Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 212 (2006) 6–24
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Accuracy limitations and the measurement of errors in the stochastic simulation of chemically reacting systems

Yang Cao *, Linda Petzold *,[1]

Department of Computer Science, University of California, Santa Barbara, CA, United States

Received 11 November 2004; received in revised form 9 May 2005; accepted 21 June 2005
Available online 18 August 2005

Abstract

This paper introduces the concept of distribution distance for the measurement of errors in exact and approximate methods for stochastic simulation of chemically reacting systems. Two types of distance are discussed: the Kolmogorov distance and the histogram distance. The self-distance, an important property of Monte-Carlo methods that quantifies the accuracy limitation at a given resolution for a given number of realizations, is defined and studied. Estimation formulas are established for the histogram and the Kolmogorov self-distance. These formulas do not depend on the distribution of the samples, and thus show a property of the Monte-Carlo method itself. Numerical results demonstrate that the formulas are very accurate. Application of these results to two problems of current interest in the simulation of biochemical systems is discussed.

1. Introduction

In microscopic systems formed by living cells, the small numbers of reactant molecules can result in dynamical behavior that is discrete and stochastic rather than continuous and deterministic [1–9]. The stochasticity (often called biochemical noise by biologists) in microscopic systems has been implicated in the lysis/lysogeny decision of the bacteria λ-phage [3] and the loss of synchrony of Circadian clocks [4]. To study the stochasticity in microscopic systems, engineered gene circuits have been designed and implemented in the laboratory. The effects of stochasticity have been observed in biological experiments [6–9].

* Tel.: +1 805 893 5362; fax: +1 805 893 5435.
E-mail address: petzold@engineering.ucsb.edu (L. Petzold).
[1] Department of Mechanical and Environmental Engineering, University of Santa Barbara, Santa Barbara, CA 93106-5070, United States.

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2005.06.012

Detailed models [1,2,10,11] for the expression of a single gene and gene networks have been proposed to explain these experiments.

The comparison between models and experiments is verified through Monte-Carlo simulations. Such simulations are based on Gillespie's stochastic simulation algorithm (SSA) [12,13], which yields an exact stochastic simulation for well-stirred chemically reacting systems. However, it can be prohibitively expensive for realistic biochemical simulation. Approximate simulation methods have been proposed, such as the τ-leaping method [14], implicit τ-method [15] and hybrid methods [16,17]. These methods can achieve greater efficiency and give a close approximation to the SSA method. Two natural questions are concerned here. One is: How should we measure the difference between the experiments and the Monte-Carlo simulations? The other is: How should we measure the accuracy of approximate methods? We seek a quantitative measurement. Typically, what is of interest from an experiment or a simulation are the stochastic properties of the solution variables or of some function of the solution values, as opposed to the values from one simulation. One possibility for measuring the error is to compute the errors in solution moments such as the mean and variance. However, often the problems for which stochastic simulation makes a big difference have a bistable distribution. A simple example of this type is given by the Schlögl [18] reaction. Some well-known problems from biology [3,7,19] also have this property. For such problems, the low-order moments such as mean and variance do not have much relevance. Rather, we need to know how well the detailed model captures the probability distribution, or how well approximate methods capture the 'exact' probability distributions for the variables and properties of interest. To describe this error, in this paper we adopt the concept of *distribution distance*. Two types of distribution distance are considered here. One is the Kolmogorov distance [20,21], defined to measure the distance between cumulative distribution functions (cdf). The other is the density distance area, defined as the L_1 distance between the probability density functions (pdf).

With the distribution distance, we can measure the accuracy of the distributions given by different approximation formulas. There are very few systems for which we can analytically solve for the distribution. Thus, for most problems, we must collect a large number of samples by experiments or by simulation methods such as SSA. The distribution is then estimated by the empirical distribution function [22] (edf), or the histogram of the samples. For a sufficiently large number of samples, the edf is close to the cdf, while the histogram is close to the pdf. However, in both experiment and computation the number of samples is always limited. Thus, such a process is subject to an inherent error due to the randomness of the variables of interest, which is traditionally called "*statistical fluctuation*" in the literature of Monte-Carlo simulations. In practice, we can measure this statistical fluctuation by the distance between two sets of independent samples with the same distribution. We call it "*self distance*". This concept is similar to that of the round-off error in classical numerical analysis. In that context, due to the limited length of the storage for each variable, there is a "*round-off error*". In stochastic simulation a simple limitation also exists, on the number of samples. For example, if there are two sets of samples for which the distribution distance between them is smaller than their self distance, we cannot tell whether or not these two sets of samples represent different distributions unless we increase the number of samples (the analogous solution in the situation of round off error is to increase the resolution of the floating point number representation). In numerical analysis, the round-off error is very small (2.22×10^{-16} in Matlab). Thus, in many cases, we do not need to worry about it. But in stochastic simulation, the self distance is usually much larger, so we must be aware of it. For example, in numerical studies of the convergence of τ-leaping methods [23], we have observed that regardless of how small the stepsize is, we cannot obtain a set of samples arbitrarily close to the samples generated by the SSA method. This is due to the self distance. In this paper, we introduce the concept of the self distance and discuss its quantitative estimation. An important feature of this estimation is that it does not depend on the distribution of the problem. This shows that self distance is actually an inherent property of the Monte-Carlo method itself.

The main contribution of this paper has two parts. First, we propose the use of distribution distance to measure the error in Monte-Carlo simulations. Second, we give estimation formulas for the self distances. The important feature is that these formulas are independent of problem distributions. We are aware of the shortcoming of our measurement. The two distribution distances discussed here are based on the distribution of a scalar variable. Although the density area distance can also be generalized to measure the distribution distance between two random vectors, as we will see later, the histogram self distance increases very rapidly as the dimension of the vector increases. However, we can always measure the distribution distance for each component of the vector. Moreover, in many practical problems, one is concerned not with the states themselves, but with a low dimensional "output function" of the states. For these problems, the measurement and error estimation discussed in this paper will still be very useful. We note that the distribution distance discussed in this paper is limited to the distance between distributions at stationary states. Thus, it is not directly applicable to the measurement of dynamical changes in the distribution. But it is not difficult to generalize the concept of distribution distance to handle the latter situation. We will briefly discuss this point in the last section.

The outline of this paper is as follows. In Section 2, we review the background of the SSA, τ-leaping methods, the Schlögl reaction, distribution functions and Kolmogorov distance. In Section 3, we introduce the concepts of the density distance area and the histogram distance, along with their applications to the study of the accuracy of τ-leaping methods. In Section 4, we discuss the self distance and present the estimate of the self distance. In Section 5, we present some numerical experiments. Finally, in Section 6, we briefly discuss the application of the distribution distance in two related areas and its generalization to measure the accuracy in the dynamical behavior.

2. Background

2.1. SSA and the τ-leaping method

We are concerned with a chemically reacting system with N species $\{S_1,\ldots,S_N\}$ and M reaction channels $\{R_1,\ldots,R_M\}$. The dynamical state of the system is denoted by $X=(X_1(t),\ldots,X_N(t))$, where $X_i(t)$ is the number of S_i molecules at time t. The dynamics of reaction channel R_j is characterized by the *propensity function* a_j and by the *state change vector* $v_j=(v_{1j},\ldots,v_{Nj})$: $a_j(x)\,\mathrm{d}t$ gives the probability that one R_j reaction will occur in the next infinitesimal time interval $[t,t+\mathrm{d}t)$, and v_{ij} gives the change in the S_i molecular population induced by one R_j reaction.

The dynamics of the system obeys the *chemical master equation* (CME)

$$\frac{\partial P(x,t|x_0,t_0)}{\partial t}=\sum_{j=1}^{M}[a_j(x-v_j)P(x-v_j,t|x_0,t_0)-a_j(x)P(x,t|x_0,t_0)], \tag{1}$$

where the function $P(x,t|x_0,t_0)$ denotes the probability that $X(t)$ will be x given that $X(t_0)=x_0$. The CME is hard to solve, both theoretically and numerically. An equivalent simulation method is the SSA [12,13]. SSA generates numerical realizations of $X(t)$. Both the CME and the SSA are exact consequences of the foregoing dynamical assumptions, so in spite of the difference in their descriptive thrusts, they are logically equivalent to each other.

