Error analysis of approximation methods for stochastically modeled chemical reaction systems

David F. Anderson*

with Arnab Ganguly and Thomas Kurtz

*anderson@math.wisc.edu

Department of Mathematics

University of Wisconsin - Madison

Stochastic Processes and their Applications

Berlin, Germany

July 27th, 2009

The model and notation

The model:

- A chemical reaction system consisting of d chemical species $\{S_1, S_2, \ldots, S_d\}$ undergoing a series of reactions, indexed by k.

- Each reaction requires some number of the species as inputs and provides some number of the species as outputs.

  Example: $S_1 \longrightarrow 2S_2$ requires one molecule of S_1 for the input and provides two molecules of S_2 for the output.

- If reaction k occurs at time t, then the state of the system, $X(t)$, is updated via addition of the reaction vector $\nu_k \in \mathbb{Z}^d$, representing the net change in the abundances of the underlying species:

$$X(t) = X(t-) + \nu_k.$$

  In the previous example: the associated reaction vector for the reaction $S_1 \longrightarrow 2S_2$ would be

$$\nu_k = [-1, 2, 0, \ldots, 0]^T, \qquad \left[\nu_k^s = [1, 0, \ldots, 0]^T, \quad \nu_k^p = [0, 2, 0, \ldots, 0]^T\right].$$

The model and notation

The model:

- The waiting times for the reactions are exponentially distributed with intensity (propensity) functions $\lambda_k : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$:

$$P(\, X(t+\Delta t) = X(t) + \nu_k \,|\, X(t)\,) = \lambda_k(X(t))\Delta t + o(\Delta t).$$

- The standard intensity function chosen is mass-action kinetics:

$$\lambda_k(x) = \kappa_k \prod_{i=1}^{d} (\nu_{ki}^s!) \binom{x_i}{\nu_{ki}^s} = \kappa_k \prod_{i=1}^{d} \frac{x_i!}{(x_i - \nu_{ki}^s)!}.$$

Example: If $S_1 \to$ anything, then $\lambda_k(x) = \kappa_k x_1$.

Example: If $S_1 + S_2 \to$ anything, then $\lambda_k(x) = \kappa_k x_1 x_2$.

Example: If $S_1 + 2S_2 \to$ anything, then $\lambda_k(x) = \kappa_k x_1 x_2 (x_2 - 1)$.

The Model

This model is a continuous time Markov chain in $\mathbb{Z}^d_{\geq 0}$ with generator

$$(\mathcal{A}f)(x) = \sum_k \lambda_k(x)(f(x+\nu_k) - f(x)),$$

and one intuitive representation for path-wise solutions is given by a random time-change

$$X(t) = X(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X(s))ds \right) \nu_k,$$

where the Y_k are independent, unit-rate Poisson processes.

- This type of model is used ubiquitously for cellular biochemical processes and is filling the pages of *Science* and *PNAS*.
- Models can be simple (single Gene-mRNA-Protein) or extremely complicated (E. coli heat shock response)

Methods of investigation: numerical simulation

$$X(t) = X(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X(s))ds \right) \nu_k,$$

(GOOD NEWS) There are a number of numerical methods that produce statistically exact sample paths:

1. Gillespie's algorithm.
2. The first reaction method.
3. The next reaction method.

For each step of these methods one must find :

(i) the amount of time that passes until the next reaction takes place: Δ_n. (the minimum of exponential RVs)

(ii) which reaction takes place at that time.

(BAD NEWS) If $\sum_k \lambda_k(X(t)) \gg 0$, then

$$\Delta_n \approx \frac{1}{\sum_k \lambda_k(X(t))} \ll 1$$

and the time needed to produce a single exact sample path over an interval $[0, T]$ can be prohibitive.

Tau-leaping

Standard "τ-leaping" [1] was developed by Dan Gillespie in an effort to overcome the problem that Δ_n may be prohibitively small.

