \rangle &= -\frac{\rho}{N} \sum_i (R_{lim} \langle x_i x_l \rangle + R_{mil} \langle x_i x_m \rangle) \\ &+ \sum_{i,j} \{ R_{ijl} \langle x_i x_j x_m \rangle + R_{ijm} \langle x_i x_j x_l \rangle - (R_{lij} + R_{mij}) \\ &\quad \langle x_i x_l x_m \rangle \} \\ &+ \frac{\varepsilon}{M} \left\{ \frac{N-1}{N} \rho (\langle x_l \rangle + \langle x_m \rangle) - 2M \langle x_l x_m \rangle \right\} \\ &\quad (1 \leq l < m \leq M). \end{aligned} \quad (19a)$$

On the other hand, twice differentiating both sides of the GFE (Equation 11) by z_l , substituting $\mathbf{z} = \mathbf{1}$, and

writing $x_i = n_i/V$, then after some simplification, the following alternative TESM is obtained:

$$\begin{aligned} \frac{d}{dt} \langle x_l(x_l - \frac{\rho}{N}) \rangle &= 2 \sum_{i,j} (R_{ijl} \langle x_i x_j x_l \rangle - R_{lij} \langle x_i x_l(x_l - \frac{\rho}{N}) \rangle) \\ &+ \frac{2\varepsilon}{M} \left\{ \frac{N+M-1}{N} \rho \langle x_l \rangle - M \langle x_l^2 \rangle \right\} \\ &(1 \leq l \leq M). \end{aligned} \tag{19b}$$

The expression of TESMs (Equation 19) consists of $M(M+1)/2$ equations containing the third-order moments $\langle x_i x_j x_k \rangle(t)$. Therefore, the TESMs are not effective unless the systems are restricted such that the third-order moments vanish (e.g., 2-molecules systems). If $N = 2$, because $\partial_{z_i} \partial_{z_j} \partial_{z_k} \phi = 0$ (ϕ is a second-order polynomial), one can calculate the 2-molecules version of the TESMs (2mTESMs);

$$\begin{aligned} \frac{d}{dt} \langle x_l x_m \rangle &= \frac{\rho}{2} \sum_i \{ R_{iml} \langle x_i x_m \rangle + R_{ilm} \langle x_i x_l \rangle \\ &- (R_{lmi} + R_{mli}) \langle x_l x_m \rangle \} \\ &+ \frac{\rho\varepsilon}{2M} (\langle x_l \rangle + \langle x_m \rangle) - 2\varepsilon \langle x_l x_m \rangle \\ &(1 \leq l < m \leq M), \end{aligned} \tag{20a}$$

$$\frac{d}{dt} \langle x_l(x_l - \frac{\rho}{N}) \rangle = \frac{2\varepsilon}{M} \left(\frac{M+1}{2} \rho \langle x_l \rangle - M \langle x_l^2 \rangle \right) \quad (1 \leq l \leq M). \tag{20b}$$

Treatment of the 2mTESMs (Equation 20) to demonstrate their effectiveness appears in a later section.

2.4. Steady-State Solutions of GFE

If the GFE (Equation 11) can be solved, this would enable us to obtain all the statistics of the catalytic reaction networks. However, it is generally difficult to solve. Here, we focus on the steady-state solutions of the GFE and consider the case that the ε -term in the GFE can be ignored. Through the following discussion, we see that the approximation is effective only if the system is ergodic.

2.4.1. The Case of Non-catalytic Reactions Only

First, we consider the steady-state solutions of non-catalytic reactions only as an introduction. The PGF $\phi_*^{nc}(z)$, corresponding to the steady state in the case of non-catalytic reactions only, should satisfy the following equation:

$$0 = \sum_{i,j} (z_i - z_j) \frac{\partial}{\partial z_j} \phi_*^{nc}(z). \tag{21}$$

By exchanging the subscripts i and j in the second term, one obtains

$$0 = \sum_{i,j} z_i \left(\frac{\partial \phi_*^{nc}}{\partial z_j} - \frac{\partial \phi_*^{nc}}{\partial z_i} \right). \tag{22}$$

Because the coefficients of variables z_i must be zero, ϕ_*^{nc} must satisfy the following equations:

$$0 = \sum_j \left(\frac{\partial \phi_*^{nc}}{\partial z_j} - \frac{\partial \phi_*^{nc}}{\partial z_i} \right) \quad (\forall i), \tag{23}$$

equivalently,

$$\frac{\partial \phi_*^{nc}}{\partial z_i} = \frac{1}{M} \sum_{j=1}^M \frac{\partial \phi_*^{nc}}{\partial z_j} \quad (\forall i). \tag{24}$$

Equation (24) implies that all of $\partial_{z_i} \phi_*^{nc}$ ($i = 1, \dots, M$) are equal to each other, i.e.,

$$\frac{\partial \phi_*^{nc}}{\partial z_i} = \frac{\partial \phi_*^{nc}}{\partial z_j} \quad (1 \leq i < j \leq M). \tag{25}$$

Considering that the PGF $\phi_*^{nc}(z)$ is an N th order polynomial of z_i and must satisfy the condition $\phi_*^{nc}(1) = 1$ by definition, the following solution of Equation (25) can be found:

$$\phi_*^{nc}(z) = \left(\frac{z_1 + z_2 + \dots + z_M}{M} \right)^N. \tag{26}$$

Therefore, we obtain the following stationary distribution in the case of non-catalytic reactions only:

$$P_*^{nc}(\mathbf{n}) = \frac{1}{M^N} \binom{N}{n_1, n_2, \dots, n_M} \quad (\mathbf{n} \in W), \tag{27}$$

where $\binom{N}{n_1, n_2, \dots, n_M} := \frac{N!}{n_1! n_2! \dots n_M!}$ are multinomial coefficients. Furthermore, Equation (13) can also be used to derive the marginal distribution of the i -th species:

$$p_{i*}^{nc}(n) = \binom{N}{n} \left(\frac{1}{M} \right)^n \left(1 - \frac{1}{M} \right)^{N-n} \quad (0 \leq n \leq N), \tag{28}$$

which is the binomial distribution with parameters N and $1/M$.

If we suppose the ergodicity $\langle n_i \rangle = \bar{n}_i$, the following statistics can be calculated from Equation (12):

$$\bar{x}_i = \frac{\rho}{M} \quad (i = 1, \dots, M), \tag{29a}$$

$$\overline{x_i x_j} = \left(\frac{\rho}{M} \right)^2 \left(1 - \frac{1}{M} \right) \quad (i \neq j), \tag{29b}$$

$$\text{Var}[x_i] = \left(\frac{\rho}{M} \right)^2 \frac{M-1}{N} \quad (i = 1, \dots, M), \tag{29c}$$

where $\text{Var}[x_i] := \overline{x_i^2} - \bar{x}_i^2$ is the variance of the concentration $x_i = n_i/V$.

2.4.2. The Case of Catalytic Reactions Only

Next, we consider the steady-state solutions of catalytic reactions only, assuming that the ε -term in the GFE (Equation 11) can be ignored. The steady-state solutions are assumed to have a form similar to Equation (26), including undetermined coefficients (λ_i) deriving from the network structure (R_{ijk}).

2.4.2.1. A condition for finding the steady state: λ -condition

The PGF $\phi_*^c(\mathbf{z})$, corresponding to the steady state in the case of catalytic reactions only, should satisfy the following equation;

$$0 = \sum_{i,j,k} R_{ijk}(z_k - z_i)z_j \frac{\partial}{\partial z_i} \frac{\partial}{\partial z_j} \phi_*^c(\mathbf{z}). \quad (30)$$

Construction of the steady-state solution requires us to find a particular solution of Equation (30) as bases of linear space consisting of N -th order polynomials. Accordingly, let us assume the following extended form of Equation (26) by introducing parameters $\{\lambda_i\}_{i=1}^M$, which eventually correspond to the concentrations (per total density ρ) of each species in the continuous limit $N \rightarrow \infty$ as shown in Section 2.4.2.3:

$$\phi_*^c(\mathbf{z}) = (\lambda_1 z_1 + \lambda_2 z_2 + \dots + \lambda_M z_M)^N, \quad (31)$$

where

$$\sum_{i=1}^M \lambda_i = 1 \quad (0 \leq \lambda_1, \lambda_2, \dots, \lambda_M \leq 1), \quad (32)$$

because $\phi_*(1) = 1$. Substituting Equation (31) into Equation (30) and setting the coefficients of variables $z_i z_j$ as zero, gives the following condition for $\{\lambda_i \lambda_j\}$:

$$\sum_{k=1}^M \{R_{kij} \lambda_k \lambda_i + R_{kji} \lambda_k \lambda_j - (R_{ijk} + R_{jik}) \lambda_i \lambda_j\} = 0 \quad (1 \leq i < j \leq M). \quad (33)$$

The λ -condition (Equation 33) represents $M(M - 1)/2$ homogeneous equations for $\lambda_i \lambda_j$; therefore, λ_i can be calculated by combining the λ -condition (Equation 33) with Equation (32). The λ -condition has trivial solutions:

$$(\lambda_1, \dots, \lambda_i, \dots, \lambda_M) = (0, \dots, 0, 1, 0, \dots, 0) \quad \exists i \in [1, M]. \quad (34)$$

These solutions represent the states for which the i -th species take all molecules (the *winner-takes-all state*). On the other hand, there is also a non-trivial solution. In the following paragraph, we treat three-species systems ($M = 3$) to demonstrate that non-trivial solutions do exist. The following procedures can easily be extended to those for arbitrary M species systems.

2.4.2.2. Demonstration for non-trivial solutions of the λ -condition

The λ -condition (Equation 33) can be rewritten in matrix form; i.e., in the case of $M = 3$ for example,

$$\begin{bmatrix} -R_{123} - R_{213} & R_{312} & R_{321} \\ R_{213} & -R_{132} - R_{312} & R_{231} \\ R_{123} & R_{132} & -R_{231} - R_{321} \end{bmatrix} \begin{bmatrix} \lambda_1 \lambda_2 \\ \lambda_1 \lambda_3 \\ \lambda_2 \lambda_3 \end{bmatrix} = 0. \quad (35)$$

The 3×3 matrix (A) on the left-hand side can be rewritten in terms of its column vectors:

$$A = [-R_{123} \mathbf{e}_1 - R_{213} \mathbf{e}_2, R_{312} \mathbf{e}_2 - R_{132} \mathbf{e}_3, R_{321} \mathbf{e}_1 + R_{231} \mathbf{e}_3], \quad (36)$$

where $\mathbf{e}_1^T = (1, 0, -1)$, $\mathbf{e}_2^T = (1, -1, 0)$, and $\mathbf{e}_3^T = (0, 1, -1)$. According to the determinant property, i.e., $\det[\mathbf{a}, \mathbf{b}, \mathbf{c}] = -\det[\mathbf{b}, \mathbf{a}, \mathbf{c}]$, one has

$$\det(A) = 0. \quad (37)$$

Therefore, the non-trivial solution of Equation (35), i.e., the eigenvector corresponding to the eigenvalue 0 of the matrix A , certainly exists and it has the following expression:

$$\begin{bmatrix} \lambda_1 \lambda_2 \\ \lambda_1 \lambda_3 \\ \lambda_2 \lambda_3 \end{bmatrix} \propto \begin{bmatrix} R_{312} R_{231} + R_{321} R_{132} + R_{312} R_{321} \\ R_{213} R_{321} + R_{231} R_{123} + R_{213} R_{231} \\ R_{123} R_{312} + R_{132} R_{213} + R_{123} R_{132} \end{bmatrix} \equiv \begin{bmatrix} \Lambda_3 \\ \Lambda_2 \\ \Lambda_1 \end{bmatrix}. \quad (38)$$

The proportionality constant (> 0) can be determined by the condition (Equation 32), and thus the desired non-trivial solution of the λ -condition (Equation 33) for $M = 3$ is obtained:

$$\lambda_1 = \frac{\Lambda_2 \Lambda_3}{\Lambda_1 \Lambda_2 + \Lambda_1 \Lambda_3 + \Lambda_2 \Lambda_3}, \quad (39a)$$

$$\lambda_2 = \frac{\Lambda_1 \Lambda_3}{\Lambda_1 \Lambda_2 + \Lambda_1 \Lambda_3 + \Lambda_2 \Lambda_3}, \quad (39b)$$

$$\lambda_3 = \frac{\Lambda_1 \Lambda_2}{\Lambda_1 \Lambda_2 + \Lambda_1 \Lambda_3 + \Lambda_2 \Lambda_3}. \quad (39c)$$

Note that it is not guaranteed that the above solution always represents a non-trivial solution of the λ -condition (Equation 33). For example, the case of $\Lambda_1 = 0$ (where Λ_2 and Λ_3 are not zero) implies the existence of a trivial solution $(\lambda_1, \lambda_2, \lambda_3) = (1, 0, 0)$. In another example, the case of $\Lambda_1 = \Lambda_2 = 0$ implies that the denominator becomes zero; thus, the expression becomes indefinite.

