The Annals of Applied Probability
2014, Vol. 24, No. 2, 721–759
DOI: 10.1214/13-AAP934

CENTRAL LIMIT THEOREMS AND DIFFUSION APPROXIMATIONS FOR MULTISCALE MARKOV CHAIN MODELS[1]

BY HYE-WON KANG, THOMAS G. KURTZ[2] AND LEA POPOVIC[3]

Ohio State University, University of Wisconsin and Concordia University

Ordinary differential equations obtained as limits of Markov processes appear in many settings. They may arise by scaling large systems, or by averaging rapidly fluctuating systems, or in systems involving multiple timescales, by a combination of the two. Motivated by models with multiple timescales arising in systems biology, we present a general approach to proving a central limit theorem capturing the fluctuations of the original model around the deterministic limit. The central limit theorem provides a method for deriving an appropriate diffusion (Langevin) approximation.

1. Introduction. There are two classical kinds of Gaussian limit theorems associated with continuous time Markov chains as well as more general Markov processes. The first of these considers a sequence $\{X^N\}$ of Markov chains that converges to a deterministic function X and gives a limit for the rescaled deviations $U^N = r_N(X^N - X)$; see, for example, Kurtz (1971), Kurtz (1977/78), van Kampen (1961). The second considers an ergodic Markov process Y with stationary distribution π and gives a limit for

$$Z^N(t) = \frac{1}{\sqrt{N}} \int_0^{Nt} h\big(Y(s)\big)\, ds = \sqrt{N} \int_0^t h\big(Y(Ns)\big)\, ds$$

for h satisfying $\int h\, d\pi = 0$; see, for example, Bhattacharya (1982) for a general result of this type.

There are many proofs for theorems like these. In particular, results of both types can be proved using the martingale central limit theorem (Theorem A.1). For example, in the first case, there is typically a sequence of functions F^N such that

$$M^N(t) = X^N(t) - X^N(0) - \int_0^t F^N\big(X^N(s)\big)\, ds$$

Received August 2012; revised March 2013.
[1]Supported in part by NSF FRG Grant DMS 05-53687.
[2]Supported in part by NSF Grant DMS 11-06424.
[3]Supported by NSERC Discovery grant.
MSC2010 subject classifications. 60F05, 92C45, 92C37, 80A30, 60F17, 60J27, 60J28, 60J60.
Key words and phrases. Reaction networks, central limit theorem, martingale methods, Markov chains, scaling limits.

is a martingale, $F^N \to F$, $\dot{X} = F(X)$ and $F^N(x^N) - F(x) \approx \nabla F(x)(x^N - x)$, for x^N converging to x. If the martingale central limit theorem gives $r_N M^N \Rightarrow M$ and $U^N(0) \Rightarrow U(0)$, then (ignoring technicalities) U^N should converge to the solution of

$$U(t) = U(0) + M(t) + \int_0^t \nabla F\big(X(s)\big)U(s)\,ds. \tag{1.1}$$

In the second case, the assumption that $\int h\,d\pi = 0$ suggests that there should be a solution of the Poisson equation $Af = -h$, where A is the generator for Y, and then

$$\begin{aligned} Z^N(t) = \frac{1}{\sqrt{N}}\Big(f\big(Y(Nt)\big) - f\big(Y(0)\big) - \int_0^{Nt} Af\big(Y(s)\big)\,ds\Big) \\ - \frac{1}{\sqrt{N}}\big(f\big(Y(Nt)\big) - f\big(Y(0)\big)\big). \end{aligned}$$

The first term on the right is a martingale and the second should go to zero, so if the martingale central limit theorem applies to the first, then Z^N should converge.

This paper addresses situations of the first type [$V_0^N \Rightarrow V_0$ for a deterministic V_0, and we want to verify convergence of $U^N = r_N(V_0^N - V_0)$] in which both approaches are required. Specifically, the function F^N giving the martingale, M^N, depends not only on V_0^N but also on another process V_1^N [think $V_1^N(t) = V_1(Nt)$], so

$$M^{N,1}(t) = V_0^N(t) - V_0^N(0) - \int_0^t F^N\big(V_0^N(s), V_1^N(s)\big)\,ds$$

is a martingale, F^N "averages" to F in the sense that

$$\int_0^t \big(F^N\big(V_0^N(s), V_1^N(s)\big) - F\big(V_0^N(s)\big)\big)\,ds \to 0,$$

(V_0^N, V_1^N) is Markov with generator A_N and there exist H_N such that $A_N H_N \approx (F^N - F)$. [Note that H_N will be a vector of functions in the domain of A_N, $\mathcal{D}(A_N)$.] Assuming that

$$\begin{aligned} M^{N,2}(t) = H_N\big(V_0^N(t), V_1^N(t)\big) - H_N\big(V_0^N(0), V_1^N(0)\big) \\ - \int_0^t A_N H_N\big(V_0^N(s), V_1^N(s)\big)\,ds \end{aligned}$$

is a martingale, and again ignoring all the technicalities, we have

$$\begin{aligned} &r_N\big(V_0^N(t) - V_0(t)\big) \\ &\quad = r_N\big(V_0^N(0) - V_0(0)\big) + r_N M^{N,1}(t) - r_N M^{N,2}(t) \\ &\qquad + \int_0^t r_N\big(F\big(V_0^N(s)\big) - F\big(V_0(s)\big)\big)\,ds \end{aligned} \tag{1.2}$$

$$+ r_N\big(H_N\big(V_0^N(t), V_1^N(t)\big) - H_N\big(V_0^N(0), V_1^N(0)\big)\big)$$
$$+ r_N \int_0^t \big(F^N\big(V_0^N(s), V_1^N(s)\big) - F\big(V_0^N(s)\big) - A_N H_N\big(V_0^N(s), V_1^N(s)\big)\big)\, ds.$$

If the last two terms on the right go to zero, the martingale terms converge,

$$r_N M^{N,1} - r_N M^{N,2} \Rightarrow M$$

and F is smooth, then we again should have $U^N \Rightarrow U$ satisfying (1.1).

The work to be done to obtain theorems of this type is now clear. We need to identify F^N and F, find an approximate solution to the Poisson equation $A_N H_N \approx F^N - F$, verify that the martingales satisfy the conditions of the martingale central limit theorem, and verify that the error terms [the last two terms in (1.2)] converge to zero. We will make this analysis more specific in stages. We are essentially considering situations in which the process V_1^N is evolving on a faster time scale than V_0^N and "averages out" to give the convergence of V_0^N to V_0. But V_1^N itself may evolve on more than one time scale. In the first stage of our development, we will replace V_1^N by (V_1^N, V_2^N) with V_1^N and V_2^N evolving on different (fast) time scales. Once the analysis for two fast time scales is carried out, the extension of the general results to more than two fast time scales should be clear. In the second stage, we consider multiply scaled, continuous-time Markov chains of a type that arises naturally in models of chemical reaction networks. For these models, many of the conditions simplify, but the notation becomes more complex.

Outline: In Section 2 we state and prove the functional central limit Theorem 2.11, and specify a sequence of Conditions 2.1–2.10 that need to be verified for it to apply. In Section 3 we additionally give a diffusion approximation implied by Theorem 2.11. Our aim is to apply these results to Markov chain models for chemical reactions. In Section 4 we identify specific aspects of the multi-scale behavior of a reaction network that one needs in order to apply Theorem 2.11 to the chemical species with a deterministic limit on the slowest time scale. Section 5 provides several examples of chemical networks (the first two evolving on two, the last one on three time-scales), and shows how to verify the conditions and obtain a diffusion approximation.

2. A central limit theorem for a system with deterministic limit and three time scales. We identify a set of conditions on a three time-scale process $V^N = (V_0^N, V_1^N, V_2^N)$ that guarantee $U^N = r_N(V_0^N - V_0)$ converges to a diffusion. As suggested earlier, we write U^N in the form

$$U^N(t) = U^N(0) + r_N\big(M^{N,1}(t) - M^{N,2}(t)\big) + r_N \int_0^t \big(\overline{F}\big(V_0^N(s)\big) - \overline{F}\big(V_0(s)\big)\big)\, ds \tag{2.1}$$

$$+ r_N \int_0^t \big(F^N\big(V^N(s)\big) - F\big(V^N(s)\big)\big)\,ds$$
$$+ r_N \int_0^t \big(F\big(V^N(s)\big) - \overline{F}\big(V_0^N(s)\big) - A_N H_N\big(V^N(s)\big)\big)\,ds$$
$$+ r_N\big(H_N\big(V^N(t)\big) - H_N\big(V^N(0)\big)\big),$$

where $M^{N,1}$, $M^{N,2}$ are martingales, V_0 is the deterministic limit of the process V_0^N, $\overline{F}$ is its infinitesimal drift and H_N is an approximate solution to a Poisson equation. Our conditions insure each individual term has a well behaved limit.

We assume that V_i^N takes values in $\mathbb{E}_i^N \subset \mathbb{R}^{d_i}$, $i = 0, 1, 2$, and that $\mathbb{E}_i^N$ converges in the sense that there exists $\mathbb{E}_i \subset \mathbb{R}^{d_i}$ such that $\mathbb{E}_i^N \subset \mathbb{E}_i$ and for each compact $K \subset \mathbb{R}^{d_i}$,

$$\lim_{N\to\infty} \sup_{x\in\mathbb{E}_i\cap K} \inf_{y\in\mathbb{E}_i^N} |x - y| = 0.$$

We will refer to A_N as the "generator" for the process $V^N = (V_0^N, V_1^N, V_2^N)$, but all we require is that A_N is a linear operator on some space $\mathcal{D}(A_N)$ of measurable functions on $\mathbb{E}^N \equiv \mathbb{E}_0^N \times \mathbb{E}_1^N \times \mathbb{E}_2^N$ and that for $h \in \mathcal{D}(A_N)$,

$$h\big(V^N(t)\big) - h\big(V^N(0)\big) - \int_0^t A_N h\big(V^N(s)\big)\,ds$$

is a local martingale.

We first identify the *time scales of the process* V^N with two sequences of positive numbers $\{r_{1,N}\}$, $\{r_{2,N}\}$, and introduce a sequence of scaling parameters $\{r_N\}$ for U^N with the following properties.

CONDITION 2.1 (Scaling parameters). *The scaling parameters* $r_N \to \infty$ *and* $\{r_{1,N}\}$, $\{r_{2,N}\}$ *are sequences of positive numbers satisfying*

$$\lim_{N\to\infty} \frac{r_N}{r_{1,N}} = 0, \qquad \lim_{N\to\infty} \frac{r_{1,N}}{r_{2,N}} = 0. \tag{2.2}$$

We next identify the "generators" for the *effective dynamics of* V_0^N, V_1^N *and* V_2^N on time scales t, $tr_{1,N}$, and $tr_{2,N}$, respectively. L_0, L_1, L_2 will be linear operators defined on sufficiently large domains, $\mathcal{D}(L_0) \subset M(\mathbb{E}_0)$, $\mathcal{D}(L_1) \subset M(\mathbb{E}_0 \times \mathbb{E}_1)$ and $\mathcal{D}(L_2) \subset M(\mathbb{E}_0 \times \mathbb{E}_1 \times \mathbb{E}_2)$, and taking values in $M(\mathbb{E}_0 \times \mathbb{E}_1 \times \mathbb{E}_2)$. The requirements that determine what is meant by "sufficiently large" will become clear, but we will assume that the domains contain all C^∞ functions having compact support in the appropriate space. We will use the notation $\mathbb{E} = \mathbb{E}_0 \times \mathbb{E}_1 \times \mathbb{E}_2$.

CONDITION 2.2 (Multiscale convergence). *For each compact* $K \subset \mathbb{R}^{d_0+d_1+d_2}$,

$$\lim_{N\to\infty} \sup_{v\in K\cap\mathbb{E}^N} \left|A_N h(v) - L_0 h(v)\right| = 0, \qquad h \in \mathcal{D}(L_0),$$

$$\lim_{N\to\infty} \sup_{v\in K\cap\mathbb{E}^N} \left|\frac{1}{r_{1,N}} A_N h(v) - L_1 h(v)\right| = 0, \qquad h \in \mathcal{D}(L_1)$$

and

$$\lim_{N\to\infty} \sup_{v\in K\cap\mathbb{E}^N} \left|\frac{1}{r_{2,N}} A_N h(v) - L_2 h(v)\right| = 0, h \in \mathcal{D}(L_2).$$

REMARK 2.3. Similar conditions are considered in Ethier and Nagylaki (1980). See also Ethier and Kurtz (1986), Section 1.7. There may be only two time-scales, in which case $d_2 = 0$, $L_2 h = 0$ and $\mathbb{E} = \mathbb{E}_0 \times \mathbb{E}_1$ (equivalently, $\mathbb{E}_2$ consists of a single point) in what follows.