Tau-leaping is essentially an Euler approximation of $\int_0^t \lambda_k(X(s))ds$:

$$\begin{aligned} Z(h) &= Z(0) + \sum_k Y_k \left(\int_0^h \lambda_k(Z(s))\, ds \right) \nu_k \\ &\approx Z(0) + \sum_k Y_k \left(\lambda_k(Z(0))\, h \right) \nu_k \\ &\stackrel{d}{=} Z(0) + \sum_k \text{Poisson}(\lambda_k(Z(0))\, h) \nu_k . \end{aligned}$$

[1] D. T. Gillespie, J. Chem. Phys., **115**, 1716 – 1733.

Tau-leaping

One representation for $Z(t)$ generated by τ-leaping is

$$Z(t) = X(0) + \sum_k Y_k \left(\int_0^t \lambda_k(Z \circ \eta(s)) ds \right) \nu_k,$$

where $\eta(s) = t_n$, if $t_n \leq s < t_{n+1}$ and the Y_k are as before.

We define the operator

$$(\mathcal{B}_z f)(x) \stackrel{\text{def}}{=} \sum_k \lambda_k(z)(f(x + \nu_k) - f(x)),$$

so that for $t_n \leq t < t_{n+1}$ the processes $Z(t)$ satisfies

$$\mathbb{E} f(Z(t)) = \mathbb{E} f(Z(t_n)) + \mathbb{E} \int_{t_n}^t (\mathcal{B}_{Z(t_n)} f)(Z(s)) ds$$

Another algorithm: the midpoint method

For a time discretization $0 = t_0 < t_1 < \cdots < t_N = T$, with $h = t_n - t_{n-1}$, let

$$\rho(z) = z + \frac{1}{2} h \sum_k \lambda_k(z) \nu_k,$$

("deterministic" midpoint approximation)

and let $\mathcal{Z}(t)$ solve:

$$\mathcal{Z}(t) = X(0) + \sum_k Y_k \left(\int_0^t \lambda_k \circ \rho(\mathcal{Z} \circ \eta(s)) ds \right) \nu_k,$$

where $\eta(s) = t_n$, if $t_n \leq s < t_{n+1}$ and the Y_k are as before.

Recall the operator

$$(\mathcal{B}_z f)(x) \stackrel{\text{def}}{=} \sum_k \lambda_k(z)(f(x + \nu_k) - f(x)).$$

For $t_n \leq t < t_{n+1}$ the processes $\mathcal{Z}(t)$ satisfies

$$\mathbb{E} f(\mathcal{Z}(t)) = \mathbb{E} f(\mathcal{Z}(t_n)) + \mathbb{E} \int_{t_n}^t (\mathcal{B}_{\rho(\mathcal{Z}(t_n))} f)(\mathcal{Z}(s)) ds$$

Previous error analysis

Under the scaling $h \to 0$:

1. Rathinam, Petzold, Cao, and Gillespie[2] showed that tau-leaping:
   - (*i*) has a weak local error of $O(h^2)$ for all moments.
   - (*ii*) is first order accurate in a weak sense if the intensity functions λ_k are linear.

2. Li[3] extended these results by showing that standard tau-leaping has a strong error (in the L^2 norm) of order $1/2$ and a weak error of order one:

$$\sup_{n \leq N} \sqrt{\mathbb{E}\left|Z(t_n) - X(t_n)\right|^2} \leq Ch^{1/2}$$

$$\left|\mathbb{E}f(Z(T)) - \mathbb{E}f(X(T))\right| \leq Ch,$$

   where $0 = t_0 < t_1 < \cdots < t_N = T$ is a partition of $[0, T]$.