2.4.2.3. PGF without winner-takes-all states

We are interested in those states in which any species does not take all molecules, because the actual simulations are performed by using the initial states excluding the winner-takes-all states. The PGF without the winner-takes-all states is represented by a linear summation of winner-takes-all states z_i^N and Equation (31) if the λ -condition has a non-trivial solution like Equation (39);

$$\phi_*^c(\mathbf{z}) = \sum_{i=1}^M b_i z_i^N + b_{M+1} \left(\sum_{i=1}^M \lambda_i z_i \right)^N, \quad (40)$$

where $b_1 + \dots + b_{M+1} = 1$. The *take-all exclusion conditions* are that the coefficients of z_i^N are zero;

$$b_i + b_{M+1} \lambda_i^N = 0 \quad (i = 1, 2, \dots, M). \quad (41)$$

Therefore, we obtain the desired PGF without the winner-takes-all states (PGFwoWTAS);

$$\phi_*^c(\mathbf{z}) = \frac{\left(\sum_{i=1}^M \lambda_i z_i\right)^N - \sum_{i=1}^M (\lambda_i z_i)^N}{1 - \sum_{i=1}^M \lambda_i^N},$$

$$\sum_{i=1}^M \lambda_i = 1, \quad 0 \leq \lambda_i < 1, \quad (42)$$

which immediately implies the following stationary distribution in the case of catalytic reactions only;

$$P_*^c(\mathbf{n}) = \begin{cases} \binom{N}{n_1, n_2, \dots, n_M} \frac{\prod_{i=1}^M \lambda_i^{n_i}}{1 - \sum_{i=1}^M \lambda_i^N} & (\mathbf{n} \in W \setminus I), \\ 0 & (\mathbf{n} \in I), \end{cases} \quad (43)$$

where $\binom{N}{n_1, n_2, \dots, n_M} = \frac{N!}{n_1! n_2! \dots n_M!}$ are multinomial coefficients. Furthermore, we can calculate the marginal distribution of the i -th species from Equation (13);

$$p_{i*}^c(n) = \begin{cases} \frac{(1 - \lambda_i)^N + \lambda_i^N - \sum_{l=1}^M \lambda_l^N}{1 - \sum_{l=1}^M \lambda_l^N} & (n = 0), \\ \binom{N}{n} \frac{\lambda_i^n (1 - \lambda_i)^{N-n}}{1 - \sum_{l=1}^M \lambda_l^N} & (1 \leq n \leq N - 1), \\ 0 & (n = N) \end{cases} \quad (44)$$

If we suppose the ergodicity $\langle n_i \rangle = \bar{n}_i$, the time-averaged concentrations can be derived from Equation (12);

$$\bar{x}_i = \frac{\lambda_i - \lambda_i^N}{1 - \sum_{i=1}^M \lambda_i^N} \rho \quad (i = 1, \dots, M). \quad (45a)$$

The above equations indicate that λ_i means the concentration per total density ρ in the continuous limit $N \rightarrow \infty$, that is, λ_i should be the solution of the classical rate equation. We can

also calculate the second-order moments $\overline{x_i x_j}$ and the variances $\text{Var}[x_i] = \overline{x_i^2} - \bar{x}_i^2$ from Equation (12);

$$\overline{x_i x_j} = \frac{\lambda_i \lambda_j}{1 - \sum_{k=1}^M \lambda_k^N} \left(1 - \frac{1}{N}\right) \rho^2 \quad (i \neq j). \quad (45b)$$

$$\text{Var}[x_i] = \left[\frac{\lambda_i^2 - \lambda_i^N - \frac{1}{N} \lambda_i (\lambda_i - 1)}{1 - \sum_{i=1}^M \lambda_i^N} - \left(\frac{\lambda_i - \lambda_i^N}{1 - \sum_{i=1}^M \lambda_i^N} \right)^2 \right] \rho^2$$

$$(i = 1, \dots, M), \quad (45c)$$

In the continuous limit $N \rightarrow \infty$, the above Equations (45b) and (45c) become $\lambda_i \lambda_j$ and 0, respectively, that is, the concentrations become mutually independent without fluctuating variables.

We compare our formulas with simulation results that are obtained by applying the Gillespie algorithm (Gillespie, 1977) to the following three-species system (Figure 1A) as an example:

$$R_{123} = R_{132} = R_{213} = R_{231} = R_{312} = 1, \quad R_{321} = 0; \quad \rho = 1, \quad (46)$$

which implies the following by Equation (39):

$$\lambda_1 = 2/11, \quad \lambda_2 = 3/11, \quad \lambda_3 = 6/11. \quad (47)$$

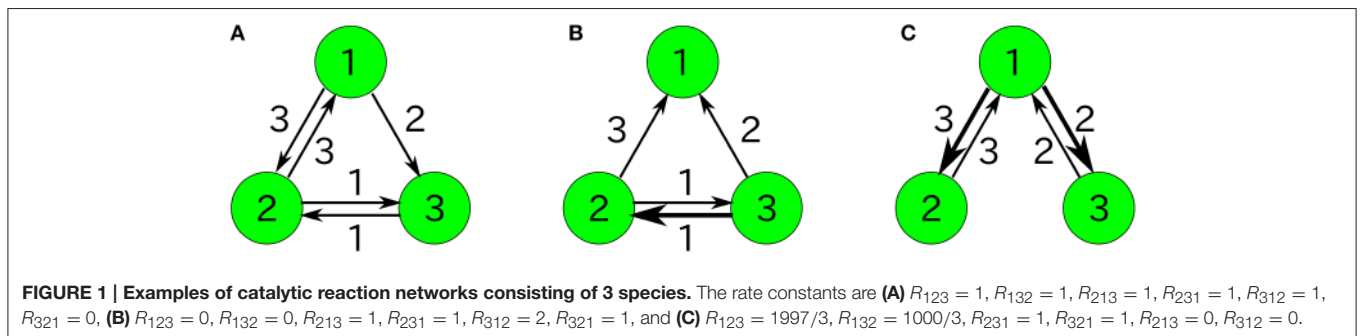
The marginal distributions are shown in Figure 2. Our formula Equation (44) is entirely in agreement with the simulation results. When the total number of molecules N is large, the marginal distributions can be approximated by normal distributions. Similarly, the formulas of the concentrations and the variances, Equations (45a) and (45c), are completely in agreement with the simulation results as shown in Figure 3. One can see that the variances monotonically decrease as N becomes larger.

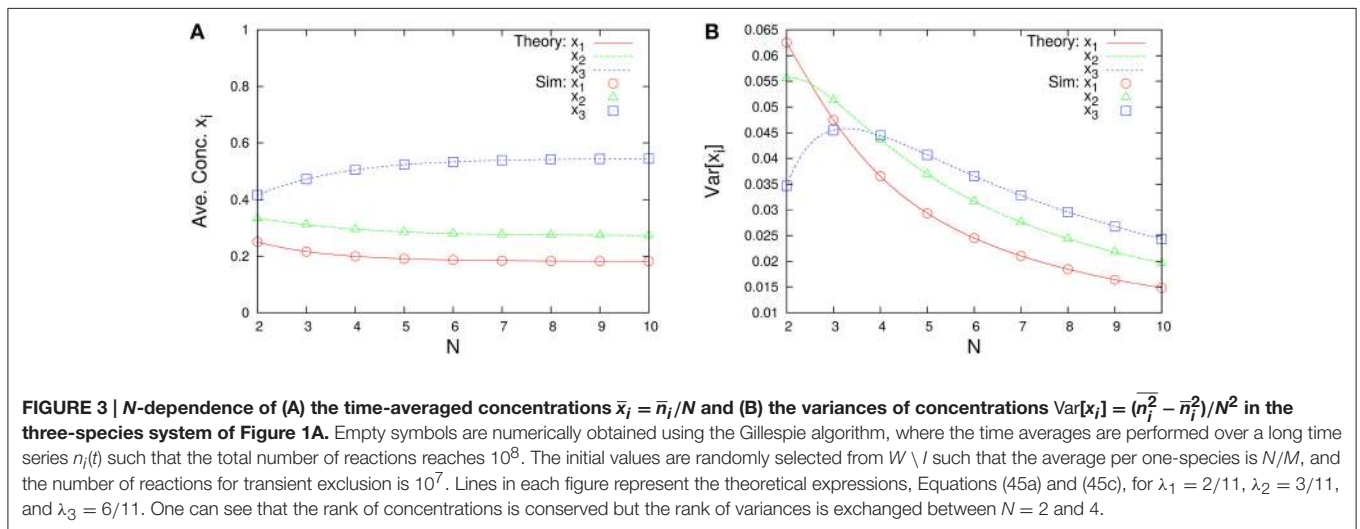
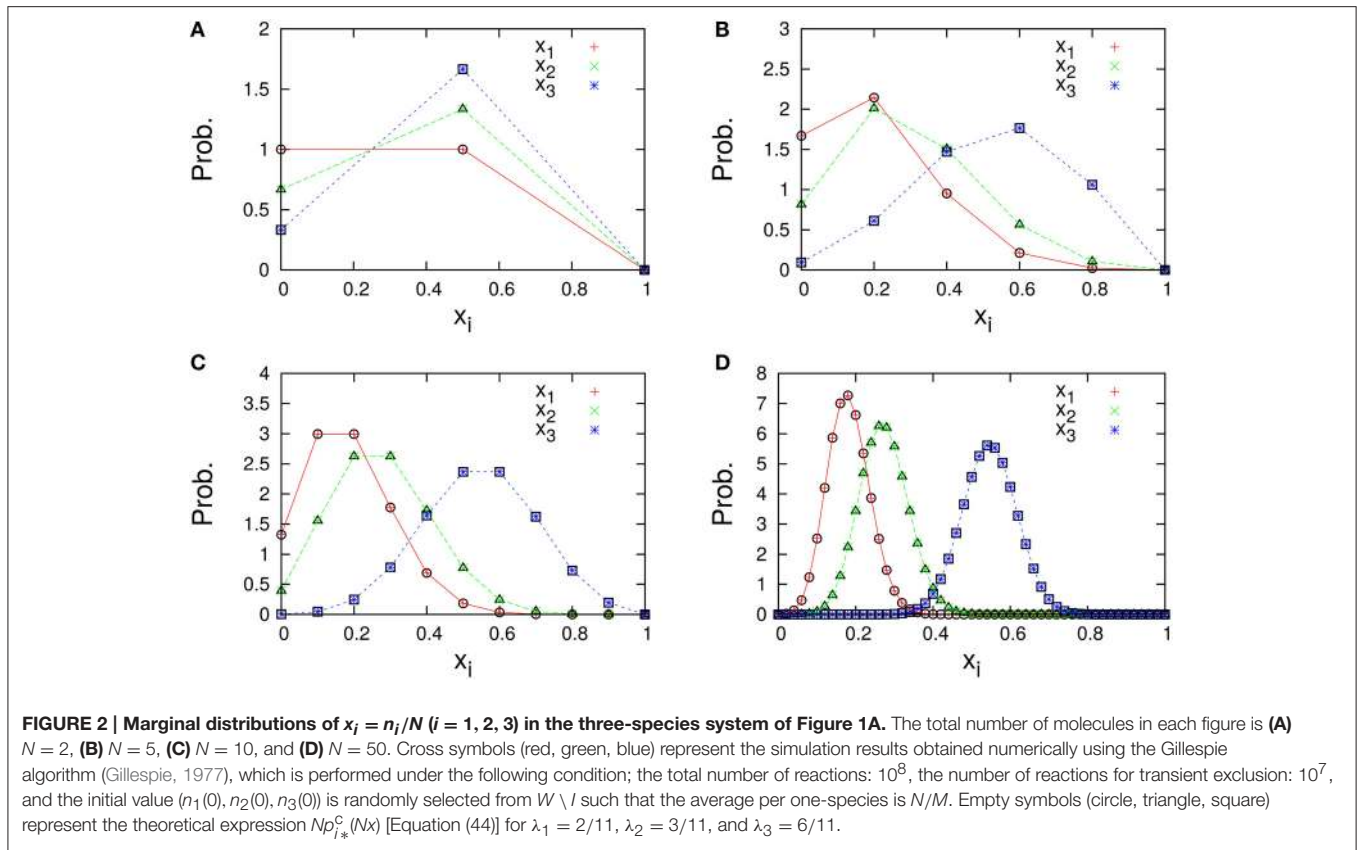
Note that if the λ -condition (Equation 33) does not have a non-trivial solution like Equation (39), the expression for the PGF (Equation 42) cannot be applied. Such special cases are treated in the following section.

