J Math Chem (2013) 51:316–337
DOI 10.1007/s10910-012-0085-8

ORIGINAL PAPER

A numerical characteristic method for probability generating functions on stochastic first-order reaction networks

Chang Hyeong Lee · Jaemin Shin · Junseok Kim

Received: 31 July 2012 / Accepted: 7 September 2012 / Published online: 16 September 2012

Abstract We propose an efficient and accurate numerical scheme for solving probability generating functions arising in stochastic models of general first-order reaction networks by using the characteristic curves. A partial differential equation derived by a probability generating function is the transport equation with variable coefficients. We apply the idea of characteristics for the estimation of statistical measures, consisting of the mean, variance, and marginal probability. Estimation accuracy is obtained by the Newton formulas for the finite difference and time accuracy is obtained by applying the fourth order Runge–Kutta scheme for the characteristic curve and the Simpson method for the integration on the curve. We apply our proposed method to motivating biological examples and show the accuracy by comparing simulation results from the characteristic method with those from the stochastic simulation algorithm.

Keywords First-order reaction network · Characteristic method · Monte Carlo method · First-order partial differential equation

1 Introduction

Various biological processes occur in a cell of living organisms. The mechanism of those processes can be described by modeling a reaction network and analyzing its dynamics with mathematical tools such as differential equations and stochastic pro-

C. H. Lee
Ulsan National Institute of Science and Technology (UNIST), Ulsan 689-798, Republic of Korea
e-mail: chlee@unist.ac.kr

J. Shin · J. Kim (✉)
Department of Mathematics, Korea University, Seoul 136-713, Republic of Korea
e-mail: cfdkim@korea.ac.kr

cesses. As researches have been recently focused on relatively small biological systems such as gene regulatory networks, stochastic modeling has been used to capture random properties such as fluctuations or noises, which are intrinsic phenomena for small systems characterized by molecular interactions [1]. In the stochastic modeling, the dynamics of a reaction network is described by the chemical master equation

$$\frac{\partial}{\partial t} p(\mathbf{n}, t) = \sum_{k=1}^{r} [a_k(\mathbf{n} - V_k) p(\mathbf{n} - V_k, t) - a_k(\mathbf{n}) p(\mathbf{n}, t)], \tag{1}$$

where $\mathbf{n}$ is the random vector of the number of molecules of species, a_k is so-called a propensity function that is the probability of kth reaction occurring per unit time, and V_k is the kth column vector of the stoichiometric matrix V [2]. The propensity a_k is determined by the mass-action or other kinetics. It is not possible to find the solution of Eq. (1) due to high dimensionality of variables except the simple cases.

One of important chemical processes in biochemical reaction networks is the catalytic reaction and many essential biochemical reactions can be considered as catalytic reactions. For example, the transcription and translation in a gene regulatory network have been modeled as catalytic reactions [3]. The analysis of the catalytic reaction is the essential step towards understanding the dynamics of complex reaction networks. The catalytic reaction can be modeled as a first-order reaction by

$$\phi \overset{kE}{\rightarrow} P, \tag{2}$$

where ϕ denotes a source outside the system, E and P denote the enzyme and the product, respectively, and k is the reaction rate constant [3,4].

A general first-order reaction network consists of first-order reactions such as conversion, production from sources outside, degradation, and catalytic production from sources [4]. For first-order reaction networks of conversions, the probability solution as well as the mean and variance can be found by analytic methods under certain conditions [4,5]. However, if a first-order reaction network has catalytic reactions such as (2), the analytic solution of only mean and variance are known under strongly-connected assumption on the network [4].

For general first reaction networks, it is very difficult, if not impossible, to find the analytical solution of the probability as well as the mean and variance. In this case, researchers rely on computational methods such as the stochastic simulation algorithm (SSA). The SSA is a Monte Carlo type algorithm, which was firstly proposed by Gillespie [6,7]. The SSA computes which reaction occurs at what time using the two random numbers under the Markovian assumption. At each iteration, time to next reaction and reaction index are randomly drawn, and then the state vectors and propensity functions are updated [6–8].

One of the shortcomings of the SSA is that the random time step would be usually very small in case that there are various species, many numbers of molecules or fast reactions. Moreover, for finding important statistical data such as the mean and variance, a considerable number of realizations should be performed. Thus, intensive and expensive computations are required when the SSA or other Monte Carlo

type algorithm are applied for simulating reasonably complex reaction networks. To overcome such shortcomings of the SSA, there have been a variety of works for the improvement of the SSA based on approximation of the chemical master equation [9–19].

In this paper, we present a novel numerical scheme to find the computational solutions of the probability equation as well as the mean and variance of general first-order reaction networks with catalytic reactions by solving the probability generating function (PGF) with characteristic curves, instead of solving or approximating the chemical master equation. Our proposed method improves the accuracy of the solution compared with the SSA and the computational cost of the method can be remarkably lower.

This paper is organized as follows. In Sect. 2, we introduce the PGF and PDE for stochastic reaction networks and present motivating biological examples. In Sect. 3, we describe the method of characteristics which will be applied to the PGF. In Sect. 4, we describe the numerical method for solving the PDE and the details of the treatment of the scheme using the characteristic curves. In Sect. 5, the numerical results of the Monte Carlo and characteristic methods are described for showing the robustness and superiority of the characteristic method. Conclusions are presented in Sect. 6.

2 Probability generating functions for first-order reactions

In this section, we first introduce the definition and properties of the PGF and then we present examples with catalytic reactions. We derive the PDEs of the PGF for the examples.

2.1 Definition and properties of the PGF

We suppose a chemical reaction network has s distinct species. The probability generating function (PGF) for the chemical reaction network is defined as

$$G(\mathbf{x}, t) = \sum_{\mathbf{m}=\underline{0}}^{\infty} \mathbf{x}^{\mathbf{m}} p(\mathbf{n} = \mathbf{m}, t), \tag{3}$$

where $\mathbf{x}^{\mathbf{m}} = x_1^{m_1} x_2^{m_2} \dots x_s^{m_s}$, $\mathbf{x} = (x_1, x_2, \dots, x_s)$, $\mathbf{m} = (m_1, m_2, \dots, m_s)$, and $x_i \in [-1, 1]$. Using the PGF, one can find the probability distribution as well as the mean and variance as follows [4]: One can obtain a PDE for G by differentiating Eq. (3) with respect to t,

$$G_t(\mathbf{x}, t) = \sum_{\mathbf{m}=\underline{0}}^{\infty} \mathbf{x}^{\mathbf{m}} \frac{\partial}{\partial t} p(\mathbf{n} = \mathbf{m}, t). \tag{4}$$

After we substitute the chemical master equation (1) for the term $\partial p/\partial t$ in Eq. (4), we obtain a PDE

$$G_t = H\left(x_1, x_2, \ldots, x_s, G, DG, D^2G, \ldots, D^\ell G\right), \tag{5}$$

where G_t denotes $\partial G/\partial t$ and $D^k G, k = 1, 2, \ldots, \ell$ denote any kth order partial derivatives of G,

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

for $1 \le i_j \le s,\ j = 1, 2, \ldots, k$. If all reactions are first-order, Eq. (5) is written as

$$G_t = H(x_1, x_2, \ldots, x_s, G, G_1, G_2, \ldots, G_s), \tag{6}$$

where G_i denotes the partial derivative $\partial G/\partial x_i$.

Initial and boundary conditions on G:

The following three conditions hold for any $\mathbf{x}$:

$$\begin{aligned} &G(\mathbf{x}, t = 0) = \mathbf{x}^{\mathbf{n}_0} \text{ if the initial condition is } \mathbf{n}(0) = \mathbf{n}_0, \\ &G(\mathbf{x} = \underline{0}, t) = p(\mathbf{n} = \underline{0}, t), \\ &G(\mathbf{x} = \mathbf{1}, t) = \sum_{\mathbf{n}} p(\mathbf{n}, t) = 1. \end{aligned}$$

Probability distribution:

$$P_i(k, t) = \frac{1}{k!} \frac{\partial^k G(\mathbf{x}, t)}{\partial x_i^k}\bigg|_{x_i = 0, x_{j \ne i} = 1},$$

where $P_i(k, t)$ denotes the marginal probability that $n_i = k$ at time t.