The next condition ensures the *uniqueness of the conditional equilibrium distributions of the fast components* V_2^N *and* V_1^N, whose "generators" are L_2 and L_1.

CONDITION 2.4 (Averaging condition). *For each* $(v_0, v_1) \in \mathbb{E}_0 \times \mathbb{E}_1$, *there exists a unique* $\mu_{v_0,v_1} \in \mathcal{P}(\mathbb{E}_2)$ *such that* $\int L_2 h(v_0, v_1, v_2)\mu_{v_0,v_1}(dv_2) = 0$ *for every* $h \in \mathcal{D}(L_2) \cap B(\mathbb{E})$. *For each* $v_0 \in \mathbb{E}_0$, *there exists a unique* $\mu_{v_0} \in \mathcal{P}(\mathbb{E}_1)$ *such that* $\int L_1 h(v_0, v_1, v_2)\mu_{v_0,v_1}(dv_2)\mu_{v_0}(dv_1) = 0$ *for every* $h \in \mathcal{D}(L_1) \cap B(\mathbb{E}_0 \times \mathbb{E}_1)$.

With this condition in mind, we define

$$\overline{L}_1 h(v_0, v_1) = \int L_1 h(v_0, v_1, v_2)\mu_{v_0,v_1}(dv_2).$$

Our first convergence condition insures that the *slow component* V_0^N *has a deterministic limit*. Essentially it implies that its "generator" $L_0 h = F \cdot \nabla h$, for $h \in C_c^\infty(\mathbb{E}_0)$. It also identifies the *intrinsic fluctuations of the slow component* via a martingale $M^{N,1}$. For an $\mathbb{R}^{d_0}$-valued process Y, we use $[Y]_t$ to denote the matrix of covariations $[Y_i, Y_j]_t$.

CONDITION 2.5 (First convergence condition). *There exist* $F^N \in M(\mathbb{E}^N, \mathbb{R}^{d_0})$ *and* $F, G_0 \in C(\mathbb{E}, \mathbb{R}^{d_0})$ *such that*

$$M^{N,1}(t) = V_0^N(t) - V_0^N(0) - \int_0^t F^N\big(V^N(s)\big)\,ds \tag{2.3}$$

is a local martingale, $[V_0^N]_t \Rightarrow 0$, *and for each compact* $K \subset \mathbb{E}$,

$$\lim_{N\to\infty} \sup_{v\in K\cap\mathbb{E}^N} \left|r_N\big(F^N(v) - F(v)\big) - G_0(v)\right| = 0. \tag{2.4}$$

We next turn to the relevant *Poisson equations* based on the conditional equilibrium distributions of the fast components and the limiting drift of the slow component. Suppose that there exist $h_1 \in \mathcal{D}(L_1)^{d_0}$ and $h_2, h_3 \in \mathcal{D}(L_2)^{d_0}$ such that

$$\begin{aligned}\overline{L}_1 h_1(v_0, v_1) &= \int F(v_0, v_1, v_2)\mu_{v_0,v_1}(dv_2) \\ &\quad - \iint F(v_0, v_1, v_2)\mu_{v_0,v_1}(dv_2)\mu_{v_0}(dv_1), \\ L_2 h_2(v_0, v_1, v_2) &= F(v_0, v_1, v_2) - \int F(v_0, v_1, v_2)\mu_{v_0,v_1}(dv_2), \\ L_2 h_3(v_0, v_1, v_2) &= \overline{L}_1 h_1(v_0, v_1) - L_1 h_1(v_0, v_1, v_2).\end{aligned} \tag{2.5}$$

Define

$$\begin{aligned}\overline{F}_1(v_0, v_1) &= \int F(v_0, v_1, v_2)\mu_{v_0,v_1}(dv_2), \\ \overline{F}(v_0) &= \iint F(v_0, v_1, v_2)\mu_{v_0,v_1}(dv_2)\mu_{v_0}(dv_1)\end{aligned}$$

and

$$H_N = \frac{1}{r_{1,N}} h_1 + \frac{1}{r_{2,N}}(h_2 + h_3). \tag{2.6}$$

Note that for H_N of this form

$$A_N H_N \approx L_1 h_1 + L_2(h_2 + h_3) = F - \overline{F}.$$

In what follows, H_N does not have to be given by (2.6). That form simply suggests the possibility of finding H_N with the desired properties. Specifically, we assume the existence of $H_N \in \mathcal{D}(A_N)$ satisfying the following convergence condition.

CONDITION 2.6 (Second convergence condition). *Assume that there exists* $G_1 \in C(\mathbb{E}, \mathbb{R}^{d_0})$ *such that for each compact* $K \subset \mathbb{E}$,

$$\lim_{N\to\infty} \sup_{v \in K \cap \mathbb{E}^N} \left| r_N\big(F(v) - \overline{F}(v_0) - A_N H_N(v)\big) - G_1(v) \right| = 0. \tag{2.7}$$

REMARK 2.7. The critical requirements for H_N are (2.7), (2.10) and (2.11). In fact, because of the possibility of large fluctuations by V_1^N and V_2^N, even if h_1, h_2 and h_3 satisfying Condition 2.5 can be found, it may be necessary to define H_N using a sequence of truncations of h_1, h_2 and h_3.

This now identifies the *fluctuations of the slow component due to convergence of the fast components* to their conditional equilibrium distributions via a martingale $M^{N,2}$. For $V_0(0) \in \mathbb{R}^{d_0}$, let V_0 satisfy

$$V_0(t) = V_0(0) + \int_0^t \overline{F}\big(V_0(s)\big)\, ds, \tag{2.8}$$

and define

$$M^{N,2}(t) = H_N\big(V^N(t)\big) - H_N\big(V^N(0)\big) - \int_0^t A_N H_N\big(V^N(s)\big)\,ds.$$

The following condition is then needed for *application of the martingale central limit theorem*, Theorem A.1, to the process $M^{N,1} - M^{N,2}$, composed of $M^{N,2}$ above and $M^{N,1}$ from (2.3). Essentially it says that the jumps of both the slow component and solutions to the Poisson equations are appropriately small, and that the quadratic variation of $M^{N,1} - M^{N,2}$ converges.

CONDITION 2.8 (Converegence of covariation). *There exists* $G \in C(\mathbb{E}, \mathbb{M}^{d_0\times d_0})$ *such that for each* $t > 0$,

$$\lim_{N\to\infty} E\Big[\sup_{s\le t} r_N \big|V_0^N(s) - V_0^N(s-)\big|\Big] = 0, \tag{2.9}$$

$$\sup_{s\le t} r_N H_N\big(V^N(s)\big) \Rightarrow 0 \tag{2.10}$$

and

$$(r_N)^2\big[V_0^N - H_N \circ V^N\big]_t - \int_0^t G\big(V^N(s)\big)\,ds \Rightarrow 0. \tag{2.11}$$

We can now *account for all the terms* in the expansion (2.1) of $U^N = r_N(V_0^N - V_0)$,

$$\begin{aligned} U^N(t) = U^N(0) &+ r_N\big(M^{N,1}(t) - M^{N,2}(t)\big) \\ &+ r_N \int_0^t \big(\overline{F}\big(V_0^N(s)\big) - \overline{F}\big(V_0(s)\big)\big)\,ds \\ &+ r_N \int_0^t \big(F^N\big(V^N(s)\big) - F\big(V^N(s)\big)\big)\,ds \\ &+ r_N \int_0^t \big(F\big(V^N(s)\big) - \overline{F}\big(V_0^N(s)\big) - A_N H_N\big(V^N(s)\big)\big)\,ds \\ &+ r_N\big(H_N\big(V^N(t)\big) - H_N\big(V^N(0)\big)\big). \end{aligned}$$

Conditions 2.1–2.8 insure that all the terms will have a limit as $N \to \infty$. The limit of the second term on the right is guaranteed by (2.9), (2.10) and (2.11) in Condition 2.8. Assuming that $\overline{F}$ is smooth, the third term on the right is asymptotic to $\int_0^t \nabla\overline{F}(V_0(s)) \cdot U^N(s)\,ds$. The fourth term is controlled by (2.4) of Condition 2.5, the fifth by (2.7) of Condition 2.6. The order of time scale parameters (2.2) from Condition 2.1 and the form of H_N in (2.6) suggests that the sixth term goes to zero, but we will explicitly assume that in the statement of the theorem.

Finally, we now only need a condition to *ensure relative compactness* of the sequence. If $\mathbb{E}$ is unbounded, let $\psi : \mathbb{E} \to [1, \infty)$ be locally bounded and satisfy

$\lim_{v\to\infty}\psi(v)=\infty$, *or let* $\psi:\mathbb{E}\to[1,\infty)$ *be such that* $\forall M<\infty$, $\{v\in\mathbb{E}:\psi(v)\le M\}$ *is relatively compact in* $\mathbb{E}$, and let D_ψ denote the collection of continuous functions f satisfying

$$\sup_{v\in\mathbb{E}}\frac{|f(v)|}{\psi(v)}<\infty,\qquad \lim_{k\to\infty}\sup_{v\in\mathbb{E},|v|>k}\frac{|f(v)|}{\psi(v)}=0.$$

For sequences of space–time random measures, the notion of convergence that we will use is that discussed in Kurtz (1992).

LEMMA 2.9. *Let* V^N *be a sequence of* $\mathbb{E}$*-valued processes, and define the occupation measure*

$$\Gamma_N(D\times[0,t])=\int_0^t \mathbf{1}_D(V^N(s))\,ds. \tag{2.12}$$

Suppose that for each $t>0$

$$\sup_N E\left[\int_0^t \psi(V^N(s))\,ds\right]<\infty. \tag{2.13}$$

Then $\{\Gamma_N\}$ *is relatvely compact, and if* $\Gamma_N\Rightarrow\Gamma$, *then for* $f_1,\dots,f_m\in D_\psi$,

$$\left(\int_0^\cdot f_1(V^N(s))\,ds,\dots,\int_0^\cdot f_m(V^N(s))\,ds\right)$$
$$\Rightarrow\left(\int_{\mathbb{E}} f_1(v)\Gamma(dv\times[0,\cdot]),\dots,\int_{\mathbb{E}} f_m(v)\Gamma(dv\times[0,\cdot])\right)$$

in $C_{\mathbb{R}^m}[0,\infty)$.

PROOF. Relative compactness of $\{\Gamma_N\}$ follows from Lemma 1.3 of Kurtz (1992). Relative compactness in $C_{\mathbb{R}^m}[0,\infty)$ follows from relative compactness of each component. To see that for $f\in D_\psi$, the sequence $X^N=\int_0^\cdot f(V^N(s))\,ds$ is relatively compact, it is enough to approximate the sequence by sequences known to be relatively compact. For $\varepsilon>0$, there exists a compact $K_\varepsilon\subset\mathbb{E}$ and $C>0$, such that $|f|\le(C\mathbf{1}_{K_\varepsilon}+\varepsilon)\psi$. Define $X_\varepsilon^N=\int_0^\cdot \mathbf{1}_{K_\varepsilon}(V^N(s))f(V^N(s))\,ds$. Note that X_ε^N is Lipschitz with Lipschitz constant $\sup_{v\in K_\varepsilon}|f(v)|$, so $\{X_\varepsilon^N\}$ is relatively compact. For $\delta>0$,

$$\sup_N P\left\{\sup_{s\le t}|X^N(s)-X_\varepsilon^N(s)|\ge\delta\right\}\le\frac{\varepsilon}{\delta}\sup_N E\left[\int_0^t\psi(V^N(s))\,ds\right],$$

and relative compactness of $\{X^N\}$ follows; see Problem 3.11.18 of Ethier and Kurtz (1986).

Assuming that $\Gamma_N\Rightarrow\Gamma$, the convergence of $\int_0^\cdot f(V^N(s))\,ds$ to $\int_{\mathbb{E}\times[0,\cdot]}f(v)\times\Gamma(dv\times ds)$ follows by the same type of approximation. □

The final condition insures relative compactness of V^N.

CONDITION 2.10 (Tightness). *If $\mathbb{E}$ is unbounded, there exists a locally bounded $\psi : \mathbb{E} \to [1, \infty)$ satisfying $\lim_{v\to\infty} \psi(v) = \infty$ such that for each $t > 0$,*

$$\sup_N E\left[\int_0^t \psi\big(V^N(s)\big)\,ds\right] < \infty \tag{2.14}$$

and all of the following functions are in D_ψ: $\sup_N |F^N|$, $\sup_N |r_N(F^N - F)|$, $\sup_N |r_N(F - \overline{F} - A_N H_N)|$, $|G|$, $\sup_N |A_N h|$ for $h \in \mathcal{D}(L_0) \cap B(\mathbb{E}_0)$, $\sup_N |\frac{1}{r_{1,N}} A_N h|$ for $h \in \mathcal{D}(L_1) \cap B(\mathbb{E}_0 \times \mathbb{E}_1)$, and $\sup_N |\frac{1}{r_{2,N}} A_N h|$ for $h \in \mathcal{D}(L_2) \cap B(\mathbb{E})$.