2.4.2.4. Ergodicity as a sufficient condition for applying our PGF

The PGFwoWTAS (Equation 42) would be applicable if the catalytic reaction network was “entirely ergodic,” which means the following in this paper (it is reminiscent of the ω -limit set):

$$\omega(\mathbf{n}_0; \xi) = W \setminus I \quad (\forall \xi), \quad (48)$$





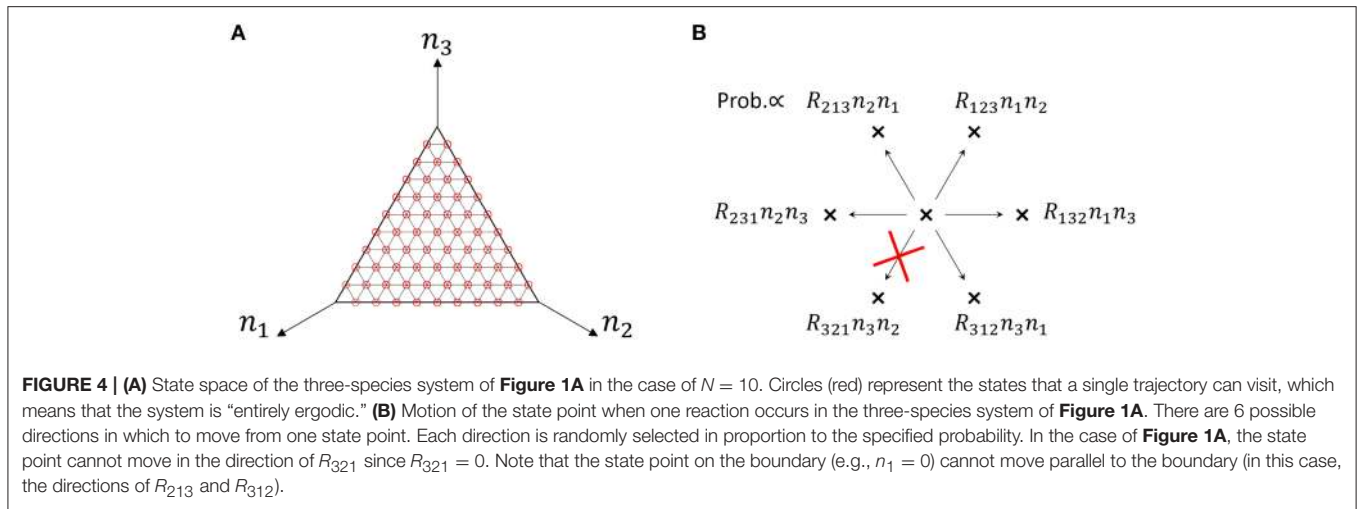
where

$$\omega(\mathbf{n}_0; \xi) := \bigcap_{t \geq 0} \{ \mathbf{n}(\tau; \xi) : \tau \geq t \}, \quad (49)$$

and ξ represents one of the trials of the stochastic process $\mathbf{n}(t)$ according to the catalytic reaction network.

We use a specific three-species system to intuitively illustrate what Equation (48) means. In the case of the three-species

system of Figure 1A, the ergodic condition (Equation 48) can be rewritten in terms of simplified conditions. As previously described, the state space of catalytic reaction networks consists of $\binom{M}{N} = \frac{(M+N-1)!}{(M-1)!N!}$ points. Therefore, in the case of $M = 3$ for example, there are $(N + 2)/(2N)!$ points. The state space of the three-species system forms a regular triangle in three-dimensional Euclidean space (see Figure 4A). The possible motion from each state point is shown in Figure 4B, where



there potentially exist six possible directions, but the network structure forbids the direction of R_{321} in the case of the network of **Figure 1A**. Note that the state point on the boundary cannot move in a direction parallel to the boundary. In this case obviously, a trajectory starting from an initial point $\mathbf{n}_0 \in W \setminus I$ visits every point in $W \setminus I$ as shown by the circles in **Figure 4A** (the set I is equivalent to the vertexes of a triangle). Therefore, as seen in **Figure 4B**, the necessary and sufficient condition to hold the ergodic condition (Equation 48) for $M = 3$ is:

$$\sum_{i,j} R_{ijk} > 0 \quad \& \quad \sum_{i,j} R_{kij} > 0 \quad (\forall k), \quad (50)$$

where the first condition represents that at least one direction for moving away from the boundary $n_k = 0$ is allowed, and the second condition represents that at least one direction for approaching the boundary $n_k = 0$ is allowed. Note that the condition (Equation 50) is no longer a sufficient condition for entire ergodicity in the case of four-species systems. In fact, in the following four-species system (**Figure 5A**), the condition (Equation 50) does not imply entire ergodicity;

$$R_{132} > 0, \quad R_{143} > 0, \quad R_{231} > 0, \quad R_{243} > 0, \quad R_{314} > 0, \\ R_{413} > 0, \quad (\text{others } 0). \quad (51)$$

This four-species system obviously satisfies the condition (Equation 50), but most initial states [in particular, $n_k(0) > 0 (\forall k)$] eventually fall into any of 2-species winners-take-all states, i.e., $\{n_1 = n_2 = 0, n_3 + n_4 = N\}$.

In the case of M -species systems, a necessary and sufficient condition for entire ergodicity is difficult to derive, although we may consider a sufficient condition instead. We discuss the sufficient condition for entire ergodicity by introducing the collection of all l -species winners-take-all states ($l = 1, 2, \dots, M$):

$$I_{M,N}^{(l)} := \bigcup_{1 \leq i_1 < \dots < i_l \leq M} I_{M,N}^{(l)}(i_1, i_2, \dots, i_l), \quad (52)$$

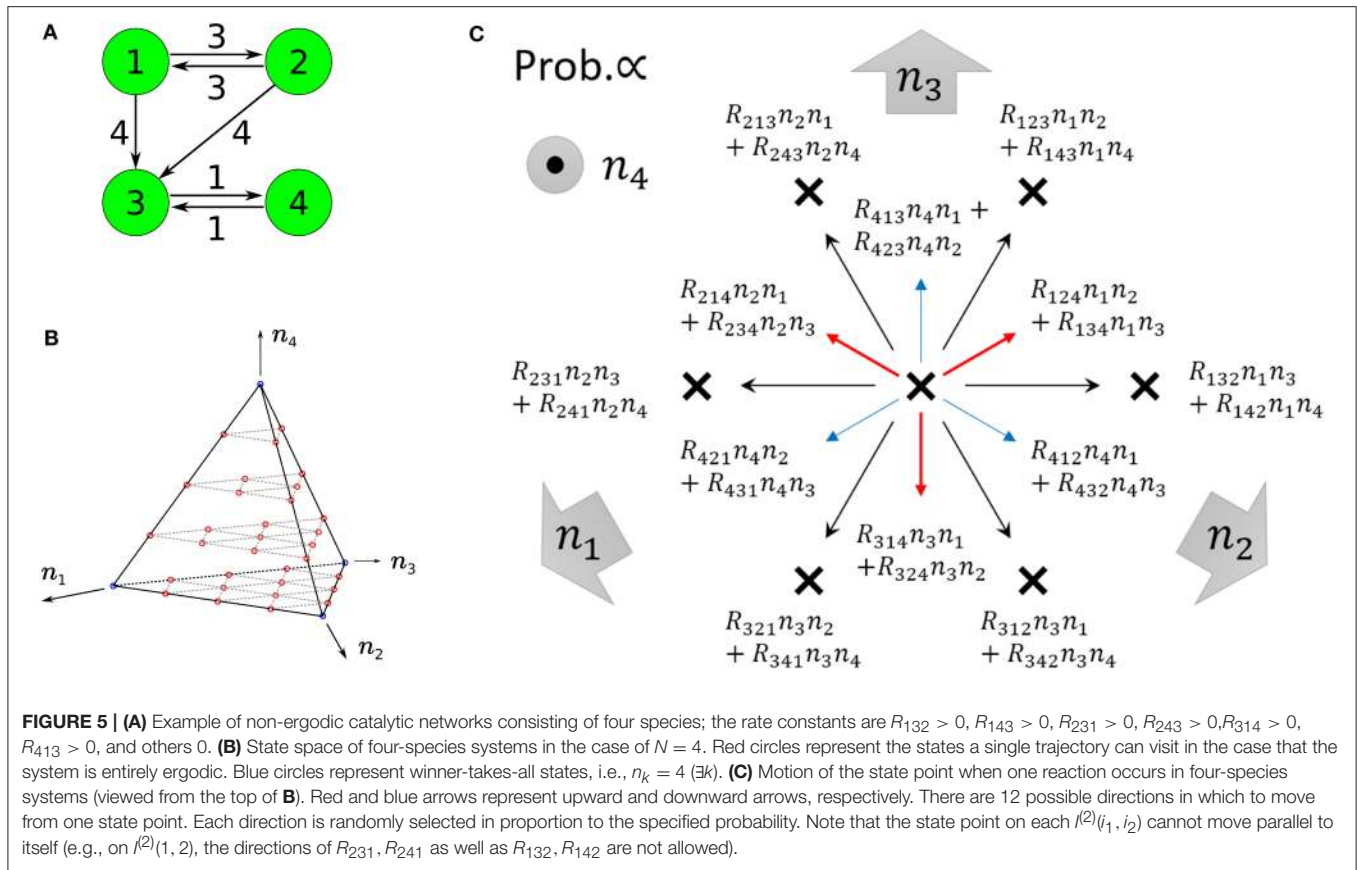
where

$$I_{M,N}^{(l)}(i_1, i_2, \dots, i_l) := \{\mathbf{n} \in W_{M,N} \mid n_{i_1} + n_{i_2} + \dots + n_{i_l} = N \\ \& n_k > 0 (\forall k)\}. \quad (53)$$