Mean and covariance:

$$\begin{aligned} \mu_i(t) &= G_i(\mathbf{x} = \mathbf{1}, t) = E[n_i(t)], \\ \mathrm{Cov}_{ij}(t) &= G_{ij}(\mathbf{x} = \mathbf{1}, t) = \begin{cases} E[n_i n_j(t)], & \text{if } i \ne j, \\ E[n_i^2(t)] - E[n_i(t)], & \text{if } i = j. \end{cases} \end{aligned}$$

Here, $E[n_i(t)]$ denotes the expectation of a random variable n_i at time t.

2.2 Definition and properties of the PDE

The first-order PDE (6) is represented as

$$G_t = rG + v_1 G_1 + v_2 G_2 + \cdots + v_s G_s. \tag{7}$$

Hereafter we denote the gradient vector by $\nabla G = (G_1, G_2, \ldots, G_s)$, velocity vector by $\mathbf{v} = (v_1, v_2, \ldots, v_s)$, and spatial vector by $\mathbf{x} = (x_1, x_2, \ldots, x_s)$ where v_i is a coefficient function depending on $\mathbf{x}$, i.e., $v_i = v_i(\mathbf{x})$. The coefficient of function G is also depending on $\mathbf{x}$, $r = r(\mathbf{x})$. Then Eq. (7) is represented as follows:

$$G_t = rG + \mathbf{v} \cdot \nabla G. \tag{8}$$

We wish to calculate the following statistical measures for $t \geq 0$.

1. Mean:

$$E(n_i, t) = G_i(\mathbf{x} = 1, t).$$

2. Variance:

$$\mathrm{V}(n_i, t) = \left(G_{ii} + G_i - G_i^2\right)(\mathbf{x} = 1, t).$$

3. Probability distributions:

$$P_i(k, t) = \frac{1}{k!} \frac{\partial^k G(\mathbf{x}, t)}{\partial x_i^k}\bigg|_{x_i=0, x_{j\neq i}=1}.$$

2.3 Examples of first-order reaction

In this section, we study motivating examples of first-order reaction networks with catalytic reactions such as a single gene model, a three-species model, and a gene transcription model.

Single gene model

We consider a single gene model proposed in [3]:

$$\phi \xrightarrow{c_1} R,\ \phi \xrightarrow{c_2 R} P,\ R \xrightarrow{c_3} \phi,\ P \xrightarrow{c_4} \phi,$$

where R and P denote mRNA and protein in the single gene, respectively. If we denote the number of molecules of R and P by n_1 and n_2, respectively, we can write the master equation

$$\begin{aligned}\frac{\partial p(\mathbf{n}, t)}{\partial t} &= c_1 p(n_1 - 1, n_2) + c_2 n_1 p(n_1, n_2 - 1) + c_3(n_1 + 1)p(n_1 + 1, n_2) \\ &\quad + c_4(n_2 + 1)p(n_1, n_2 + 1) - (c_1 + c_2 n_1 + c_3 n_1 + c_4 n_2)p(n_1, n_2),\end{aligned}$$

where $\mathbf{n} = (n_1, n_2)$. From the master equation, we can derive the PDE of $G(\mathbf{x}, t)$ where $\mathbf{x} = (x_1, x_2)$:

$$\begin{aligned}G_t &= (x_1 - 1)(c_1 G - c_3 G_1) + (x_2 - 1)(c_2 x_1 G_1 - c_4 G_2) \\ &= c_1(x_1 - 1)G + \Big[c_2 x_1(x_2 - 1) + c_3(1 - x_1)\Big]G_1 - c_4(x_2 - 1)G_2.\end{aligned} \tag{9}$$

Here, we assume the parameters $c_1 = 10$, $c_2 = 10$, $c_3 = 5$, $c_4 = 0.1$ (arbitrary units) and the initial condition $G(\mathbf{x}, 0) = x_1^2 x_2^4$ [20,22].

A three-species model with catalytic reactions

We consider a three-species model in which each species catalyzes the production of another species;

$$\phi \xrightarrow{c_1 A_1} A_2,\ \phi \xrightarrow{c_2 A_2} A_3,\ \phi \xrightarrow{c_3 A_3} A_1.$$

If we denote the number of molecules of A_1, A_2, and A_3 by n_1, n_2, and n_3, respectively and $\mathbf{n} = (n_1, n_2, n_3)$, we can write the master equation

$$\begin{aligned}\frac{\partial p(\mathbf{n}, t)}{\partial t} &= c_1 n_1 p(n_1, n_2 - 1, n_3) + c_2 n_2 p(n_1, n_2, n_3 - 1) \\ &\quad + c_3 n_3 p(n_1 - 1, n_2, n_3) - (c_1 n_1 + c_2 n_2 + c_3 n_3) p(n_1, n_2, n_3).\end{aligned}$$

From the master equation, we can derive the PDE of $G(\mathbf{x}, t)$ where $\mathbf{x} = (x_1, x_2, x_3)$:

$$G_t = c_1 x_1 (x_2 - 1) G_1 + c_2 x_2 (x_3 - 1) G_2 + c_3 x_3 (x_1 - 1) G_3, \tag{10}$$

where $c_1 = 1$, $c_2 = 2$, and $c_3 = 0.1$ (arbitrary units) and the initial condition is $G(\mathbf{x}, 0) = x_1 x_2 x_3$.

Gene Transcription Model

We consider a gene transcription model [20,21]

$$D_i \to D_i^*,\ D_i^* \to D_i,\ D_i^* \to D_i^* + R,\ 1 \le i \le m$$

and

$$R \to R + P,\ R \to \phi,\ P \to \phi,$$

where D_i and D_i^* denote the ith gene copy in its inactive and active states, respectively, and R and P denote mRNA and protein, respectively. The third reaction represents the active ith gene producing mRNA and has the form of catalytic production from a source. The fourth reaction also involves catalytic production from a source; in this case mRNA causes protein to be produced. The fifth and sixth reactions model the degradation of mRNA and protein, respectively. We assume $m = 3$ as in [20]. Thus, the model is described as

$$\begin{aligned}&D_1 \xrightarrow{c_1} D_1^*,\ D_1^* \xrightarrow{c_2} D_1,\ D_2 \xrightarrow{c_3} D_2^*,\ D_2^* \xrightarrow{c_4} D_2, \\ &D_3 \xrightarrow{c_5} D_3^*,\ D_3^* \xrightarrow{c_6} D_3,\ D_1^* \xrightarrow{c_7} D_1^* + R,\ D_2^* \xrightarrow{c_8} D_2^* + R, \\ &D_3^* \xrightarrow{c_9} D_3^* + R,\ R \xrightarrow{c_{10}} R + P,\ R \xrightarrow{c_{11}} \phi,\ P \xrightarrow{c_{12}} \phi.\end{aligned}$$

Here, we denote the number of $D_1,\ D_1^*,\ D_2,\ D_2^*,\ D_3,\ D_3^*,\ R,\ P$ by $n_1,\ n_2,\ n_3,\ n_4,\ n_5,\ n_6,\ n_7,\ n_8$, respectively. We can obtain the master equation

$$\begin{aligned}\frac{\partial p}{\partial t}(\mathbf{n},t) = {} & c_1(n_1+1)p(\mathbf{n}+\mathbf{e}_1-\mathbf{e}_2,t)+c_2(n_2+1)p(\mathbf{n}-\mathbf{e}_1+\mathbf{e}_2,t)\\ & +c_3(n_3+1)p(\mathbf{n}+\mathbf{e}_3-\mathbf{e}_4,t)+c_4(n_4+1)p(\mathbf{n}-\mathbf{e}_3+\mathbf{e}_4,t)\\ & +c_5(n_5+1)p(\mathbf{n}+\mathbf{e}_5-\mathbf{e}_6,t)+c_6(n_6+1)p(\mathbf{n}-\mathbf{e}_5+\mathbf{e}_6,t)\\ & +c_7n_2p(\mathbf{n}-\mathbf{e}_7,t)+c_8n_4p(\mathbf{n}-\mathbf{e}_7,t)+c_9n_6p(\mathbf{n}-\mathbf{e}_7,t)\\ & +c_{10}n_7p(\mathbf{n}-\mathbf{e}_8,t)+c_{11}(n_7+1)p(\mathbf{n}+\mathbf{e}_7,t)+c_{12}(n_8+1)p(\mathbf{n}+\mathbf{e}_8,t)\\ & -(c_1n_1+c_2n_2+c_3n_3+c_4n_4+c_5n_5+c_6n_6+c_7n_2+c_8n_4+c_9n_6\\ & +c_{10}n_7+c_{11}n_7+c_{12}n_8)\,p(\mathbf{n},t),\end{aligned}$$

where $\mathbf{e}_i,\ i=1,2,\ldots,8$ denote the 8×1 unit vector containing 1 at the ith entry and 0 elsewhere. From the above master equation, we obtain the PDE of $G(\mathbf{x},t)$ where $\mathbf{x}=(x_1,x_2,\ldots,x_8)$;

$$\begin{aligned}G_t = {} & c_1(x_2-x_1)G_1+\Big[c_2(x_1-x_2)+c_7x_2(x_7-1)\Big]G_2\\ & +c_3(x_4-x_3)G_3+\Big[c_4(x_3-x_4)+c_8x_4(x_7-1)\Big]G_4\\ & +c_5(x_6-x_5)G_5+\Big[c_6(x_5-x_6)+c_9x_6(x_7-1)\Big]G_6\\ & +\Big[c_{10}x_7(x_8-1)+c_{11}(1-x_7)\Big]G_7+c_{12}(1-x_8)G_8,\end{aligned}$$

where we assume $c_1=c_2=c_3=c_9=c_{11}=c_{12}=0.1$, $c_4=c_5=c_6=c_7=c_8=c_{10}=1$ (arbitrary units) and the initial condition $G(\mathbf{x},0)=x_1x_3x_5$.

