A probability generating function method for stochastic reaction networks

Pilwon Kim and Chang Hyeong Lee

Citation: J. Chem. Phys. **136**, 234108 (2012); doi: 10.1063/1.4729374
View online: http://dx.doi.org/10.1063/1.4729374
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v136/i23
Published by the AIP Publishing LLC.

A probability generating function method for stochastic reaction networks

Pilwon Kim[a] and Chang Hyeong Lee[b]
Ulsan National Institute of Science and Technology (UNIST), Ulsan Metropolitan City 689-798, South Korea

(Received 29 October 2011; accepted 31 May 2012; published online 20 June 2012)

In this paper we present a probability generating function (PGF) approach for analyzing stochastic reaction networks. The master equation of the network can be converted to a partial differential equation for PGF. Using power series expansion of PGF and Padé approximation, we develop numerical schemes for finding probability distributions as well as first and second moments. We show numerical accuracy of the method by simulating chemical reaction examples such as a binding-unbinding reaction, an enzyme-substrate model, Goldbeter-Koshland ultrasensitive switch model, and G_2/M transition model. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4729374]

I. INTRODUCTION

In recent years stochastic modeling has been used for analyzing biochemical reaction networks. In the stochastic modeling, the system is described by so-called chemical master equation,

$$\frac{d}{dt}p(\mathbf{n},t) = \sum_k Q_k(\mathbf{n}-V_k)\cdot p(\mathbf{n}-V_k,t) - \sum_k Q_k(\mathbf{n})\cdot p(\mathbf{n},t), \quad (1)$$

for $\mathbf{n}(t) = (n_1(t), n_2(t), \ldots, n_s(t))$, where $n_i(t)$ is the molecular number of ith species at time t. Here Q_k is a propensity function for the kth reaction and V_k is the kth column of the stoichiometric matrix V whose (i, j)th entry represents changes of stoichiometric amount of the ith species when the jth reaction occurs. The propensity can be computed by mass-action kinetics or other kinetics.[1] Since the master equation (1) is linear, the large size of the system makes it very difficult, if not impossible, to find solutions except for simple cases. The finite state projection (FSP) method deals with the linear expression of (1), $\frac{dp}{dt} = Ap$, where p is the vector of probability of each state and A is the transition rate matrix of states or the state reaction matrix.[2] The FSP method proposes to truncate the original linear ODE (ordinary differential equation) system to a reduced system and approximate the exact solution by finding the solution of the reduced system. If the system has a large number of reactions and species, the transition matrix A is very huge, and in this case, it is a challenging problem to approximate the solution of the master equation by applying the FSP method. The situation does not differ for numerical approximation. Accordingly, researchers often turn their attention to finding the first and second moments instead and try to describe stochastic properties such as average and fluctuations. But even the two moments are not simple to find in most cases.[3] To overcome these difficulties in handling master equations, we suggest a probability generating function (PGF) approach in this paper.

The PGF approach gives an alternative way of finding probability distribution as well as mean and variance.[4] While PGF has been used for the analysis of stochastic kinetics since 1960s,[5,6] it has not been directly applied to numerical computation of probability distributions to the best of the authors' knowledge.

Introducing PGF, one can convert the master equation (1) into a partial differential equation (PDE). Such PDE is parabolic and the statistical information that we want to derive, such as the probability distributions and the mean, can be found as a combination of derivatives of the solution of the PDE.

For the case of the one state variable, both initial and boundary conditions for the PDE can be properly assigned and conventional numerical schemes such as a finite difference method become available. If there are more than two independent state variables, one has to solve an initial value problem on the open domain which is no more trivial for numerical simulation. We develop a symbolic method based on power series expansion and Padé approximation in order to handle such higher dimensional cases.

An outline of the paper is as follows. In Sec. II, we study the properties of PDEs that PGF should satisfy. Section III presents the PGF method that approximate solutions of such PDEs. In what follows, we illustrate numerical accuracy of the method by simulating examples such as a binding/unbinding reaction, an enzyme-substrate reaction model, the Goldbeter-Koshland ultrasensitive switch model, and the G_2/M transition model.

II. PROBABILITY GENERATING FUNCTION

For $\mathbf{z} = (z_1, \ldots, z_s)$ and $\mathbf{n} = (n_1, n_2, \ldots, n_s)$, let us denote

$$\mathbf{z}^{\mathbf{n}} = z_1^{n_1} z_2^{n_2} \cdots z_s^{n_s}.$$

Probability generating function is defined as

$$G(\mathbf{z},t) = \sum_{\mathbf{n}=\mathbf{0}}^{\infty} \mathbf{z}^{\mathbf{n}} p(\mathbf{n},t), \quad (2)$$

[a]Email: pwkim@unist.ac.kr.
[b]Author to whom correspondence should be addressed. Electronic mail: chlee@unist.ac.kr.

where $z_i \in [-1, 1]$. We differentiate (2) with respect to t. After application of Eq. (1) and manipulation of the index, one can derive a PDE

$$G_t = F(z_1, \ldots, z_s, G, DG, D^2G, \ldots, D^mG). \tag{3}$$

Here D^kG denote any kth order partial derivatives of G as

$$D^kG = G_{i_1 \ldots i_k} = \frac{\partial}{\partial z_{i_1}} \frac{\partial}{\partial z_{i_2}} \cdots \frac{\partial}{\partial z_{i_k}} G,$$

for $1 \le i_j \le s, j = 1, \ldots, k$. We can easily confirm from Eq. (1) that the resulted PDE is linear and has the same order as the reaction order, m. The dimension of the domain is s, the number of the species.