We omit the subscripts M and N provided there is no concern for confusion. The above collections are not empty sets if $l \leq N$. Note that $I = I^{(1)}$ and $W = \bigcup_{l=1}^M I^{(l)}$. One can see that the following diagram holds if the catalytic network is fully connected [i.e., $R_{ijk} > 0$ for all i, j, k ($i \neq j \neq k \neq i$)];

$$I^{(1)} \not\rightleftharpoons I^{(2)} \rightleftharpoons I^{(3)} \rightleftharpoons \dots \rightleftharpoons I^{(M-1)} \rightleftharpoons I^{(M)} \quad (54)$$

where each arrow represents the direction in which a state point can move by one reaction. Each $I^{(l)}(i_1, \dots, i_l)$ is a $(l - 1)$ -dimensional object in the whole $(M - 1)$ -dimensional state space W . For example, the state space of four-species systems forms a regular tetrahedron $W_{M=4,N}$ (see **Figure 5B**); $I^{(4)}(1, 2, 3, 4)$ is the interior of the regular tetrahedron; $I^{(3)}(1, 2, 3)$, $I^{(3)}(1, 2, 4)$, $I^{(3)}(1, 3, 4)$, and $I^{(3)}(2, 3, 4)$ are regular triangles that form the boundaries of $I^{(4)}$; and $I^{(2)}(1, 2)$, $I^{(2)}(1, 3)$, $I^{(2)}(1, 4)$, $I^{(2)}(2, 3)$, $I^{(2)}(2, 4)$, and $I^{(2)}(3, 4)$ are line segments that form the boundaries of $I^{(3)}$. Note that $I^{(2)}(i_1, i_2) \rightarrow I^{(2)}(j_1, j_2)$ for $(i_1, i_2) \neq (j_1, j_2)$ is possible but $I^{(2)}(i_1, i_2) \rightarrow I^{(2)}(i_1, i_2)$ is always impossible as seen from **Figure 5C**. By considering the diagram (Equation 54) and keeping in mind **Figures 5B,C**, we can expect the following fact: *if entire ergodicity holds in the restricted system on $I^{(l)}$ for $l \geq 3$, then entire ergodicity holds in the whole state space W* . In other words, in systems consisting of M -species, the sufficient condition under which the ergodic condition (Equation 48) holds can be geometrically described as follows: *on each $I^{(l)}(i_1, \dots, i_l)$ for $l \geq 3$, at least one direction for moving away from each boundary $I^{(l-1)}(j_1, \dots, j_{l-1})$ should be allowed, and simultaneously at least one direction for approaching each boundary $I^{(l-1)}(j_1, \dots, j_{l-1})$ should also be allowed, where $\{j_1, \dots, j_{l-1}\} \subset \{i_1, \dots, i_l\}$.*



2.4.3. The Case of Catalytic-Noncatalytic Mixed Reactions

We could not obtain the steady-state solution of the general GFE (Equation 11) in the case of catalytic-noncatalytic mixed reactions ($\varepsilon > 0$), but we expect our PGF (Equation 42) to be a good approximation for mixed-reaction systems if ε is sufficiently small ($\varepsilon \ll \min\{R_{ijk} > 0\}$). More specifically, we expect the PGF (Equation 42) to be robust against non-catalytic reactions if the catalytic reaction system constituting the mixed reaction system has an ergodic component spread across the entire state space (*entire ergodicity*). Otherwise, if the catalytic reaction system has several ergodic components in the entire state space, the non-catalytic reactions may imply that a certain highly stable ergodic component attracts all possible trajectories, which obviously means that our PGF is not applicable to such a mixed reaction system. Although we were unable to prove this mathematically, we used numerical simulations to determine whether our PGF is a good approximation in the case of “entirely ergodic.”

Figure 6 represents the non-catalytic reaction rate constant ε dependencies of time-averaged concentrations x_i for the three-species system of Figures 1A,B. (As shown later, the system shown in Figure 1B is not entirely ergodic, which is the reason why the systems shown in Figures 1A,B are compared here.) It can be seen that the differences of x_i in Figure 6A between $\varepsilon = 0$ and $\varepsilon = 0.1$ are smaller than those in Figure 6B. We consider the robustness in the case of Figure 6A to originate

from the entire ergodicity of Figure 1A, which means that the ergodic component spreads on the entire state space except for winner-takes-all states (see Figure 4A).

3. APPLICATIONS

The starting point of our analysis is the GFE (Equation 11), from which several useful formulas are derived, namely the PRE (Equation 15), SME (Equation 18), TESMs (Equation 19) and (Equation 20), the λ -condition (Equation 33), and the PGFwoWTAS (Equation 42). In this section, we reveal the effectiveness of these formulas by showing important applications for several catalytic reaction networks.

3.1. Rank Conservation Law for Concentrations

We show that the rank of concentrations is conserved even if the total number of molecules changes in catalytic reaction networks (excluding non-catalytic and auto-catalytic reactions).

Suppose the concentration of the i th-species is expressed by Equation (45a) in the state corresponding to the PGFwoWTAS. When determining the rank conservation, it suffices to confirm that the relation between the amount of two arbitrary species is unchanged if the total number of molecules is changed. Let λ_1, λ_2 be the concentrations per total density in the continuous limit such that $\lambda_1 < \lambda_2$. Because $\sum_{i=1}^M \lambda_i = 1$, the inequality $\lambda_1 + \lambda_2 <$

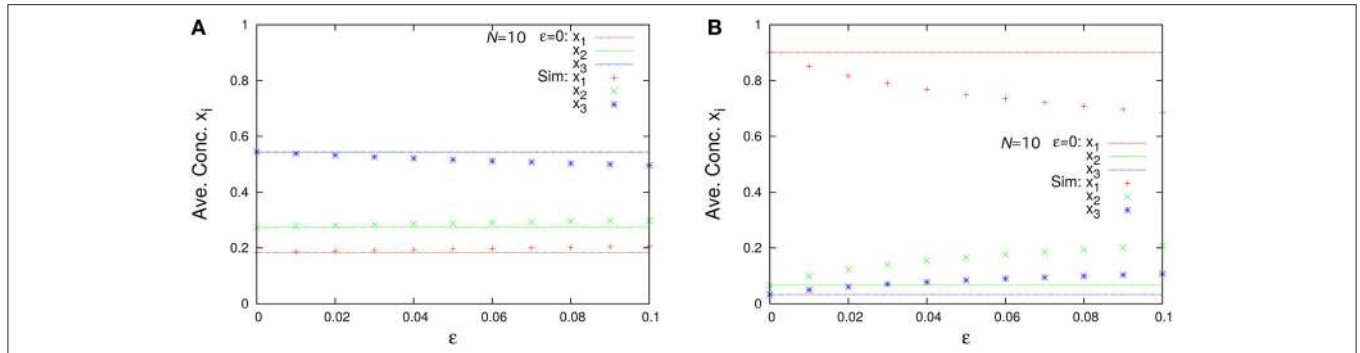


FIGURE 6 | ϵ -dependence of the time-averaged concentrations x_i ($N = 10$, $\rho = 1$) in the case of (A) the three-species system of Figure 1A with non-catalytic reactions and (B) that of Figure 1B with non-catalytic reactions. Points in the figures are numerically obtained using the Gillespie algorithm in the same way as Figure 3. The catalytic reaction network of Figure 1A is entirely ergodic; thus, the non-catalytic reactions may be treated as perturbation. On the other hand, that of Figure 1B is non-ergodic; thus, the non-catalytic reactions introduce relatively large differences compared with the case of $\epsilon = 0$.

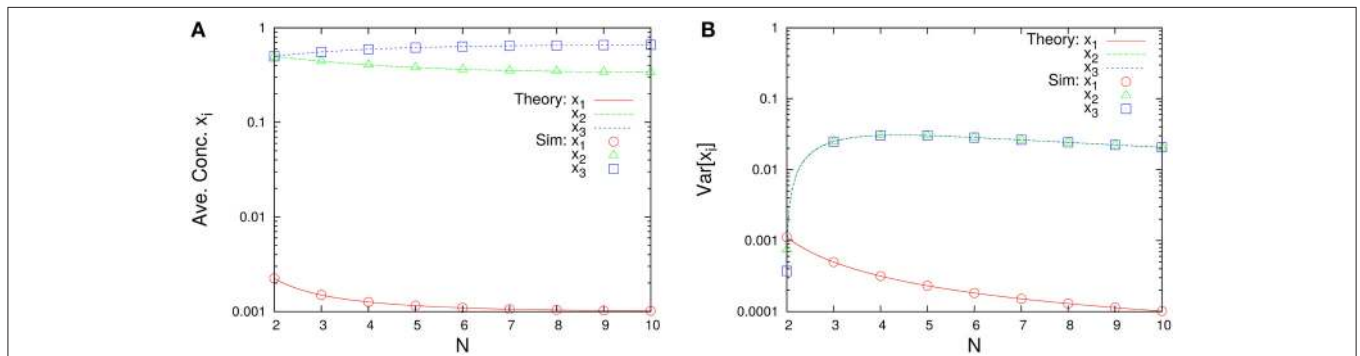


FIGURE 7 | N -dependence of (A) the time-averaged concentrations $\bar{x}_i = \bar{n}_i/N$ and (B) the variances of concentrations $\text{Var}[x_i] = (\overline{n_i^2} - \bar{n}_i^2)/N^2$ in the three-species system of Figure 1C. Empty symbols are numerically obtained using the Gillespie algorithm in the same way as Figure 3. Lines in each figure represent the theoretical expressions, Equations (45a) and (45c), for $\lambda_1 = 1/1000$, $\lambda_2 = 1/3$, and $\lambda_3 = 1997/3000$. It is clear that the rank of time-averaged concentrations is conserved.

1 must be satisfied. Therefore, the following evaluation holds:

$$\begin{aligned} \bar{x}_2 - \bar{x}_1 &= \frac{(\lambda_2 - \lambda_2^N) - (\lambda_1 - \lambda_1^N)}{1 - \sum_{i=1}^M \lambda_i^N} \rho \\ &= \frac{(\lambda_2 - \lambda_1)\rho}{1 - \sum_{i=1}^M \lambda_i^N} \left[1 - (\lambda_2 + \lambda_1)^{N-1} \right. \\ &\quad \left. + \sum_{k=1}^{N-2} \left\{ \binom{N-1}{k} - 1 \right\} \lambda_2^{N-1-k} \lambda_1^k \right] \\ &> 0 \quad (\forall N \in \mathbb{N}_{\geq 2}). \end{aligned} \tag{55}$$

This is the rank conservation law of concentrations.

Note that the rank of the variances of concentrations is generally not conserved when the total number of molecules changes. For example, let us consider the following three-species system (Figure 1C);

$$\begin{aligned} R_{123} &= 1997/3, \quad R_{132} = 1000/3, \quad R_{231} = R_{321} = 1, \\ R_{213} &= R_{312} = 0; \quad \rho = 1. \end{aligned} \tag{56}$$

As shown in Figure 7, this system depends on the total number of molecules for the time-averaged concentrations \bar{x}_i and the variances of concentrations $\text{Var}[x_i]$, represented by Equations (45a) and (45c) with $\lambda_1 = 1/1000$, $\lambda_2 = 1/3$, and $\lambda_3 = 1997/3000$. The rank of time-averaged concentrations is always conserved, but the rank of variances is exchanged at certain N (see Figure 8).