3 Method of characteristics

In this section, we present the characteristic method to solve the first-order PDE (8) at a given point $\mathbf{x}_0$. We define the non-linear ordinary differential equations

$$\frac{dx_1}{dt}=v_1(\mathbf{x}),\ \frac{dx_2}{dt}=v_2(\mathbf{x}),\ldots,\ \frac{dx_s}{dt}=v_s(\mathbf{x}). \tag{11}$$

For simplicity of exposition, hereafter Eq. (11) are written as $\mathbf{x}_t=\mathbf{v}(\mathbf{x})$. In order to describe the application of the characteristic curve, we first assume that the coefficient of G is zero, i.e., $r=0$. We represent the PDE (8) as a typical transport equation [23]

$$\begin{aligned}G_t-\mathbf{v}\cdot\nabla G=0,\ \text{in } R^s\times(0,\infty),\\ G=g,\ \text{on } R^s\times\{t=0\},\end{aligned} \tag{12}$$

where the initial condition $g = g(\mathbf{x})$ is a continuous function in R^s. Let us suppose the curve is described parametrically by the function

$$\mathbf{x}(\xi) = (x_1(\xi), x_2(\xi), \ldots, x_s(\xi)),$$

where the parameter ξ is in R. We define

$$z(\xi) := G(\mathbf{x}(\xi), t - \xi). \tag{13}$$

We then differentiate $z(\xi)$ with respect to ξ,

$$\begin{aligned}\frac{d}{d\xi}z(\xi) &= \frac{\partial G}{\partial x_1}\frac{dx_1}{d\xi} + \frac{\partial G}{\partial x_2}\frac{dx_2}{d\xi} + \cdots + \frac{\partial G}{\partial x_s}\frac{dx_s}{d\xi} - \frac{\partial G}{\partial t} \\ &= \mathbf{v}\cdot\nabla G(\mathbf{x}(\xi), t-\xi) - G_t(\mathbf{x}(\xi), t-\xi) = 0.\end{aligned}$$

Thus $z(\xi)$ is a constant function of ξ and consequently for each point $(\mathbf{x}, t)$. It implies that G is also constant function on the curve with the direction $\mathbf{v}(\mathbf{x}(\xi)) \in R^s$ on $\mathbf{x}(\xi)$. Because the initial value of G is given at any point on each curve, we can find the value of $\mathbf{x}(\xi)$ everywhere in $R^s \times (0, \infty)$.

Note that, in the general expression of the method of characteristics, the coefficients of the gradient term of Eq. (12) are positive. However, by focusing on the fixed point $\mathbf{x}_0$, we want to directly get the value of $G(\mathbf{x}_0, t)$ for $t \geq 0$. This setting is for going backward to the characteristic curve. Hereafter we say the curve $\mathbf{x}(t)$ by the backward characteristic curve with the initial $\mathbf{x}(0)$.

Since G is a constant function on the curve and from Eq. (13), we have

$$z(0) = z(t),$$

for any value $t \in R$. Hence we deduce

$$G(\mathbf{x}_0, t) = G(\mathbf{x}(0), t) = G(\mathbf{x}(t), 0) = g(\mathbf{x}(t), 0), \tag{14}$$

for some initial position $\mathbf{x}_0 = \mathbf{x}(0)$, $\mathbf{x}(t) \in R^s$, and $t \geq 0$. Now, we consider the case $r \neq 0$.

$$\begin{aligned}G_t - \mathbf{v}\cdot\nabla G &= rG, \text{ in } R^s \times (0, \infty), \\ G &= g, \text{ on } R^s \times \{t = 0\}.\end{aligned}$$

Using the same definition of $z(\xi)$ with Eq. (13), we get

$$\begin{aligned}\frac{d}{d\xi}z(\xi) &= \mathbf{v}\cdot\nabla G(\mathbf{x}(\xi), t-\xi) - G_t(\mathbf{x}(\xi), t-\xi) \\ &= -r(\mathbf{x}(\xi))G(\mathbf{x}(\xi), t-\xi) = -r(\mathbf{x}(\xi))z(\xi).\end{aligned}$$

We solve the ordinary differential equation by using the integrating factor.

$$z(\xi) = z(0) \exp\left(-\int_0^{\xi} r\left(\mathbf{x}(\xi)\right) d\xi \right).$$

Finally, substituting $\xi = t$ and organizing with respect to $z(0)$, we obtain

$$z(0) = z(t) \exp\left(\int_0^{t} r\left(\mathbf{x}(t)\right) dt \right).$$

Moreover, by the definition (13) of z, we have

$$G(\mathbf{x}_0, t) = g(\mathbf{x}(t), 0) \exp\left(\int_0^{t} r\left(\mathbf{x}(t)\right) dt \right). \tag{15}$$

4 Numerical methods

In this section, we propose a numerical method for solving the probability generating functions by using the characteristic curves. We can directly solve Eq. (8) if we have the closed-form solution of system (11). However, if the closed-form solution is not available, then we need to use numerical techniques to approximate the solution.

We present a method to estimate the mean, variance, and probabilities for each variable n_i. Because these statistical measures are obtained by using the derivatives of the function G, we need a numerical differentiation. For a numerical differentiation, we use the Newton forward and backward formulas [24] which are the standard finite difference methods. We apply the backward difference formula to the mean and variance and apply the forward difference formula to the probability. For the sake of simplicity, we use the notation $\mathbf{x} = \mathbf{y}(i) = a$ to denote that a vector $\mathbf{x}$ is equal to $\mathbf{y}$ except for ith component and ith component of $\mathbf{x}$ is a.

Mean: Set a vector $\mathbf{x}_0 = \mathbf{1}$ and define $\mathbf{x}_1 = \mathbf{x}_0(i) = 1 - h$. Then we approximate the first derivative of G with respect to x_i

$$\nabla_i G(\mathbf{x}_0, t) := \frac{G(\mathbf{x}_0, t) - G(\mathbf{x}_1, t)}{h},$$

where h is the spatial step size. We estimate the mean as $E(n_i, t) = \nabla_i G(\mathbf{x}_0, t)$.

Variance: We define $\mathbf{x}_2 = \mathbf{x}_0(i) = 1 - 2h$. Then we approximate the second derivative of G with respect to x_i

$$\nabla_i^2 G(\mathbf{x}_0, t) := \frac{G(\mathbf{x}_0, t) - 2G(\mathbf{x}_1, t) + G(\mathbf{x}_2, t)}{h^2}.$$

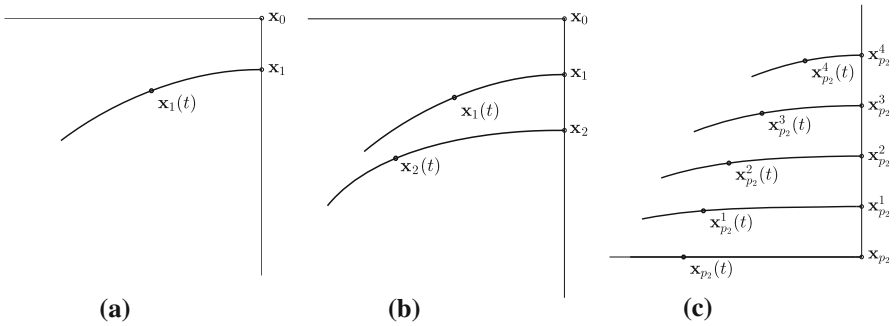

Fig. 1 Schematics for estimating the **a** mean, **b** variance, and **c** probability

We estimate the variance, $V(n_i, t) = \left(\nabla_i^2 G + \nabla_i G - \nabla_i G^2\right)(\mathbf{x}_0, t)$, using the first and second derivatives.