3. The midpoint method is no more accurate than standard tau-leaping

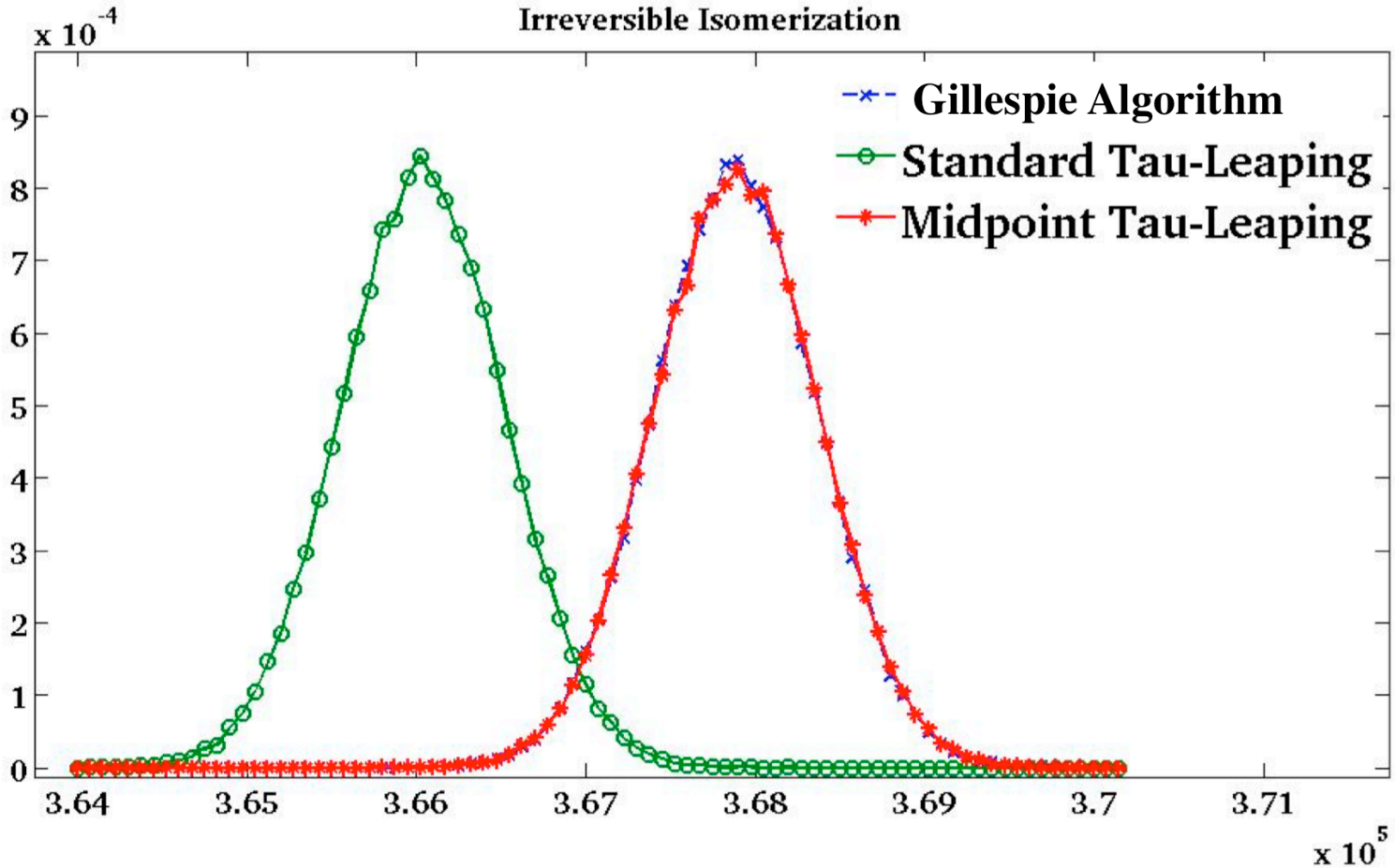
[2] M. Rathinam et al., SIAM Multi. Model. Simul., **4**, 2005, 867 – 895.
[3] T. Li, SIAM Multi. Model. Simul., **6**, 2007, 417 – 436.