Assuming the above conditions and defining

$$\overline{G}(v_0) = \iint G(v_0, v_1, v_2)\mu_{v_0,v_1}(dv_2)\mu_{v_0}(dv_1), \tag{2.15}$$

and similarly for $\overline{G}_0$ and $\overline{G}_1$, we have the following functional central limit theorem.

THEOREM 2.11. *Under the above conditions, suppose that $\lim_{N\to\infty} U^N(0) = U(0)$, that $\overline{F}$ is continuously differentiable and that the solution (necessarily unique) of (2.8) exists for all time. Then for each $t > 0$,*

$$\sup_{s\le t} \big|V_0^N(s) - V_0(s)\big| \Rightarrow 0,$$

$r_N(M^{N,1} - M^{N,2}) \Rightarrow M$, where M has Gaussian, mean-zero, independent increments with

$$E\big[M(t)M^T(t)\big] = \int_0^t \overline{G}\big(V_0(s)\big)\,ds, \tag{2.16}$$

and $U^N \Rightarrow U$ satisfying

$$U(t) = U(0) + M(t) + \int_0^t \big(\nabla\overline{F}\big(V_0(s)\big)U(s) + \overline{G}_0\big(V_0(s)\big) + \overline{G}_1\big(V_0(s)\big)\big)\,ds.$$

Assuming $\overline{G} = \sigma\sigma^T$, we can write

$$\begin{aligned} U(t) = U(0) &+ \int_0^t \sigma\big(V_0(s)\big)\,dW(s) \\ &+ \int_0^t \big(\nabla\overline{F}\big(V_0(s)\big)U(s) + \overline{G}_0\big(V_0(s)\big) + \overline{G}_1\big(V_0(s)\big)\big)\,ds. \end{aligned} \tag{2.17}$$

REMARK 2.12. As noted above, the corresponding theorem for systems with two time-scales is obtained by assuming $\mathbb{E}_2$ consists of a single point so $L_2 f \equiv 0$.

PROOF OF THEOREM 2.11. Let Γ_N be the occupation measure defined as in (2.12). Then by Lemma 2.9, $\{\Gamma_N\}$ is relatively compact. Assume, for simplicity that $\Gamma_N \Rightarrow \Gamma$. We will show that Γ is uniquely determined.

Condition 2.5, equation (2.9) and the martingale central limit theorem, Theorem A.1, imply $M^{N,1} \Rightarrow 0$, and Lemma 2.9 then implies $V_0^N \Rightarrow V_0^\infty$, where

$$V_0^\infty(t) = V_0(0) + \int_{\mathbb{E}\times[0,t]} F(v)\Gamma(dv \times ds). \tag{2.18}$$

Condition 2.10, the definition of L_2, and Lemma 2.9 imply

$$\frac{1}{r^{2,N}}\Big(h\big(V^N(t)\big) - h\big(V^N(0)\big) - \int_0^t A_N h\big(V^N(s)\big)\,ds\Big) \Rightarrow \int_{\mathbb{E}\times[0,t]} L_2 h(v)\Gamma(dv \times ds)$$

for every $h \in C_c^\infty(\mathbb{E})$. The uniform integrability implied by (2.14) implies that the limit is a continuous martingale with sample paths of finite variation and hence is identically zero. Condition 2.4 then implies [see Example 2.3 of Kurtz (1992)] that Γ can be written

$$\Gamma(dv \times ds) = \mu_{v_0,v_1}(dv_2)\Gamma^{0,1}(dv_0 \times dv_1 \times ds).$$

A similar argument gives

$$\begin{aligned} 0 &= \int_{\mathbb{E}\times[0,t]} L_1 h(v)\Gamma(dv \times ds) \\ &= \int_{\mathbb{E}_0\times\mathbb{E}_1\times[0,t]} \overline{L}_1 h(v_0, v_1)\Gamma^{0,1}(dv_0 \times dv_1 \times ds), \end{aligned}$$

which implies

$$\Gamma^{0,1}(dv_0 \times dv_1 \times ds) = \mu_{v_0}(dv_1)\Gamma^0(dv_0 \times ds).$$

But the convergence of V_0^N to V_0^∞ implies $\Gamma^0(dv_0 \times ds) = \delta_{V_0^\infty(s)}(dv_0)\,ds$.

Now (2.18) can be rewritten

$$V_0^\infty(t) = V_0(0) + \int_0^t \overline{F}\big(V_0^\infty(s)\big)\,ds, \tag{2.19}$$

and it follows that $V_0^\infty = V_0$.

Similarly, (2.11) now becomes

$$(r_N)^2\big[V_0^N - H_N \circ V^N\big]_t \Rightarrow \int_0^t \overline{G}\big(V_0(s)\big)\,ds,$$

and it follows that $r_N(M^{N,1} - M^{N,2}) \Rightarrow M$ as desired.

Finally, the uniform integrability implied by (2.14) and Condition 2.10 allows interchange of limits and integrals in the expansion of U^N given in (2.1), and the convergence of U^N to U follows. □

3. Diffusion approximation. The functional central limit theorem, Theorem 2.11, suggests approximating V_0^N by $V_0 + \frac{1}{r_N}U$. In turn, that observation and (2.17) suggest approximating V_0^N by a diffusion process given by the Itô equation

$$\begin{aligned} D^N(t) &= V_0^N(0) + \frac{1}{r_N}\int_0^t \sigma\big(D^N(s)\big)\,dW(s) \\ &\quad + \int_0^t \left(\overline{F}\big(D^N(s)\big) + \frac{1}{r_N}\overline{G}_0\big(D^N(s)\big) + \frac{1}{r_N}\overline{G}_1\big(D^N(s)\big)\right) ds. \end{aligned} \tag{3.1}$$

The approximation

$$V_0^N \approx \widehat{D}^N \equiv V_0 + \frac{1}{r_N}U$$

is, of course, justified by Theorem 2.11. Justification for the approximation $V_0^N \approx D^N$ is less clear, since D^N is not produced as a limit. Noting, however, that

$$\begin{aligned} \widehat{D}^N(t) &= V_0^N(0) + \frac{1}{r_N}\int_0^t \sigma\big(V_0(s)\big)\,dW(s) \\ &\quad + \int_0^t \left(\overline{F}\big(V_0(s)\big) + \frac{1}{r_N}\nabla\overline{F}\big(V_0(s)\big)U(s) + \frac{1}{r_N}\overline{G}_0\big(V_0(s)\big)\right. \\ &\qquad \left. + \frac{1}{r_N}\overline{G}_1\big(V_0(s)\big)\right) ds, \end{aligned}$$

assuming smoothness of $\overline{F}$, $\overline{G}_0$ and $\overline{G}_1$, we see that $r_N^2(D^N - \widehat{D}^N)$ converges to $\widehat{U}$ satisfying

$$\begin{aligned} \widehat{U}(t) &= \int_0^t \nabla\sigma\big(V_0(s)\big)U(s)\,dW(s) \\ &\quad + \int_0^t \left(\nabla\overline{F}\big(V_0(s)\big)\widehat{U}(s) + \frac{1}{2}U^T(s)\partial^2\overline{F}\big(V_0(s)\big)U(s)\right. \\ &\qquad \left. + \big(\nabla\overline{G}_0\big(V_0(s)\big) + \nabla\overline{G}_1\big(V_0(s)\big)\big)U(s)\right) ds, \end{aligned}$$

and since the central limit theorem demonstrates that the fluctuations of V^N are of order $O(r_N^{-1})$, we see that the difference between the two approximations D^N and $\widehat{D}^N$ is negligible compared to these fluctuations.

4. Markov chain models for chemical reactions. A reaction network is a chemical system involving multiple reactions and chemical species. The kind of stochastic model for a network that we will consider treats the system as a continuous time Markov chain whose state X is a vector giving the number of molecules X_i of each species of type $i \in \mathcal{I}$ present. Each reaction is modeled as a possible

transition for the state. The model for the kth reaction, for each $k \in \mathcal{K}$, is determined by a vector of inputs ν_k specifying the numbers of molecules of each chemical species that are consumed in the reaction, a vector of outputs ν_k' specifying the numbers of molecules of each species that are produced in the reaction, and a function of the state $\lambda_k(x)$ that gives the rate at which the reaction occurs as a function of the state. Specifically, if the kth reaction occurs at time t, the change in X is a vector of integer values $\zeta_k = \nu_k' - \nu_k$.

Let $R_k(t)$ denote the number of times that the kth reaction occurs by time t. Then R_k is a counting process with intensity $\lambda_k(X(t))$ (called the *propensity* in the chemical literature) and can be written as

$$R_k(t) = Y_k\biggl(\int_0^t \lambda_k\bigl(X(s)\bigr)\,ds\biggr),$$

where the Y_k are independent unit Poisson processes. The state of the system at time t can be written as

$$X(t) = X(0) + \sum_k \zeta_k R_k(t) = X(0) + \sum_k \zeta_k Y_k\biggl(\int_0^t \lambda_k\bigl(X(s)\bigr)\,ds\biggr).$$

In the stochastic version of the law of mass action, the rate function is proportional to the number of ways of selecting the molecules that are consumed in the reaction, that is,

$$\lambda_k(x) = \kappa_k' \prod_i \nu_{ik}! \prod_i \binom{x_i}{\nu_{ik}} = \kappa_k' \prod_i x_i(x_i - 1)\cdots(x_i - \nu_{ik} + 1).$$

Of course, physically, $|\nu_k| = \sum_i \nu_{ik}$ is usually assumed to be less than or equal to two, but that does not play a significant role in the analysis that follows.

A reaction network may exhibit behavior on multiple scales due to the fact that some species may be present in much greater abundance than others, and the rate functions may vary over several orders of magnitude. Following Kang and Kurtz (2013), we embed the model of interest in a sequence of models indexed by a scaling parameter N. The model of interest corresponds to a particular value of the scaling parameter N_0. For each species $i \in \mathcal{I} = \{1, \ldots, s\}$, we specify a parameter $\alpha_i \geq 0$ and normalize the number of molecules by $N_0^{\alpha_i}$ defining $Z_i^{N_0}(t) = N_0^{-\alpha_i} X_i(t)$ so that it is of $O(1)$. For each reaction $k \in \mathcal{K}$, we specify another parameter β_k and normalize the reaction rate constant as $\kappa_k' = \kappa_k N_0^{\beta_k}$ so that κ_k is of $O(1)$. One can observe this model on different time scales as well, by replacing t by tN_0^{γ}, for some $\gamma \in \mathbb{R}$. The model then becomes a Markov chain on $\mathbb{E}^{N_0} = N_0^{-\alpha_1}\mathbb{Z}_+ \times \cdots \times N_0^{-\alpha_s}\mathbb{Z}_+$ which, when $N = N_0$, evolves according to

$$Z_i^N(t) = Z_i^N(0) + \sum_k N^{-\alpha_i} \zeta_{ik} Y_k\biggl(\int_0^t N^{\nu_k\cdot\alpha+\beta_k+\gamma} \lambda_k^N\bigl(Z^N(s)\bigr)\,ds\biggr)$$

with

$$\lambda_k^N(z) = \kappa_k \prod_i z_i\big(z_i - N^{-\alpha_i}\big)\cdots\big(z_i - (\nu_{ik}-1)N^{-\alpha_i}\big).$$

If for some i, $\alpha_i > 0$ and $\nu_{ik} > 1$, then λ_k^N varies with N but converges as $N \to \infty$. To simplify notation, we will write $\lambda_k(z)$ rather than λ_k^N, but one should check that the N-dependence is indeed negligible in the analysis that we do. Defining $\Lambda_N = \operatorname{diag}(N^{-\alpha_1}, \ldots, N^{-\alpha_s})$, so $Z^N = \Lambda_N X$, let

$$A_N f(z) = \sum_k N^{\rho_k} \lambda_k(z)\big(f(z + \Lambda_N \zeta_k) - f(z)\big),$$

where $\rho_k = \nu_k \cdot \alpha + \beta_k + \gamma$. Since the change of time variable from t to tN^γ is equivalent to scaling the generator by a factor of N^γ, we initially take γ to be zero. We subsequently consider the behaviour of Z^N on different time-scales $Z^N(\cdot N^\gamma)$.

To be precise regarding the domain of A_N, note that because the jumps of Z^N are uniformly bounded, if we define $\tau_r^N = \inf\{t : |Z^N(t)| \geq r\}$, then for every continuous function f,

$$f\big(Z^N\big(t \wedge \tau_r^N\big)\big) - f\big(Z^N(0)\big) - \int_0^{t\wedge\tau_r^N} A_N f\big(Z^N(s)\big)\,ds$$

is a martingale.