Probability: We define the vectors $\mathbf{x}_{p_i} = \mathbf{x}_0(i) = 0$ and $\mathbf{x}^k_{p_i} = \mathbf{x}_0(i) = 1 - kh$ for $k = 1, 2, \ldots, n$. To calculate high-order differentiations, we use the forward difference with respect to x_i

$$\Delta_i^n G(\mathbf{x}_{p_i}, t) := \sum_{k=0}^{n} (-1)^{n-k} \, {}_nC_k \; G\left(\mathbf{x}^k_{p_i}, t\right),$$

where ${}_nC_k = \frac{n!}{k!(n-k)!}$. We estimate the probability by $P_i(k, t) = \Delta_i^n G(\mathbf{x}_{p_i}, t)/n!$

Figure 1 shows the schematics for estimating the mean, variance and probability for n_2. The locations of what we want to calculate are at the corner of the unit domain. For estimating the mean or variance, we need only one or two backward characteristic curves, as shown in Fig. 1a and b. On the other hand, in terms of the probability, we calculate the curves as many as the order of differentiation. For example, see Fig. 1c, we need five curves to get the marginal probability that $n_2 = 4$. In particular, we also use these curves to calculate marginal probabilities that $n_2 < 4$. The suggested difference methods are just of first order accuracy. However, because we can choose the sufficiently small value for the spatial step size h, we can maximize the accuracy of numerical solutions.

Next, we present the method to calculate $\mathbf{x}_1(t)$, $\mathbf{x}_2(t)$, and $\mathbf{x}_{p_i}(t)$ with the initial condition $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_{p_i}$, respectively. If $r = 0$, we solve the ordinary differential system (11) using the forth order Runge–Kutta method [24] for Eq. (14). Here $\mathbf{x}^n$ is defined as the approximation of $\mathbf{x}(n\Delta t)$ where Δt is a temporal step size. If we set an initial condition $\mathbf{x}^0 = \mathbf{x}(0)$, for given vector $\mathbf{x}^n$, we approximate the next time vector $\mathbf{x}^{n+1}$ as follows: We calculate four intermediate vectors $\mathbf{k}_1 = \mathbf{v}(\mathbf{x}^n)$, $\mathbf{k}_2 = \mathbf{v}(\mathbf{x}^n + 0.5\Delta t\mathbf{k}_1)$, $\mathbf{k}_3 = \mathbf{v}(\mathbf{x}^n + 0.5\Delta t\mathbf{k}_2)$, $\mathbf{k}_4 = \mathbf{v}(\mathbf{x}^n + \Delta t\mathbf{k}_3)$, then $\mathbf{x}^{n+1} = \mathbf{x}^n + \Delta t\,(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4)/6$. Finally, we calculate the value of function G of $\mathbf{x}^0$ using the given initial state $\mathbf{x}^0$.

$$G(\mathbf{x}^0, n\Delta t) = G(\mathbf{x}^n, 0).$$

If $r \neq 0$, we calculate the numerical integration of Eq. (15) using the Simpson method [24,25] on the characteristic curve, which have been solved by the Runge–Kutta method

$$G\left(\mathbf{x}^0, n\Delta t\right) = G(\mathbf{x}^n, 0)\exp\left(\frac{\Delta t}{3}\sum_{k=1}^{m}\left\{r\left(\mathbf{x}^{k-2}\right) + 4r\left(\mathbf{x}^{k-1}\right) + r\left(\mathbf{x}^k\right)\right\}\right),$$

where $n = 2m$ for $m = 1, 2, \ldots$.

In addition, we apply the multiple precision to the estimation of the marginal probability. When we estimate the marginal probability for a large number of species n_i, we need high-order differentiations. In this case, it is possible that unreasonable results can be obtained because the precision, which is the number of correct digits in some quantity, is limited. As a result of the finite number of digits used by computer machines, numbers are stored inexactly and operations are carried out inaccurately [25].

5 Numerical experiments

We perform numerical experiments containing two, three, and eight dimensional simulations consisting of a single gene model, a three-species model, and a gene transcription model.

5.1 Simulation of the single gene model

Figure 2 shows backward characteristic curves of the single gene model. We place starting points (∘) around the boundary. The temporal step size $\Delta t = 0.001$ is used to draw the curves. A dashed curve is associated with an initial point located on the line $x_1 = 0$ and a solid line is related to a corresponding circle on the line $x_1 = 1$. Likewise, one curve associated with an initial point located on the line $x_2 = 1$ cannot distinguish from the other curve. All curves which are associated with an initial point on the line $x_2 = 1$ may only move on the line $x_2 = 1$ and toward $\mathbf{x} = \mathbf{1}$.

The information of a point on a corresponding curve is transported to the circle along the curve. The single gene model has a critical curve and it is identical to a zero contour of the coefficient of G_1 term, that is $c_2x_1(x_2 - 1) + c_3(1 - x_1)$. Since the information direction of x_1 spreads out from the zero contour, there exists a critical curve. Curves approach the critical curve, and then these converge to $\mathbf{x} = \mathbf{1}$. The starting points are not used to estimate stochastic measures, but the figure is just illustrated for visualizing the information on the unit domain.

Figure 3 shows the evolution of the single gene model on the unit domain at $t = 0, 1, 10$. The mesh size is 33×33 and the temporal step size is $\Delta t = 0.001$. The initial condition is $G(\mathbf{x}, 0) = x_1^2x_2^4$ as shown in Fig. 3a. The value of function G is decreased by the damping term $c_1(x_1 - 1)G$. The parameter c_4 which determines the speed of information direction of x_2 is quite lower than c_1 which determines the damping rate. Figure 3b shows the progress that the values near the position $(0, 1)$

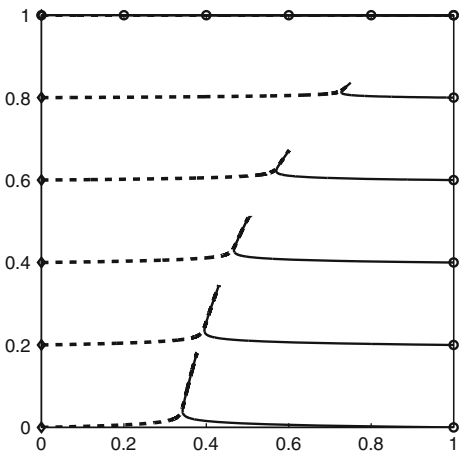

Fig. 2 Backward characteristic curves up to $t = 2$ of the single gene model

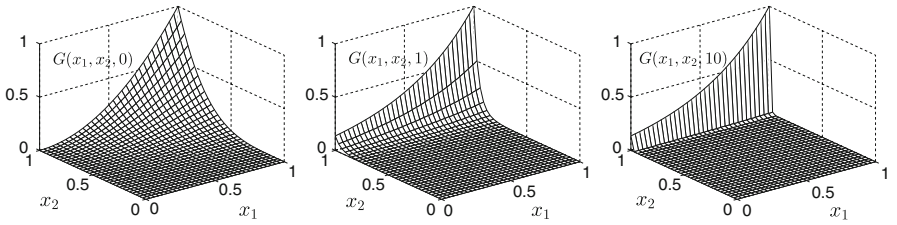

Fig. 3 Evolution of the single gene model. **a** $t = 0$, **b** $t = 1$, **c** $t = 10$

is developed and the gradient of surface is steeper than one of the initial surface in Fig. 3a. Finally, there is a region where the gradient is steep near $x_2 = 1$, see Fig. 3c.

Figure 4 shows results for the mean and variance of the single gene model. The solid and dashed lines represent the results of the Monte Carlo and characteristic methods, respectively. The estimation of the mean and variance is based on 50,000 realizations for the Monte Carlo method. The numerical parameters for the characteristic method are as follows: The spatial step size is $h = 5\text{E}{-5}$ and the temporal step size is $\Delta t = 5\text{E}{-3}$.