An example that needs explaining: part I

Consider the simple example

$$X \xrightarrow{1} 0, \qquad X(0) = 10^6.$$

Letting $h = 1/10^2$ and simulating $100,000$ sample paths with each method yields the following approximate distributions



Problem with the $h \to 0$ scaling

Recall, tau-leaping methods are only useful if $h \gg \Delta_n$, for otherwise an exact method would be performed. Therefore, we should require that

$$h \gg \frac{1}{\sum_k \lambda_k(X(t))} \approx \Delta_n \qquad \text{or} \qquad h \sum_k \lambda_k(X(t)) \gg 1.$$

The "classical scaling"

τ-leaping is only useful if

$$h\sum_k \lambda_k(X(t)) \gg 1 \quad \text{while} \quad \frac{1}{\sum_k \lambda_k(X(t))} \ll 1.$$

We will make the following natural assumptions on our model:

(*i*) $X(0) = Vx_0$ for some $x_0 \in \mathbb{R}^d_{>0}$ and $V \gg 1$.

(*ii*) The rates of specific reactions scale with V in the following way:

$$\emptyset \to *, \qquad S_1 \to *, \qquad S_1 + S_2 \to *, \qquad 2S_1 \to *$$

have corresponding respective intensities

$$\lambda_k(x) = d_k V, \quad \lambda_k(x) = d_k x_1, \quad \lambda_k(x) = \frac{d_k}{V} x_1 x_2, \quad \lambda_k(x) = \frac{d_k}{V} x_1(x_1 - 1).$$

- This scaling is the so called "classical scaling" and arises naturally by thinking of V as the volume.

A key property of this scaling

For each k and $x \in \mathbb{R}^d$

$$\lambda_k(Vx) = VA_k^V(x),$$

where $A_k^V(x) = a_k(x) + \frac{1}{V} b_k(x)$ for functions a_k, b_k.

Note: $a_k(x)$ will always be classical *deterministic* mass action kinetics.

Examples:

- For the reaction $S_1 + S_2 \to *$,

$$\lambda_k(Vx) = \kappa_k(Vx_1)(Vx_2) = \frac{d_k}{V} V^2 x_1 x_2 = V(d_k x_1 x_2) = VA_k^V(x).$$

- For the reaction $2S_1 \to *$,

$$\lambda_k(Vx) = \kappa_k(Vx_1)(Vx_1 - 1) = \frac{d_k}{V} V^2 (x_1^2 - \frac{1}{V} x_1) = V(d_k x_1^2 - \frac{1}{V} d_k x_1) = VA_k^V(x).$$

Limitations of this scaling

- It is clear that the choice of scaling laid out above will not explicitly cover all cases of interest (i.e. multiple scales).
- However, the purpose of this analysis is not to handle every conceivable case.
- Our purpose is to give a more accurate picture of how tau-leaping methods approximate the exact solution, both strongly and weakly, in at least one plausible setting.
- We believe that error analyses conducted under different modeling assumptions can be carried out similarly.

time-step

Recall: τ-leaping is only useful if

$$h\sum_k \lambda_k(X(t)) \gg 1 \quad \text{while} \quad \frac{1}{\sum_k \lambda_k(X(t))} \ll 1.$$

We choose a time step

$$h = 1/V^\beta,$$

for $\beta \in (0,1)$.

This scaling satisfies the above requirements:

$$h\sum_k \lambda(X(t)) = V^{-\beta}\left(V\sum_k A_k^V\left(\frac{X(t)}{V}\right)\right) = O(V^{1-\beta}) \gg 1$$

$$\left(\sum_k \lambda_k(X(t))\right)^{-1} = \left(V\sum_k A_k^V\left(\frac{X(t)}{V}\right)\right)^{-1} = O(V^{-1}) \ll 1,$$

and so this is a reasonable scaling.