The SSA [12,13] is based on the *next-reaction density function* $p(\tau,j|x,t)$ which is defined as the probability, given $X(t)=x$, that the next reaction in the system will occur in the infinitesimal time interval $[t+\tau,t+\tau+\mathrm{d}t)$ and will be an R_j reaction. It follows from the definition of a_j that

$$p(\tau,j|x,t)=a_j(x)\exp(-a_0(x)\tau)\quad(\tau\geqslant 0;\ j=1,\ldots,M), \tag{2}$$

where $a_0(x) = \sum_{j=1}^{M} a_j(x)$. The SSA generates τ and j according to (2) and proceeds with time τ and the state vector by

$$X(t+\tau) = X(t) + v_j.$$

The SSA is exact in the sense that it generates the same distribution given by the CME. But it is also very time-consuming because in each step, the simulation can only proceed with one reaction.

The τ-leaping method [14] tries to accelerate the simulation by asking a different question: How many times does each reaction channel fire in a subinterval of given length τ? In each step, the τ-leaping method can proceed with many reactions. This is achieved at the cost of some accuracy. Defining

$$\begin{aligned} K_j(\tau; x, t) = {} & \text{the number of times, given } X(t) = x, \\ & \text{that reaction channel } R_j \text{ will fire in the time interval } [t, t+\tau) \quad (j = 1, \ldots, M), \end{aligned} \tag{3}$$

the τ-leaping method begins by assuming the Leaping Condition: Require τ to be small enough that the change in the state during $[t,t+\tau)$ will be so slight that no propensity function will suffer an appreciable change in its value. Then $K_j(\tau;x,t)$ is given by the Poisson random variable

$$K_j(\tau; x, t) = P(a_j(x), \tau) \quad (j = 1, \ldots, M), \tag{4}$$

where $P(a,t)$ is the Poisson random variable with mean and variance at. The basic τ-leaping method is: Choose a value for τ that satisfies the Leaping Condition. Generate for each $j = 1,\ldots,M$ a sample value k_j of the Poisson random variable $P(a_j(x),\tau)$, and update the state by

$$X(t+\tau) = x + \sum_j k_j v_j. \tag{5}$$

Numerical experiments [14] have demonstrated the advantage of the τ-leaping method. A convergence analysis [23] has shown that this method is of order 1 for the mean and variance. But for a system with bistable states, the mean and variance do not have a physically significant meaning. If we want to study the convergence properties of the τ-leaping method for this type of problem, we should focus on the distribution rather than the first few moments. The Schlögl reaction is just such an example.

2.2. Schlögl reaction

The Schlögl model is famous for its bistable distribution. It is comprised of a set of coupled chemical reactions:

$$\begin{aligned} & B_1 + 2X \underset{c_2}{\overset{c_1}{\rightleftharpoons}} 3X, \\ & B_2 \underset{c_4}{\overset{c_3}{\rightleftharpoons}} X, \end{aligned} \tag{6}$$

where B_1 and B_2 denote buffered species whose respective molecular populations N_1 and N_2 are assumed to remain essentially constant over the time intervals of interest. The state change vectors are $v_1 = v_3 = 1$, $v_2 = v_4 = -1$. The propensity functions are:

$$\begin{aligned} a_1(x) &= \frac{c_1}{2} N_1 x(x-1), \\ a_2(x) &= \frac{c_2}{6} x(x-1)(x-2), \\ a_3(x) &= c_3 N_2, \\ a_4(x) &= c_4 x. \end{aligned} \tag{7}$$

For some parameter values, this model has two stable states. This is the case for the parameter set that we used in our simulation,

$$c_1 = 3 \times 10^{-7}, \quad c_2 = 10^{-4}, \quad c_3 = 10^{-3}, \quad c_4 = 3.5; \quad N_1 = 1 \times 10^5, \quad N_2 = 2 \times 10^5. \tag{8}$$

The SSA or the τ-leaping method can be applied to simulate the Schlögl model. The histogram generated from SSA is shown in Fig. 1.

2.3. Distribution functions

A number of basic concepts and functions are important in the study of distributions. In the following discussion, we focus on a scalar random variable X. The cumulative distribution function (cdf) is defined as

$$F_X(x) = P(X \leqslant x). \tag{9}$$

Another important function is the probability density function (pdf). For a continuous distribution, the pdf is defined as

$$p_X(x)\,\mathrm{d}x = P(x \leqslant X < x + \mathrm{d}x), \tag{10}$$

while for a discrete distribution, the pdf is the δ-function given by

$$p_X(x) = \sum_x P(X = x)\delta(X - x). \tag{11}$$

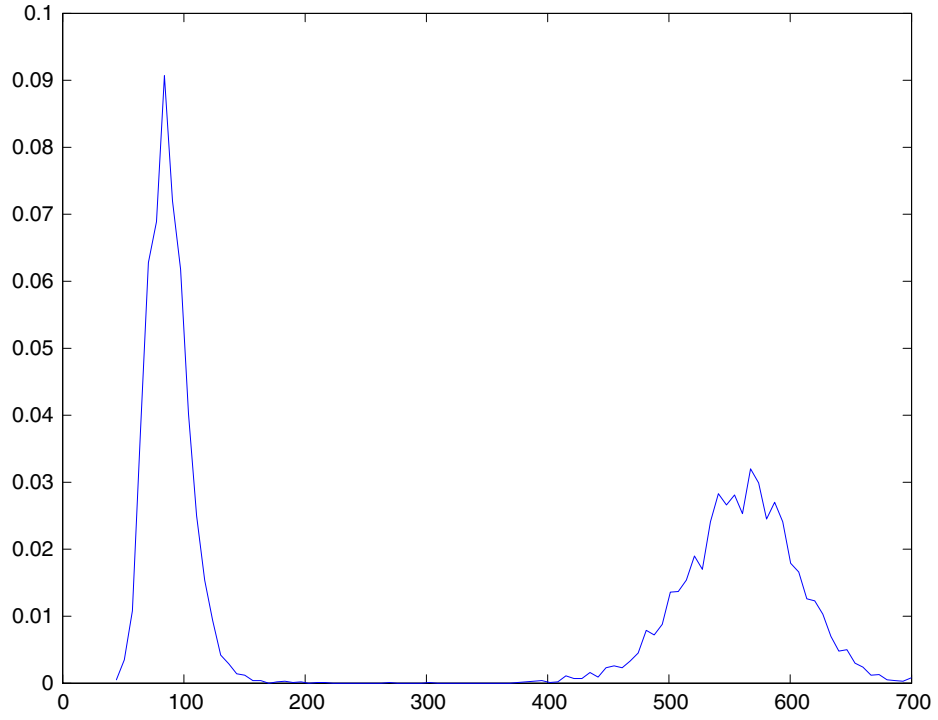

Fig. 1. Histogram (10,000 samples) of X solved by the SSA method for the Schlögl reaction with $x(0) = 250$ and final time $T = 4$.

It follows that

$$F_X(x) = \int_{-\infty}^{x} p_X(s)\,\mathrm{d}s. \tag{12}$$

In many practical problems, it is difficult or impossible to obtain an analytic distribution. Instead we obtain samples from Monte-Carlo simulations or observations. With those samples, the empirical distribution function (edf) is used to measure the cdf, and the histogram is used to measure the pdf. Let $x_1,x_2,\ldots,x_N$ be independent random variables each having the same distribution as X. Defining the sign function

$$\kappa(x) = \begin{cases} 1, & x \geqslant 0, \\ 0, & x < 0, \end{cases} \tag{13}$$

the empirical distribution function is defined as

$$F_N(x) = \frac{1}{N}\sum_{j=1}^{N} \kappa(x - x_j). \tag{14}$$

The sum in (14) gives the number of points that are smaller than x. When it is divided by N, we obtain the fraction of points smaller than x, which approximates the cdf. Suppose that all the sample values are bounded in the interval $I = [x_{\min},x_{\max})$. Let $L = x_{\max} - x_{\min}$. Divide the interval I into K subintervals and denote the subintervals by $I_i = [x_{\min} + \frac{(i-1)L}{K}, x_{\min} + \frac{iL}{K})$. We define the characteristic function $\chi(x,I_i)$ as

$$\chi(x, I_i) = \begin{cases} 1 & \text{if } x \in I_i, \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

Then the pdf p_X can be approximated by the histogram function h_X computed from

$$h_X(I_i) = \frac{K}{NL}\sum_{j=1}^{N} \chi(x_j, I_i). \tag{16}$$

The sum in (16) gives the number of points falling into the interval I_i. When that sum is divided by N we get the fraction of the points inside that interval, which approximates the probability of a sample point lying inside that interval. We divide this by the interval length, L/K, to approximate the probability density. Thus, $h_X(I_i)$ measures the average density function of X in the interval I_i. When K tends to infinity, the length of I_i reduces to 0. Then I_i is close to a point and h_X is close to p_X at that point.