Let us take an example of a system of two species. If the involved reaction is at most second-order, the function F in Eq. (3) contains as much second-order derivatives and therefore can be written as

$$G_t = F(z_1, z_2, G, G_1, G_2, G_{11}, G_{12}, G_{22}). \tag{4}$$

Now, the initial condition of the PDE (3) is given as

$$G(\mathbf{z}, t = 0) = \mathbf{z}^{\mathbf{n}_0},$$

where $\mathbf{n}_0$ is the initial condition of $\mathbf{n}$. There are some additional conditions that always hold for z

$$G(\mathbf{z} = \mathbf{0}, t) = p(\mathbf{n} = \mathbf{0}, t), \text{ and } G(\mathbf{z} = \mathbf{1}, t) = \sum_{\mathbf{n}} p(\mathbf{n}, t) = 1. \tag{5}$$

Once the solution G to the PDE (3) is found, we can derive the statistical information for the reaction networks.[4]

Mean and variance:

$$M_k(t) = G_k(\mathbf{z} = \mathbf{1}, t) = E[n_k(t)], \tag{6}$$

$$V_{ij}(t) = G_{ij}(\mathbf{z} = \mathbf{1}, t) = \begin{cases} E[n_i n_j] & \text{if } i \neq j \\ \left(E\left[n_i^2(t)\right] - E[n_i(t)]\right) & \text{if } i = j, \end{cases} \tag{7}$$

where $E[\cdot]$ denotes the expectation of a random variable.

Probability distribution:

$$p_{n_i}(k, t) = \frac{1}{k!} \frac{\partial^k G(z, t)}{\partial z_i^k}\bigg|_{z_i = 0, z_j = 1, j \neq i}, \tag{8}$$

where $p_{n_i}(k,t)$ denotes the marginal probability of n_i at time t.

III. PGF APPROXIMATION METHOD

The difficulties of numerical simulation of the PDE (3) mainly lie in two areas; complicated form of the PDE and lack of boundary conditions. Although the resulted PDE is always linear, it has variable coefficients, which often makes numerical approximation poor. Moreover, conventional schemes are not directly applicable since there are generally no appropriate boundary conditions.

It is notable that there is an exception for boundary conditions. In some cases we can reduce the number of the state variables, using conservation laws. We may use the above condition (5) as a boundary condition of G, especially if G has only one state variable; Let $\mathbf{n} = \mathbf{n}_0 + V\mathbf{m}$, where $\mathbf{m}$ is the extent of reactions.[7] Here components of the extent are always non-negative integers. If $V\mathbf{m} = -\mathbf{n}_0$ has a solution, then $p(\mathbf{n} = \mathbf{0}, t)$ must have a positive probability at certain time t. Otherwise, $p(\mathbf{n} = \mathbf{0}, t) = 0$. That is, if $V\mathbf{m} = -\mathbf{n}_0$ has no non-negative solutions, then we get

$$G(\mathbf{z} = \mathbf{0}, t) = p(\mathbf{n} = \mathbf{0}, t) = 0,$$

which constitutes a boundary condition with $G(\mathbf{z} = 1, t) = 1$.

In order to handle the difficulties mentioned above, we take a semi-analytic approach. It is based on the power series expansion and Padé approximation. Let us explain its mathematical framework.

In this paper, we consider bounded stochastic systems. We say the stochastic system described in (1) is bounded if there is a constant positive integer vector $\mathbf{M}$ such that $p(\mathbf{n}, t) = 0$ for any $\mathbf{n}$ with $\|\mathbf{n}\| \ge \|\mathbf{M}\|$ and any $t \ge 0$. (Here the norm $\|\cdot\|$ is defined as $\|\mathbf{n}\| = \sqrt{n_1^2 + \cdots + n_s^2}$ for $\mathbf{n} = (n_1, \ldots, n_s)$.) For example, any closed reaction networks are bounded stochastic systems. It is clear that, for bounded systems, the probability function $p(\mathbf{n}, t)$ is analytic for any $t \ge 0$, since bounded systems have finite number of states and the corresponding probability function $p(\mathbf{n}, t)$ is the solution of $\frac{dp}{dt} = Kp$. One can find the formal solution $p(\mathbf{n}, t) = e^{Kt}p(0)$ analytic with respect to t, and therefore, so is $G(\mathbf{z}, t)$ from (3).

This observation enables us to set $G(\mathbf{z}, t)$ as a power expansion with respect to t as

$$G(\mathbf{z}, t) = \sum_{n=0}^{\infty} f_n(\mathbf{z}) t^n, \tag{9}$$

where f_n, $n = 0, 1, \ldots$, are polynomials of $\mathbf{z}$. Note that the initial condition determines the first coefficient as $f_0(\mathbf{z}) = \mathbf{z}^{\mathbf{n}_0}$. By plugging (9) into both the sides of (3) and comparing both the power series, one can derive a recursive relation between $f_{n-1} = f_{n-1}(\mathbf{z})$ and $f_n = f_n(\mathbf{z})$ as

$$f_n = H_n(z_1, \ldots, z_s, f_{n-1}, Df_{n-1}, D^2 f_{n-1}, \ldots, D^m f_{n-1}), \tag{10}$$

$$n = 1, 2, \ldots,$$

where m is the highest reaction order. Note that the recursive equation H_n depends on n as well. Now, we apply the relation (10) and construct coefficient functions $f_1(\mathbf{z}), f_2(\mathbf{z}), \ldots, f_N(\mathbf{z})$ from $f_1(\mathbf{z}) = \mathbf{z}_0^{\mathbf{n}}$ for sufficiently large N. It is important in a computational aspect that all $f_n(\mathbf{z})$ generated from (10) are polynomials, since the initial function $f_0(\mathbf{z})$ and all every coefficient of the PDE (2) are polynomials. This reduces computational cost dramatically, compared to general recursive computations. One may expect that the above procedure leads to an approximated solution

$$\widetilde{G}(\mathbf{z}, t) = \sum_{n=0}^{N} f_n(\mathbf{z}) t^n. \tag{11}$$

By applying this to Eqs. (6)–(8), we can obtain approximations for the corresponding statistical quantities. One of the examples is the approximated mean of the ith species along

time

$$M_i(t) = \widetilde{G}_i(\mathbf{z} = \mathbf{1}, t). \tag{12}$$

However, the power series expansion such as (12) is generally slow in convergence and takes more computational cost to obtain an approximation at a desirable level, especially when the solution is expected to converge an equilibrium. To speed up the convergence and reduce the cost, we use Padé approximation instead. In general, Taylor expansion can be accelerated quite greatly or even turned from divergent to convergent, by being rearranged into a ratio of two series expansions in Padé approximation. In the formal form of power series, we write Padé approximation[8] as

$$\sum_{k=0}^{\infty} c_k t^k = \frac{\sum_{k=0}^{L} a_k t^k}{\sum_{k=0}^{M} b_k t^k} + O(t^{L+M+1}). \tag{13}$$