The results from the characteristic method are comparable to those from the Monte Carlo method. It is shown at the Appendix that the mean of n_1 is 2 for all time and the variance of n_1 rapidly increase and asymptotically approach 2. The mean and variance of n_1 from the characteristic method fit in well with the exact solutions. However, in contrast, the mean and variance of n_1 from the Monte Carlo method slightly fluctuate near 2. In terms of the mean of n_2, the results of two method are almost identical. From the numerical result of the characteristic method, we can reasonably predict that the asymptotic value is about 583.4.

Figure 5 shows marginal probabilities of the single gene model using the Monte Carlo and characteristic methods. The Monte Carlo and characteristic methods are associated with the solid and dashed lines, respectively. In the case of n_1, the circle

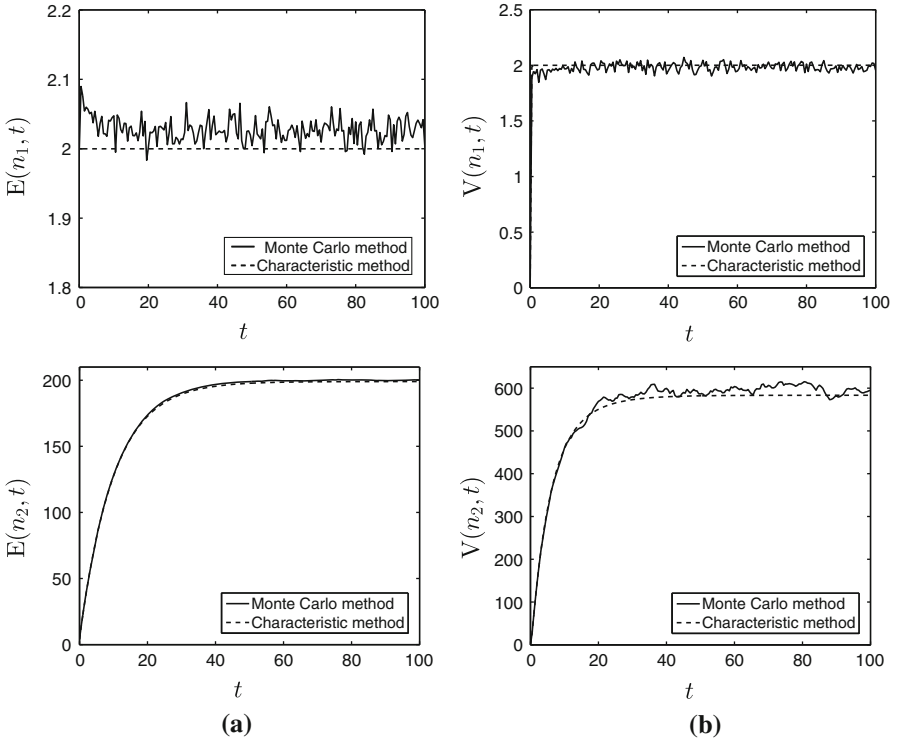

Fig. 4 Means and variables of the single gene model: first and second rows are with respect to n_1 and n_2, respectively. **a** Means, **b** Variances

(○), right triangle (▷), and diamond (◇) symbols are corresponding to marginal probabilities that $n_1 = 2$, 3, and 4, respectively. Meanwhile, in the case of n_2, the circle (○), right triangle (▷), and diamond (◇) symbols are corresponding to marginal probabilities that $n_2 = 4$, 6, and 8, respectively. The results of the Monte Carlo method are based on 50,000 realizations. On the other hand, for the characteristic method, the spatial step size $h = 1\text{E}{-10}$ and temporal step size $\Delta t = 0.01$ are used.

All marginal probabilities of n_1 go to steady state on the time interval [0, 1]. In terms of n_2, the marginal probability that $n_2 = 4$ monotonically decreases and approaches zero. On the other hand, the marginal probability that $n_2 > 4$ increases until a certain time and monotonically decreases and vanishes after some time. Two methods have similar behavior, but the results of the characteristic method are more smooth.

Figure 6 shows marginal probabilities that $n_2 = k$, where the value k is relatively larger than one of Fig. 5b. For long time simulation and higher order differentiation to estimate marginal probability, the numerical parameters for the characteristic method are chosen as follows: The number of digit is 2^{12}, the spatial step size is $h = 1\text{E}{-10}$, and the temporal step size is $\Delta t = 0.01$. The results from Monte Carlo method are based on 50,000 realizations. As shown in Fig. 4, the mean of n_2 increases up to the value 199. The results from the characteristic method show that the probability monotonically increases over time if the value of k is greater than 199.

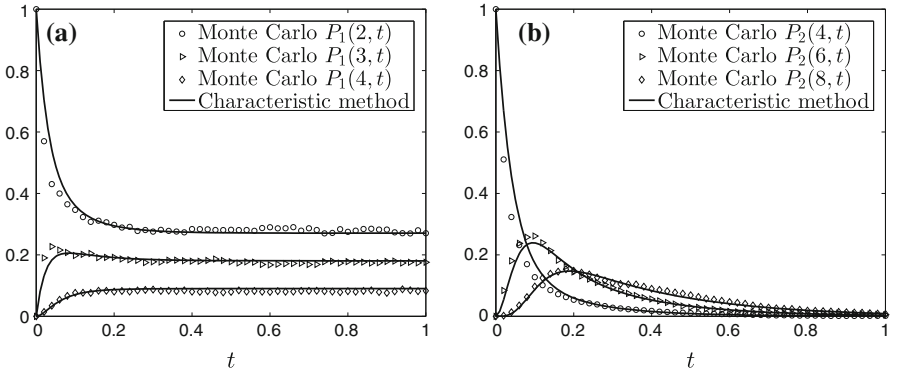

Fig. 5 The *symbols* and *solid lines* are probabilities of the single gene model using the Monte Carlo and characteristic method, respectively. **a** $P_1(k,t)$, **b** $P_2(k,t)$

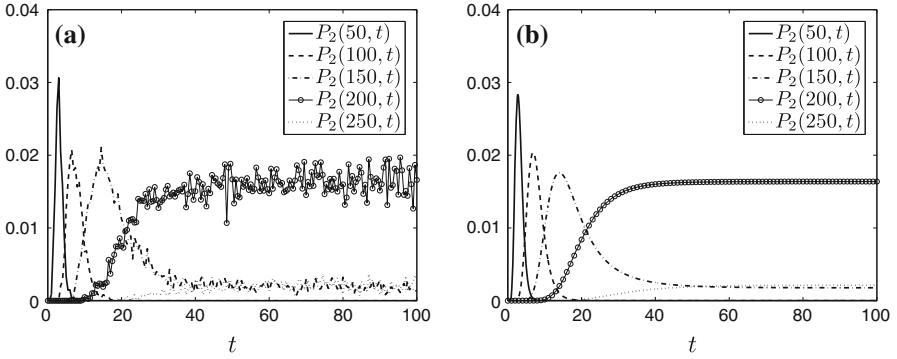

Fig. 6 Probabilities of the single gene model over long time. **a** Monte Carlo method, **b** Characteristic method

Figure 7a and b are the summation of probabilities $P_1(k,t)$ from $k=1$ to $k=10$ and the summation of probabilities $P_2(k,t)$ from $k=1$ to $k=300$, respectively. We are concerned about the probability that n_1 is from 0 to 10, because there is the case that the number of species n_1 is zero and the variance is about 2. On the other hand, we are interested in the probability that n_2 is from 1 to 300, because the mean is about 200 and the standard deviation distribution is bounded by the value 25. For both cases of n_1 and n_2, the summation is almost 1. The small decrease in the case of n_2 is due to the increase of the mean and variance of n_2.

5.2 Simulation of a three-species model

Figure 8 shows backward characteristic curves of the three-species model, whose the initial points are represented by circles (∘). To show the traveling speed of each curve, a curve is drawn by the dots (·) and solid line (-) which show the time evolution up to $t=1$ and $t=8$, respectively. In order to show backward characteristic curves, the temporal step size $\Delta t = 0.001$ is used.