3.2. A Special Case: The Connecting State to the Winner-Takes-All State

There exists an N -molecular state, which connects to the winner-takes-all state in the continuous limit $N \rightarrow \infty$. As is later shown, such a special case is not the “weakly reversible” case (Anderson et al., 2010).

We show this by taking the following limit in the PGFwoWTAS (Equation 42):

$$\lambda_1 \rightarrow 1, \quad \lambda_2, \dots, \lambda_M \rightarrow 0, \tag{57}$$

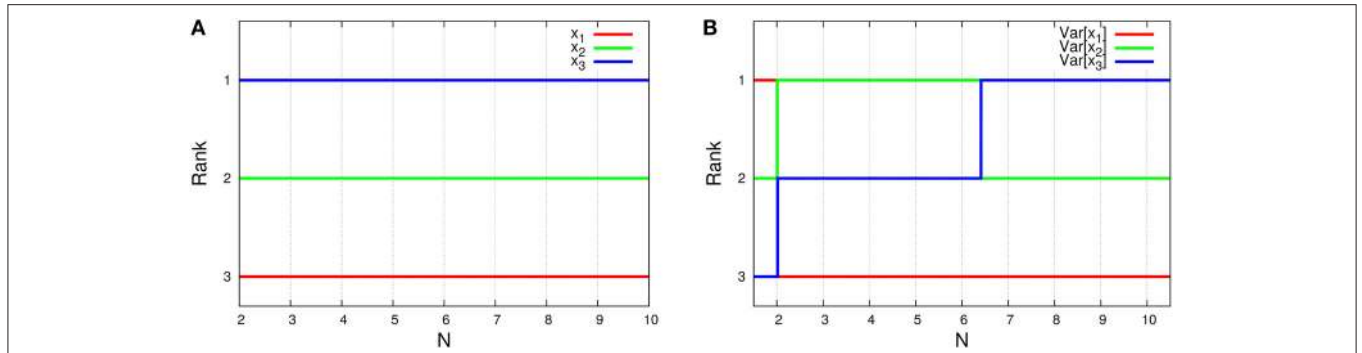


FIGURE 8 | N-dependence of (A) the rank of time-averaged concentrations and (B) the rank of its variances in the three-species system of Figure 1C. The ranks (lines) in the figures are depicted using the formulas (Equations 45a and 45c). Clearly, the rank of concentrations is conserved, although the rank of variances is exchanged between $N = 2$ and 3, and also $N = 6$ and 7.

where we suppose the following constraints are always satisfied:

$$\frac{\lambda_i}{\lambda_2 + \lambda_3 + \dots + \lambda_M} = \kappa_i \quad (i = 2, 3, \dots, M), \quad (58)$$

where $\{\kappa_i\}$ are positive constants, which are determined from the network structure $\{R_{ijk}\}$ (it is explained later). Evidently, the following holds;

$$\sum_{i=2}^M \kappa_i = 1. \quad (59)$$

The PGFwoWTAS (Equation 42) has the following limiting expression:

$$\phi_*^c(z) = \frac{\left(\frac{\lambda_1}{1-\lambda_1}z_1 + \sum_{i=2}^M \kappa_i z_i\right)^N - \left(\frac{\lambda_1}{1-\lambda_1}z_1\right)^N - \sum_{i=2}^M (\kappa_i z_i)^N}{\left(\frac{1}{1-\lambda_1}\right)^N - \left(\frac{\lambda_1}{1-\lambda_1}\right)^N - \sum_{i=2}^M \kappa_i^N}, \quad (60)$$

$$\xrightarrow{\lambda_1 \rightarrow 1} z_1^{N-1} (\kappa_2 z_2 + \dots + \kappa_M z_M) \equiv \phi_*^{cs}(z).$$

The state corresponding to the PGF (Equation 60) is the connecting state to the winner-takes-all state (CStoWTAS). The stationary distribution corresponding to the CStoWTAS is immediately obtained:

$$P_*^{cs}(\mathbf{n}) = \begin{cases} \kappa_i & (\mathbf{n} \in J), \\ 0 & (\mathbf{n} \in W \setminus J), \end{cases} \quad (61)$$

where $J_{M,N}$ (abbr. J) is defined as

$$J = \{\mathbf{n} \in W_{M,N} \mid n_1 = N - 1 \ \& \ \exists i \in [2, M] \text{ s.t. } n_i = 1\}. \quad (62)$$

Furthermore, the marginal distributions of the i -th species can be derived by Equation (13):

$$p_{i*}^{cs}(n) = \begin{cases} \delta_{n,N-1} & \text{for } i = 1, \\ (1 - \kappa_i)\delta_{n,0} + \kappa_i\delta_{n,1} & \text{for } i = 2, \dots, M, \end{cases} \quad (0 \leq n \leq N), \quad (63)$$

where δ_{ij} is the Kronecker delta, $\delta_{ij} = 0$ (if $i \neq j$) or 1 (if $i = j$). The time-averaged concentrations calculated by Equation (12) are

$$\bar{x}_1 = \left(1 - \frac{1}{N}\right)\rho, \quad (64a)$$

$$\bar{x}_i = \frac{\kappa_i}{N}\rho \quad (i = 2, \dots, M), \quad (64b)$$

and the variances of the time-averaged concentrations calculated by Equation (12) are

$$\text{Var}[x_1] = 0, \quad (65a)$$

$$\text{Var}[x_i] = \frac{\kappa_i(1 - \kappa_i)}{N^2}\rho^2 \quad (i = 2, \dots, M). \quad (65b)$$

Next, we derive the relation between the positive constants $\{\kappa_i\}$ and the network structure $\{R_{ijk}\}$. The λ -condition (Equation 33) can be converged to conditions for $\{\kappa_i\}$ (the κ -condition) by taking the limit $\lambda_1 \rightarrow 1$ as follows. Divide the λ -condition (Equation 33) into two groups, of which one group is the case of $2 \leq i < j \leq M$,

$$R_{1ij}\lambda_1\kappa_i + R_{1ji}\lambda_1\kappa_j - (R_{ij1} + R_{ji1})\kappa_i\kappa_j(1 - \lambda_1) + (1 - \lambda_1) \sum_{k=2}^M \{R_{kij}\kappa_k\kappa_i + R_{kji}\kappa_k\kappa_j - (R_{ijk} + R_{jik})\kappa_i\kappa_j\} = 0, \quad (66a)$$

and the other group is the case of $i = 1, 2 \leq j \leq M$,

$$\sum_{k=2}^M \{R_{k1j}\kappa_k\lambda_1 + R_{k1l}\kappa_k\kappa_j(1 - \lambda_1) - (R_{1jk} + R_{j1k})\lambda_1\kappa_j\} = 0, \quad (66b)$$

where we substituted $\lambda_i = \kappa_i(1 - \lambda_1)$, $i \geq 2$ and divided by $1 - \lambda_1$. Then, taking the limit $\lambda_1 \rightarrow 1$ in Equation (66a),

$$\kappa_i R_{1ij} + \kappa_j R_{1ji} = 0 \quad \text{i.e., } R_{1ij} = 0 \quad (\forall i, j), \quad (67a)$$

and taking that in Equation (66b),

$$\sum_{k=2}^M (\kappa_k R_{k1j} - \kappa_j R_{j1k}) = 0 \quad (2 \leq j \leq M). \quad (67b)$$

The condition (Equation 67a) expresses that the 1st-species cannot be a substrate, and the condition (Equation 67b) represents the desired κ -condition. Note that the CStoWTAS must be a limiting state corresponding to the PGFwoWTAS [see Equation (60)]. Therefore, the network structure $\{R_{ijk}\}$ must have definite $\{\lambda_i\}$.

For example, let us consider the following three-species system (Figure 1B):

$$R_{123} = R_{132} = 0, R_{213} = R_{231} = R_{321} = 1, R_{312} = 2; \rho = 1, \quad (68)$$

which implies $\kappa_2 = 2/3$ and $\kappa_3 = 1/3$. Obviously, this system is not weakly reversible, which means that the theorems (Theorem 4.1 and 4.2) in the previous study (Anderson et al., 2010) cannot be applied. This system should have a CStoWTAS because $\lambda_1, \lambda_2,$ and λ_3 are definite from Equation (39) with $\Lambda_1 = 0, \Lambda_2 = 2,$ and $\Lambda_3 = 4$. As shown in Figure 9, this system certainly has the time-averaged concentrations \bar{x}_i represented by Equation (64) and the variances $\text{Var}[x_i]$ represented by Equation (65). In this case also, the rank conservation law of concentrations holds.

3.3. M-Species 2-Molecules Systems with Non-catalytic Reactions

The 2mTESM (Equation 20) becomes the closed equation of the second-order moments $\langle x_l x_m \rangle$ if the first-order moments $\langle x_i \rangle$ are substituted by the second-order moments according to the PRE (Equation 15). In this subsection, we consider catalytic-non-catalytic mixed reaction systems of M species, consisting of only 2 molecules in total.

We first focus on the second formula in the 2mTESMs (Equation 20). It can be seen that each variance

of time-averaged concentration $\text{Var}[x_i] = \overline{x_i^2} - \bar{x}_i^2$ in the steady state depends solely on its concentration:

$$\text{Var}[x_i] = \left(\frac{M+1}{2M} \rho - \bar{x}_i \right) \bar{x}_i, \quad (69)$$

where we supposed the ergodicity $\langle x_i \rangle = \bar{x}_i$. The above equation expresses that the fluctuation of the concentration x_i becomes larger as its time-averaged concentration approaches

$$\bar{x} = \frac{M+1}{4M} \rho, \quad (70)$$

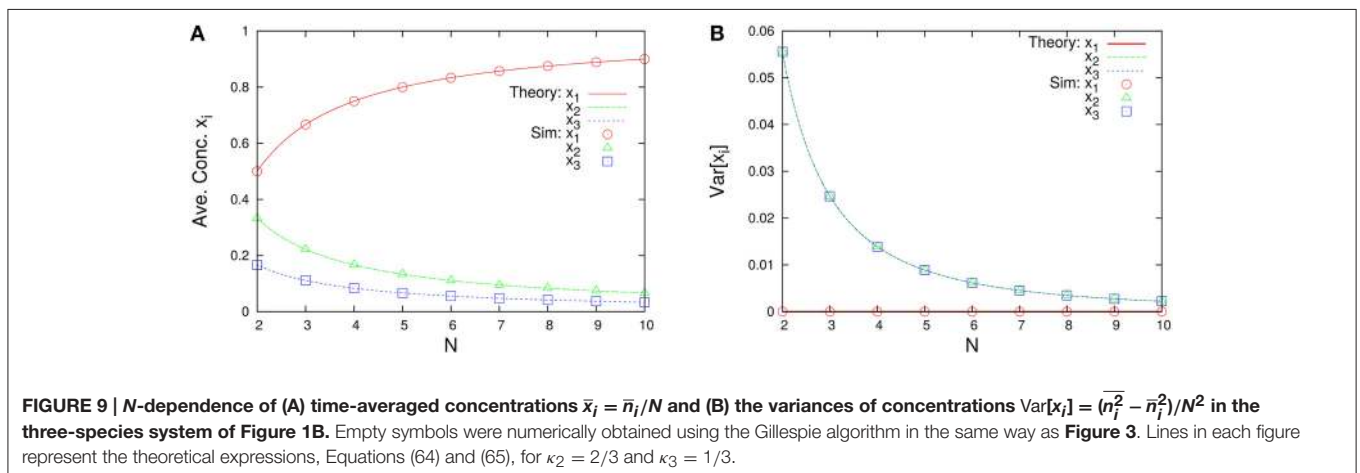
at which the fluctuation takes the largest value $\left(\frac{M+1}{4M} \rho\right)^2$. Furthermore, the time-averaged concentrations potentially have the maximum value because the variance must be positive;