That is,

$$\left(\sum_{k=0}^{\infty} c_k t^k\right)\left(\sum_{k=0}^{M} b_k t^k\right) = \sum_{k=0}^{L} a_k t^k + O(t^{L+M+1}). \tag{14}$$

Equating the coefficients of $1, t, t^2, \ldots, t^L$ and setting $b_0 = 1$ in (14), we find

$$\begin{aligned} a_0 &= c_0, \\ a_1 &= c_1 + c_0 b_1, \\ a_2 &= c_2 + b_1 c_1 + b_2 c_0, \\ &\vdots = \vdots \\ a_L &= c_L + \sum_{i=1}^{\min(L,M)} b_i c_{L-i}. \end{aligned} \tag{15}$$

Also, equating the coefficients of $t^{L+1}, \ldots, t^{L+M}$ in (14), we obtain a system

$$\begin{bmatrix} c_{L-M+1} & c_{L-M+2} & \cdots & c_L \\ c_{L-M+2} & c_{L-M+3} & \cdots & c_{L+1} \\ \vdots & \vdots & \vdots & \vdots \\ c_L & c_{L+1} & \cdots & c_{L+M-1} \end{bmatrix} \begin{bmatrix} b_M \\ b_{M-1} \\ \vdots \\ b_1 \end{bmatrix} = - \begin{bmatrix} c_{L+1} \\ c_{L+2} \\ \vdots \\ c_{L+M} \end{bmatrix}, \tag{16}$$

where we define, for consistency, $c_k = 0$ for negative k. Thus, we can determine the coefficient of the Padé approximation by using Eqs. (15) and (16).

In the examples in Sec. IV, we use the Padé approximation order $L = M = N/2$ for some even number N, from an expectation that the systems eventually converge to constant equilibria.

Convergence of the PGF method is guaranteed by the Taylor's theorem. In practice, to determine an appropriate order N, one may observe differences between results from consecutive even orders. Let us use

$$d_i^N(t) = G_i^N(t) - G_i^{N-2}(t),$$

where $G_i^N(t)$ and $G_i^{N-2}(t)$ are the approximations of the ith mean of order N and $N-2$, respectively. Note that computation of an higher order approximation includes those of lower orders and therefore evaluating d_i^N requires only a simple subtraction. Considering that Cauchy convergence implies convergence in complete metric spaces, one can use d_i^N as a good criterion for error estimation and choice of N as well. Figure 1 illustrates the graph of $d_1^N(t)$ for some values of N in a binding reaction in Sec. IV A.

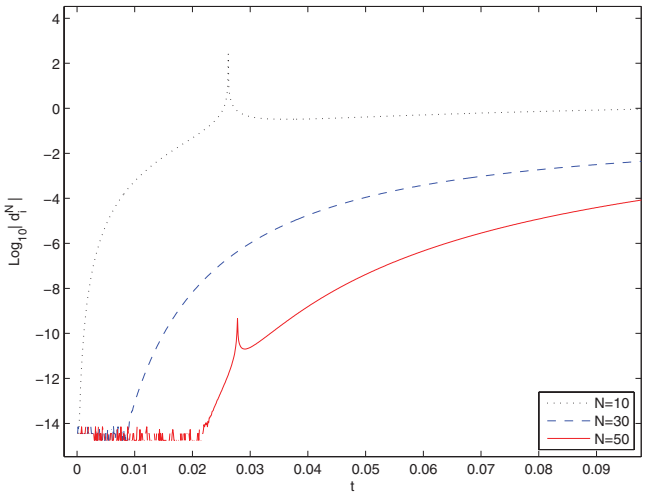

FIG. 1. The graphs of $d_1^N(t)$ for a binding reaction in Sec. IV A are illustrated. $d_1^N(t)$ denotes difference between the approximations of consecutive orders N and $N-2$. It gradually diminishes as N increases.

IV. APPLICATIONS

A. Binding reaction

We consider a binding/unbinding reaction

$$A + B \underset{c_{-1}}{\overset{c_1}{\rightleftarrows}} C.$$

Let $\mathbf{n} = (n_1, n_2, n_3)$, where n_1, n_2, n_3 denote the number of molecules of species A, B, C, respectively. Set the initial condition $\mathbf{n}(0) = (a_0, b_0, c_0)$. Since $n_1(t) + n_3(t) = a_0 + c_0$ and $n_2(t) + n_3(t) = b_0 + c_0$ for all $t > 0$ by the conservation law, there is only one independent variable among n_1, n_2, n_3. We choose n_1 for one independent variable and the random vector $\mathbf{n}$ can be represented by only n_1. The master equation for species A is

$$\begin{aligned} \frac{dp_a(t)}{dt} &= c_1[(a+1)(b_0 - a_0 + a + 1)p_{a+1}(t) \\ &\quad - a(b_0 - a_0 + a)p_a(t)] \\ &\quad + c_{-1}[(a_0 + c_0 - a + 1)p_{a-1}(t) \\ &\quad - (a_0 + c_0 - a)p_a(t)], \end{aligned} \tag{17}$$

where $p_a(t) = \text{Prob}\{n_1(t) = a\}, a = 0, 1, 2, \ldots$. Let $G(z, t) = \sum_a z^a p_a(t)$ be the PGF of $p_a(t)$. Using (17), we can derive a PDE for $G(z, t)$ as follows:

$$\begin{aligned} G_t &= c_1 z(1-z)G_{zz} - (c_{-1}z^2 + [c_1(b_0 - a_0 + 1) - c_{-1}]z \\ &\quad - c_1(b_0 - a_0 + 1))G_z \\ &\quad + c_{-1}(a_0 + c_0)(z-1)G\,. \end{aligned} \tag{18}$$

We find the initial and boundary conditions for (18) as follows:

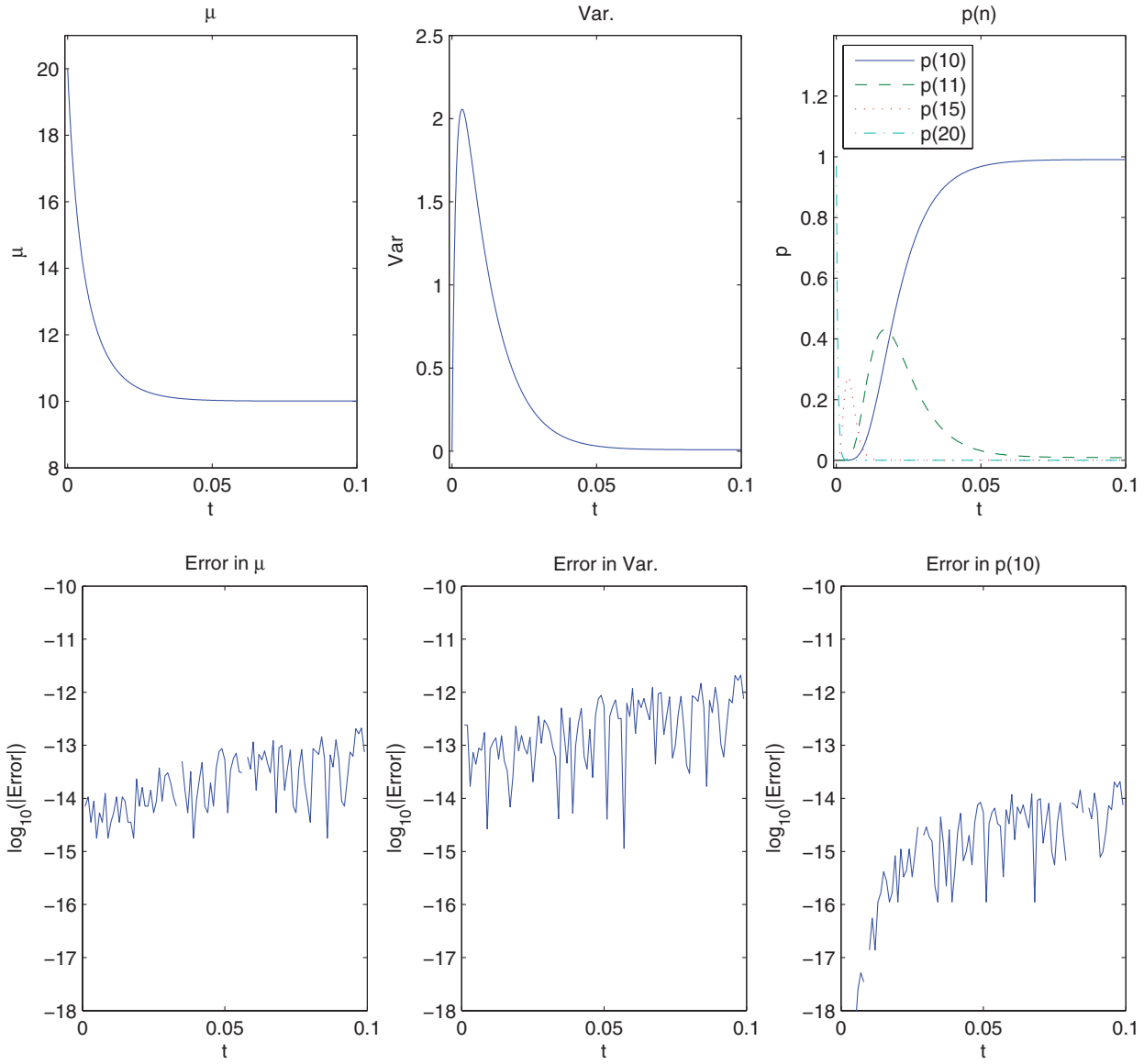

FIG. 2. Upper three figures are mean, variance, and marginal probability of the number of species A obtained from the PGF method for binding/unbinding model. In the figure of probability distribution p, each $p(i)$ denotes the marginal probability that the number of A is i. Lower three figures illustrate the errors in $\log_{10}$ generated from the PGF method, that is, $\log_{10}$ |exact solution − approximate solution|. The initial condition $a_0 = 20$, $b_0 = 10$, $c_0 = 0$, and parameters $c_1 = 1$, $c_{-1} = 0.1$ (arbitrary units) are assumed. The exact probability, mean, and variance are obtained directly by solving Kolmogorov equation $dp/dt = Kp$.[3] Some parts of the error graphs are not properly marked because the corresponding errors are lower than the machine precision.

- Initial condition: $G(z, t = 0) = z^{n_1(0)}$.
- Boundary condition: $G(z = 1, t) = 1$ and $G(z = 0, t) = 0$.

Figure 2 illustrates the result of the PGF method with $N = 100$. The comparison with the exact solution is done for the mean, the variance, and the probability distributions. One can confirm that the method reproduces the solution at a high precision level.

B. Enzyme kinetics

The enzyme-substrate reaction is one of the fundamental and important biochemical reactions, in that most real biochemical systems contain such reactions. In this section we consider an enzyme-substrate system

$$E + S \underset{c_{-1}}{\overset{c_1}{\rightleftarrows}} ES, \quad ES \overset{c_2}{\rightarrow} E + P,$$

where E, S, ES, and P denotes enzyme,substrate, enzyme-substrate complex, and product, respectively. Let n_1, n_2, n_3, and n_4 be the molecular numbers of E, S, ES, and P, respectively. The governing equation of the stochastic enzyme-substrate system is

$$\frac{dp(n_1, n_2, n_3, n_4, t)}{dt}$$
$$= c_1(n_1 + 1)(n_2 + 1)p(n_1 - 1, n_2 - 1, n_3 + 1, n_4, t)$$