2.4. Kolmogorov distance

For systems with a bistable distribution, such as the Schlögl reaction, the mean and variance do not have a physically significant meaning. Instead we are more concerned with the distribution. Thus, we need to measure the distance between distributions. One such measurement defined in information theory is the Kolmogorov distance [20].

The Kolmogorov distance is the ∞-distance of the cdf. For two random variables X and Y with cdf F_X and F_Y, the Kolmogorov distance is defined as

$$K(X, Y) = \max_{-\infty<x<\infty} |F_X(x) - F_Y(x)|. \tag{17}$$

Note that the Kolmogorov distance is a function distance, as opposed to a variable distance. It has the following properties:

$$\begin{aligned} &\text{Scaling property}: \quad K(\lambda X, \lambda Y) = K(X, Y), \\ &\text{Bound property}: \quad K(X, Y) \leqslant 1. \end{aligned} \tag{18}$$

The Kolmogorov distance is used in the Kolmogorov–Smirnov test [21,22] for two classical problems in statistics: the goodness-of-fit problem and the two-sample problem. Let $X_1, X_2, \ldots, X_N$ be independent random variables following the same distribution $U(x) = Pr(X_i < x)$. The *goodness-of-fit* problem is to devise a test of the hypothesis

$$H_0: \quad U(x) = F(x), \tag{19}$$

where $F(x)$ is a given distribution function. Let $Y_1, Y_2, \ldots, Y_M$ be independent random variables with the common distribution $V(x) = Pr(Y_i < x)$. The *two-sample* problem is to devise a test of the hypothesis

$$H_0': \quad U(x) = V(x). \tag{20}$$

Note that the two-sample problem is related to our problem of measuring the accuracy of the approximation formula. Thus, the Kolmogorov distance is a natural candidate for our purposes. In practice, we measure the Kolmogorov distance of the edf

$$K_{M,N}(X, Y) = \max_{-\infty<x<\infty} |F_{N,X}(x) - F_{M,Y}(x)|, \tag{21}$$

where $F_{N,X}$ and $F_{M,Y}$ are the corresponding edf for the two groups of samples X_i and Y_j. Smirnov [24,25] proved that, when $M \to \infty$, $N \to \infty$,

$$S_{mn} = \sqrt{\frac{MN}{M+N}} K_{M,N}(X, Y) \tag{22}$$

has a limiting distribution

$$\Phi(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k \mathrm{e}^{-2k^2\lambda^2}. \tag{23}$$

A short table [24] and an amplified table [26] are given for Φ values by Smirnov. For a given probability tolerance, if S_{mn} is larger than a particular value in the table, H_0' is rejected. Otherwise, it is accepted.

3. Density distance area

The Kolmogorov distance is a distance measure for the cdf. When the pdf is concerned, we need a distance measure for it. There are several definitions of distribution distance for the pdf in the literature, for example the Kullback–Leibler distance [27], the total variation distance [28] and the χ^2 distance [29]. Suppose X and Y have probability density functions p_X and p_Y. We define the *density distance* between X and Y as

$$D(X, Y) = \int |p_X(s) - p_Y(s)| \, \mathrm{d}s. \tag{24}$$

When X and Y are integers, (24) becomes

$$D(X, Y) = \sum_n (|P(X = n) - P(Y = n)|). \tag{25}$$

To simplify the derivation, we focus on (24), which gives the L_1 distance of the pdf. We call this the density difference area. According to Rosenthal [28], the density difference area is twice the total variation distance. But this form is numerically easier to calculate. It is easy to derive some properties of the density distance area. In particular, we have:

$$\begin{aligned} &\text{Scaling property}: \quad D(\lambda X, \lambda Y) = D(X, Y), \\ &\text{Bound property}: \quad D(X, Y) \leqslant 2. \end{aligned} \tag{26}$$

Since the Kolmogorov distance can also be written as

$$K(X,Y) = \max_{-\infty<x<\infty} \left| \int_{-\infty}^{x} p_X(s) - p_Y(s)\, \mathrm{d}s \right|, \tag{27}$$

it is obvious that

$$K(X,Y) \leqslant D(X,Y). \tag{28}$$

Both the Kolmogorov distance and the density distance area can be used to measure accuracy. A notable difference between them is that the density distance area takes the absolute value of the density difference into account, while the Kolmogorov distance takes the sign into account. Because the signed differences may cancel each other, the Kolmogorov distance may underestimate the difference.

In Monte-Carlo simulation, the pdf $p(x)$ is approximated by the histogram $h(x)$. For two groups of samples X_i and Y_j, we have the *histogram distance*

$$D_K(X,Y) = \sum_{i=1}^{K} \frac{|h_X(I_i) - h_Y(I_i)|L}{K}. \tag{29}$$

Substituting (16) into (29), we obtain

$$D_K(X,Y) = \sum_{i=1}^{K} \left| \frac{\sum_{j=1}^{N} \chi(x_j, I_i)}{N} - \frac{\sum_{j=1}^{M} \chi(y_j, I_i)}{M} \right|. \tag{30}$$

$D_K(X,Y)$ varies depending on the value of K. When $K = 1$ there is only one subinterval and we cannot tell the difference between X and Y. When K becomes larger we obtain more detailed information about the difference, and $D_K(X,Y)$ will increase. When K is very large we must generate a large number of samples, otherwise there will not be enough data falling into each subinterval and there will be a large measurement error. When K, N and M are sufficiently large, the histogram distance $D_K(X,Y)$ is close to the density distance area $D(X,Y)$,

$$D_K(X,Y) \to D(X,Y) \quad \text{as } K,N,M \to \infty. \tag{31}$$

It can be verified that the histogram distance inherits the scaling and bound properties of the density distance area:

$$\begin{aligned} &\text{Scaling property}: \quad D_K(\lambda X, \lambda Y) = D_K(X,Y), \\ &\text{Bound property}: \quad D_K(X,Y) \leqslant 2. \end{aligned} \tag{32}$$

With the Kolmogorov distance and the histogram distance, we can measure the accuracy of an approximation formula. We applied the τ-leaping method with different τ's to the Schlögl reaction and compared the distribution distance to the samples generated from the SSA method. Table 1 shows the corresponding results with 10,000 samples for each τ.

In Table 1, both the Kolmogorov distance and the histogram distance decrease when τ decreases. This numerically demonstrates that the solution obtained by the τ-leaping method converges to that obtained by the SSA as $\tau \to 0$, which agrees with the convergence analysis of the τ-leaping method [15]. We note two observations from Table 1. First, the Kolmogorov distance at $\tau = 0.4$ is smaller than the Kolmogorov distance at $\tau = 0.04$, while for the histogram distance the distance at $\tau = 0.04$ is smaller. This is due to the cancellation effect of the Kolmogorov distance. Fig. 2 shows the plot of the edf's given by SSA and the τ-leaping method with $\tau = 0.4$. We can see that the edf difference has a sign change in the middle. As a result, the Kolmogorov distance underestimates the difference, while the histogram distance gives a more meaningful estimate for our purposes. But the histogram distance depends on the bin number K. It increases as K increases.