For notational simplicity, assume that the α_i satisfy $0 \leq \alpha_1 \leq \cdots \leq \alpha_s$, and let $d_\circ \geq 0$ satisfy $\alpha_i = 0$, $i \leq d_\circ$ and $\alpha_i > 0$, $i > d_\circ$.

To apply the results of Section 2, we identify $r_N, r_{1,N}, r_{2,N}$ from the reaction network and the parameters $\{\alpha_i\}, \{\beta_k\}$ as follows. Let

$$m_2 = \max\{\rho_k - \alpha_i : \zeta_{ik} \neq 0\},$$

and define $r_{2,N} = N^{m_2}$. Then there exists a linear operator L_2 such that for each compact $K \subset \mathbb{R}^s$,

$$\lim_{N\to\infty} \sup_{z\in K\cap\mathbb{E}^N} \left|\frac{1}{r_{2,N}} A_N h(z) - L_2 h(z)\right| = 0, \qquad h \in \mathcal{D}(L_2) = C^1\big(\mathbb{R}^s\big).$$

Depending on the relationship between ρ_k and α_i for $\zeta_{ik} \neq 0$ and the time-scale parameter γ, the limiting operator L_2 is either the generator for a Markov chain, a differential operator, or a combination of the two, which would be the generator for a piecewise deterministic Markov process (PDMP). We classify the reactions by defining

$$\mathcal{K}_{2,\circ} = \{k \in \mathcal{K} : \rho_k = m_2\}$$

and

$$\mathcal{K}_{2,\bullet} = \{k \in \mathcal{K} : \rho_k - \alpha_i = m_2 \text{ for some } i \text{ with } \alpha_i > 0, \zeta_{ik} \neq 0\}.$$

For each $k \in \mathcal{K}_{2,\circ} \cup \mathcal{K}_{2,\bullet}$ define

$$\zeta_{2,k} = \lim_{N\to\infty} N^{\rho_k - m_2} \Lambda_N \zeta_k \in \mathbb{Z}.$$

Note that throughout the paper $\zeta_{2,k}$ will denote the limiting reaction vector, not to be confused with the single matrix entry ζ_{ik}. Then, for $h \in C^1(\mathbb{R}^s)$

$$L_2 h(z) = \sum_{k\in\mathcal{K}_{2,\circ}} \lambda_k(z)\big(h(z+\zeta_{2,k}) - h(z)\big) + \sum_{k\in\mathcal{K}_{2,\bullet}} \lambda_k(z)\nabla h(z)\cdot\zeta_{2,k}. \tag{4.1}$$

Note that, although $\lambda_k(z)$ depends on all species types, the dynamics defined by L_2 makes changes only due to reactions $\mathcal{K}_{2,\circ} \cup \mathcal{K}_{2,\bullet}$. In other words, only the subnetwork defined by reactions $\mathcal{K}_{2,\circ} \cup \mathcal{K}_{2,\bullet}$ is relevant on the time-scale corresponding to $\gamma = -m_2$. If $\mathcal{K}_{2,\bullet}$ is empty, the process corresponding to L_2 is a Markov chain, and if $\mathcal{K}_{2,\circ}$ is empty, the process is just the solution of an ordinary differential equation. If both are nonempty, the process is piecewise deterministic in the sense of Davis (1993).

The process corresponding to L_2 can be obtained as the solution of

$$V_2(t) = V_2(0) + \sum_{k\in\mathcal{K}_{2,\circ}} \zeta_{2,k} Y_k\left(\int_0^t \lambda_k\big(V_2(s)\big)\,ds\right) + \sum_{k\in\mathcal{K}_{2,\bullet}} \zeta_{2,k}\int_0^t \lambda_k\big(V_2(s)\big)\,ds,$$

and assuming that V_2 does not hit infinity in finite time, $Z^N(\cdot N^{-m_2}) \Rightarrow V_2$.

The central limit theorem in Section 2 assumes that the state space is a product space and that the fast process "averages out" one component. The state space on which functions in the domain of L_1 in Condition 2.2 are defined is such that every function on it is contained in the kernal of L_2. In order to separate the state space in this way, we need to identify the combinations of species variables whose change on the fastest time-scale $\gamma = -m_2$ is less than $O(1)$. This can be done with a change of basis of the original state space as follows.

Let $S_{\mathcal{K}}$ be a matrix whose columns are $\zeta_k, k \in \mathcal{K}$ for some subset $\mathcal{K} \subset \mathcal{K}$. Then $S_{\mathcal{K}}$ is the *stoichiometric matrix* associated with the reaction subnetwork $\mathcal{K}$. For the species types whose behavior is discrete, $S_{\mathcal{K}}$ gives the possible jumps, while for the species whose behavior evolves continuously, $S_{\mathcal{K}}$ determines the possible paths. We will let $\mathcal{R}(S_{\mathcal{K}}) = \operatorname{span}\{\zeta_k, k \in \mathcal{K}\} \subset \mathbb{R}^s$ denote the range of $S_{\mathcal{K}}$, called the *stoichiometric subspace* of the chemical reaction subnetwork $\mathcal{K}$, and we will let

$$\mathcal{N}(S_{\mathcal{K}}^T) = \left\{\theta \in \mathbb{R}^s : \sum_{i\in\mathcal{I}} \theta_i \zeta_{ik} = 0\ \forall k \in \mathcal{K}\right\}$$

denote the null space of $S_{\mathcal{K}}^T$ which is the othogonal complement of $\mathcal{R}(S_{\mathcal{K}})$. For each initial value z_0 of the reaction system, $z_0 + \mathcal{R}(S_{\mathcal{K}})$ defines the *stoichiometric compatibility class* of the system. Then both stochastically and deterministically evolving components of the system must remain in the stoichiometric compatibility class for all time $t > 0$. The linear combinations of the species $\theta \cdot X$ for

$\theta \in \mathcal{N}(S_{\mathcal{K}}^T)$ are conserved quantities; that is, they are constant along the trajectories of the evolution of the reaction subnetwork $\mathcal{K}$.

On the time scale $\gamma = -m_2$, the fast subnetwork determined by L_2 has the stoichiometric matrix S_2 whose columns are $\{\zeta_{2,k}, k \in \mathcal{K}_{2,\circ} \cup \mathcal{K}_{2,\bullet}\}$. Define $\mathcal{N}(S_2^T)$ as above, and note that $\theta \cdot V_2$, $\theta \in \mathcal{N}(S_2^T)$, are conserved quantities for the fast subnetwork, that is, $\theta \cdot V_2(t)$ does not depend on t. Let s_2 denote the dimension of $\mathcal{R}(S_2)$, and $s_1' = s - s_2$ be the dimension of $\mathcal{N}(S_2^T)$. We now replace the natural state space of the process by $\mathcal{N}(S_2^T) \times \mathcal{R}(S_2)$, mapping the original processes onto this product space by the orthogonal projection $\Pi_{\mathcal{N}(S_2^T)} \times \Pi_{\mathcal{R}(S_2)}$, that is,

$$\big(V_1'^N(t), V_2^N(t)\big) = \big(\Pi_{\mathcal{N}(S_2^T)} Z^N(t), \Pi_{\mathcal{R}(S_2)} Z^N(t)\big).$$

Note that the original coordinates have different underlying state spaces $N^{-\alpha_i}\mathbb{Z}$; however, the change of basis will combine only those coordinates with the same scaling parameter α_i. To see that this is the case, note that by the definition of $\zeta_{2,k}$, $\zeta_{2,ik} \neq 0$ and $\zeta_{2,jk} \neq 0$ implies $\alpha_i = \alpha_j$. It follows that there is a basis $\theta_1, \ldots, \theta_{s_1'}$ for $\mathcal{N}(S_2^T)$ such that $\theta_{il} \neq 0$ and $\theta_{jl} \neq 0$ implies $\alpha_i = \alpha_j$, and we can take this basis to be orthonormal. We denote the common scaling parameter by α_{θ_l}. Let Θ_1 be the matrix with rows $\theta_1^T, \ldots, \theta_{s_1'}^T$ so that $(\Theta_1 z)^T = (\theta_1 \cdot z, \ldots, \theta_{s_1'} \cdot z)^T$ and the orthogonal projection is given by

$$\Pi_{\mathcal{N}(S_2^T)} = \Theta_1^T \Theta_1 = \sum_{l=1}^{s_1'} \theta_l \theta_l^T.$$

On the next time scale we only need to consider the dynamics of the projection of the original process that is unaffected by the fast subnetwork $V_1'^N = \Pi_{\mathcal{N}(S_2^T)} \Lambda_N X$. Since $\Pi_{\mathcal{N}(S_2^T)} \Lambda_N = \Lambda_N \Pi_{\mathcal{N}(S_2^T)}$, we have

$$V_1'^N(t) = \Pi_{\mathcal{N}(S_2^T)} Z^N(0) + \Lambda_N \sum_k \Pi_{\mathcal{N}(S_2^T)} \zeta_k Y_k\bigg(N^{\rho_k} \int_0^t \lambda_k^N\big(Z^N(s)\big)\, ds\bigg).$$

Note that $\Pi_{\mathcal{R}(S_2)} \zeta_k$ is not necessarily equal to $\zeta_{2,k}$, nor is the other projection $\Pi_{\mathcal{N}(S_2^T)} \zeta_k = \zeta_k - \Pi_{\mathcal{R}(S_2)} \zeta_k$ necessarily equal to $\zeta_k - \zeta_{2,k}$. To identify the next time scale let

$$m_1 = \max\{\rho_k - \alpha_{\theta_l} : \theta_l \cdot \zeta_k \neq 0\} = \max\big\{\rho_k - \alpha_i : (\Pi_{\mathcal{N}(S_2^T)} \zeta_k)_i \neq 0\big\},$$

and define $r_{1,N} = N^{m_1}$. Note that $m_1 < m_2$. Then there exists a linear operator L_1 such that for each compact $K \subset \mathbb{R}^{s_1'}$,

$$\lim_{N\to\infty} \sup_{z \in K \cap \mathbb{E}^N} \bigg| \frac{1}{r_{1,N}} A_N h(z) - L_1 h(z) \bigg| = 0, \tag{4.2}$$

where $h \in \mathcal{D}(L_1)$ satisfies $h(z) = f(\theta_1 \cdot z, \ldots, \theta_{s_1'} \cdot z)$ for $f \in C^1(\mathbb{R}^{s_1'})$. Define

$$\mathcal{K}_{1,\circ} = \Big\{k \in \mathcal{K} : \rho_k = m_1, \max_l |\theta_l \cdot \zeta_k| > 0\Big\}$$

and

$$\mathcal{K}_{1,\bullet} = \{k \in \mathcal{K} : \rho_k - \alpha_{\theta_l} = m_1 \text{ for some } l \text{ with } \alpha_{\theta_l} > 0, \theta_l \cdot \zeta_k \neq 0\}.$$

Let $\Lambda_N^{\Theta_1} = \operatorname{diag}(N^{-\alpha_{\theta_1}}, \ldots, N^{-\alpha_{\theta_{s_1'}}})$, and for each $k \in \mathcal{K}_{1,\circ} \cup \mathcal{K}_{1,\bullet}$ define

$$\zeta_{1,k}^{\theta} = \lim_{N\to\infty} N^{\rho_k - m_1} \Lambda_N^{\Theta_1} \Theta_1 \zeta_k = \lim_{N\to\infty} N^{\rho_k - m_1} \big(N^{-\alpha_{\theta_1}} \theta_1 \cdot \zeta_k, \ldots, N^{-\alpha_{\theta_{s_1'}}} \theta_{s_1'} \cdot \zeta_k\big)^T.$$

Then for $h(z) = f(\Theta_1 z)$ with $f \in C^1(\mathbb{R}^{s_1'})$

$$L_1 h(z) = \sum_{k\in\mathcal{K}_{1,\circ}} \lambda_k(z)\big(f\big(\Theta_1 z + \zeta_{1,k}^{\theta}\big) - f(\Theta_1 z)\big) + \sum_{k\in\mathcal{K}_{1,\bullet}} \lambda_k(z) \nabla f(\Theta_1 z) \cdot \zeta_{1,k}^{\theta}.$$

If V_1 denotes the process corresponding to L_1, then assuming that V_1 does not hit infinity in finite time, $V_1'^N(\cdot N^{-m_1}) = \Pi_{\mathcal{N}(S_2^T)} Z^N(\cdot N^{-m_1}) \Rightarrow V_1$.