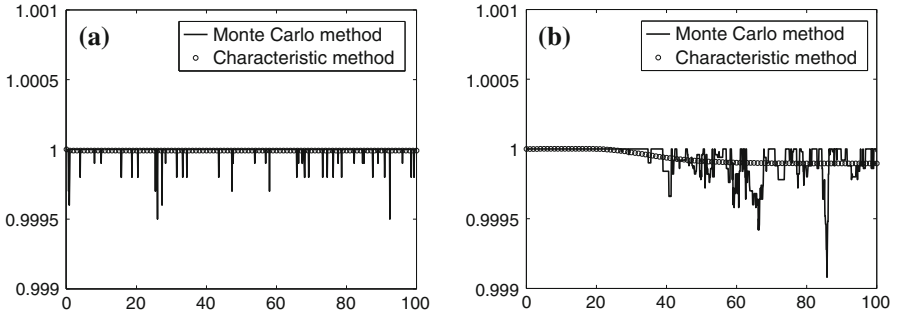

Fig. 7 Summations of probabilities of **a** the first 10 states with zero state for n_1 and **b** the first 300 states for n_2 in the single gene model **a** n_1, **b** n_2

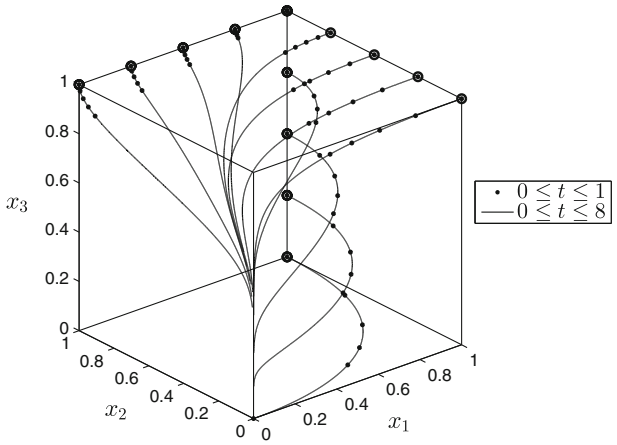

Fig. 8 Backward characteristic curves of the three-species model: *Dots* and *solid lines* are up to $t = 1$ and $t = 8$, respectively

Unlike the single gene model, backward characteristic curves are sucked in the point $\mathbf{x} = \underline{0}$ such as the spiral. The speed of curve is the faster, the closer the dots are on the solid line. The parameter c_3 is the coefficient of G_3 term of Eq. (10), and it is related to the speed of information to x_3 direction. In particular, a left curve corresponding to (0, 1, 1) is more sensitive to c_3 than a right curve corresponding to (1, 0, 1). Thus, the left curve is down to $\underline{0}$ more faster than the right curve. We can deduce that all values except $\mathbf{x} = \mathbf{1}$ will be zero. Therefore, there is a singularity at $\mathbf{x} = \mathbf{1}$.

For the three-species model, we illustrate the means and variances for n_1 and n_2, see Fig. 9. The Monte Carlo and characteristic methods have almost identical performance. The results from Monte Carlo method are based on 50,000 realizations. The numerical parameters for estimating the means and variance using the characteristic method are as follows: The spacial step size is $h = 1\text{E}{-4}$ and temporal step size is $\Delta t = 1\text{E}{-3}$. Being based on the spiral and singularity of Fig. 8, we can deduce that there is no steady state solution. All means and variances are exponentially increasing. The mean and variance of n_2 is more steeply increasing than those of n_1.

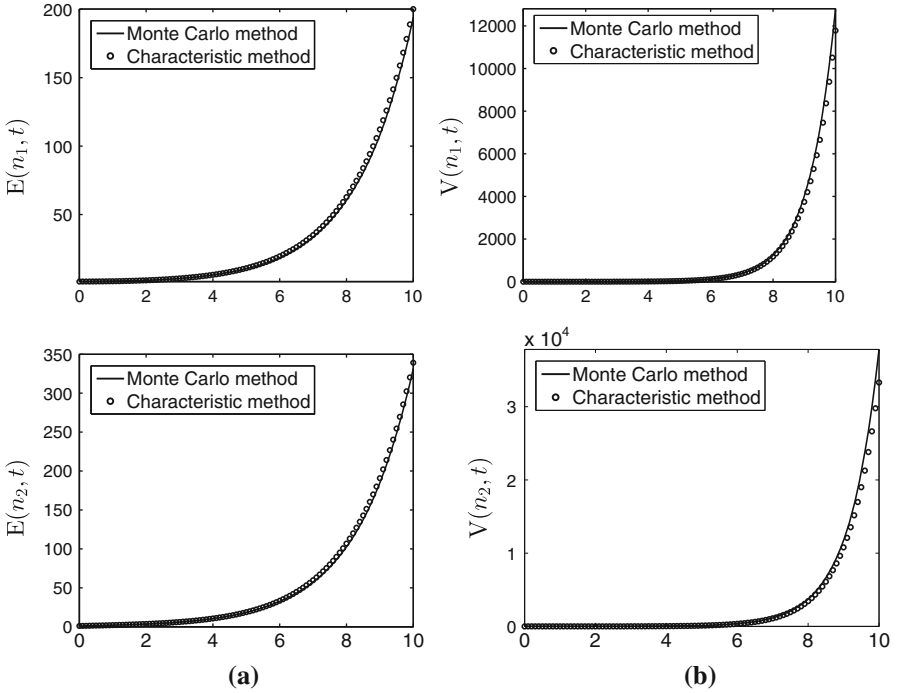

Fig. 9 **a** Means and **b** variances of n_1 and n_2 in the three-species model

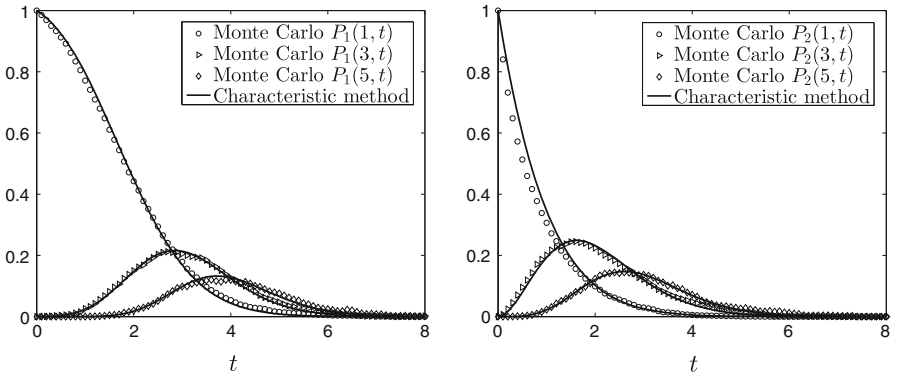

Fig. 10 Probabilities of the three-species model using Monte Carlo and characteristic methods

Figures 10 and 11 sfhow probabilities of the three-species model. In Fig. 10, the circle (∘), right triangle (▷), and diamond (◇) symbols are corresponding to marginal probabilities that $n_1 = 1,\ 3$, and 5, respectively. Likewise, in Fig. 11, the circle (∘), right triangle (▷), and diamond (◇) symbols are corresponding to marginal probabilities that $n_2 = 50,\ 100$, and 200, respectively. The results of the Monte Carlo method are based on 50,000 realizations. To estimate the result from the characteristic method,

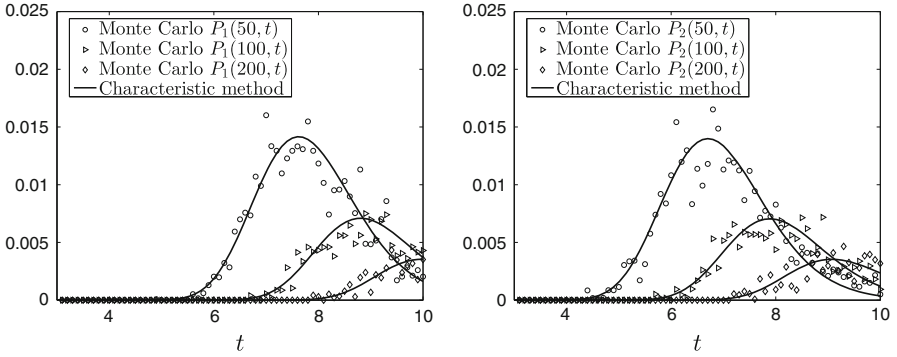

Fig. 11 Probabilities of the three-species model using the Monte Carlo and characteristic methods

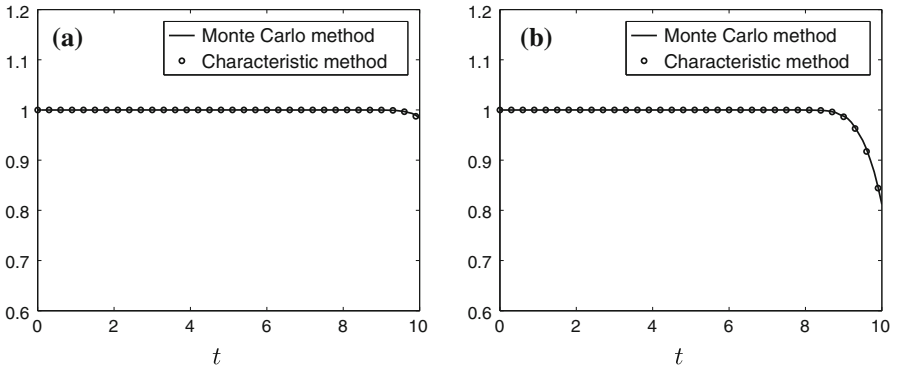

Fig. 12 Summations of probabilities of the first 500 states for both **a** n_1 and **b** n_2 in the three-species model

the following parameters are used. The number of digit is 2^{12}, the spatial step size is $h = 1\text{E}{-}10$, and the temporal step size is $\Delta t = 0.01$.