Table 1
Kolmogorov distance and histogram distance comparison (10,000 samples) between the distributions given by the SSA and the τ-leaping method applied to the Schlögl reaction

Kolmogorov distance between SSA and the τ-leaping method							
τ	0.4	0.1	0.04	0.01	0.004	0.001	0.0001
	0.0681	0.0222	0.0714	0.0095	0.0083	0.0097	0.0117
Histogram distance between SSA and the τ-leaping method							
K/τ	0.4	0.1	0.04	0.01	0.004	0.001	0.0001
50	0.2856	0.0976	0.1886	0.0712	0.0632	0.0570	0.0588
100	0.2888	0.1142	0.1954	0.0904	0.0878	0.0860	0.0874
200	0.3056	0.1542	0.2192	0.1242	0.1212	0.1356	0.1272

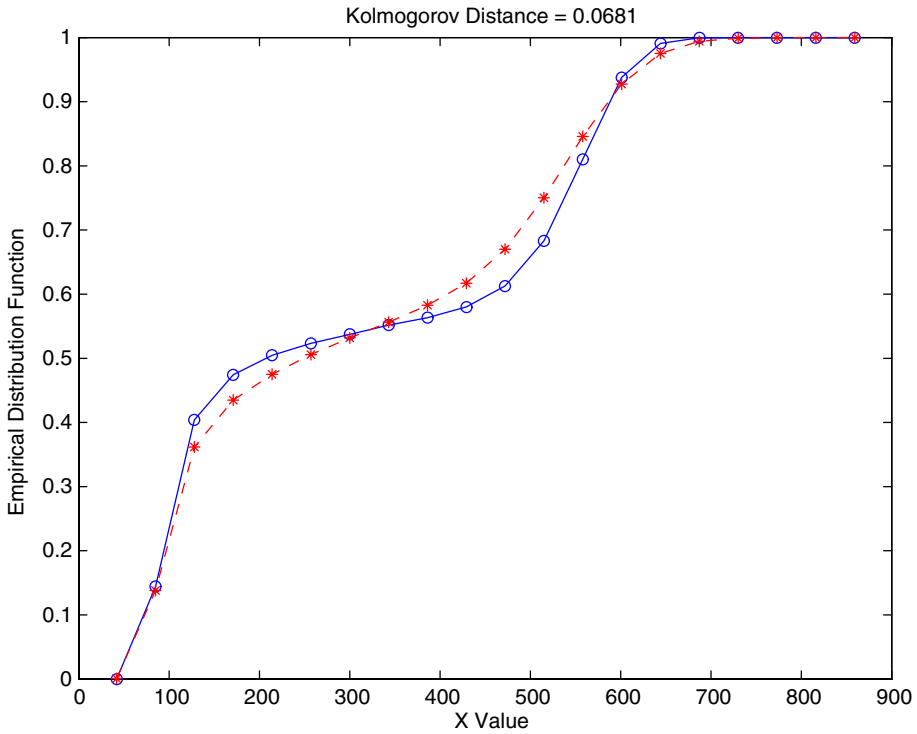

Fig. 2. The edf given by SSA (solid line with '$\circ$') and the τ-leaping method (dashed line with '$*$') with $\tau = 0.4$. for the Schlögl reaction with $x(0) = 250$ and final time $T = 4$.

Second, since the τ-leaping method converges to the SSA method, in the deterministic case one would expect the distance to decrease to 0 as $\tau \to 0$. But for this example, when $\tau < 0.01$, these distribution distances do not decrease further as $\tau \to 0$. This is related to an importance feature when we measure the distribution difference of Monte-Carlo methods. We will call this property the self distance.

4. Self distance

Monte-Carlo methods are based on a limited number of simulations. Although the theory states that the simulation converges to the real distribution when the number of samples tends to infinity, a limited num-

ber of samples will always result in some difference from the real distribution. Moreover, any set of samples has some distance from any other set of samples. For example, compare one set of 10,000 Monte-Carlo samples with another set of 10,000 Monte-Carlo samples. For each set, we can generate a histogram or an edf. To simplify the discussion, in the following we focus on the histogram. Due to the randomness, the two histograms are not the same. We call the distribution distance between the two sets of samples the *self distance*.

Definition 4.1. For two sets of samples that independently follow the same distribution, $X_N = \{x_1, \ldots, x_N\}$ and $Y_M = \{y_1, \ldots, y_M\}$, the self distance is defined as the distribution distance between the two samples X_N and Y_M. The Kolmogorov self distance is $K(X_N, Y_M)$ and the histogram self distance is $D_K(X_N, Y_M)$.

Self distance reflects the nature of stochastic simulation. It plays a similar role to that of round-off error in classical numerical analysis. Also it is related to the classical *two-sample* problem in statistics. When the distribution distance of the two samples is less than the self distance of one sample, we would accept the hypothesis H_0'. Since the two sets of samples are random, the self distance is also random. The Kolmogorov and Smirnov series of work [20,21,24–26] have shown that the Kolmogorov self distance has a distribution which is distribution free (in the sense that it does not depend on the distribution of the samples). In this section, we will show that a similar result applies to the histogram self distance. In the following, unless specified, the self distance means the histogram self distance.

From Definition 4.1, the self distance depends on K, N, M and the distribution of the two sets of samples. In particular, we might expect a large difference in the self distance for different distributions. But for sufficiently large N and M, the self distance for different distributions follows a common bound. We have the following theorem.

Theorem 4.1. *For sufficiently large N and M, the mean of the histogram self distance is bounded by $\sqrt{\frac{2K}{\pi}\left(\frac{1}{N}+\frac{1}{M}\right)}$. The variance of the self distance is bounded by $\frac{(\pi-2)K}{\pi}\left(\frac{1}{N}+\frac{1}{M}\right)$.*

Proof. We first need the following two lemmas.

Lemma 4.2. *For a random variable X that follows the normal distribution $N(0, \sigma^2)$, the mean value of the absolute value $|X|$ satisfies*

$$E(|X|) = \sqrt{\frac{2}{\pi}}\sigma. \tag{33}$$

The variance of $|X|$ satisfies

$$\mathrm{Var}(|X|) = \frac{\pi - 2}{\pi}\sigma^2. \tag{34}$$

Lemma 4.3. *For the sum of random variables, we have*:

$$E\left(\sum_{i=1}^{K} X_i\right) = \sum_{i=1}^{K} E(X_i), \tag{35}$$

$$\sqrt{\mathrm{Var}\left(\sum_{i=1}^{K} X_i\right)} \leqslant \sum_{i=1}^{K} \sqrt{\mathrm{Var}(X_i)}. \tag{36}$$

These two lemmas can be easily proved (see Appendix).

We begin from a detailed analysis of the histogram method. The domain is divided into K subintervals I_i, $i = 1, \ldots, K$. On each subinterval, the probability that the random variable X will be in this subinterval is $p_i = \int_{\frac{(i-1)L}{K}}^{\frac{iL}{K}} p_X(s)\,\mathrm{d}s$. In the histogram method, this is approximated by

$$q_i = \frac{1}{N}\sum_{j=1}^{N} \chi(x_j, I_i). \tag{37}$$

We note that p_i is a deterministic variable, while q_i is a random variable. The self distance arises due to the randomness of q_i.