To separate the state space in terms of the next time scale (if there is one), define

$$\zeta_{1,k} = \lim_{N\to\infty} N^{\rho_k - m_1} \Lambda_N \Pi_{\mathcal{N}(S_2^T)} \zeta_k.$$

In other words, $\zeta_{1,k} = \Theta_1^T \zeta_{1,k}^{\theta}$ is embedded in the original space, and $\zeta_{1,k}^{\theta} = \Theta_1 \zeta_k^1$. On the time scale $\gamma = -m_1$, the subnetwork determined by L_1 has the stoichiometric matrix S_1 with columns $\{\zeta_{1,k} k \in \mathcal{K}_{1,\circ} \cup \mathcal{K}_{1,\bullet}\}$. Define the subspace $\mathcal{N}(S_1^T)$ as before, and let s_1 denote the dimension of $\mathcal{R}(S_1)$ and $s_0 = s_1' - s_1$ be the dimension of $\mathcal{N}(S_1^T)$. As before we need to map the processes $V_1'^N$ onto this product space by the orthogonal projection $\Pi_{\mathcal{N}(S_1^T)} \times \Pi_{\mathcal{R}(S_1)}$. Since $\zeta_{1,k} \in \mathcal{N}(S_2^T) = \operatorname{span}(\theta_1, \ldots, \theta_{s_1'})$, we can assume that the θ_l are selected so that

$$\mathcal{R}(S_1) = \operatorname{span}(\theta_{s_0+1}, \ldots, \theta_{s_1'}) = \operatorname{span}(\zeta_{1,1}, \ldots, \zeta_{1,s_1}).$$

Define

$$\Pi_0 = \sum_{l=1}^{s_0} \theta_l \theta_l^T = \Pi_{\mathcal{N}(S_1^T)}, \qquad \Pi_1 = \sum_{l=s_0+1}^{s_1'} \theta_l \theta_l^T = \Pi_{\mathcal{R}(S_1)} \quad \text{and} \quad \Pi_2 = \Pi_{\mathcal{R}(S_2)}.$$

On the next time scale we need only consider the projection $\Pi_0 Z^N$ of the original process which is unaffected by either of the faster subnetworks. To identify the next time scale, let

$$m_0 = \max\{\rho_k - \alpha_{\theta_l} : \theta_l \cdot \zeta_k \neq 0, 1 \le l \le s_0\} = \max\big\{\rho_k - \alpha_i : (\Pi_0 \zeta_k)_i \neq 0\big\},$$

and define $r_{0,N} = N^{m_0}$. Note that if $1 \le s_2, 1 \le s_1, 1 \le s_0$ $(s_0 + s_1 + s_2 = s)$, $m_0 < m_1 < m_2$. Without loss of generality, we can assume that time is scaled so that

$m_0 = 0$. Then, there exists a linear operator L_0 such that for each compact $K \subset \mathbb{R}^{s_0}$,

$$\lim_{N\to\infty} \sup_{z\in K\cap\mathbb{E}^N} \bigl|A_N h(z) - L_0 h(z)\bigr| = 0,$$

where $h \in \mathcal{D}(L_0)$ satisfies $h(z) = f(\theta_1 \cdot z, \ldots, \theta_{s_0} \cdot z)$ for $f \in C^1(\mathbb{R}^{s_0})$. Define

$$\mathcal{K}_{0,\circ} = \Bigl\{k \in \mathcal{K} : \rho_k = 0, \max_l |\theta_l \cdot \zeta_k| > 0\Bigr\}$$

and

$$\mathcal{K}_{0,\bullet} = \{k \in \mathcal{K} : \rho_k - \alpha_{\theta_l} = 0 \text{ for some } l \text{ with } \alpha_{\theta_l} > 0, \theta_l \cdot \zeta_k \neq 0, 1 \le l \le s_0\}.$$

As before, let Θ_0 be the matrix with rows $\theta_1^T, \ldots, \theta_{s_0}^T$, and let $\Lambda_N^{\Theta_0} = \operatorname{diag}(N^{-\alpha_{\theta_1}}, \ldots, N^{-\alpha_{\theta_{s_0}}})$, so that $\Pi_0 = \Pi_{\mathcal{N}(S_1^T)} = \Theta_0^T\Theta_0$ and for each $k \in \mathcal{K}_{0,\circ} \cup \mathcal{K}_{0,\bullet}$ define

$$\zeta_k^{\theta,0} = \lim_{N\to\infty} N^{\rho_k}\Lambda_N^{\Theta_0}\Theta_0\zeta_k = \lim_{N\to\infty} N^{\rho_k}\bigl(N^{-\alpha_{\theta_1}}\theta_1 \cdot \zeta_k, \ldots, N^{-\alpha_{\theta_{s_0}}}\theta_{s_0} \cdot \zeta_k\bigr)^T.$$

For $h(z) = f(\Theta_0 z)$ with $f \in C^1(\mathbb{R}^{s_0})$

$$L_0 h(z) = \sum_{k\in\mathcal{K}_{0,\circ}} \lambda_k(z)\bigl(f\bigl(\Theta_0 z + \zeta_k^{\theta,0}\bigr) - f(\Theta_0 z)\bigr) + \sum_{k\in\mathcal{K}_{0,\bullet}} \lambda_k(z)\nabla f(\Theta_0 z) \cdot \zeta_k^{\theta,0}.$$

To relate the above calculations to the results of Section 2, we assume that $\mathcal{K}_{0,\circ} = \varnothing$ so that

$$L_0 h(z) = \sum_{k\in\mathcal{K}_{0,\bullet}} \lambda_k(z)\nabla f(\Theta_0 z) \cdot \zeta_k^{\theta,0}.$$

Let $V^N = TZ^N \equiv (\Pi_0 Z^N, \Pi_1 Z^N, \Pi_2 Z^N)$, so $V^N = (V_0^N, V_1^N, V_2^N) \in \mathcal{N}(S_1^T) \times \mathcal{R}(S_1) \times \mathcal{R}(S_2)$, and note that T is invertible so that the intensities can be written as functions of $v \in \mathcal{N}(S_1^T) \times \mathcal{R}(S_1) \times \mathcal{R}(S_2)$, that is, $\lambda_k(T^{-1}v)$. Since $\Pi_0 z = \sum_{l=1}^{s_0}(\theta_l \cdot z)\theta_l$ and $\Pi_1 z = \sum_{l=s_0+1}^{s_1'}(\theta_l \cdot z)\theta_l$, the process $V_0^N = \Pi_0 Z^N$ is the embedding of $\Theta_0 Z^N$, and similarly $(V_0^N, V_1^N) = (\Pi_0 Z^N, \Pi_1^N Z^N)$ is just the embedding of $\Theta_1 Z^N$. Let $\mathbb{E}_0$, $\mathbb{E}_1$, and $\mathbb{E}_2$ denote the limit of the state spaces for V_0^N, V_1^N and V_2^N.

The function F^N in (2.3) is given by

$$F^N(v) = \sum_k N^{\rho_k}\Lambda_N^{\Theta_0}\lambda_k\bigl(T^{-1}v\bigr)\Theta_0\zeta_k \tag{4.3}$$

and

$$F(v) = \lim_{N\to\infty} F^N(v) = \sum_{k\in\mathcal{K}_{0,\bullet}} \lambda_k\bigl(T^{-1}v\bigr)\zeta_k^{\theta,0}.$$

To satisfy Condition 2.4 we will assume that L_2 is such that for each $(v_0, v_1) \in \mathbb{E}_0 \times \mathbb{E}_1$ there exists a unique conditional equilibrium distribution $\mu_{v_0,v_1}(dv_2) \in \mathcal{P}(\mathbb{E}_2)$ for L_2. Then $\overline{L}_1 h(v_0, v_1) = \int L_1 h(v_0, v_1, u_2)\mu_{v_0,v_1}(du_2)$ is

$$\begin{aligned}\overline{L}_1 h(v_0, v_1) = &\sum_{k\in\mathcal{K}_{1,\circ}} \overline{\lambda}_k(v_0, v_1)\big(f\big((v_0, v_1) + \zeta^{\theta}_{1,k}\big) - f(v_0, v_1)\big)\\ &+ \sum_{k\in\mathcal{K}_{1,\bullet}} \overline{\lambda}_k(v_0, v_1)\nabla f(v_0, v_1)\cdot\zeta^{\theta}_{1,k},\end{aligned} \tag{4.4}$$

where $\overline{\lambda}_k(v_0, v_1) = \int \lambda_k(T^{-1}(v_0, v_1, v_2))\mu_{v_0,v_1}(dv_2)$. For Condition 2.4 to be met, we also need to assume that for each $v_0 \in \mathbb{E}_0$ there exists a unique conditional equilibrium distribution $\mu_{v_0}(dv_1) \in \mathcal{P}(\mathbb{E}_1)$ for $\overline{L}_1$.

We further need to assume that there are functions $h_1 \in \mathcal{D}(L_1) : \mathbb{E}_0 \times \mathbb{E}_1 \mapsto \mathbb{R}^{|\mathbb{E}_0|}$ and $h_2, h_3 \in \mathcal{D}(L_2) : \mathbb{E} \mapsto \mathbb{R}^{|\mathbb{E}_0|}$ that solve the following Poisson equations:

$$\overline{L}_1 h_1 = \overline{F}_1 - \overline{F}, \qquad L_2 h_2 = F - \overline{F}_1, \qquad L_2 h_3 = \overline{L}_1 h_1 - L_1 h_1,$$

where

$$\overline{F}_1(v_0, v_1) = \int F(v_0, v_1, u_2)\mu_{v_0,v_1}(du_2), \qquad \overline{F}(v_0) = \int \overline{F}_1(v_0, u_1)\mu_{v_0}(du_1)$$

in order for Condition 2.6 to be met. We refer the reader to Glynn and Meyn (1996) for results on sufficient conditions for the existence of solutions to a Poisson equation for a general class of Markov processes. For the class of general piecewise deterministic processes see also Costa and Dufour (2003). For the examples considered in Section 5, we were able to explicitly compute the desired functions. In general, however, explicit computation may not be possible, so results that ensure the existence of these functions may be useful.

We now need to identify r_N, which will be of the form $r_N = N^p$, for some $0 < p < m_1$. Assuming that there is no cancellation among the terms in the sum in (4.3), for (2.4) to hold, we must have

$$p \le \max\{\alpha_{\theta_l} - \rho_k : \theta_l \cdot \zeta_k \ne 0, \rho_k < \alpha_{\theta_l}, 1 \le l \le s_0\}. \tag{4.5}$$

Then

$$\theta_l \cdot G_0(v) = \lim_{N\to\infty} r_N \theta_l \cdot \big(F^N(v) - F(v)\big) = \sum_{k\,:\,\alpha_{\theta_l} - \rho_k = p} \lambda_k\big(T^{-1}v\big)\theta_l \cdot \zeta_k$$

and

$$G_0(v) = \sum_{l=1}^{s_0} \sum_{k:\alpha_{\theta_l} - \rho_k = p} \lambda_k\big(T^{-1}v\big)\theta_l \cdot \zeta_k \theta_l.$$

Now let $H^N = r^{-1}_{1,N} h_1 + r^{-1}_{2,N}(h_2 + h_3)$. To ensure that the limit in (2.7) exists, with reference to the definition of L_2, we must have

$$p \le \min\{\alpha_i + m_2 - \rho_k : \zeta_{ik} \ne 0, \alpha_i + m_2 - \rho_k > 0\} \tag{4.6}$$

and

$$p \le \min\{2\alpha_i + m_2 - \rho_k : \zeta_{ik} \neq 0, \alpha_i > 0\}, \tag{4.7}$$

and with reference to the definition of L_1, we must have

$$p \le \min\{\alpha_{\theta_l} + m_1 - \rho_k : \theta_l \cdot \zeta_k \neq 0, \alpha_{\theta_l} + m_1 - \rho_k > 0\} \tag{4.8}$$

and

$$p \le \min\{2\alpha_{\theta_l} + m_1 - \rho_k : \theta_l \cdot \zeta_k \neq 0, \alpha_{\theta_l} > 0\}. \tag{4.9}$$

Note that (4.5) implies the minimum in (4.8) and (4.9) only needs to be taken over $s_0 + 1 \le l \le s_1'$.