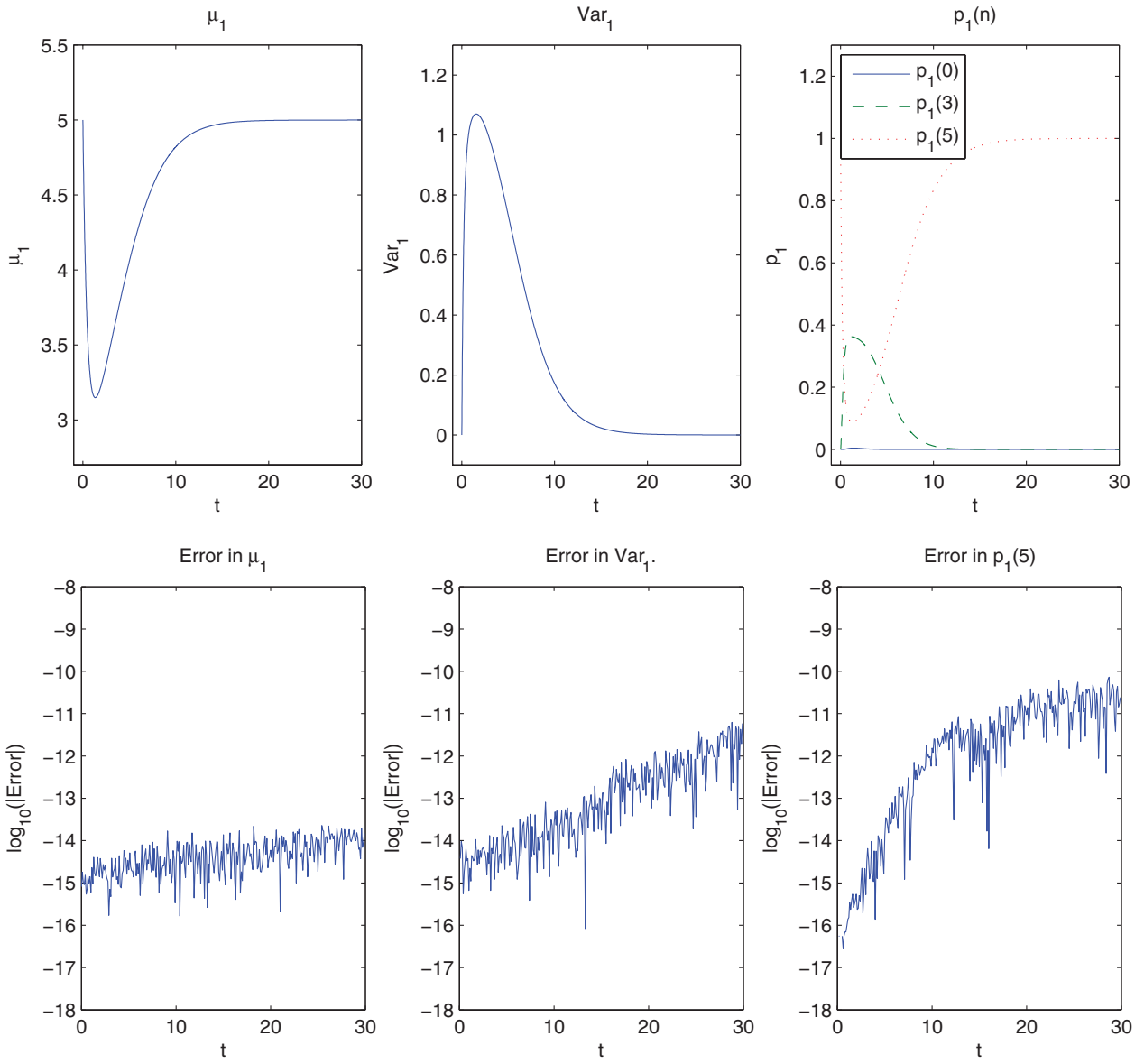

FIG. 3. Upper three figures are mean, variance, and marginal probability of the number of enzyme E obtained from the PGF method for the enzyme-substrate model. In the figure of probability distribution p, each $p(i)$ denotes the marginal probability that the number of enzyme is i. Lower three figures illustrate the errors in $\log_{10}$ generated from the PGF method, that is, $\log_{10}$ |exact solution − approximate solution|. In the figures, the condition $n_1(0) = 5$, $n_2(0) = 10$, and $a_0 = 5$, and parameters $c_1 = c_2 = 0.1$, $c_3 = 1$ are assumed. The exact probability, mean and variance are obtained directly by solving Kolmogorov equation $dp/dt = Kp$. According to the figures, the error gets bigger as the system approaches its steady state, but it is still very small, compared to the magnitudes of mean, variance, and probability.

$$+ c_{-1}(n_3+1)p(n_1-1, n_2-1, n_3+1, n_4, t)$$
$$+ c_2(n_3+1)p(n_1-1, n_2, n_3+1, n_4-1, t)$$
$$- (c_1 n_1 n_2 + c_{-1} n_3 + c_2 n_3)p(n_1, n_2, n_3, n_4, t). \quad (19)$$

Note that the stoichiometric matrix

$$V = \begin{bmatrix} -1 & 1 & 1 \\ -1 & 1 & 0 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

has rank 2, which implies that there are two stoichiometrically dependent variables. We choose n_3, n_4 as two dependent variables and remove the dependent variables in Eq. (19). The corresponding PDE of $G(z_1, z_2, t)$ is found as

$$G_t = c_1(1 - z_1 z_2)G_{12} + \left(-c_2 z_1^2 z_2 - c_3 z_1^2 + (c_2 + c_3)z_1\right)G_1 + (c_2 a_0 z_1 z_2 + c_3 a_0 z_1 - (c_2 + c_3)a_0)G,$$

where $a_0 = n_1(0) + n_3(0)$. The initial condition on G is $G(z_1, z_2, t = 0) = z_1^{\mathbf{n}(0)_1} z_2^{\mathbf{n}(0)_2}$. Since a complete boundary condition is hardly found, we use the power series expansion followed by Padé approximation. In Figure 3, we show the mean and the variance obtained from the PGF method with $N = 200$ and their error.