Let i be fixed. Because samples $x_1, \ldots, x_N$ are independent of each other, for each j, $\chi(x_j, I_i)$ is an independent random variable which follows the distribution $P(\chi(x_j, I_i) = 1) = p_i$, $P(\chi(x_j, I_i) = 0) = 1 - p_i$. Thus, the sum $B_i = \sum_{j=1}^{N} \chi(x_j, I_i)$ has a binomial distribution $B(p_i, N)$ [30]. The mean and variance of B_i are given by the classical result [30] for binomial distributions,

$$E(B_i) = Np_i, \quad \mathrm{Var}(B_i) = Np_i(1 - p_i). \tag{38}$$

From (38), we can derive the mean and variance for $q_i = B_i/N$,

$$E(q_i) = p_i, \quad \mathrm{Var}(q_i) = \frac{p_i(1 - p_i)}{N}. \tag{39}$$

On the other hand, in (37), q_i is the average value of N independent random variables $\chi(x_j, I_i)$ satisfying the same distribution. From the central limit theorem, when N is sufficiently large, q_i follows the normal distribution $N\left(p_i, \frac{p_i(1-p_i)}{N}\right)$. For the same reason, for another independent set of realizations $\hat{x}_1, \ldots, \hat{x}_M$, which also obey the distribution of X, we have $\hat{q}_i = \frac{1}{M}\sum_{j=1}^{M} \chi(\hat{x}_j, I_i)$. Thus, when M is sufficiently large, $\hat{q}_i$ follows the normal distribution $N\left(p_i, \frac{p_i(1-p_i)}{M}\right)$. To consider the self distance, we study the difference $d_i = q_i - \hat{q}_i$. Since q_i and $\hat{q}_i$ are independent random variables with normal distributions, the difference d_i also has the normal distribution. The corresponding mean value is $E(d_i) = E(q_i) - E(\hat{q}_i) = 0$. The variance is $p_i(1 - p_i)\left(\frac{1}{N} + \frac{1}{M}\right)$. Thus, d_i satisfies the normal distribution $N\left(0, p_i(1 - p_i)\left(\frac{1}{N} + \frac{1}{M}\right)\right)$. From Lemma 4.2, the mean value of $|d_i|$ is

$$E(|d_i|) = \sqrt{\frac{2}{\pi} p_i(1 - p_i)\left(\frac{1}{N} + \frac{1}{M}\right)} \tag{40}$$

and the variance of $|d_i|$ is

$$\mathrm{Var}(|d_i|) = \frac{\pi - 2}{\pi} p_i(1 - p_i)\left(\frac{1}{N} + \frac{1}{M}\right). \tag{41}$$

The self distance is defined as $\sum_{i=1}^{K} |d_i|$. Thus, the mean of the self distance can be bounded by

$$\begin{aligned} E\left(\sum_{i=1}^{K} |d_i|\right) &= \sum_{i=1}^{K} E(|d_i|) = \sum_{i=1}^{K} \sqrt{\frac{2}{\pi} p_i(1 - p_i)\left(\frac{1}{N} + \frac{1}{M}\right)} \leqslant \sum_{i=1}^{K} \sqrt{\frac{2}{\pi}\left(\frac{1}{N} + \frac{1}{M}\right)}\sqrt{p_i} \\ &\leqslant \sqrt{\frac{2}{\pi}\left(\frac{1}{N} + \frac{1}{M}\right)}\sqrt{K\sum_{i=1}^{K} p_i} \quad \text{(by Cauchy inequality)} \leqslant \sqrt{\frac{2K}{\pi}\left(\frac{1}{N} + \frac{1}{M}\right)}. \end{aligned} \tag{42}$$

Because the d_i's are not independent to each other, we do not have a sharp bound for the variance. Applying the inequality (36). we obtain the inequality

$$\mathrm{Var}\left(\sum_{i=1}^{K}|d_i|\right) \leqslant \left(\sum_{i=1}^{K}\sqrt{\mathrm{Var}(|d_i|)}\right)^2 = \left(\sum_{i=1}^{K}\sqrt{\frac{\pi-2}{\pi}p_i(1-p_i)\left(\frac{1}{N}+\frac{1}{M}\right)}\right)^2$$
$$\leqslant \frac{(\pi-2)K}{\pi}\left(\frac{1}{N}+\frac{1}{M}\right). \qquad \square \tag{43}$$

Remark 4.1. In our experience, the bound on the variance (43) is much larger than the observed variance. The variance can be estimated much more accurately in the following. Practically when both N and M are large, d_i may be treated as independent. Thus, the variance can be estimated by

$$\mathrm{Var}\left(\sum_{i=1}^{K}|d_i|\right) \approx \sum_{i=1}^{K}\mathrm{Var}(|d_i|) = \sum_{i=1}^{K}\frac{\pi-2}{\pi}p_i(1-p_i)\left(\frac{1}{N}+\frac{1}{M}\right) \leqslant \frac{(\pi-2)}{\pi}\left(\frac{1}{N}+\frac{1}{M}\right). \tag{44}$$

Then we have the following conjecture.[2]

Conjecture 4.1. *For sufficiently large N and M, the variance of the self distance can be estimated by* $\frac{\pi-2}{\pi}\left(\frac{1}{N}+\frac{1}{M}\right)$.

Remark 4.2. For a practical estimate of the self distance, we use only the bound for the mean $\sqrt{\frac{2K}{\pi}\left(\frac{1}{N}+\frac{1}{M}\right)}$. When $N = M$, the bound becomes $\sqrt{\frac{4K}{\pi N}}$. Note that the self distance increases when K increases. Thus, if we want to know more detailed information about a distribution, we will introduce a larger self distance. On the other side, the self distance is proportional to the inverse of the square root of N. Thus, if we want to obtain an r times more accurate distribution, we need r^2 times more samples.

Remark 4.3. In this paper, we focus on scalar random variables. The histogram distance can be generalized to measure the distribution distance of multi-dimension random vectors, In this case, K will be the product of K_i, where K_i is the number of splitting on the ith dimension. It is easy to see that as the dimension increases, K increases exponentially with the dimension. According to (42), the self distance may increase dramatically as well. Thus, high dimensional histogram distance is not practical due to the self distance.

Remark 4.4. This bound explains Table 1. For $\tau \geqslant 0.01$, the error is due to the error of the τ-leaping method. Around $\tau = 0.01$, the distance is within the self distance of the SSA method. Thus, when we further decrease τ, the self distance always remains. We cannot obtain a smaller distance unless we increase the number of samples. This is very similar to the situation in classical numerical analysis of ordinary differential equations. Usually, we are only concerned with the truncation error of an integration formula. But when the stepsize is very small, the truncation error may have the same magnitude as the round-off error. In that case, we cannot improve the accuracy further unless we change the arithmetic precision. The difference here is only that the self distance is relatively much larger than round-off error.

For the Kolmogorov self distance, we can also estimate the mean value. We have the following theorem.

Theorem 4.4. *For a continuous distribution function and sufficiently large N and M, the mean of the Kolmogorov self distance is bounded by* $\sqrt{\frac{\pi}{2}}\log 2\sqrt{\frac{1}{N}+\frac{1}{M}}$. *The variance of the Kolmogorov self distance is bounded by* $\left(\pi^2/12-\frac{\pi}{2}\log^2 2\right)\left(\frac{1}{N}+\frac{1}{M}\right)$

Proof. According to Kolmogorov [20,21,25], for two groups of samples X_i and Y_j with the same continuous distribution function and sufficiently large N and M, $S_{mn} = \sqrt{\frac{MN}{M+N}}K_{M,N}(X,Y)$ has the limiting distribution (23), which can be rewritten as

[2] This is called a conjecture because the derivation of this formula is based on the assumption that d_i may be treated as independent. This assumption is actually not valid. But when both N and M are large, it is often assumed in Monte-Carlo practice.

$$\Phi(\lambda) = 1 - 2\sum_{k=1}^{+\infty}(-1)^{k-1}\mathrm{e}^{-2k^2\lambda^2}. \tag{45}$$

Thus, the mean of S_{mn} can be calculated by

$$E(S_{mn}) = \int_0^{+\infty} \lambda \,\mathrm{d}\Phi(\lambda) = -2\sum_{k=1}^{\infty}(-1)^{k-1}\int_0^{+\infty} \lambda \,\mathrm{d}(\mathrm{e}^{-2k^2\lambda^2}). \tag{46}$$

For each k, we have

$$\int_0^{+\infty} \lambda \mathrm{d}(\mathrm{e}^{-2k^2\lambda^2}) = \lambda \mathrm{e}^{-2k^2\lambda^2}|_0^{\infty} - \int_0^{+\infty} \mathrm{e}^{-2k^2\lambda^2}\,\mathrm{d}\lambda = -\frac{1}{\sqrt{2}k}\int_0^{\infty} \mathrm{e}^{-\lambda^2}\,\mathrm{d}\lambda = -\frac{1}{\sqrt{2}k}\frac{\sqrt{\pi}}{2}. \tag{47}$$