Assuming that h_1, h_2 and h_3 are sufficiently smooth, these assumptions insure that there exists $G_1 : \mathbb{E} \mapsto \mathbb{R}^{|\mathbb{E}_0|}$

$$\begin{aligned} G_1(v) &= \lim_{N\to\infty} \left(r_N\left(\frac{A^N}{r_{2,N}} - L_2\right)(h_2 + h_3) + r_N\left(\frac{A^N}{r_{1,N}} - L_1\right)h_1 \right) \\ &= G_{12}(v) + G_{11}(v). \end{aligned}$$

To identify G_{12}, define

$$\begin{aligned} \widetilde{\zeta}_{2,k} &= \lim_{N\to\infty} N^p \big(N^{\rho_k - m_2} \Lambda_N \zeta_k - \zeta_{2,k}\big), \\ \widetilde{\xi}_{2,kij} &= \lim_{N\to\infty} N^{p+\rho_k - m_2 - \alpha_i - \alpha_j} \zeta_{ik}\zeta_{kj} \end{aligned}$$

and

$$\begin{aligned} \mathcal{K}_{2,\circ}^p &= \{k \in \mathcal{K} : \theta_l \cdot \zeta_k \neq 0 \text{ for some } l \text{ with } \alpha_{\theta_l} = 0, m_2 - \rho_k = p\}, \\ \mathcal{K}_{2,\bullet}^p &= \{k \in \mathcal{K} \setminus \mathcal{K}_{2,\bullet} : \theta_l \cdot \zeta_k \neq 0 \text{ for some } l \text{ with } \alpha_{\theta_l} > 0, m_2 - \rho_k + \alpha_{\theta_l} = p\}. \end{aligned}$$

Then setting $h(z) = h_2(Tz) + h_3(Tz)$, $G_{12}(v) = H_{12}(T^{-1}v)$, where

$$\begin{aligned} H_{12}(z) = &\sum_{k\in\mathcal{K}_{2,\circ}} \lambda_k(z) \nabla h(z + \zeta_{2,k}) \cdot \widetilde{\zeta}_{2,k} + \sum_{k\in\mathcal{K}_{2,\bullet}\cup\mathcal{K}_{2,\bullet}^p} \lambda_k(z) \nabla h(z) \cdot \widetilde{\zeta}_{2,k} \\ &+ \sum_{k\in\mathcal{K}_{2,\bullet}} \lambda_k(z) \frac{1}{2} \sum_{ij} \partial_{z_i}\, \partial_{z_j} h(z) \widetilde{\xi}_{2,kij} + \sum_{k\in\mathcal{K}_{2,\circ}^p} \lambda_k(z) \big(h(z + \widetilde{\zeta}_{2,k}) - h(z)\big). \end{aligned}$$

Similarly, to identify G_{11}, define

$$\begin{aligned} \widetilde{\zeta}_{1,k}^{\theta} &= \lim_{N\to\infty} N^p \big(N^{\rho_k - m_1} \Lambda_N^{\Theta_1} \Theta_1 \zeta_k - \zeta_{1,k}^{\theta}\big), \\ \widetilde{\xi}_{1,kll'}^{\theta} &= \lim_{N\to\infty} N^{p+\rho_k - m_1 - \alpha_{\theta_l} - \alpha_{\theta_{l'}}} \zeta_{1,kl}^{\theta} \zeta_{1,kl'}^{\theta} \end{aligned}$$

and

$$\mathcal{K}_{1,\circ}^{p} = \{k \in \mathcal{K} : \theta_l \cdot \zeta_k \neq 0 \text{ for some } l \text{ with } \alpha_{\theta_l} = 0, m_1 - \rho_k = p\},$$

$$\mathcal{K}_{1,\bullet}^{p} = \{k \in \mathcal{K} \setminus \mathcal{K}_{\bullet}^{1} : \theta_l \cdot \zeta_k \neq 0 \text{ for some } l \text{ with } \alpha_{\theta_l} > 0, m_1 - \rho_k + \alpha_{\theta_l} = p\}$$

Then $G_{11}(v) = H_{11}(T^{-1}v)$, where

$$\begin{aligned} H_{11}(z) = {} & \sum_{k \in \mathcal{K}_{1,\circ}} \lambda_k(z) \nabla h_1\big(\Theta_1 z + \zeta_{1,k}^{\theta}\big) \cdot \widetilde{\zeta}_{1,k}^{\theta} \\ & + \sum_{k \in \mathcal{K}_{1,\bullet} \cup \mathcal{K}_{1,\bullet}^{p}} \lambda_k(z) \nabla h_1(\Theta_1 z) \cdot \widetilde{\zeta}_{1,k}^{\theta} \\ & + \sum_{k \in \mathcal{K}_{1,\bullet}} \lambda_k(z) \frac{1}{2} \sum_{ij} \partial_l \, \partial_{l'} h_1(\Theta_1 z) \widetilde{\xi}_{1,kll'}^{\theta} \\ & + \sum_{k \in \mathcal{K}_{1,\circ}^{p}} \lambda_k(z) \big(h_1\big(\Theta_1 z + \widetilde{\zeta}_{1,k}^{\theta}\big) - h_1(\Theta_1 z)\big). \end{aligned}$$

We now need to identify $G : \mathbb{E} \to \mathbb{M}^{|\mathbb{E}_0| \times |\mathbb{E}_0|}$ satisfying (2.11) in Condition 2.8. Let

$$R_k^N(t) = Y_k\left(N^{\rho_k} \int_0^t \lambda_k\big(Z^N(s)\big)\, ds\right)$$

and $\widetilde{H}_N(V^N) = \Theta_0 Z^N - H_N(V^N) = V_0^N - H_N(V^N)$. Then denoting $z^{\otimes 2} = zz^T$,

$$\begin{aligned} & N^{2p}\big[\widetilde{H}_N(V^N)\big]_t \\ & \quad = \sum_k N^{2p} \int_0^t \big(\widetilde{H}_N\big(V^N(s-) + T\Lambda_N \zeta_k\big) - \widetilde{H}_N\big(V^N(s-)\big)\big)^{\otimes 2}\, dR_k^N(s), \end{aligned}$$

which is asymptotic to

$$\sum_k N^{2p+\rho_k} \int_0^t \big(\widetilde{H}_N\big(V^N(s-) + T\Lambda_N \zeta_k\big) - \widetilde{H}_N\big(V^N(s-)\big)\big)^{\otimes 2} \lambda_k\big(Z^N(s)\big)\, ds.$$

Taking the limit as $N \to \infty$ and integrating with respect to $\mu_{v_0,v_1}(dv_2)$ and $\mu_{v_0}(dv_1)$ then gives the value of $\overline{G}$.

5. Examples. We now apply the central limit theorem to several examples of chemical reaction networks with multiple scales.

5.1. *Three species viral model.* Ball et al. (2006) considered asymptotics for a model of an intracellular viral infection originally given in Srivastava et al. (2002) and studied further in Haseltine and Rawlings (2002). The model includes three time-varying species, the viral template, the viral genome and the viral structural

protein, involved in six reactions,

$$\begin{array}{lccc}
(1) & T + \text{stuff} & \xrightarrow{\kappa_1} & T + G, \\
(2) & G & \xrightarrow{\kappa_2} & T, \\
(3) & T + \text{stuff} & \xrightarrow{\kappa_3} & T + S, \\
(4) & T & \xrightarrow{\kappa_4} & \varnothing, \\
(5) & S & \xrightarrow{\kappa_5} & \varnothing, \\
(6) & G + S & \xrightarrow{\kappa_6} & V,
\end{array}$$

whose reaction rates (*propensities*) are of mass-action kinetics form $\lambda_k(x) = \kappa'_k \prod_i x_i^{\nu_{ki}}$ with constants

$$\begin{array}{ccc}
\kappa'_1 & 1 & 1, \\
\kappa'_2 & 0.025 & 2.5N_0^{-2/3}, \\
\kappa'_3 & 1000 & N_0, \\
\kappa'_4 & 0.25 & 0.25, \\
\kappa'_5 & 2 & 2, \\
\kappa'_6 & 7.5 \times 10^{-6} & 0.75N_0^{-5/3},
\end{array}$$

here expressed in terms of $N_0 = 1000$.

We denote T, G, S as species 1, 2 and 3, respectively, and let $X_i(t)$ denote the number of molecules of species i in the system at time t. The stochastic model is

$$\begin{aligned}
X_1(t) &= X_1(0) + Y_2\biggl(\int_0^t 0.025X_2(s)\,ds\biggr) \\
&\quad - Y_4\biggl(\int_0^t 0.25X_1(s)\,ds\biggr), \\
X_2(t) &= X_2(0) + Y_1\biggl(\int_0^t X_1(s)\,ds\biggr) \\
&\quad - Y_2\biggl(\int_0^t 0.025X_2(s)\,ds\biggr) \\
&\quad - Y_6\biggl(\int_0^t 7.5 \cdot 10^{-6}X_2(s)X_3(s)\,ds\biggr), \\
X_3(t) &= X_3(0) + Y_3\biggl(\int_0^t 1000X_1(s)\,ds\biggr) \\
&\quad - Y_5\biggl(\int_0^t 2X_3(s)\,ds\biggr) \\
&\quad - Y_6\biggl(\int_0^t 7.5 \cdot 10^{-6}X_2(s)X_3(s)\,ds\biggr).
\end{aligned}$$

We take

$$\alpha_1 = 0, \qquad \alpha_2 = 2/3, \qquad \alpha_3 = 1.$$

The scaling of the rate constants gives

k	κ_k	β_k	ρ_k
1	1	0	0
2	2.5	$-2/3$	0
3	1	1	1
4	0.25	0	0
5	2	0	1
6	0.75	$-5/3$	0.

Changing time $t \to N^{2/3}t$, the normalized system becomes

$$\begin{aligned}
Z_1^N(t) &= Z_1^N(0) + Y_2\Big(\int_0^t N^{2/3}2.5Z_2^N(s)\,ds\Big) - Y_4\Big(\int_0^t N^{2/3}0.25Z_1^N(s)\,ds\Big),\\
Z_2^N(t) &= Z_2^N(0) + N^{-2/3}Y_1\Big(\int_0^t N^{2/3}Z_1^N(s)\,ds\Big)\\
&\quad - N^{-2/3}Y_2\Big(\int_0^t N^{2/3}2.5Z_2^N(s)\,ds\Big)\\
&\quad - N^{-2/3}Y_6\Big(\int_0^t N^{2/3}0.75Z_2^N(s)Z_3^N(s)\,ds\Big),\\
Z_3^N(t) &= Z_3^N(0) + N^{-1}Y_3\Big(\int_0^t N^{5/3}Z_1^N(s)\,ds\Big) - N^{-1}Y_5\Big(\int_0^t N^{5/3}2Z_3^N(s)\,ds\Big)\\
&\quad - N^{-1}Y_6\Big(\int_0^t N^{2/3}0.75Z_2^N(s)Z_3^N(s)\,ds\Big).
\end{aligned}$$

We assume that the initial value for Z_2^N is chosen to satisfy $Z_2(0) = \lim_{N\to\infty} Z_2^N(0) \in (0,\infty)$.

In this model, there are only two time-scales, so we set

$$m_1 = \max\{\rho_k - \alpha_i : \zeta_{ik} \neq 0\} = \max\{\tfrac{2}{3} - 0, \tfrac{2}{3} - \tfrac{2}{3}, \tfrac{5}{3} - 1, \tfrac{2}{3} - 1\} = \tfrac{2}{3},$$

and we have $r_{1,N} = N^{2/3}$. We have $\zeta_{1,1} = 0, \zeta_{1,2} = e_1, \zeta_{1,3} = e_3, \zeta_{1,4} = -e_1$, $\zeta_{1,5} = -e_3, \zeta_{1,6} = 0$. The operator $L_1 = \lim_{N\to\infty} N^{-2/3}A_N$ is given by

$$\begin{aligned}
L_1h(z) &= \lambda_2(z)\big(h(z+e_1) - h(z)\big) + \lambda_4(z)\big(h(z-e_1) - h(z)\big)\\
&\quad + \big(\lambda_3(z) - \lambda_5(z)\big)\partial_{z_3}h(z)
\end{aligned}$$

and note that for smooth h,

$$N^{-2/3}A_Nh = L_1h + O\big(N^{-2/3}\big). \tag{5.1}$$

Functions $h \in \ker(L_1)$ are functions of the coordinate z_2 only, $\mathbb{E}_1 = \mathcal{R}(S_1) = \operatorname{span}\{e_1, e_3\}$ and $\mathbb{E}_0 = \mathcal{N}(S_1^T) = \operatorname{span}\{e_2\}$. Taking $h \in \mathcal{D}(L_0) = C^1(\mathbb{E}_0)$,

$$L_0 h(z) = \lim_{N\to\infty} A_N h(z) = \big(\lambda_1(z) - \lambda_2(z) - \lambda_6(z)\big)\partial_{z_2} h(z_2).$$

Setting $V_0^N = Z_2^N$ and $V_1^N = (Z_1^N, Z_3^N)$, the compensator for V_0^N is

$$F^N(z) = \lambda_1(z) - \lambda_2(z) - \lambda_6(z),$$

so $F(z) = F^N(z)$ and $G_0(z) \equiv 0$ in Condition 2.5.