Two methods match well in the case of the lower species. However, the method of characteristics is more reasonable, because the results of the Monte Carlo method are fluctuant and numerically unstable. Using the characteristic method, we can find the time when the probability reaches its maximum, while it is difficult to find the appropriate time from the Monte Carlo method due to the fluctuations. Since the means and variances of the three-species model keep increasing, any probability eventually tends to zero.

Figure 12 shows the summation of probabilities $P_i(k,t)$ from $k = 1$ to $k = 500$ for $i = 1, 2$. In the three-species model, the increase of the mean and variance is continued, and we estimate up to $k = 500$. In both of Fig. 12a and b, the summation is decreasing in the last stage. Because the both mean keep increasing, the summation up to $k = 500$ is not sufficient to be 1 after some time.

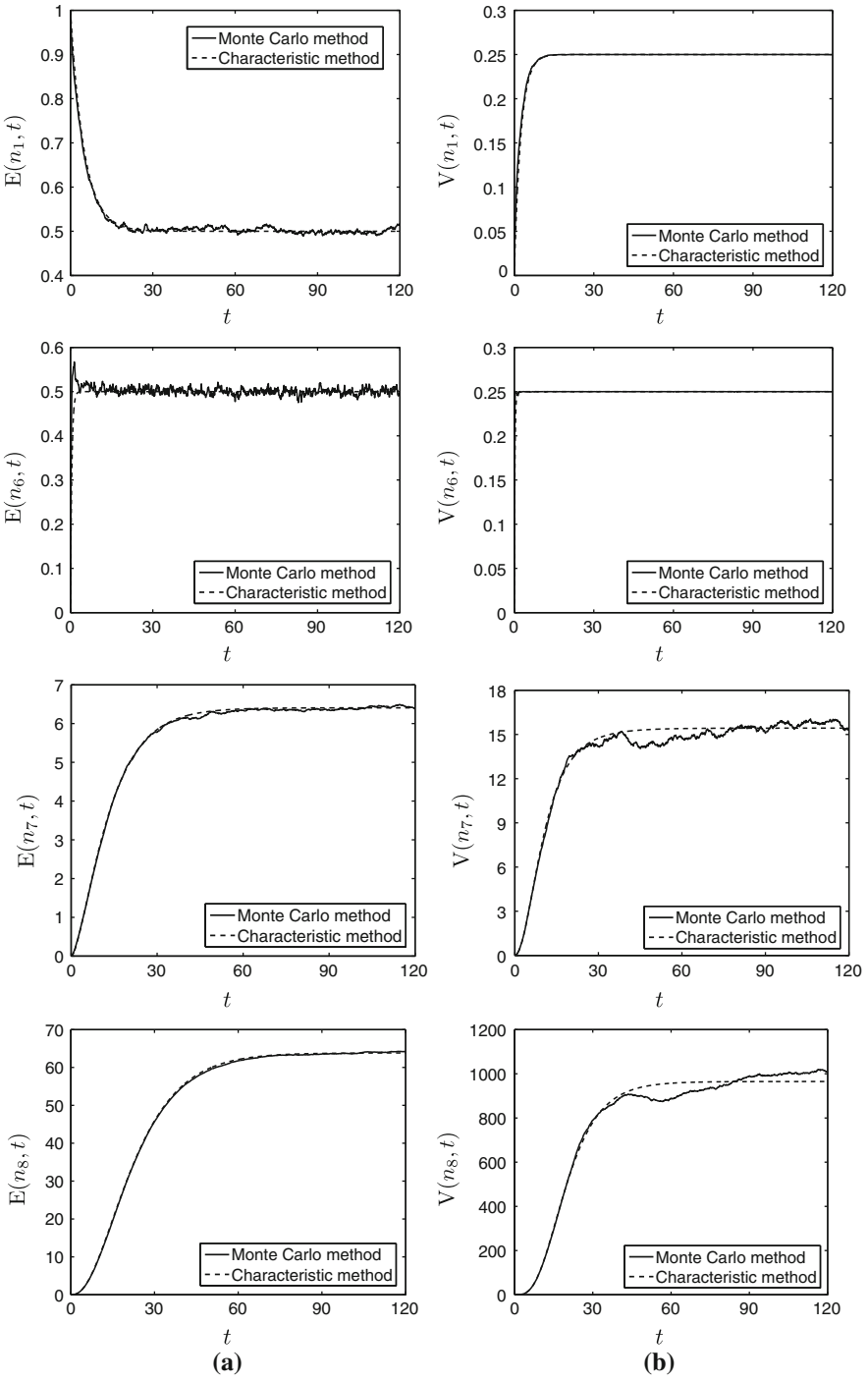

Fig. 13 **a** Means and **b** variances of the gene transcription model

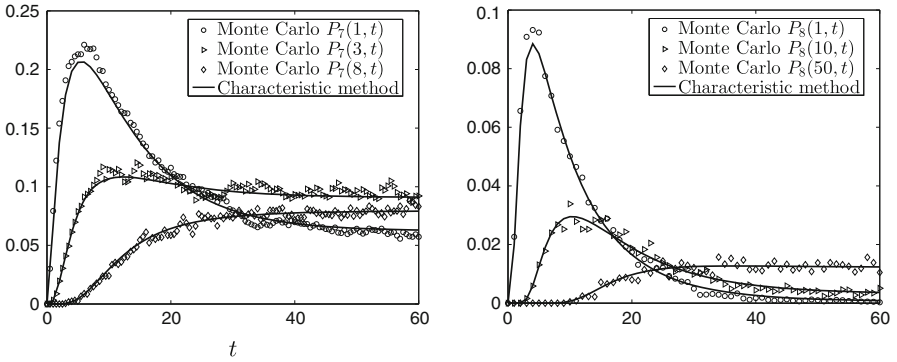

Fig. 14 Probabilities of the gene transcription model using the Monte Carlo and characteristic methods

5.3 Simulation of gene transcription model

Figure 13 shows means and variances for the gene transcription model. We illustrate the results with respect to n_1, n_6, n_7, and n_8. To estimate the mean and variance using the characteristic method, the following parameters are used: The spatial step size is $h = 0.0001$ and the temporal step size is $\Delta t = 0.1$. The results from the Monte Carlo simulation are based on 30,000 realizations. The Monte Carlo and characteristic methods have analogous results. The characteristic method is smooth and consistent, while the Monte Carlo method has fluctuations. the means of n_7 and n_8 increase up to about 6.4065 and 63.8321.

We estimate probabilities of n_7 and n_8, see Fig. 14. The results of the Monte Carlo method are based on 50,000 realizations. For the characteristic method, the following parameters are used: The number of digit is 2^{12}, the spatial step size is $h = 1\text{E}{-}10$, and the temporal step size is $\Delta t = 0.01$. In the case of n_7, the circle ($\circ$), right triangle ($\triangleright$), and diamond ($\diamond$) symbols are corresponding to the probability that $n_7 = 1$, 3, and 8, respectively. Otherwise, in the case of n_8, the circle ($\circ$), right triangle ($\triangleright$), and diamond ($\diamond$) symbols are corresponding to the probability that $n_8 = 1$, 10, and 50, respectively. Two methods have similar performance. Since both species n_7 and n_8 have no initial state, probabilities of lower populations start from 0 and increase at early time, and then they decrease after some time, and reach the steady-state. In contrast, probabilities of higher populations slowly increase and reach the steady-state.