Summing them up we obtain

$$E(S_{mn}) = \sqrt{\frac{\pi}{2}}\sum_{k=1}^{\infty}(-1)^{k-1}\frac{1}{k} = \sqrt{\frac{\pi}{2}}\log 2. \tag{48}$$

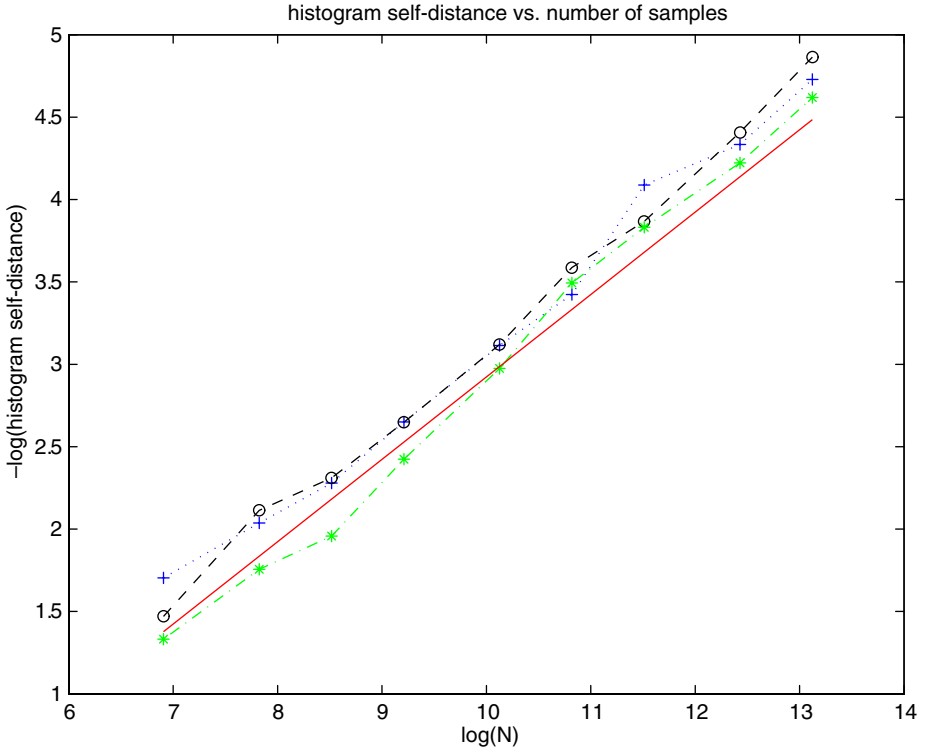

Fig. 3. The histogram self distance for different sample size N ($K = 50$ fixed) and different distributions (normal distribution '+', uniform distribution '∗', distribution from Schlögl model '○'). The solid straight line is the bound of the mean given by $\sqrt{\frac{4K}{\pi N}}$.

Similarly, the second moment of S_{mn} can be calculated by

$$E(S_{mn}^2) = \int_0^{+\infty} \lambda^2 \, \mathrm{d}\Phi(\lambda) = \sum_{k=1}^{\infty} \frac{1}{2k^2} = \frac{\pi^2}{12}. \tag{49}$$

Thus, the variance of S_{mn} is

$$\mathrm{Var}(S_{mn}) = E(S_{mn}^2) - E^2(S_{mn}) = \frac{\pi^2}{12} - \frac{\pi}{2}\log^2 2. \tag{50}$$

From the definition of S_{mn}, the mean of the Kolmogorov self distance is $\sqrt{\frac{\pi}{2}}\log 2\sqrt{\frac{1}{N}+\frac{1}{M}}$ and the corresponding variance is $\left(\frac{\pi^2}{12} - \frac{\pi}{2}\log^2 2\right)\left(\frac{1}{N}+\frac{1}{M}\right)$. □

5. Numerical experiments

To verify the theoretical results and conjectures, we generated samples for different distributions. The distributions we chose are the uniform distribution $U(0,1)$, the normal distribution $N(0,1)$ in Matlab, and the distribution generated from the Monte-Carlo simulation of the Schlögl model. For each distribution we generated two independent sets of samples for each N. To make the experiments simple, we always let $N = M$. The self distance is then measured by the histogram distance using different bin numbers K and the Kolmogorov distance of the two sets of samples. The results are plotted in Fig. 3 (−log(histogram self distance) vs. log(N)), Fig. 4 (−log(histogram self distance) vs. log(K)) and Fig. 5 (−log(Kolmogorov self

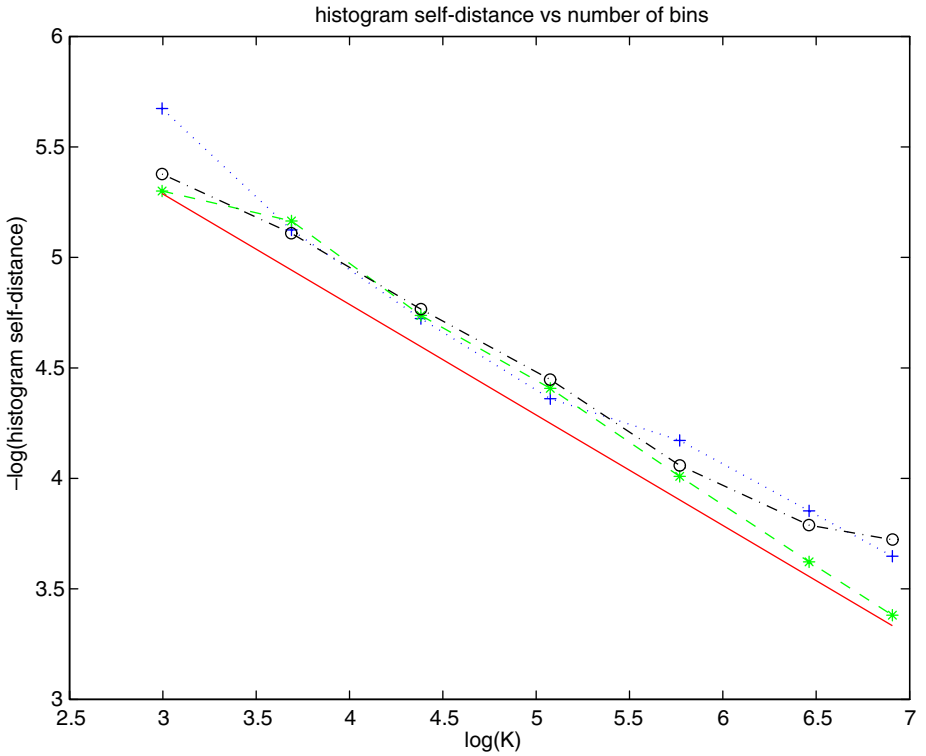

Fig. 4. The histogram self distance for different bin number K ($N = 10^6$ fixed) and different distributions (normal distribution '+', uniform distribution '∗', distribution from Schlögl model '○'). The solid straight line is the bound given by $\sqrt{\frac{4K}{\pi N}}$.