Error analysis on the scaled processes

We can consider the scaled processes $X^V(t), Z^V(t)$, and $\mathcal{Z}^V(t)$ that satisfy

$$X^V(t) \stackrel{\text{def}}{=} \frac{X(t)}{V} = X^V(0) + \frac{1}{V}\sum_k Y_k\left(V\int_0^t A_k^V(X^V(s))ds\right)\nu_k$$

$$Z^V(t) \stackrel{\text{def}}{=} \frac{Z(t)}{V} = X^V(0) + \frac{1}{V}\sum_k Y_k\left(V\int_0^t A_k^V(Z^V\circ\eta(s))ds\right)\nu_k$$

$$\mathcal{Z}^V(t) \stackrel{\text{def}}{=} \frac{\mathcal{Z}^V(t)}{V} = X^V(0) + \frac{1}{V}\sum_k Y_k\left(V\int_0^t A_k^V\circ\rho^V(\mathcal{Z}^V\circ\eta(s))ds\right)\nu_k,$$

where

$$\rho^V(z) \stackrel{\text{def}}{=} z + \frac{1}{2}h\sum_k A_k^V(z)\nu_k,$$

and prove theorems about strong (L^1) convergence,

$$\mathbb{E}|X^V(t) - Z^V(t)|, \quad \mathbb{E}|X^V(t) - \mathcal{Z}^V(t)|$$

and weak convergence,

$$|\mathbb{E}f(X^V(t)) - \mathbb{E}f(Z^V(t))|, \quad |\mathbb{E}f(X^V(t)) - \mathbb{E}f(\mathcal{Z}^V(t))|,$$

in terms of V (or $h = 1/V^\beta$).

Strong error of standard tau-leaping

Theorem (Anderson, Ganguly, Kurtz – 2009)

Let $X^V(t)$ and $Z^V(t)$ denote the scaled processes for $t \in [0, T]$ with $h = V^{-\beta}$. There exists a constant $C = C(T) > 0$ such that

$$\sup_{t \leq T} \mathbb{E}|X^V(t) - Z^V(t)| \leq \frac{C}{V^\beta}.$$

Main tools

Define the filtration for the i^{th} Process:

$$\mathcal{F}_u^k \stackrel{\text{def}}{=} \sigma\{Y_i(r), Y_k(s) \ : \ i \neq k, \ s \leq u, \ r \in [0, \infty)\}.$$

Lemma

Suppose that $X(t)$ satisfies the random time change representation with non-negative intensity functions λ_k. For $t \geq 0$ and a choice of k,

$$\tau_k(t) = \int_0^t \lambda_k(X(s))ds$$

is an $\{\mathcal{F}_u^k\}$-stopping time.

Proof.

For $u \geq 0$ let $\alpha(u)$ satisfy

$$\int_0^{\alpha(u)} \lambda_k(X(s))ds = u,$$

where we take $\alpha(u) = \infty$ if $\int_0^\infty \lambda_k(X(s))ds < u$. Then $\alpha(u)$ is adapted to $\mathcal{F}_u^k$ and $\{\tau_k(t) \leq u\} = \{t \leq \alpha(u)\} \in \mathcal{F}_u^k$. □

Main tools

Therefore, we have our “Itô Isometry”:

$$\begin{aligned}
\mathbb{E}\left|Y_k\left(\int_0^t \lambda_{k,1}(X(r))dr\right) - Y_k\left(\int_0^t \lambda_{k,2}(Z(r))dr\right)\right| &= \\
&= \mathbb{E}\left[Y_k\left(\tau_{k,1}(t) \vee \tau_{k,2}(t)\right) - Y_k\left(\tau_{k,1}(t) \wedge \tau_{k,2}(t)\right)\right] \\
&= \mathbb{E}\left(\tau_{k,1}(t) \vee \tau_{k,2}(t) - \tau_{k,1}(t) \wedge \tau_{k,2}(t)\right) \\
&= \mathbb{E}\left|\int_0^t \lambda_{k,1}(X(r))dr - \int_0^t \lambda_{k,2}(Z(r))dr\right|.
\end{aligned}$$