The process corresponding to L_1 is piecewise deterministic with Z_1 discrete and Z_3 continuous. For fixed z_2, with reference to Condition 2.4, the conditional equilibrium distribution satisfies

$$\begin{aligned}\int \Big[& 2.5 z_2 \big(g(z_1+1, z_3) - g(z_1, z_3)\big) \\ & + 0.25 z_1 \big(g(z_1 - 1, z_3) - g(z_1, z_3)\big) \\ & + (z_1 - 2z_3)\frac{\partial g}{\partial z_3}(z_1, z_3)\Big] \mu_{z_2}(dz_1, dz_3) = 0.\end{aligned} \tag{5.2}$$

Note that the marginal for Z_1 is Poisson$(10z_2)$, so

$$\int z_1 \mu_{z_2}(dz_1, dz_3) = 10 z_2.$$

Taking $g(z_1, z_3) = z_3$ in (5.2), we see

$$\int z_3 \mu_{z_2}(dz_1, dz_3) = 5 z_2.$$

These calculations imply that the averaged value for the drift F is

$$\overline{F}(z_2) = \int \big(\lambda_1(z) - \lambda_2(z) - \lambda_6(z)\big)\mu_{z_2}(dz_1, dz_3) = 7.5 z_2 - 3.75 z_2^2,$$

with $\nabla \overline{F}(z_2) = 7.5 - 7.5 z_2$. For the current example, we will see that $\overline{F}$ and $\overline{G}$ in (2.15) can be obtained without explicitly computing with μ_{z_2}.

With reference to (2.5), we look for a solution h_1 to the Poisson equation

$$L_1 h_1(z) = (z_1 - 2.5 z_2 - 0.75 z_2 z_3) - \big(7.5 z_2 - 3.75 z_2^2\big). \tag{5.3}$$

Trying h_1 of the form $h_1(z) = z_1 u_1(z_2) + z_3 u_3(z_2)$, we have

$$L_1 h_1(z) = u_1(z_2)(2.5 z_2 - 0.25 z_1) + u_3(z_2)(z_1 - 2 z_3)$$

and equating the factors multiplying z_1 and z_3, we get $u_1(z_2) = 1.5 z_2 - 4$ and $u_3(z_2) = 0.375 z_2$. Thus $h_1(z) = z_1(1.5 z_2 - 4) + z_3(0.375 z_2)$ and $H^N(z) = N^{-2/3} h_1(z)$.

Since the solution of (5.3) is exact and (as we shall see) $r_N = N^{1/3}$, by (5.1), we have $G_1 = 0$ in Condition 2.6. With reference to Condition 2.8, (2.9) and (2.10) are immediate.

The only restriction that remains to determine r_N is the asymptotic behavior of the quadratic variation of $Z_2^N - H^N(Z^N) = Z_2^N - N^{-2/3}h_1(Z^N)$. Direct calculation shows that to get a nontrivial G in (2.11) we must take $r_N = N^{1/3}$. We then have

$$\begin{aligned}
&N^{2/3}\bigl[Z_2^N - H^N(Z^N)\bigr]_t \\
&\quad= \sum_{k=1}^{6} N^{-2/3} \int_0^t \bigl(\zeta_{2k} + h_1\bigl(Z^N(s-)\bigr) - h_1\bigl(Z^N(s-) + \Lambda_N \zeta_k\bigr)\bigr)^2 \, dR_k^N(s) \\
&\quad\approx \int_0^t Z_1^N(s)\,ds + \int_0^t \bigl(-1 - 1.5Z_2^N(s) + 4\bigr)^2 2.5 Z_2^N(s)\,ds \\
&\qquad+ \int_0^t \bigl(1.5Z_2^N(s) - 4\bigr)^2 0.25 Z_1^N(s)\,ds + \int_0^t 0.75 Z_2^N(s) Z_3^N(s)\,ds,
\end{aligned}$$

where we observe that jumps by R_3^N and R_5^N do not contribute to the limit. Dividing the equation for Z_1^N by $N^{2/3}$, we observe that

$$\int_0^t Z_1^N(s)\,ds \approx \int_0^t 10 Z_2^N(s)\,ds.$$

Similarly, dividing the equation for Z_3^N by $N^{2/3}$ we see that

$$\int_0^t Z_3^N(s)\,ds \approx \frac{1}{2}\int_0^t Z_1^N(s)\,ds \approx \int_0^t 5 Z_2^N(s)\,ds,$$

which in turn implies

$$\int_0^t Z_2^N(s) Z_3^N(s)\,ds \approx \int_0^t 5 Z_2^N(s)^2\,ds.$$

It follows that $\overline{G}(z_2)$ is

$$\begin{aligned}
&10z_2 + (3 - 1.5z_2)^2 2.5 z_2 + (4 - 1.5z_2)^2 2.5 z_2 + 3.75 z_2^2 \\
&\quad= 72.5 z_2 - 48.75 z_2^2 + 11.25 z_2^3.
\end{aligned}$$

Let Z_2 be the solution of

$$Z_2(t) = Z_2(0) + \int_0^t \bigl(7.5 Z_2(s) - 3.75 Z_2^2(s)\bigr)\,ds$$

and $U^N = N^{1/3}(Z_2^N - Z_2)$. Then

$$\sup_{s \le t} \bigl|Z_2^N(s) - Z_2(s)\bigr| \Rightarrow 0 \quad \text{and} \quad U^N \Rightarrow U,$$

where, for W a standard Brownian motion, U satisfies

$$U(t) = U(0) + \int_0^t \sqrt{72.5Z_2(s) - 48.75Z_2(s)^2 + 11.25Z_2(s)^3}\, dW(s) + \int_0^t \big(7.5 - 7.5Z_2(s)\big)U(s)\, ds.$$

The corresponding diffusion approximation is

$$D^N(t) = Z_2^N(0) + N^{-1/3} \int_0^t \sqrt{72.5D^N(s) - 48.75D^N(s)^2 + 11.25D^N(s)^3}\, dW(s) + \int_0^t \big(7.5D^N(s) - 3.75D^N(s)^2\big)\, ds.$$

We compare simulations for the original value of the amount of genome $X_2(\cdot)$ with the approximations given by the Gaussian approximation $N^{2/3}Z_2(\cdot N^{-2/3}) + N^{1/3}U(\cdot N^{-2/3})$, and the diffusion approximation $N^{2/3}D^N(\cdot N^{-2/3})$. For comparison we also give the deterministic value given by $N^{2/3}Z_2(\cdot N^{-2/3})$. We use $N = 1000$ and a time interval on the scale $\gamma = 2/3$. The initial values are set to $X_1(0) = X_3(0) = 0$, $X_2(0) = 10$ and 500 realizations are performed for each of the three stochastic processes. Figure 1 shows the mean and one standard deviation above and below the mean for each of the three processes, and Figure 2 shows five trajectories for the three processes.

For the diffusion process, these plots use only sample paths that hit one $(= 100/N_0^{2/3})$ before they hit zero. For small initial values, the diffusion approxi-

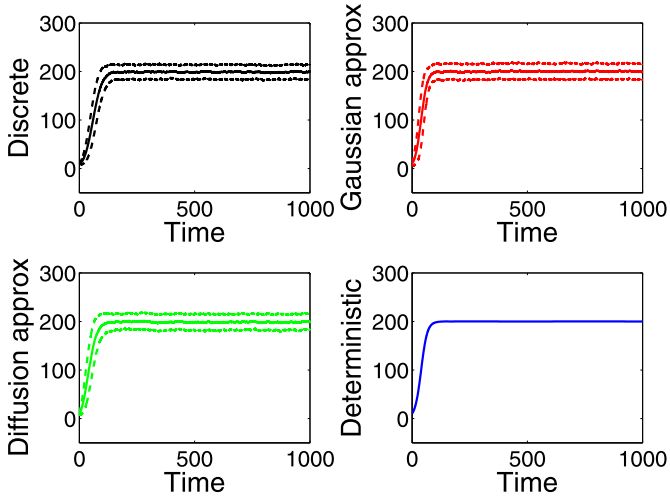

FIG. 1. *Mean and standard deviation of the amount of genome in the three species model* [500 *simulations with parameters* $N_0 = 1000$, $\gamma = 2/3$, $X_1(0) = 0$, $X_2(0) = 10$, $X_3(0) = 0$].

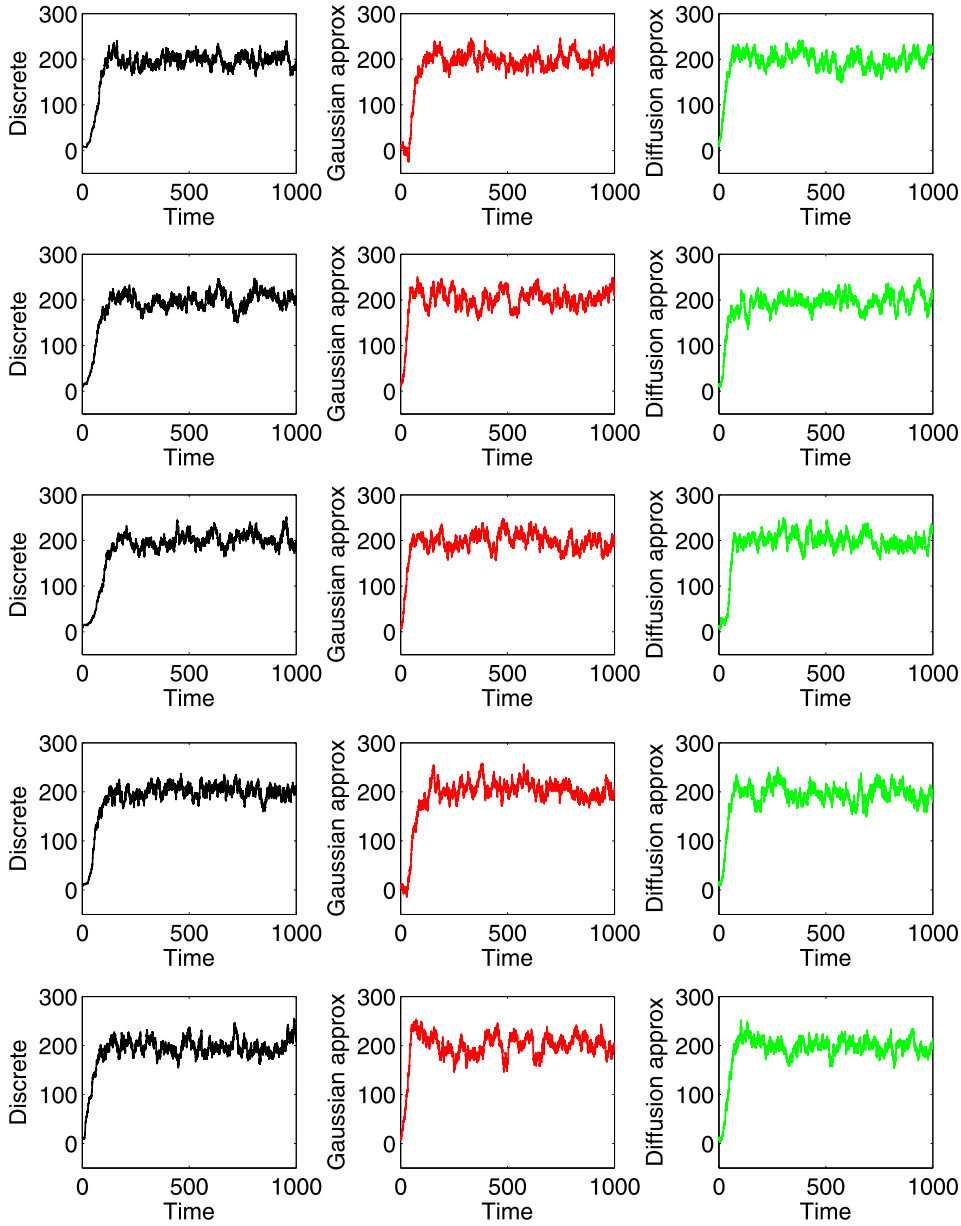

FIG. 2. *Five trajectories of the amount of genome in the three species model* (*same parameters as in Figure* 1).

mation does not give a good approximation of the probability of hitting zero (and hence absorbing at zero), before (e.g.) hitting one. Let

$$\tau_Z^N = \inf\{t > 0 : Z_2^N(t) = 0 \text{ or } Z_2^N(t) \geq 1\}$$

and

$$\tau_D^N = \inf\{t > 0 : D^N(t) = 0 \text{ or } D^N(t) \geq 1\}.$$

It is shown in Ball et al. (2006) that

$$\lim_{N\to\infty} P\{Z^N(\tau_Z^N) = 0 | Z^N(0) = N^{-2/3}k\} = 4^{-k}$$

while a standard calculation for the diffusion process gives

$$\lim_{N\to\infty} P\{D^N(\tau_D^N) = 0 | D^N(0) = N^{-2/3}k\} = e^{-(6/29)k}.$$

5.2. *Michaelis–Menten enzyme model.* A basic model for an enzymatic reaction includes three time-varying species, the substrate, the free enzyme and the substrate-bound enzyme, involved in three reactions:

$$\begin{aligned}
&(1) \quad S + E \xrightarrow{\kappa_1'} SE,\\
&(2) \quad SE \xrightarrow{\kappa_2'} S + E,\\
&(3) \quad SE \xrightarrow{\kappa_3'} P + E,
\end{aligned}$$

with mass-action kinetics and with rate constants such that $\kappa_2', \kappa_3' \gg \kappa_1'$. To be precise, let $\kappa_2' = \kappa_2 N$, $\kappa_3' = \kappa_3 N$, and $\kappa_1' = \kappa_1$.