C. Goldbeter-Koshland (GK) ultrasensitive switch

The GK ultrasensitive switch is the well-known enzymatic reaction system in which a substrate S produces

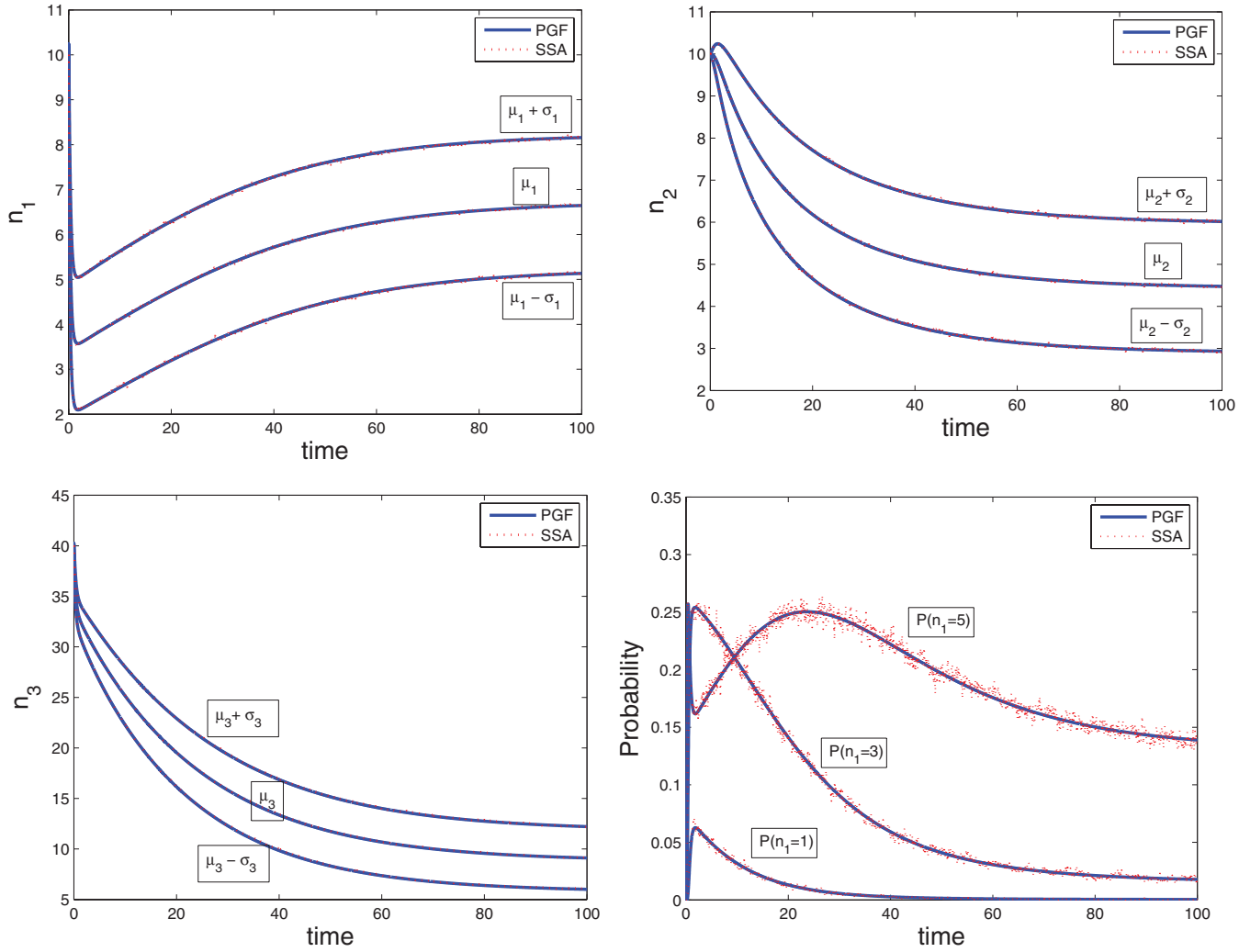

FIG. 4. Mean, standard deviation, and probability obtained from the PGF method and the SSA (stochastic simulation algorithm) under the initial condition $n_1(0) = 10$, $n_2(0) = 10$, $n_3(0) = 40$, $n_4(0) = n_5(0) = n_6(0) = 0$. μ_i and σ_i, $i = 1, 2, 3$ denote the mean and standard deviation of n_1, n_2, and n_3, respectively. In the figure of probability, each curve $p(n_1 = i)$ denotes the time-dependent probability solution that $n_1 = i$, $i = 1, 3, 5$. The results by the SSA are based on 30 000 realizations.

a product P by an enzyme E, and P produces S by an enzyme D (Ref. 9),

$$D + P \underset{c_{-1}}{\overset{c_1}{\rightleftarrows}} DP \xrightarrow{c_2} D + S, \quad E + S \underset{c_{-3}}{\overset{c_3}{\rightleftarrows}} ES \xrightarrow{c_4} E + P,$$

where $c_1 = 0.05555$, $c_{-1} = 0.83$, $c_2 = 0.17$, $c_3 = 0.05$, $c_{-3} = 0.8$, and $c_4 = 0.1$. Let $n_1, n_2, n_3, n_4, n_5, n_6$ be the number of D, E, P, S, DP, ES, respectively. One can obtain the master equation,

$$\begin{aligned}\frac{dp}{dt} &= c_1(n_1+1)(n_3+1)p(\mathbf{n}+\mathbf{e}_1+\mathbf{e}_3-\mathbf{e}_5,t)\\ &+ c_{-1}(n_5+1)p(\mathbf{n}-\mathbf{e}_1-\mathbf{e}_3+\mathbf{e}_5,t)\\ &+ c_2(n_5+1)p(\mathbf{n}-\mathbf{e}_1-\mathbf{e}_4+\mathbf{e}_5,t)\\ &+ c_3(n_2+1)(n_4+1)p(\mathbf{n}+\mathbf{e}_2+\mathbf{e}_4-\mathbf{e}_6,t)\\ &+ c_{-3}(n_6+1)p(\mathbf{n}-\mathbf{e}_2-\mathbf{e}_4+\mathbf{e}_6,t)\\ &+ c_4(n_6+1)p(\mathbf{n}-\mathbf{e}_2-\mathbf{e}_3+\mathbf{e}_6,t)\\ &- [c_1n_1n_3+(c_{-1}+c_2)n_5+c_3n_2n_4+(c_{-3}+c_4)n_6]p(\mathbf{n},t),\end{aligned}$$

where $\mathbf{e}_i$, $i = 1, 2, \ldots, 6$ denote the 6×1 unit vector containing 1 at ith entry and 0 elsewhere.

From the above master equation, we derive a PDE of the moment generating function $G(z_1, \ldots, z_6, t)$ as follows:

$$\begin{aligned}G_t &= G_5[c_{-1}(z_1z_3 - z_5) + c_2(z_1z_4 - z_5)]\\ &+ G_6[c_{-3}(z_2z_4 - z_6) + c_4(z_2z_3 - z_6)]\\ &+ G_{13}c_1(z_5 - z_1z_3) + G_{24}c_3(z_6 - z_2z_4),\end{aligned}$$

where $G_i = \dfrac{\partial G}{\partial z_i}$, $i = 1, 2, \ldots, 6$ and $G_{ij} = \dfrac{\partial^2 G}{\partial z_i \partial z_j}$, $i, j = 1, 2, \ldots, 6$.