Sharper bound for standard tau-leaping with $\beta < 1/2$

Theorem (Anderson, Ganguly, Kurtz – 2009)

Let $X^V(t)$ and $Z^V(t)$ denote the scaled processes for $t \in [0, T]$ with $h = V^{-\beta}$. There exists a constant $C = C(T) > 0$ such that

$$\mathbb{E} \sup_{t \le T} |X^V(t) - Z^V(t)| \le \frac{C}{V^\beta} + \frac{C}{V^{1/2}}.$$

Why the $V^{-1/2}$? Let $\tilde{Y}_k(x) = Y_k(x) - x$ be the centered process, we have terms like

$$\frac{|\nu_k|}{V} \mathbb{E} \sup_{t \le T} \left| \tilde{Y}_k \left(V \int_0^t A_k^V(X^V(s)) ds \right) \right|.$$

Central limit theorem says this term is $O(V^{-1/2})$.

We can do better with new coupling of the processes.

A new coupling of $X^V(t)$ and $Z^V(t)$

Now let

$$X^V(t) = X^V(0) + \frac{1}{V}\sum_k \Bigg[Y_{k,1}\left(V\int_0^t A_k^V(X^V(s)) \wedge A_k^V(Z^V(s)) \wedge A_k^V(Z^V \circ \eta(s))\, ds\right)$$
$$+ Y_{k,2}\left(V\int_0^t A_k^V(X^V(s)) - A_k^V(X^V(s)) \wedge A_k^V(Z^V(s))\, ds\right)$$
$$+ Y_{k,3}\left(V\int_0^t A_k^V(X^V(s)) \wedge A_k^V(Z^V(s)) - A_k^V(X^V(s)) \wedge A_k^V(Z^V(s)) \wedge A_k^V(Z^V \circ \eta(s))\, ds\right)\Bigg]\nu_k$$

$$Z^V(t) = X^V(0) + \frac{1}{V}\sum_k \Bigg[Y_{k,1}\left(V\int_0^t A_k^V(X^V(s)) \wedge A_k^V(Z^V(s)) \wedge A_k^V(Z^V \circ \eta(s))\, ds\right)$$
$$+ Y_{k,4}\left(V\int_0^t A_k^V(Z^V(s)) \wedge A_k^V(Z^V \circ \eta(s)) - A_k^V(X^V(s)) \wedge A_k^V(Z^V(s)) \wedge A_k^V(Z^V \circ \eta(s))\, ds\right)$$
$$+ Y_{k,5}\left(V\int_0^t A_k^V(Z^V \circ \eta(s)) - A_k^V(Z^V(s)) \wedge A_k^V(Z^V \circ \eta(s))\, ds\right)\Bigg]\nu_k .,$$

where the $Y_{k,i}$ are mutually independent Poisson processes.

Note: distribution of the "marginal processes" $X^V(t)$ and $Z^V(t)$ are unchanged.

Sharper bound for standard tau-leaping with $\beta \geq 1/2$

Theorem (Anderson, Ganguly, Kurtz – 2009)

Let $X^V(t)$ and $Z^V(t)$ satisfy the new coupling for $t \in [0, T]$ with $h = V^{-\beta}$. Then there exists a constant $C = C(T) > 0$ such that

$$\mathbb{E} \sup_{t \leq T} |X^V(t) - Z^V(t)| \leq \frac{C}{V^\beta}.$$

Exact asymptotics of error process

Let

$$\mathcal{E}^V(t) \stackrel{\text{def}}{=} V^\beta(X^V(t) - Z^V(t)) \tag{1}$$

denote the scaled error process. Also, let $\mathcal{E}(t)$ satisfy the deterministic initial value problem

$$\mathcal{E}(t) = \int_0^t DF(x(s,x_0))\mathcal{E}(s)ds + \frac{1}{2}\int_0^t DF(x(s,x_0))F(x(s,x_0))ds, \quad \mathcal{E}(0) = 0,$$

where $x(t,x_0)$ satisfies the corresponding deterministic initial value problem

$$\frac{d}{dt}x(t,x_0) = F(x(s,x_0)), \quad x(0,x_0) = x_0.$$

Theorem (Anderson, Ganguly, Kurtz – 2009)