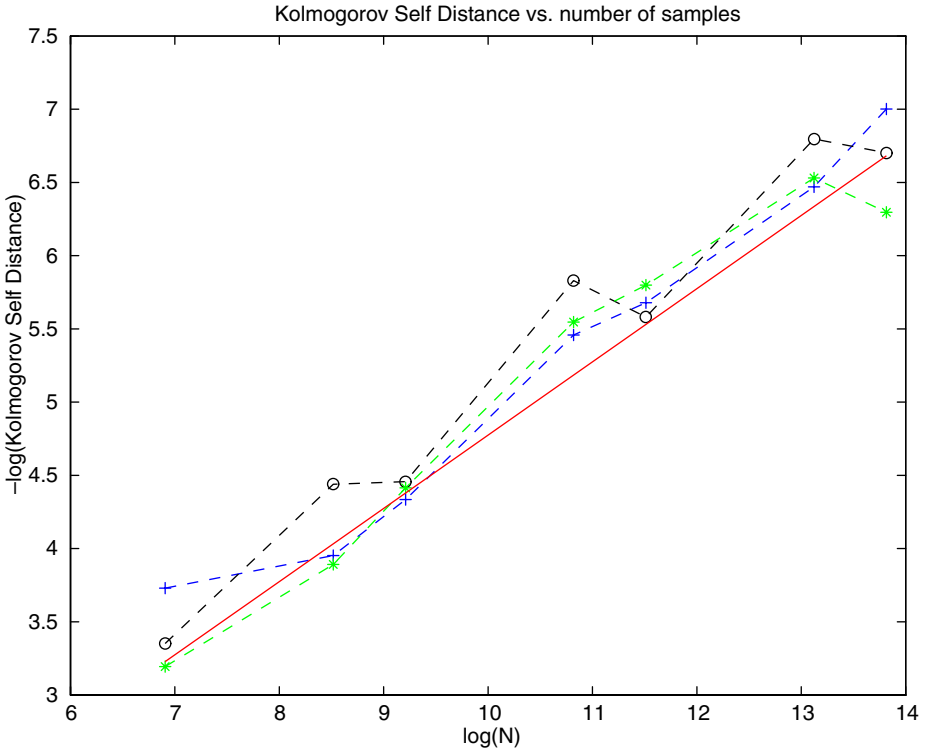

Fig. 5. The Kolmogorov self distance for different sample number N and different distributions (normal distribution '+', uniform distribution '*', distribution from Schlögl model 'O'). The solid straight line is the bound of the mean given by $\log 2\sqrt{\frac{\pi}{N}}$.

distance) vs. log(N)). In all of the figures, we also plot the bound of the mean value: $\sqrt{\frac{4K}{\pi N}}$ for the histogram self distance and $\log 2\sqrt{\frac{\pi}{N}}$ for the Kolmogorov self distance. In a few cases the self distance exceeds the bound. This is because the bound is for the mean. A single sample may generate a value away from the mean. But in most cases the formulas give quite accurate estimates of the self distances. An important observation from these figures is that, for a fairly large number of samples, the fact that the self distances are close to our estimates is independent of their original distributions. We obtained similar results from

Table 2
Mean and variance of histogram self distance for different N, K and distributions

Distribution	N	K	Mean	$\sqrt{\frac{4K}{\pi N}}$	Var	$\frac{\pi-2}{\pi}\frac{2}{N}$
Normal	10,000	10	0.0253	0.0357	7.26×10^{-5}	7.27×10^{-5}
	10,000	20	0.0377	0.0505	7.57×10^{-5}	7.27×10^{-5}
	10,000	40	0.0546	0.0714	7.72×10^{-5}	7.27×10^{-5}
Uniform	10,000	10	0.0339	0.0357	6.93×10^{-5}	7.27×10^{-5}
	10,000	20	0.0488	0.0505	7.35×10^{-5}	7.27×10^{-5}
	10,000	40	0.0701	0.0714	6.79×10^{-5}	7.27×10^{-5}
Schlögl reaction	1000	10	0.0914	0.1128	6.64×10^{-4}	7.27×10^{-4}
	1000	20	0.1341	0.1596	7.20×10^{-4}	7.27×10^{-4}
	1000	40	0.1910	0.2257	7.66×10^{-4}	7.27×10^{-4}

Table 3
Mean and variance of Kolmogorov self distance for different N and distributions

Distribution	N	Mean	$\log 2\sqrt{\frac{\pi}{N}}$	Var	$\left(\frac{\pi^2}{6}-\pi\log^2 2\right)\frac{1}{N}$
Normal	10,000	0.0121	0.0123	1.3392×10^{-5}	1.355×10^{-5}
Uniform	10,000	0.0122	0.0123	1.2895×10^{-5}	1.355×10^{-5}
Schlögl reaction	1000	0.0373	0.0389	1.3696×10^{-4}	1.355×10^{-4}

Monte-Carlo simulations of chemically reacting systems other than the Schlögl reactions. The experiments demonstrated two facts. The first is that the self distance is a feature of Monte-Carlo simulation independent of problem distribution. The second is that our estimate formulas are quite accurate.

To further verify the bound for the variance of the self distance, we generated 1000 independent self distances from each set of different N, K and distributions. We computed the mean and variance and compared it with the proved bound $\sqrt{\frac{4K}{\pi N}}$ for the mean value of the histogram self distance and $\log 2\sqrt{\frac{\pi}{N}}$ for the mean value of the Kolmogorov self distance, the variance estimation $\frac{\pi-2}{\pi}\frac{2}{N}$ for the variance of the histogram self distance and the variance bound $\left(\frac{\pi^2}{6}-\pi\log^2 2\right)\frac{1}{N}$ for the Kolmogorov self distance. The results are listed in Tables 2 and 3. We can see that the means are always bounded by our estimates. The Kolmogorov self distance is always very close to our estimates. The variances of the histogram self distance is close to the estimated values. When $N = 1000$, the variances of the Kolmogorov self distance is a little larger than the variance bound but when $N = 10{,}000$, the variance of the Kolmogorov self distance is smaller than the bound.

6. Conclusion and discussion

In this paper, we have introduced the concept of distribution distance to measure the errors in exact and approximate methods for stochastic simulation of chemically reacting systems. Self distance, an important property of the Monte-Carlo method, was defined and studied. We derived estimation formulas for the bounds of the mean and variance values of two kinds of self distance, Numerical experiments demonstrated the accuracy of our estimation formulas and the fact that the estimation formulas are independent of the problem distribution.

Recently, we have applied the distribution distance and the self distance to two related Monte-Carlo applications. One direct application involves the use of stochastic sensitivity analysis to study the robustness [31] of biochemical systems. In this work, the distribution distance between an original distribution and a perturbed distribution was used to measure the sensitivity of the distribution with respect to parameter perturbations. The perturbation to the parameters should be relatively small according to the definition of sensitivity. But it should not be so small that the distribution distance between the original distribution and the perturbed one is less than the corresponding self distance. The estimation formulas for the self distances proved to be a useful tool in selection of the perturbation size.

In a second application, we used the distribution distance to measure the convergence rate of τ-leaping methods [23]. Fig. 6 shows the distribution distances between SSA and the τ-leaping method with different stepsizes, applied to the Schlögl model. This figure suggests a linear convergence rate for the τ-leaping method. In this plot, we have $K = 50$ and $N = 1{,}000{,}000$. According to our estimation formula of the self distance, we know that the self distance is around 0.008. The histogram errors we show in the plot are at least 0.03, which is away from the self distance region. Thus, we can be sure that the distribution errors we measured are mostly due to the τ-leaping approximation rather than the "statistical fluctuation". We did not plot the distribution error at stepsize 0.025 because that value was too close to the self distance to be considered accurate.

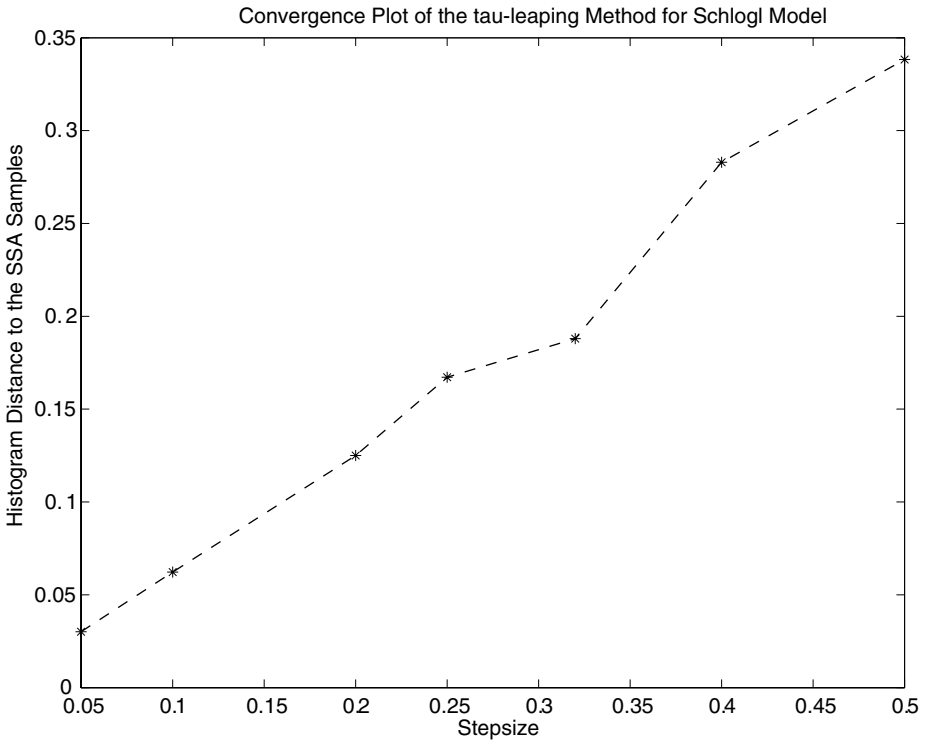

Fig. 6. The histogram distances corresponding to different stepsizes. The histogram distances are measured between 1,000,000 samples, each given by SSA and the τ-leaping method with different stepsizes applied to Schlögl model.