In Figure 4, we compare the simulation results obtained by the PGF method and the SSA.

D. G_2/M transition model

The G_2/M transition network is the model that describes the dynamics of regulators of the G2-to-mitosis phase transition in the eukaryotic cell cycle.[10] The mechanism of G_2/M

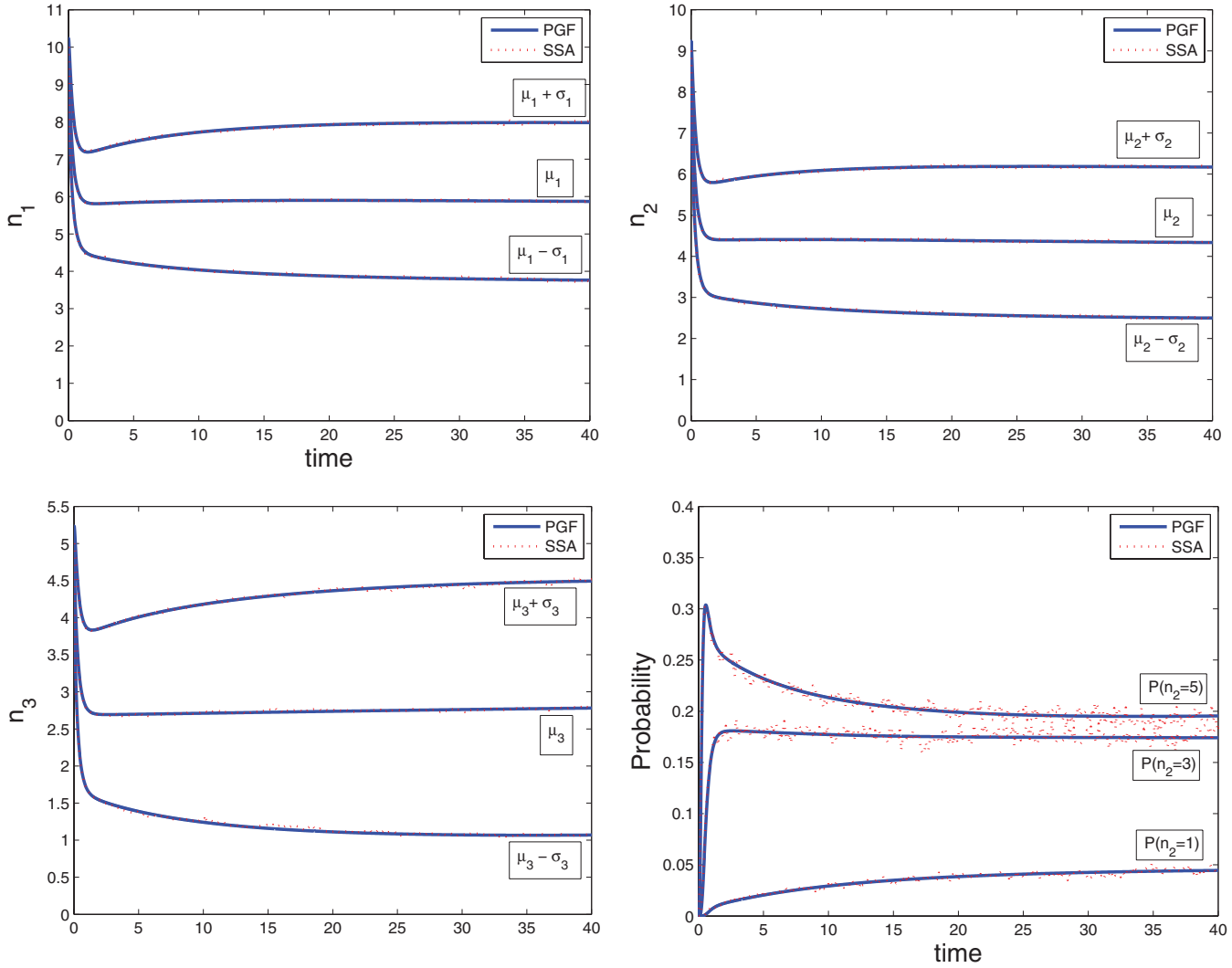

FIG. 5. Mean, standard deviation, and probability obtained from the PGF method and SSA under the initial condition $n_1(0) = 10$, $n_2(0) = 9$, $n_3(0) = 5$, $n_4(0) = 4$, $n_5(0) = 4$, $n_6(0) = 5$, $n_7(0) = n_8(0) = n_9(0) = n_{10}(0) = 0$. μ_i and σ_i, $i = 1, 2, 3$ denote the mean and standard deviation of n_1, n_2, and n_3, respectively. In the figure of probability, each curve $p(n_2 = i)$ denotes the time-dependent probability solution that $n_2 = i$, $i = 1, 3, 5$. The results by the SSA are based on 30 000 realizations.

transition network is described by the reaction scheme

$$X + Y_p \underset{c_2}{\overset{c_1}{\rightleftarrows}} C_x \overset{c_3}{\rightarrow} X_p + Y_p, \quad E_1 + X_p \underset{c_5}{\overset{c_4}{\rightleftarrows}} C_x^e \overset{c_6}{\rightarrow} X + E_1,$$

$$X_p + Y \underset{c_8}{\overset{c_7}{\rightleftarrows}} C_y \overset{c_9}{\rightarrow} X_p + Y_p, \quad E_2 + Y_p \underset{c_{11}}{\overset{c_{10}}{\rightleftarrows}} C_y^e \overset{c_{12}}{\rightarrow} Y + E_2.$$

Here we assume the stochastic parameter values $c_1 = c_4 = c_7 = c_{10} = 0.2\ \mathrm{s}^{-1}$, $c_2 = c_5 = c_8 = c_{11} = 1\ \mathrm{s}^{-1}$, $c_3 = c_6 = c_9 = c_{12} = 0.1\ \mathrm{s}^{-1}$.