Figure 15a shows the summation of probabilities $P_7(k,t)$ from $k = 1$ to $k = 20$. In the case of n_7, the standard deviation is less than $\sigma := 4$ and the mean of n_7 plus 3.2906σ is bounded by 20 over time. The probability that the population is from 1 to 20 is larger than 99.9 %. Fig. 15b shows the summation of probabilities $P_8(k,t)$ from $k = 1$ to $k = 100$. The maximum of the standard deviation is about 31.0641, but the number of probabilities for the summation is not sufficient to be one. Thus, the summation of probabilities falls down.

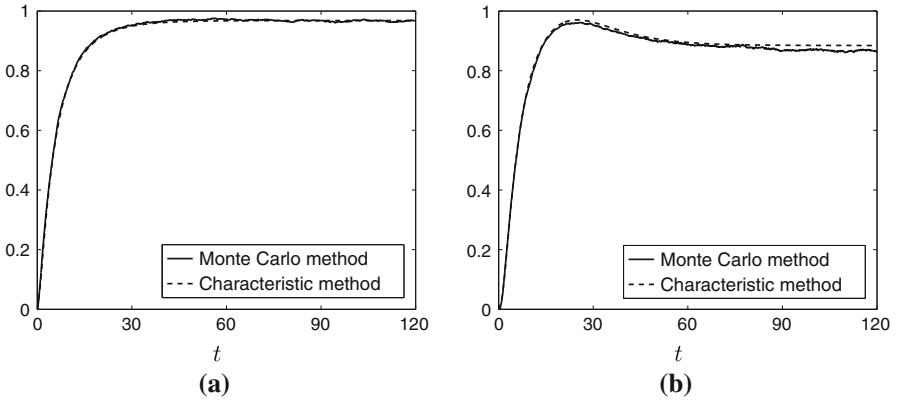

Fig. 15 Summation of probabilities of **a** the first 20 states for n_7 and **b** the first 100 states for n_8 in the gene transcription model

6 Conclusions

In this paper, we presented an accurate and efficient numerical algorithm for solving stochastic first-order reaction networks. The characteristic method is applied to solve the given governing equation. Because the first-order reaction PDE can be derived to the nonlinear ordinary differential system, the proposed method is quite fast even the high dimensional cases. If a first-order reaction network includes catalytic reactions and many molecules are involved in the dynamics of the network, the analytic solution of probability distribution cannot be found and the stochastic simulation algorithm needs heavy computations for finding the computational solution. The method presented in this paper gives an efficient and accurate way of finding the probability as well as the mean and variance for first-order reaction networks with catalytic reactions.

The characteristic method has three merits: (1) efficiency; The computational time is quite short because we just consider one or two points at which we need to estimate the mean and variance. Moreover, it is not affected by the additional dimension. (2) accuracy; Because we can use sufficiently small spatial step size h, we can control and obtain the high accuracy of the space. Finally (3) independence of the calculation of characteristic curves which makes the parallelism easy.

Acknowledgments The first author (Chang Hyeong Lee) was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0024849).

7 Appendix

We present the solution of the single gene equation (9) at $x_2 = 1$. The information direction of x_2 is downward, and it means that the solution on the line $x_2 = 1$ is independent to the value of the unit domain. The Eq. (9) is modified by

$$G_t = k_1(x_1 - 1)G - k_3(x_1 - 1)G_1,$$

where $k_1 = 10$ and $k_3 = 5$, and the closed-form solution is

$$G(x_1, 1, t) = \left((x_1 - 1)e^{-k_3 t} + 1\right)^2 e^{-\frac{k_1}{k_3}(x_1-1)e^{-k_3 t} + \frac{k_1}{k_3}(x_1-1)}.$$

In particular, the steady state solution is $G(x_1, 1, \infty) = \exp(k_1(x_1 - 1)/k_3)$. We evaluate the partial derivative with respect to x_1, and substitute the value 1 for x_1, then

$$\frac{\partial}{\partial x_1} G(x_1, 1, t)\bigg|_{x_1=1} = \frac{k_1}{k_3} = 2.$$

The second derivative can be calculated as follows:

$$\begin{aligned}\frac{\partial^2}{\partial x_1^2} G(x_1, 1, t)\bigg|_{x_1=1} &= \left(\frac{k_1}{k_3} - \frac{k_1}{k_3}e^{-k_3 t}\right)\left(2e^{-k_3 t} - \frac{k_1}{k_3}e^{-k_3 t} + \frac{k_1}{k_3}\right) \\ &= 4\left(1 - e^{-5t}\right).\end{aligned}$$

In particular, the value of $G_{11}(x_1, 1, t)$ asymptotically approaches 4 as t goes infinity. By the definition $E(n_1, t)$ and $V(n_1, t)$, $E(n_1, t) = 2$ and $V(n_1, t) = 2 - \exp(-5t)$ Therefore, the mean $E(n_1, t)$ is a constant function of 2, and the variance $V(n_1, t)$ rapidly increases but it is bounded by 2.

References

1. C.V. Rao, M.W. Wolf, A.P. Arkin, Control, exploitation and tolerance of intracellular noise. Nature **420**, 231–237 (2002)
2. D.T. Gillespie, A rigorous derivation of the chemical master equation. Phys. A **188**, 404–425 (1992)
3. M. Thattai, A. van Oudenaarden, Intrinsic noise in gene regulatory networks. Proc. Nat. Acad. Sci. **98**, 8614 (2001)
4. C. Gadgil, C.H. Lee, H.G. Othmer, A stochastic analysis of first-order reaction networks. Bull. Math. Biol. **67**, 901–946 (2005)
5. T. Jahnke, W. Huisinga, Solving the chemical master equation for monomolecular reaction systems analytically. J. Math. Biol. **54**, 1–26 (2007)
6. D.T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. **22**, 403–434 (1976)
7. D.T. Gillespie, Exact simulation of coupled chemical reactions. J. Phys. Chem. **81**, 2340–2361 (1977)
8. D.J. Higham, Modeling and simulating chemical reactions. SIAM Rev. **50**(2), 347–368 (2008)
9. M. Gibson, J. Bruck, Efficient exact stochastic simulation of chemical systems with many species and many channels. J. Phys. Chem. A **104**, 1876–1889 (2000)
10. D.T. Gillespie, The chemical Langevin equation. J. Chem. Phys. **113**, 297–306 (2000)
11. D.T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. **115**, 1716–1733 (2001)
12. E.L. Haseltine, J.B. Rawlings, Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. J. Chem. Phys. **117**, 6959 (2002)
13. C.V. Rao, A.P. Arkin, Stochastic chemical kinetics and the quasi-steady- state assumption: application to the Gillespie algorithm. J. Chem. Phys. **118**, 4999 (2003)
14. Y. Cao, D.T. Gillespie, L. Petzold, The slow-scale stochastic simulation algorithm. J. Chem. Phys. **122**, 014116 (2005)

15. J. Goutsias, Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. J. Chem. Phys. **122**, 184102 (2005)
16. H. Salis, Y. Kaznessis, Accurate hybrid stochastic simulation of coupled chemical or biochemical reactions. J. Chem. Phys. **122**, 0541031 (2005)
17. B. Munsky, M. Khammash, The finite state projection algorithm for the solution of the chemical master equation. J. Chem. Phys. **124**, 044104 (2006)
18. E. Weinan, D. Liu, E. Vanden-Eijnden, Nested stochastic simulation algorithm for chemical kinetic systems with multiple time scales. J. Comput. Phys. **221**, 158–180 (2007)
19. C.H. Lee, R. Lui, A reduction method for multiple time scale stochastic reaction networks. J. Math. Chem. **46**, 1292–1321 (2009)
20. J.M. Raser, E.K. O'Shea, Control of stochasticity in eukaryotic gene expression. Science **304**, 1811 (2004)
21. D.J. Higham, R. Khanin, Chemical master versus chemical langevin for first-order reaction networks. Open Appl. Math. J. **2**, 59–79 (2008)
22. S. Intep, D.J. Higham, X. Mao, Switching and diffusion models for gene regulation networks. Multiscale Model Simul. **8**(1), 30–45 (2009)
23. L.C. Evans, *Partial Differential Equations, Graduate studies in Mathematics*, 2nd edn. (American Mathematical Society, Providence, RI, 2010)
24. R. Burden, J. Faires, *Numerical Analysis* (Brooks/Cole, Boston, 2011)
25. A. Neumaier, *Introduction to Numerical Analysis* (Cambridge University Press, Cambridge, 2001)