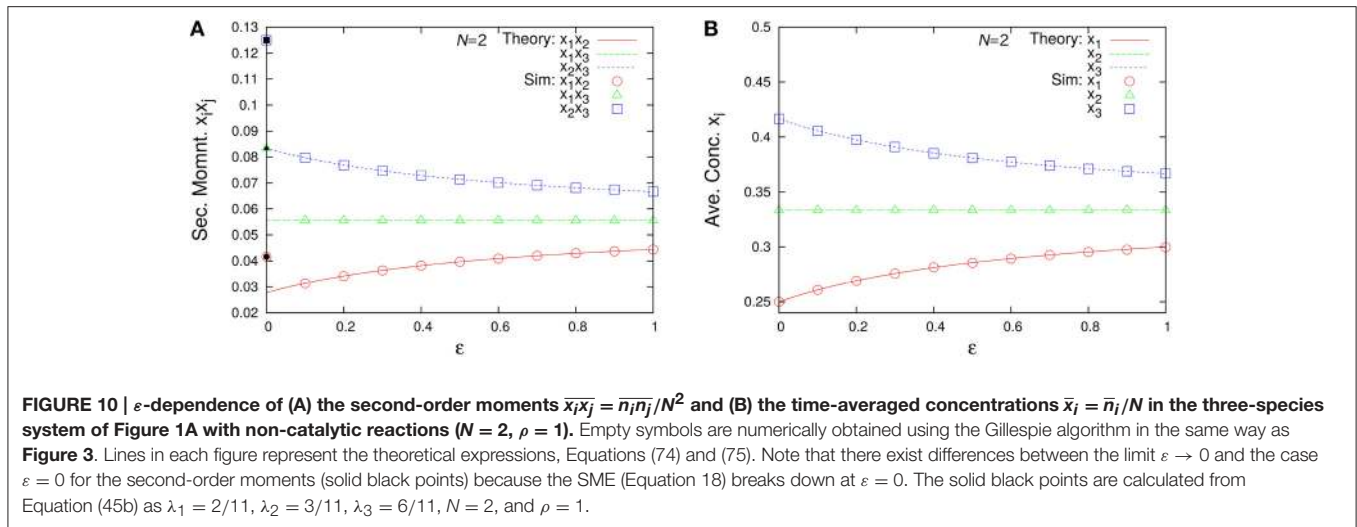
$$\bar{x}_i \leq \frac{M+1}{2M} \rho \quad (\forall i \in [1, M]). \quad (71)$$

Next, we focus on the first formula in the 2mTESMs (Equation 20). Let us consider the steady state of Equation (20a) and suppose the ergodicity $\langle x_i \rangle = \bar{x}_i$. Eliminating the first-order moments in Equation (20a) by using the SME (Equation 18) gives the determination equation of second-order moments in the 2 molecules system (2mDESM);

$$\begin{aligned} \frac{4\varepsilon}{\rho} \bar{x}_l \bar{x}_m + \sum_i \{ (R_{lmi} + R_{mli}) \bar{x}_l \bar{x}_m - R_{ilm} \bar{x}_i \bar{x}_l - R_{iml} \bar{x}_i \bar{x}_m \} \\ - \frac{1}{M} \sum_{ij} \{ (R_{jil} + R_{jim}) \bar{x}_i \bar{x}_j - R_{lij} \bar{x}_i \bar{x}_l - R_{mij} \bar{x}_i \bar{x}_m \} \\ = \frac{2\rho\varepsilon}{M^2} \quad (1 \leq l < m \leq M). \end{aligned} \quad (72)$$

The above 2mDESM can be rewritten in the $M(M-1)/2 \times M(M-1)/2$ matrix form. We demonstrate the procedure for processing





the 2mDESM by considering the case of $M = 3$. In this case, the 2mDESM (Equation 72) has a matrix of the following form:

$$\begin{bmatrix} \frac{4}{3}(R_{123} + R_{213}) + \frac{4\epsilon}{\rho} & -\frac{4}{3}R_{312} & -\frac{4}{3}R_{321} \\ -\frac{4}{3}R_{213} & \frac{4}{3}(R_{132} + R_{312}) + \frac{4\epsilon}{\rho} & -\frac{4}{3}R_{231} \\ -\frac{4}{3}R_{123} & -\frac{4}{3}R_{132} & \frac{4}{3}(R_{231} + R_{321}) + \frac{4\epsilon}{\rho} \end{bmatrix} \begin{bmatrix} \overline{x_1 x_2} \\ \overline{x_1 x_3} \\ \overline{x_2 x_3} \end{bmatrix} = \frac{2\rho\epsilon}{9} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \quad (73)$$

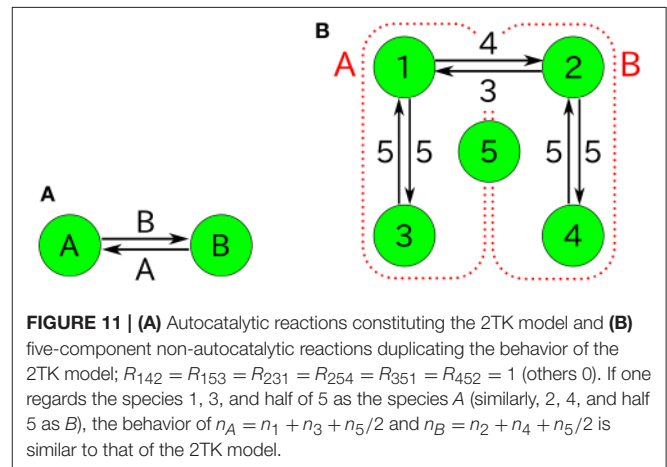
Moreover, we consider the specific three-species system of Figure 1A ($\rho = 1$) including non-catalytic reactions. Solving Equation (73) as $\epsilon > 0$ and $\rho = 1$ in the case of Figure 1A, then

$$\begin{bmatrix} \overline{x_1 x_2} \\ \overline{x_1 x_3} \\ \overline{x_2 x_3} \end{bmatrix} = \begin{bmatrix} \frac{1+3\epsilon}{18(2+3\epsilon)} \\ \frac{1}{18} \\ \frac{1+\epsilon}{6(2+3\epsilon)} \end{bmatrix}, \quad (74)$$

which is consistent with the result obtained when solving the CME (Equation 7) directly [recall $\langle x_1 x_2 \rangle = \frac{1}{4}P(n_1 = 1, n_2 = 1, n_3 = 0)$, $\langle x_1 x_3 \rangle = \dots$ and so on]. By the SME (Equation 18), the concentrations become

$$\overline{x_1} = \frac{1 + 2\epsilon}{2(2 + 3\epsilon)}, \quad \overline{x_2} = \frac{1}{3}, \quad \overline{x_3} = \frac{5 + 6\epsilon}{6(2 + 3\epsilon)}. \quad (75)$$

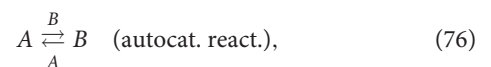
Figure 10 shows the second-order moments $\overline{x_i x_j}$ and the time-averaged concentrations $\overline{x_i}$ as functions of ϵ . Note that in the case of $\epsilon = 0$ (catalytic reactions only), we need to derive the second-order moments $\overline{x_i x_j}$ from Equation (45b) corresponding to $\lambda_1 = 2/11, \lambda_2 = 3/11$, and $\lambda_3 = 6/11$. The non-catalytic reaction rate constant ϵ seems to be a singular perturbation against the second-order moments (not the concentrations).



3.4. Non-autocatalyzation of Autocatalytic Reaction Networks

The framework we developed in this paper applies to non-autocatalytic reaction networks. However, our framework may be applicable to autocatalytic reaction networks if it were possible to convert autocatalytic to non-autocatalytic networks. Here, we show several examples of such conversions using a minimal autocatalytic reaction network “2TK model” (Ohkubo et al., 2008; Saito and Kaneko, 2015), which is a well-studied model in the context of discreteness-induced phenomena.

The 2TK model consists of only two species, and includes both autocatalytic reactions (rate const. r) and non-catalytic reactions (rate const. $\epsilon \ll r$) (see Figure 11A);



If we suppose that each of the species A and B consists of two further species, we can convert autocatalytic reactions Equation

(76) of the 2TK model to non-autocatalytic reactions as shown in **Figure 11B**. The catalysis of the species 5 between the species 1 and 3 (or, 2 and 4) is to establish an equilibrium between the concentrations of the species 1 and 3 (or, 2 and 4), that would make it possible to regard these as one species. The non-autocatalytic reactions of **Figure 11B** plus non-catalytic reactions $i \xrightarrow{\text{Prob. } 1/5} j$ is the desired five-component model, of which the CME is simply described by Equation (7) with

$$R_{142} = R_{153} = R_{231} = R_{254} = R_{351} = R_{452} = 1 \quad (\text{others } 0). \tag{78}$$

As shown in **Figure 12**, the behavior of the variables $n_A = n_1 + n_3 + n_5/2$ and $n_B = n_2 + n_4 + n_5/2$ in the five-component model is similar to that of the 2TK model (compare with Figures 1A,B in Saito and Kaneko, 2015). The λ -condition (Equation 33) of catalytic reactions in the five-component model can be written in matrix form,

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \lambda_1 \lambda_2 \\ \lambda_1 \lambda_3 \\ \lambda_1 \lambda_4 \\ \lambda_1 \lambda_5 \\ \lambda_2 \lambda_3 \\ \lambda_2 \lambda_4 \\ \lambda_2 \lambda_5 \\ \lambda_3 \lambda_4 \\ \lambda_3 \lambda_5 \\ \lambda_4 \lambda_5 \end{bmatrix} = 0, \tag{79}$$

in which there exist six non-trivial eigenvectors corresponding to the 0-eigenvalue; hence, the general eigenvector corresponding to the 0-eigenvalue can be expressed as a linear summation of those

six non-trivial eigenvectors:

$$\begin{bmatrix} \lambda_1 \lambda_2 \\ \lambda_1 \lambda_3 \\ \lambda_1 \lambda_4 \\ \lambda_1 \lambda_5 \\ \lambda_2 \lambda_3 \\ \lambda_2 \lambda_4 \\ \lambda_2 \lambda_5 \\ \lambda_3 \lambda_4 \\ \lambda_3 \lambda_5 \\ \lambda_4 \lambda_5 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_5 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + c_6 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \tag{80}$$

The above equation can be used to derive six types of non-trivial solutions $(\lambda_1, \lambda_2, \dots, \lambda_5)$, which immediately correspond to the stationary states of catalytic reactions in the five-component model by using Equation (43):

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = \begin{cases} \text{(i)} & (0, c, 0, c, 1 - 2c), \\ \text{(ii)} & (c, 0, c, 0, 1 - 2c), \\ \text{(iii)} & (2c, 1 - 2c, 0, 0, 0), \\ \text{(iv)} & (2c, 0, 1 - 2c, 0, 0), \\ \text{(v)} & (0, 2c, 0, 1 - 2c, 0), \\ \text{(vi)} & (0, 0, 2c, 1 - 2c, 0), \end{cases} \quad \forall c \in (0, 1/2), \tag{81}$$

where each case (i-vi) corresponds to (i) $c_3 = c^2, c_6 = c(1 - 2c)$ (others 0), (ii) $c_2 = c^2, c_5 = c(1 - 2c)$ (others 0), (iii) $c_1 = 2c(1 - 2c)$ (others 0), (iv) $c_2 = 2c(1 - 2c)$ (others 0), (v) $c_3 = 2c(1 - 2c)$ (others 0), and (vi) $c_4 = 2c(1 - 2c)$ (others 0). The switching behavior of the five-component model may be explained in terms of transition processes between the above six types of steady states and the five trivial steady states $\lambda_i = 1$ (others 0) of catalytic reactions in the five-component model, which are sometimes caused by non-catalytic reactions.