Let n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_8, n_9, n_{10} be the number of X_p, Y_p, X, Y, E_1, E_2, C_x, C_x^e, C_y, C_y^e. We derive the master equation

$$\begin{aligned}
\frac{dp}{dt} &= c_1(n_2+1)(n_3+1)p(\mathbf{n}+\mathbf{e}_2+\mathbf{e}_3-\mathbf{e}_7,t)\\
&\quad + c_2(n_7+1)p(\mathbf{n}-\mathbf{e}_2-\mathbf{e}_3+\mathbf{e}_7,t)\\
&\quad + c_3(n_7+1)p(\mathbf{n}-\mathbf{e}_1-\mathbf{e}_2+\mathbf{e}_7,t)\\
&\quad + c_4(n_1+1)(n_5+1)p(\mathbf{n}+\mathbf{e}_1+\mathbf{e}_5-\mathbf{e}_8,t)\\
&\quad + c_5(n_8+1)p(\mathbf{n}-\mathbf{e}_1-\mathbf{e}_5+\mathbf{e}_8,t)\\
&\quad + c_6(n_8+1)p(\mathbf{n}-\mathbf{e}_3-\mathbf{e}_5+\mathbf{e}_8,t)\\
&\quad + c_7(n_1+1)(n_4+1)p(\mathbf{n}+\mathbf{e}_1+\mathbf{e}_4-\mathbf{e}_9,t)\\
&\quad + c_8(n_9+1)p(\mathbf{n}-\mathbf{e}_1-\mathbf{e}_4+\mathbf{e}_9,t)\\
&\quad + c_9(n_9+1)p(\mathbf{n}-\mathbf{e}_1-\mathbf{e}_2+\mathbf{e}_9,t)\\
&\quad + c_{10}(n_2+1)(n_6+1)p(\mathbf{n}+\mathbf{e}_2+\mathbf{e}_6-\mathbf{e}_{10},t)\\
&\quad + c_{11}(n_{10}+1)p(\mathbf{n}-\mathbf{e}_2-\mathbf{e}_6+\mathbf{e}_{10},t)\\
&\quad + c_{12}(n_{10}+1)p(\mathbf{n}-\mathbf{e}_4-\mathbf{e}_6+\mathbf{e}_{10},t)\\
&\quad - [c_1n_2n_3+(c_2+c_3)n_7+c_4n_1n_5+(c_5+c_6)n_8+c_7n_1n_4\\
&\quad + c_8n_9c_9n_9 + c_{10}n_2n_6 + c_{11}n_{10} + c_{12}n_{10}]p(\mathbf{n},t),
\end{aligned}$$

where $\mathbf{e}_i$, $i = 1, 2, \ldots, 10$ denote the 10×1 unit vector containing 1 at ith entry and 0 elsewhere. From the master equation, we derive a PDE of the moment generating function

$G(\mathbf{z}, t)$ as follows:

$$\begin{aligned} G_t = & c_1 G_{23}(z_7 - z_2 z_3) + c_2 G_7(z_2 z_3 - z_7) + c_3 G_7(z_1 z_2 - z_7) \\ & + c_4 G_{15}(z_8 - z_1 z_5) + c_5 G_8(z_1 z_5 - z_8) + c_6 G_8(z_3 z_5 - z_8) \\ & + c_7 G_{14}(z_9 - z_1 z_4) + c_8 G_9(z_1 z_4 - z_9) + c_9 G_9(z_1 z_2 - z_9) \\ & + c_{10} G_{26}(z_{10} - z_2 z_6) + c_{11} G_{10}(z_2 z_6 - z_{10}) \\ & + c_{12} G_{10}(z_4 z_6 - z_{10}), \end{aligned}$$

where $G_i = \frac{\partial G}{\partial z_i}, i = 1, 2, \ldots, 10$ and $G_{ij} = \frac{\partial^2 G}{\partial z_i \partial z_j}, i, j = 1, 2, \ldots, 10$.

In Figure 5, we illustrate the simulation results obtained by the PGF method and compare them with results by the SSA.

V. CONCLUSION

In this work, we developed a numerical scheme based on PGF for stochastic reaction networks. Advantage of the PGF method is, rather than struggling with the chemical master equation of extremely high order, one can handle a partial differential equation of low order. Using conventional numerical schemes for PDEs and approximation methods such as the power series expansion and Padé expansion, the method enables us to evaluate the probability distribution, the mean, and the variance accurately. If a system involves fast reactions or contains many molecules, its stochastic simulation often needs heavy computational work. In that case, the PGF method can be used to reduce computational load. We expect that the method proposed in the paper will lead to a new direction to computation of stochastic reaction networks.

ACKNOWLEDGMENTS

P. Kim acknowledges that this work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0023486). C. H. Lee acknowledges that this research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0027047).

[1] D. T. Gillespie, "A rigorous derivation of the chemical master equation," Physica A **188**, 404–425 (1992).
[2] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," J. Chem. Phys. **124**, 044104 (2006).
[3] C. H. Lee, K. Kim, and P. Kim, "A moment closure method for stochastic reaction networks," J. Chem. Phys. **130**, 134107 (2009).
[4] C. Gadgil, C. H. Lee, and H. G. Othmer, "A stochastic analysis of first-order reaction networks," Bull. Math. Biol. **67**, 901–946 (2005).
[5] P. J. Gans, "Open first-order stochastic processes," J. Chem. Phys. **33**(3), 691–694 (1960).
[6] D. A. McQuarrie, "Stochastic approach to chemical kinetics," J. Appl. Probab. **4**, 413–478 (1967).
[7] E. L. Haseltine and J. B. Rawlings, "Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics," J. Chem. Phys. **117**(15), 6959–6969 (2002).
[8] G. A. Baker and P. Graves-Morris, *Padé Approximants*, 2nd ed. (Cambridge University Press, 1996).
[9] D. Barik, M. R. Paul, W. T. Baumann, Y. Cao, and J. J. Tyson, "Stochastic simulation of enzyme-catalyzed reactions with disparate timescales," Biophys. J. **95**, 3563–3574 (2008).
[10] A. Kumar and K. Josić, "Reduced models of networks of coupled enzymatic reactions," J. Theor. Biol. **278**, 87–106 (2011).