In the above discussion, we have limited ourselves to the distribution distance measured at the end time. In some other applications, the dynamical behavior is of concern. Thus, sometimes we need to measure the distribution distance not only at one time point but also on the whole time interval. This requirement can be easily achieved by collecting samples at several fixed time points during the time interval. The overall distribution distance is then measured by the sum of all the distribution distance at all time points. This generalized concept of distribution distance can be applied to study the dynamical behavior in a system.

Acknowledgments

This work was supported in part by the California Institute of Technology under DARPA Award No. F30602-01-2-0558, by the U.S. Department of Energy under DOE award No. DE-FG02-04ER25621, by the National Science Foundation under NSF award CCF-0326576 and ACI00-86061, and by the Institute for Collaborative Biotechnologies through grant DAAD19-03-D-0004 from the U.S. Army Research Office.

Appendix A. Proof of Lemma 4.2

For a random variable X that follows the normal distribution $N(0,\sigma^2)$, the probability density function is

$$p(s) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{s^2}{2\sigma^2}}. \tag{51}$$

Thus, the mean of $|X|$ is calculated by

$$E(|X|) = \int_{-\infty}^{\infty} |s| \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{s^2}{2\sigma^2}} ds = 2\int_0^{\infty} s \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{s^2}{2\sigma^2}} ds = \frac{1}{\sqrt{2\pi}\sigma} \int_0^{\infty} e^{-\frac{s^2}{2\sigma^2}} ds^2 = \frac{1}{\sqrt{2\pi}\sigma} 2\sigma^2 = \sqrt{\frac{2}{\pi}}\sigma. \tag{52}$$

For the second moment, $E(|X|^2) = E(X^2) = \sigma^2$. Thus

$$\mathrm{Var}(|X|) = E(|X|^2) - (E(|X|))^2 = \frac{\pi - 2}{\pi}\sigma^2. \tag{53}$$

Appendix B. Proof of Lemma 4.3

The linear combination theorem (p. 32 of [18]) gives:

$$E\left(\sum_{i=1}^{K} a_i X_i\right) = \sum_{i=1}^{K} a_i E(X_i), \tag{54}$$

$$\mathrm{Var}\left(\sum_{i=1}^{K} a_i X_i\right) = \sum_{i=1}^{K} a_i^2 \mathrm{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_i a_j \mathrm{cov}\{X_i, X_j\}, \tag{55}$$

and the covariance range theorem (p. 33 of [18]) yields

$$|\mathrm{cov}\{X_i, X_j\}| \leqslant \sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_i)}. \tag{56}$$

Thus

$$E\left(\sum_{i=1}^{K} X_i\right) = \sum_{i=1}^{K} E(X_i) \tag{57}$$

and

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{K} X_i\right) &= \sum_{i=1}^{K} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathrm{cov}\{X_i, X_j\} \\ &\leqslant \sum_{i=1}^{K} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)} = \left(\sum_{i=1}^{K} \sqrt{\mathrm{Var}(X_i)}\right)^2.\end{aligned} \tag{58}$$

References

[1] H.H. McAdams, A. Arkin, It's a noisy business! Trend Genet. 15 (2) (1999) 65–69.
[2] H.H. McAdams, A. Arkin, Stochastic mechanisms in gene expression, Proc. Natl. Acad. Sci. USA 94 (1997) 814–819.
[3] A. Arkin, J. Ross, H.H. McAdams, Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *E. coli* cells, Genetics 149 (1998) 1633–1648.
[4] N. Barkai, S. Leibler, Circadian clocks limited by noise, Nature 403 (1999) 267–268.
[5] N. Fedoroff, W. Fontana, Small numbers of big molecules, Science 297 (2002) 1129–1131.
[6] M. Elowitz, S. Leibler, A synthetic oscillatory network of transcriptional regulators, Nature 403 (2000) 335–338.
[7] A. Becskei, B. Seraphin, L. Serrano, Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion, EMBO 20 (2001) 2528–2535.

[8] E.M. Ozbudak, M. Thattai, I. Kurtser, A.D. Grossman, A. van Oudenaarden, Regulation of noise in the expression of a single gene, Nature Genet. 31 (2002) 69–73.
[9] J. Hasty, D. McMillen, J. Collins, Engineered gene circuits, Nature 420 (2002) 224–230.
[10] A. Kierzek, J. Zaim, P. Zielenkiewicz, The effect of transcription and translation initiation frequencies on the stochastic fluctuations in Prokaryotic gene expression, J. Bio. Chem. 276 (2001) 8165–8172.
[11] T. Kepler, T. Elston, Stochasticity in transcriptional regulation: origins, consequences and mathematical representations, Biophys. J. 81 (2001) 3116–3136.
[12] D. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys. 22 (1976) 403–434.
[13] D. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81 (1977) 2340–2361.
[14] D. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems, J. Chem. Phys. 115 (2001) 1716.
[15] M. Rathinam, L. Petzold, Y. Cao, D. Gillespie, Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method, J. Chem. Phys. 119 (2003) 12784–12794.
[16] E. Haseltine, J. Rawlings, Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics, J. Chem. Phys. 117 (2002) 6959–6969.
[17] T. Mattheyses, M. Simmons, Hybrid simulation of cellular behavior, Bioinformatics 20 (2004) 316–322.
[18] D. Gillespie, Markov Processes, An Introduction for Physical Scientists, Academic Press, New York, 1992.
[19] T. Gardner, C. Cantor, J. Collins, Construction of a genetic toggle switch in *Escherichia coli*, Nature 403 (2000) 339–342.
[20] A.N. Kolmogorov, Sulla determinazione empirica di une legge di distribuzione, Giorn. DellIstit. degli att. 4 (1933) 83–91.
[21] A.N. Kolmogorov, Confidence limits for an unknown distribution function, Ann. Math. Stat. 12 (1941) 461–463.
[22] D.A. Darling, The Kolmogorov–Smirnov, Cramer–von Mises Tests, Ann. Math. Stat. 28 (1957) 823–838.
[23] M. Rathinam, L. Petzold, Y. Cao, D. Gillespie, Consistency and stability of tau leaping schemes for chemical reaction systems, SIAM Multiscale Model. Simul. (to appear).
[24] N. Smirnov, On the estimation of the discrepancy between empirical curves of distribution for two independent samples, Bull. Math. Univ. Moscow 2 (1939) 3–14.
[25] N. Smirnov, Approximate laws of distribution of random variables from empirical data, Uspehi Matem. Nauk. 10 (1944) 179–206.
[26] N. Smirnov, Table for estimating the goodness of fit of empirical distributions, Ann. Math. Stat. 19 (1948) 279–281.
[27] S. Kullback, Information Theory and Statistics, Wiley, New York, 1959.
[28] J. Rosenthal, Convergence rates for Markov chains, SIAM Rev. 37 (1995) 387–405.
[29] W. Cochran, The χ^2 test of goodness of fit, Ann. Math. Stat. 23 (1952) 315–345.
[30] S. Ross, Introduction to Probability Models, eighth ed., Academic Press, New York, 2002.
[31] R. Gunawan, Y. Cao, L. Petzold, F. Doyle III, Sensitivity analysis of discrete stochastic systems, Biophys. J. 88 (2005) 2530–2540.