Let $X^V(t)$ and $Z^V(t)$ satisfy the new coupling for $t \in [0,T]$. Then for $0 < \beta < 1$

$$\lim_{V\to\infty} \mathbb{E}\sup_{t\le T}\left|\mathcal{E}^V(t) - \mathcal{E}(t)\right| = 0.$$

Weak error of standard tau-leaping

The following is now immediate.

Theorem (Anderson, Ganguly, Kurtz – 2009)

Let $X^V(t)$ be a process with generator

$$(\mathcal{A}^V f)(x) \stackrel{def}{=} \sum_k V A_k^V(x)(f(x+\nu_k/V) - f(x)).$$

Define the operator

$$(\mathcal{B}_z^V f)(x) \stackrel{def}{=} \sum_k V A_k^V(z)(f(x+\nu_k/V) - f(x)),$$

so that for $t_n \leq t < t_{n+1}$ the processes $Z^V(t)$ satisfies

$$\mathbb{E}f(Z^V(t)) = \mathbb{E}f(Z^V(t_n)) + \mathbb{E}\int_{t_n}^t (\mathcal{B}_{Z^V(t_n)}^V f)(Z^V(s))ds$$

Then, for any continuous differentiable function f

$$\lim_{V\to\infty} V^\beta \left(\mathbb{E}f(X^V(T)) - \mathbb{E}f(Z^V(T))\right) = \mathcal{E}(T)\cdot\nabla f(x(T,x_0)).$$

Strong error of midpoint tau-leaping

Theorem (Anderson, Ganguly, Kurtz – 2009)

Let $X^V(t)$ and $\mathcal{Z}^V(t)$ denote the scaled processes for $t \in [0, T]$ and $h = 1/V^{-\beta}$. Then there exists a constant $C = C(T) > 0$ such that

$$\sup_{t \leq T} \mathbb{E}|X^V(t) - \mathcal{Z}^V(t)| \leq \frac{C}{V^{\kappa(\beta)}}, \quad \textit{where} \quad \kappa(\beta) = \min\left\{2\beta, \frac{1+\beta}{2}\right\}.$$

- The reason for the $\frac{1+\beta}{2}$ term is another central limit theorem argument.

Weak error of midpoint tau-leaping

Theorem (Anderson, Ganguly, Kurtz – 2009)

Let $X^V(t)$ be a process with generator

$$(\mathcal{A}^V f)(x) \stackrel{def}{=} \sum_k V A_k^V(x)(f(x+\nu_k/V) - f(x)).$$

Define the operator

$$(\mathcal{B}_z^V f)(x) \stackrel{def}{=} \sum_k V A_k^V(z)(f(x+\nu_k/V) - f(x)),$$

so that for $t_n \leq t < t_{n+1}$ the processes $\mathcal{Z}^V(t)$ satisfies

$$\mathbb{E}f(\mathcal{Z}^V(t)) = \mathbb{E}f(\mathcal{Z}^V(t_n)) + \mathbb{E}\int_{t_n}^t (\mathcal{B}^V_{\rho^V(\mathcal{Z}^V(t_n))} f)(Z^V(s))ds$$

Then, for any C^3 function f, there exists a constant $C = C(f,T) > 0$ such that

$$V^{2\beta}\left|\mathbb{E}f(X^V(T)) - \mathbb{E}f(\mathcal{Z}^V(T))\right| \leq C.$$

An example that needs explaining: part II

Consider the simple example

$$X \xrightarrow{1} 0, \qquad X(0) = 10^6.$$

$V = 10^6$. Letting $h = 1/10^2 = 1/V^{1/3}$ (so $\beta = 1/3$) and simulating $100,000$ sample paths with each method yields the following approximate distributions