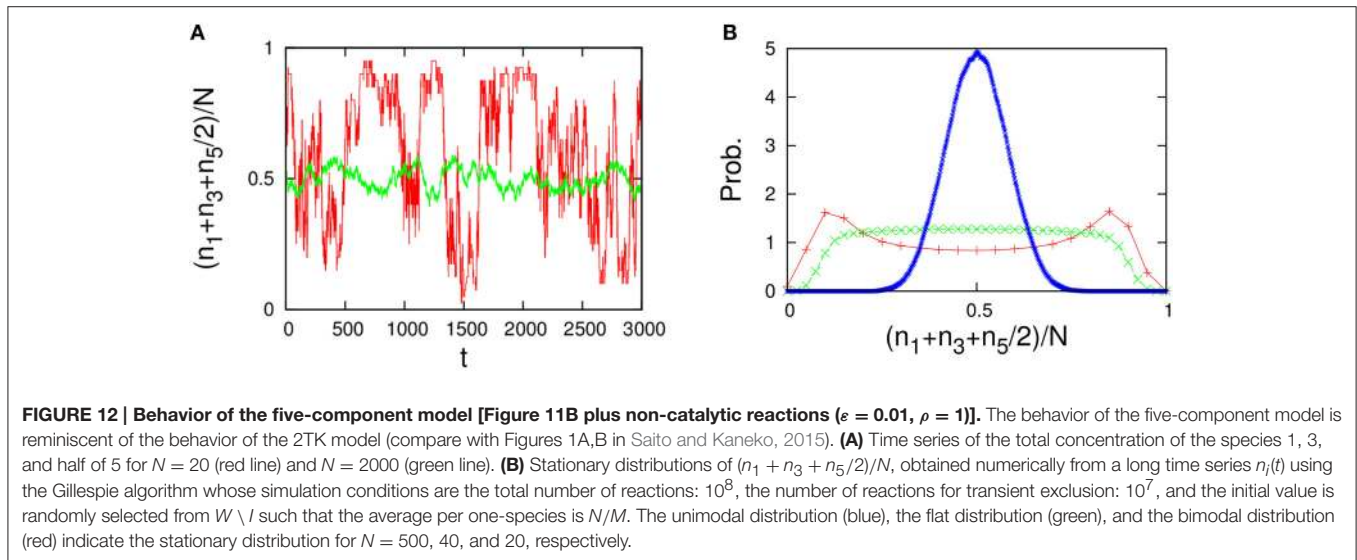


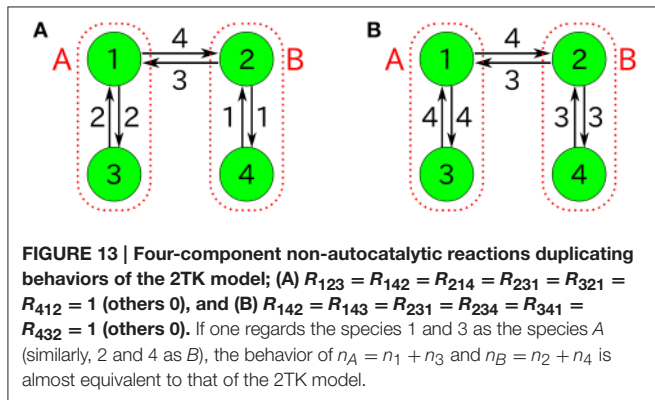
FIGURE 12 | Behavior of the five-component model [Figure 11B plus non-catalytic reactions ($\epsilon = 0.01, \rho = 1$)]. The behavior of the five-component model is reminiscent of the behavior of the 2TK model (compare with Figures 1A,B in Saito and Kaneko, 2015). **(A)** Time series of the total concentration of the species 1, 3, and half of 5 for $N = 20$ (red line) and $N = 2000$ (green line). **(B)** Stationary distributions of $(n_1 + n_3 + n_5/2)/N$, obtained numerically from a long time series $n_i(t)$ using the Gillespie algorithm whose simulation conditions are the total number of reactions: 10^8 , the number of reactions for transient exclusion: 10^7 , and the initial value is randomly selected from $W \setminus j$ such that the average per one-species is N/M . The unimodal distribution (blue), the flat distribution (green), and the bimodal distribution (red) indicate the stationary distribution for $N = 500, 40$, and 20 , respectively.

Furthermore, the marginal distribution of the species A , $p_A(n)$, can be derived by calculating the convolution of the marginal distributions $p_1(n)$, $p_3(n)$, and $p_5(n)$, if $p_1(n)$, $p_3(n)$, and $p_5(n)$ were obtained for the case including non-catalytic reactions.

Other non-autocatalytic reaction networks duplicating the 2TK model are shown in **Figure 13A**, and consist of 4 components only. Although readers would think that the four-component models do not function properly because the transitions between the species group (1, 3) [or (2, 4)] depend on the other species group (2, 4) [or (1, 3)], the four-component models actually generate much the same behavior with the 2TK model (see **Figure 14**). The switching behavior of the four-component model may also be explained as a transition process between the following three types of steady states

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \begin{cases} (0, 0, c, 1 - c), \\ (0, c, 0, 1 - c), \\ (c, 0, 1 - c, 0), \end{cases} \quad \forall c \in (0, 1), \quad (82)$$

and the four trivial steady states $\lambda_i = 1$ (others 0) of catalytic reactions in the four-component model, which are sometimes caused by non-catalytic reactions. We also confirmed

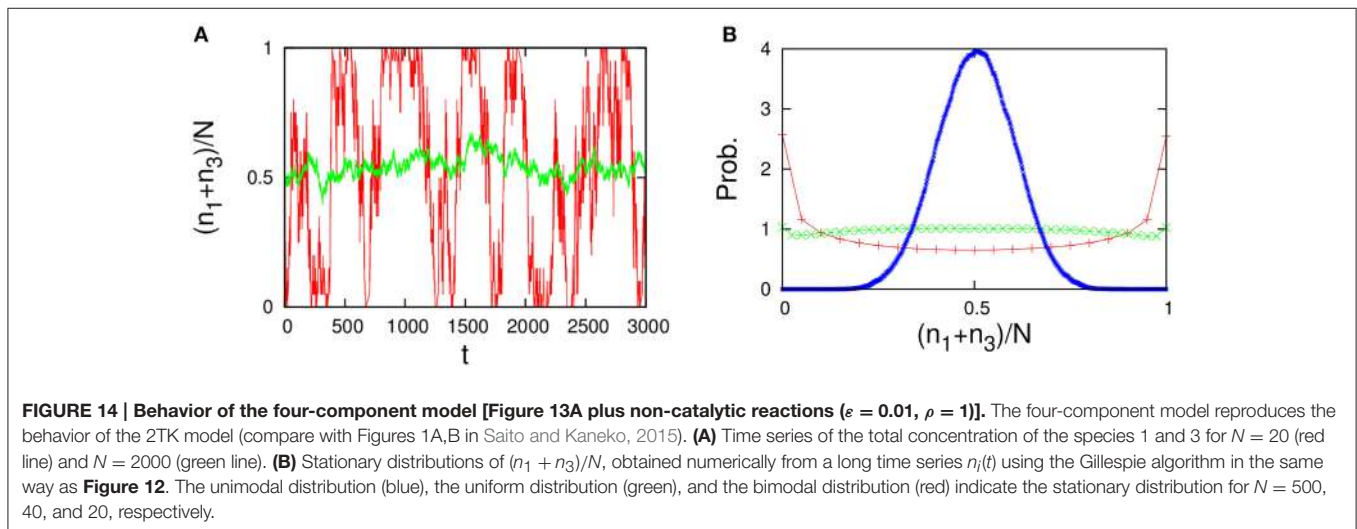


that the results are not changed even in the other four-component model of **Figure 13B**, of which the catalysis role of each the species 1 and 2 is exchanged by that of 3 and 4, respectively.

4. SUMMARY AND DISCUSSIONS

The framework we presented in this paper facilitates the prediction of the effect of the small-number issue on the concentration of each species in catalytic reaction networks. This can be described in an extreme manner by comparing the concentrations between the continuous limit ($N \rightarrow \infty$) and the case of 2 molecules ($N = 2$). If the reaction network does not include non-catalytic reactions (or, includes negligible non-catalytic reactions), we can use the formula (Equation 45a) to compare them. On the other hand, if the reaction network includes (non-negligible) non-catalytic reactions, we need to apply the formula, 2mDESM (Equation 72), with the SME (Equation 18) to obtain the concentrations in the case of 2 molecules. Although our theory has a presupposition referred to as *entire ergodicity*, the presupposition is intuitively verifiable if the system is specified, as in Section 2.4.2.4. We also demonstrated three examples for *non-autocatalyzation conversions of autocatalytic reaction networks* in Section 3.4. We consider this type of conversion to be generalized relatively easily such that our analytical framework can be applied to more general catalytic reaction networks including autocatalytic reactions.

One might think that the analysis presented in this paper can be straightforwardly extended to the case including autocatalytic reactions [in fact, the CME (Equation 7) and GFE (Equation 11) themselves hold even in the case including autocatalytic reactions]. However, if autocatalytic reactions are included (i.e., the case $R_{ikk} > 0$ is allowed), we cannot consider catalytic reactions and non-catalytic reactions to be completely separate. The reason is that, in the case including autocatalytic reactions,



the absence of non-catalytic reactions generally implies winner-takes-all steady states. Generally, solving the CME (or GFE) of catalytic-noncatalytic mixed reactions systems is more advanced and a more difficult task than that of catalytic reactions only. The proposed strategy, i.e., non-autocatalyzation conversions, is one of our ideas to address the problem.

The formulas obtained in the present work are specific and satisfactorily simple. Therefore, our theory has the capabilities to be developed into a general theory for catalytic reaction networks. On the other hand, there exists a mathematical theory for a certain class of catalytic reaction networks that are “weakly reversible” and “deficiency zero” (Anderson et al., 2010). Our formulas (Equations 27 and 43) are consistent with the main theorems (theorem 4.1 and 4.2 in Anderson et al., 2010) in the above-mentioned mathematical theory. One of the advantages of our theory compared to the above mathematical theory is understandably the probability generating function (PGF) approach, because the PGF is a major analytical tool for physicists, chemists, and mathematical biologists. Therefore, our theory is easily verifiable, and one can design a computer algorithm to calculate our analytical formulas. We also showed the extensibility of our theory by using applications (in Section 3), especially because of the CStoWTAS (Equation 61), which was not suggested by the above mathematical theory.

Actual biochemical pathways in the cell involve thousands of chemical species, and their chemical properties vary. Our theoretical framework is general and extensible to such complex reaction networks, if they can be represented by CMEs such as Equation (7). As our current model consists of simple two-body catalytic reactions, it is difficult to point out examples in actual biological systems that correspond exactly to our model. Biochemical reactions in reality may involve a number of intermediates. There are also autocatalytic processes such as autophosphorylation, and replication of templates such as DNA, in which the catalyst or template species is also a substrate or a product. Our framework is applicable to many such cases involving network conversion, as shown for simple autocatalytic cases.

Nevertheless, the reaction kinetics of each enzyme is not always simple. Enzymes are complex macromolecules and their reaction cycles may depend on their conformational states. Therefore, the prediction of biological phenomena caused by small-number effects in real biochemical reactions, would entail further analytical challenges for catalytic reaction networks including arbitrary higher-order mixed reactions (rather than first- and second-order reactions only) or internal dynamics of the enzymes (as modeled and analyzed in Togashi and Casagrande, 2015) as important issues.

Throughout this work, our primary intention is to approach small-number issues in biological systems. One might wonder how general these small-number issues appear, and how important they are, in living cells. Recently, absolute quantification of various proteins and mRNAs in the cell has become possible, and the integration of experimental results (e.g., the construction of a database Milo et al., 2010) is also underway. Taniguchi et al. investigated the copy number distribution for more than a thousand protein species in

bacteria (Taniguchi et al., 2010). Li et al. further discussed the relationship between the copy number and synthesis rate, and also the role, of proteins (Li et al., 2014). According to the result, some transcription factors, particularly activators, are rare, of the order of 0.1 to 10 molecules per genome equivalent. Although stochastic gene expression has been intensively discussed for years (McAdams and Arkin, 1997; Thattai and van Oudenaarden, 2001; Elowitz et al., 2002; Raj and van Oudenaarden, 2008; Shahrezaei and Swain, 2008), the discrete small-number nature of transcription factors has often been ignored; hence, the finding may urge us to reconsider the issue. Synthetic approaches are also becoming popular to confirm small-number effects. Ma et al. reported that an additional stable state in a genetic bistable toggle switch attributable to the small-number effect, which was predicted by stochastic simulations, was indeed observed in bacteria containing the genetically engineered switch (Ma et al., 2012). These results suggest that such rare proteins, of the order of one molecule per cell, are common and affect regulatory function in bacteria.

Although eukaryotic cells are much larger than bacteria, they have complex membrane structures and cytoskeletons inside, and the small-number issues can be particularly significant in compartments or bottlenecks (e.g., if we consider the volume of a synaptic vesicle represented by a sphere 40 nm in diameter, then 1 molecule corresponds to ca. 50 $\mu\text{mol/L}$). Rare proteins are also involved in physiologically important signaling pathways in eukaryotes. In the Wnt signaling pathway, for example, the concentration of axin is reported to be 20 pmol/L in *Xenopus* eggs (Lee et al., 2003) (though suggested to be higher in mammalian cells Tan et al., 2012). Another example is the MAP-kinase cascade, where proteins of the order of merely 10^2 molecules (e.g., Ste5) exist in a yeast cell (Thomson et al., 2011). For scaffold proteins such as axin and Ste5, specifically, localization (locally high concentrations) of other chemical species around the scaffold may drastically change the reaction behavior, as spatial discreteness of the scaffolds becomes significant (Shnerb et al., 2000; Togashi and Kaneko, 2004). Further studies in which spatial structures (cf. reaction-diffusion equations) are considered are also expected to be important.

In the presented framework, we mainly focused on the steady-state solutions of GFE. Of course, temporal courses are biologically crucial in some cases. A well-studied example is oscillatory behavior in circadian clocks (Bell-Pedersen et al., 2005). In such oscillations, if a chemical factor is depleted down to a small number in a certain phase, then, the period can be susceptible to stochastic reactions involving the factor in that phase; on the other hand, sequestration of a factor may contribute to regular oscillations (Jolley et al., 2012). Again, the internal dynamics of enzymes can also be relevant in some systems. Although stochastic simulations are indeed powerful and many attempts are currently underway, further theoretical understanding as well as the experimental quantitative observation of rare factors would be required.

Note that a chemical “species” here can also be interpreted as a specific state of a molecule; e.g., we can consider proteins or genes, with and without modification, as separate species.

Moreover, a similar interpretation is also applicable to ecology and ethology (Biancalani et al., 2014), if the laws governing the system are analogous to reactions. We remain hopeful that theoretical frameworks including ours will facilitate the exploration of small-number issues at equally higher levels of biological systems in future.

AUTHOR CONTRIBUTIONS

MN and YT conceived the research. MN performed the analysis and simulations. Both authors discussed the results and wrote the paper.

REFERENCES

- Anderson, D. F., Craciun, G., and Kurtz, T. G. (2010). Product-form stationary distributions for deficiency zero chemical reaction networks. *Bull. Math. Biol.* 72, 1947–1970. doi: 10.1007/s11538-010-9517-4
- Awazu, A., and Kaneko, K. (2007). Discreteness-induced transition in catalytic reaction networks. *Phys. Rev. E* 76:041915. doi: 10.1103/PhysRevE.76.041915
- Awazu, A., and Kaneko, K. (2009). Ubiquitous “glassy” relaxation in catalytic reaction networks. *Phys. Rev. E* 80:041931. doi: 10.1103/PhysRevE.80.041931
- Bell-Pedersen, D., Cassone, V. M., Earnest, D. J., Golden, S. S., Hardin, P. E., Thomas, T. L., et al. (2005). Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nat. Rev. Genet.* 6, 544–556. doi: 10.1038/nrg1633
- Biancalani, T., Dyson, L., and McKane, A. J. (2014). Noise-induced bistable states and their mean switching time in foraging colonies. *Phys. Rev. Lett.* 112:038101. doi: 10.1103/PhysRevLett.112.038101
- Biancalani, T., Rogers, T., and McKane, A. J. (2012). Noise-induced metastability in biochemical networks. *Phys. Rev. E* 86:010106. doi: 10.1103/PhysRevE.86.010106
- Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science* 297, 1183–1186. doi: 10.1126/science.1070919
- Gadgil, C., Lee, C. H., and Othmer, H. G. (2005). A stochastic analysis of first-order reaction networks. *Bull. Math. Biol.* 67, 901–946. doi: 10.1016/j.bulm.2004.09.009
- Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22, 403–434. doi: 10.1016/0021-9991(76)90041-3
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361. doi: 10.1021/j100540a008
- Haruna, T. (2010). Investigating the gap between discrete and continuous models of chemically reacting systems. *J. Comput. Chem. Jpn.* 9, 135–142. doi: 10.2477/jccj.H2110
- Haruna, T. (2015). Distinguishing between discreteness effects in stochastic reaction processes. *Phys. Rev. E* 91:052814. doi: 10.1103/PhysRevE.91.052814
- Houchmandzadeh, B., and Vallade, M. (2015). Exact results for a noise-induced bistable system. *Phys. Rev. E* 91:022115. doi: 10.1103/PhysRevE.91.022115
- Jolley, C. C., Ode, K. L., and Ueda, H. R. (2012). A design principle for a posttranslational biochemical oscillator. *Cell Rep.* 2, 938–950. doi: 10.1016/j.celrep.2012.09.006
- Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* 22, 437–467. doi: 10.1016/0022-5193(69)90015-0
- Kim, P., and Lee, C. H. (2012). A probability generating function method for stochastic reaction networks. *J. Chem. Phys.* 136, 234108. doi: 10.1063/1.4729374
- Krieger, I. M., and Gans, P. J. (1960). First-order stochastic processes. *J. Chem. Phys.* 32, 247–250. doi: 10.1063/1.1700909
- Lee, C. H., Kim, K. H., and Kim, P. (2009). A moment closure method for stochastic reaction networks. *J. Chem. Phys.* 130, 134107. doi: 10.1063/1.3103264

FUNDING

This work was supported by the Ministry of Education, Culture, Sports, Science, and Technology, Japan (KAKENHI 23115007 “Spying minority in biological phenomena”), and Japan Agency for Medical Research and Development (Platform for Dynamic Approaches to Living System).

ACKNOWLEDGMENTS

We would like to thank Dr. Nen Saito for valuable discussion and useful comments.

- Lee, E., Salic, A., Krüger, R., Heinrich, R., and Kirschner, M. W. (2003). The roles of APC and axin derived from experimental and theoretical analysis of the Wnt pathway. *PLoS Biol.* 1:e10. doi: 10.1371/journal.pbio.0000010
- Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624–635. doi: 10.1016/j.cell.2014.02.033
- Ma, R., Wang, J., Hou, Z., and Liu, H. (2012). Small-number effects: a third stable state in a genetic bistable toggle switch. *Phys. Rev. Lett.* 109, 248107. doi: 10.1103/PhysRevLett.109.248107
- McAdams, H. H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 94, 814–819. doi: 10.1103/PhysRevLett.109.248107
- McQuarrie, D. A. (1967). Stochastic approach to chemical kinetics. *J. Appl. Prob.* 4, 413–478. doi: 10.2307/3212214
- Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). BioNumbers — the database of key numbers in molecular and cell biology. *Nucl. Acids Res.* 38(Suppl. 1), D750–D753. doi: 10.1093/nar/gkp889
- Munsky, B., and Khammash, M. (2006). The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.* 124:044104. doi: 10.1063/1.2145882
- Ohkubo, J., Shnerb, N., and Kessler, D. A. (2008). Transition phenomena induced by internal noise and quasi-absorbing state. *J. Phys. Soc. Jpn.* 77:044002. doi: 10.1143/JPSJ.77.044002
- Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216–226. doi: 10.1016/j.cell.2008.09.050
- Saito, N., and Kaneko, K. (2015). Theoretical analysis of discreteness-induced transition in autocatalytic reaction dynamics. *Phys. Rev. E* 91:022707. doi: 10.1103/PhysRevE.91.022707
- Shahrezaei, V., and Swain, P. S. (2008). Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17256–17261. doi: 10.1073/pnas.0803850105
- Shnerb, N. M., Louzoun, Y., Bettelheim, E., and Solomon, S. (2000). The importance of being discrete: life always wins on the surface. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10322–10324. doi: 10.1073/pnas.180263697
- Tan, C. W., Gardiner, B. S., Hirokawa, Y., Layton, M. J., Smith, D. W., and Burgess, A. W. (2012). Wnt signalling pathway parameters for mammalian cells. *PLoS ONE* 7:e31882. doi: 10.1371/journal.pone.0031882
- Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., et al. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cell. *Science* 329, 533–538. doi: 10.1126/science.1188308
- Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8614–8619. doi: 10.1073/pnas.151588598
- Thomson, T. M., Benjamin, K. R., Bush, A., Love, T., Pincus, D., Resnekov, O., et al. (2011). Scaffold number in yeast signaling system sets tradeoff between system output and dynamic range. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20265–20270. doi: 10.1073/pnas.1004042108
- Togashi, Y., and Casagrande, V. (2015). Spatiotemporal patterns enhanced by intra- and inter-molecular fluctuations in arrays of allosterically regulated enzymes. *New J. Phys.* 17:033024. doi: 10.1088/1367-2630/17/3/033024

- Togashi, Y., and Kaneko, K. (2001). Transitions induced by the discreteness of molecules in a small autocatalytic system. *Phys. Rev. Lett.* 86, 2459–2462. doi: 10.1103/PhysRevLett.86.2459
- Togashi, Y., and Kaneko, K. (2004). Molecular discreteness in reaction-diffusion systems yields steady states not seen in the continuum limit. *Phys. Rev. E* 70:020901. doi: 10.1103/PhysRevE.70.020901
- Togashi, Y., and Kaneko, K. (2007). Switching dynamics in reaction networks induced by molecular discreteness. *J. Phys. Condens. Matter* 19:065150. doi: 10.1088/0953-8984/19/6/065150

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Nakagawa and Togashi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.