We denote E, S, P as species 1, 2 and 3, respectively, and let $X_i(t)$ be the number of molecules of species i in the system at time t. Note that the total number of unbound and substrate-bound enzyme molecules is conserved, and we let M denote this amount. The stochastic model is

$$\begin{aligned}
X_1(t) &= X_1(0) - Y_1\left(\int_0^t \kappa_1' X_1(s)X_2(s)\,ds\right) + Y_2\left(\int_0^t \kappa_2'(M - X_1(s))\,ds\right)\\
&\quad + Y_3\left(\int_0^t \kappa_3'(M - X_1(s))\,ds\right),\\
X_2(t) &= X_2(0) - Y_1\left(\int_0^t \kappa_1' X_1(s)X_2(s)\,ds\right) + Y_2\left(\int_0^t \kappa_2'(M - X_1(s))\,ds\right),\\
X_3(t) &= X_3(0) + Y_3\left(\int_0^t \kappa_3'(M - X_1(s))\,ds\right).
\end{aligned}$$

If the initial amount of substrate is $O(N) \gg M$, then the normalizations of the species abundances are given by

$$\alpha_1 = 0, \qquad \alpha_2 = 1, \qquad \alpha_3 = 1,$$

and the scaling exponents for the rate constants are

$$\beta_1 = 0, \qquad \beta_2 = 1, \qquad \beta_3 = 1.$$

The normalized system becomes

$$Z_1^N(t) = Z_1^N(0) - Y_1\Big(\int_0^t N\kappa_1 Z_1^N(s)Z_2^N(s)\,ds\Big) + Y_2\Big(\int_0^t N\kappa_2\big(M - Z_1^N(s)\big)\,ds\Big) + Y_3\Big(\int_0^t N\kappa_3\big(M - Z_1^N(s)\big)\,ds\Big),$$

$$Z_2^N(t) = Z_2^N(0) - N^{-1}Y_1\Big(\int_0^t N\kappa_1 Z_1^N(s)Z_2^N(s)\,ds\Big) + N^{-1}Y_2\Big(\int_0^t N\kappa_2\big(M - Z_1^N(s)\big)\,ds\Big),$$

$$Z_3^N(t) = Z_3^N(0) + N^{-1}Y_3\Big(\int_0^t N\kappa_3\big(M - Z_1^N(s)\big)\,ds\Big).$$

Again, there are only two time-scales with the fast time-scale $m_1 = 1$ giving $r_{1,N} = N$. Then $\zeta_{1,1} = -e_1$, $\zeta_{1,2} = \zeta_{1,3} = e_1$, and the operator L_1 is given by

$$L_1h(z) = \kappa_1 z_1 z_2\big(h(z - e_1) - h(z)\big) + (\kappa_2 + \kappa_3)(M - z_1)\big(h(z + e_1) - h(z)\big),$$

and for smooth h,

$$N^{-1}A_N h = L_1 h + O\big(N^{-1}\big). \tag{5.4}$$

Functions $h \in \ker(L_1)$ are functions of coordinates z_2 and z_3 only. Thus $\mathbb{E}_1 = \{z_1e_1 : z_1 = 0, \ldots, M\} \subset \mathcal{R}(S_1)$ and $\mathbb{E}_0 = \mathcal{N}(S_1^T) = \{(z_2e_2, z_3e_3) : z_2, z_3 \geq 0\}$. For $h \in \mathcal{D}(L_0) = C^1(\mathbb{E}_0)$,

$$L_0h(z) = \big(\kappa_2(M - z_1) - \kappa_1 z_1 z_2\big)\partial_{z_2}h(z) + \kappa_3(M - z_1)\partial_{z_3}h(z).$$

Taking $V_0^N = (Z_2^N, Z_3^N)$, the compensator for V_0^N in (2.3) is

$$F^N(z) = \big(\kappa_2(M - z_1) - \kappa_1 z_1 z_2, \kappa_3(M - z_1)\big)^T,$$

so $F(z) = F^N(z)$ and $G_0(z) \equiv 0$.

On the fast time-scale, the process whose generator is L_1 is a Markov chain on $\mathbb{E}_1$ describing the dynamics of an urn scheme with a total of M molecules, and for a fixed value of z_2, z_3, with transition rates $\kappa_1 z_2$ for outflow and $\kappa_2 + \kappa_3$ for inflow. Its stationary distribution $\mu_{z_2,z_3}(z_1)$ is binomial$(M, p(z_2))$ for

$$p(z_2) = \frac{\kappa_2 + \kappa_3}{\kappa_2 + \kappa_3 + \kappa_1 z_2},$$

so $\int z_1 \mu_{z_2,z_3}(dz_1) = Mp(z_2)$.

This observation implies that the averaged value for the drift F is

$$\overline{F}(z_2, z_3) = \left(-M\frac{\kappa_1\kappa_3 z_2}{\kappa_2 + \kappa_3 + \kappa_1 z_2}, M\frac{\kappa_1\kappa_3 z_2}{\kappa_2 + \kappa_3 + \kappa_1 z_2}\right)^T = -\kappa_3 M\big(1 - p(z_2)\big)\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

with

$$\nabla\overline{F} = -M\frac{\kappa_1\kappa_3(\kappa_2+\kappa_3)}{(\kappa_2+\kappa_3+\kappa_1 z_2)^2}\begin{pmatrix}1 & 0\\ -1 & 0\end{pmatrix},$$

and we need to solve the Poisson equation

$$L_1 h_1(z) = \Bigg(\kappa_2(M-z_1) - \kappa_1 z_1 z_2 + M\frac{\kappa_1\kappa_3 z_2}{\kappa_2+\kappa_3+\kappa_1 z_2},\\ \kappa_3(M-z_1) - M\frac{\kappa_1\kappa_3 z_2}{\kappa_2+\kappa_3+\kappa_1 z_2}\Bigg)^T$$

for h_1. Trying h_1 of the form $h_1(z) = (z_1 u_1(z_2), z_1 u_2(z_2))^T$, we have

$$L_1 h_1(z) = \big(-\kappa_1 z_1 z_2 u_1(z_2) + (\kappa_2+\kappa_3)(M-z_1)u_1(z_2), -\kappa_1 z_1 z_2 u_2(z_2)\\ + (\kappa_2+\kappa_3)(M-z_1)u_2(z_2)\big)^T,$$

and equating terms with the same power of z_1, we get $u_1(z_2) = (\kappa_1 z_2 + \kappa_2)/(\kappa_1 z_2 + \kappa_2 + \kappa_3)$ and $u_2(z_2) = \kappa_3/(\kappa_1 z_2 + \kappa_2 + \kappa_3)$. Note that $u_1(z_2) + u_2(z_2) = 1$. Thus

$$\begin{aligned} h_1(z) &= \left(\frac{z_1(\kappa_1 z_2+\kappa_2)}{(\kappa_1 z_2+\kappa_2+\kappa_3)}, \frac{z_1\kappa_3}{(\kappa_1 z_2+\kappa_2+\kappa_3)}\right)^T\\ &= z_1\big(u_1(z_2), 1-u_1(z_2)\big)^T, \end{aligned}$$

and $H^N(z) = N^{-1}h_1(z)$.

Examining the quadratic variation of $V_0^N - H^N \circ V^N$, we see that r_N must be $N^{1/2}$, and by (5.4), it follows that $G_1 = 0$ in (2.7).

Finally, letting $z^{\otimes 2} = zz^T$,

$$\begin{aligned} &N\big[V_0^N - H^N\circ V^N\big]_t\\ &\quad= N^{-1}\sum_{k=1}^3\int_0^t \big(\Theta_0\zeta_k + h_1\big(Z^N(s-)\big) - h_1\big(Z^N(s-)+\Lambda_N\zeta_k\big)\big)^{\otimes 2}\,dR_k^N(s)\\ &\quad\approx \int_0^t\left(-\begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}u_1\big(Z_2^N(s)\big)\\ 1-u_1\big(Z_2^N(s)\big)\end{pmatrix}\right)^{\otimes 2}\kappa_1 Z_1^N(s)Z_2^N(s)\,ds\\ &\qquad+\int_0^t\left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}u_1\big(Z_2^N(s)\big)\\ 1-u_1\big(Z_2^N(s)\big)\end{pmatrix}\right)^{\otimes 2}\kappa_2\big(M-Z_1^N(s)\big)\,ds\\ &\qquad+\int_0^t\left(\begin{pmatrix}0\\1\end{pmatrix} - \begin{pmatrix}u_1\big(Z_2^N(s)\big)\\ 1-u_1\big(Z_2^N(s)\big)\end{pmatrix}\right)^{\otimes 2}\kappa_3\big(M-Z_1^N(s)\big)\,ds\\ &\quad\approx \int_0^t\begin{pmatrix}\big(1-u_1\big(Z_2^N(s)\big)\big)^2 & -\big(1-u_1\big(Z_2^N(s)\big)\big)^2\\ -\big(1-u_1\big(Z_2^N(s)\big)\big)^2 & \big(1-u_1\big(Z_2^N(s)\big)\big)^2\end{pmatrix}\kappa_1 Z_1^N(s)Z_2^N(s)\,ds \end{aligned}$$

$$+\int_0^t \begin{pmatrix} (1-u_1(Z_2^N(s)))^2 & -(1-u_1(Z_2^N(s)))^2 \\ -(1-u_1(Z_2^N(s)))^2 & (1-u_1(Z_2^N(s)))^2 \end{pmatrix} \kappa_2 (M - Z_1^N(s))\,ds$$
$$+\int_0^t \begin{pmatrix} u_1(Z_2^N(s))^2 & -u_1(Z_2^N(s))^2 \\ -u_1(Z_2^N(s))^2 & u_1(Z_2^N(s))^2 \end{pmatrix} \kappa_3 (M - Z_1^N(s))\,ds,$$

and averaging Z_1^N gives

$$\lim_{N\to\infty} N[V_0^N - H_N \circ V^N]_t = \int_0^t \overline{G}(Z(s))\,ds$$
$$= \int_0^t \begin{pmatrix} \overline{g}(Z_2(s)) & -\overline{g}(Z_2(s)) \\ -\overline{g}(Z_2(s)) & \overline{g}(Z_2(s)) \end{pmatrix} ds,$$

where $Z = (Z_2, Z_3)$ satisfies

$$Z(t) = Z(0) + \int_0^t M \frac{\kappa_1\kappa_3 Z_2(s)}{\kappa_2 + \kappa_3 + \kappa_1 Z_2(s)} \begin{pmatrix} -1 \\ 1 \end{pmatrix} ds$$

and

$$\overline{g}(z_2) = M(1-u_1(z_2))^2(\kappa_1 p(z_2) z_2 + \kappa_2(1-p(z_2)))$$
$$+ M u_1(z_2)^2 \kappa_3 (1-p(z_2)).$$

Let $U^N = N^{1/2}(Z_2^N - Z_2, Z_3^N - Z_3)^T$. Then

$$\sup_{s\le t} |(Z_2^N(s) - Z_2(s), Z_3^N(s) - Z_3(s))| \Rightarrow 0 \quad \text{and} \quad U^N \Rightarrow U,$$

where $U = (U_2, U_3)^T$ satisfies

$$U(t) = U(0) + \int_0^t \begin{pmatrix} -1 \\ 1 \end{pmatrix} \sqrt{\overline{g}(Z_2(s))}\,dW(s)$$
$$+ \int_0^t \frac{M\kappa_1\kappa_3(\kappa_2+\kappa_3)}{(\kappa_2+\kappa_3+\kappa_1 Z_2(s))^2} U_2(s) \begin{pmatrix} -1 \\ 1 \end{pmatrix} ds$$

for W a standard scalar Brownian motion.

The corresponding diffusion approximation is

$$\begin{pmatrix} D_2^N(t) \\ D_3^N(t) \end{pmatrix} = \begin{pmatrix} Z_2^N(0) \\ Z_3^N(0) \end{pmatrix} + N^{-1/2} \int_0^t \begin{pmatrix} -1 \\ 1 \end{pmatrix} \sqrt{\overline{g}(D_2^N(s))}\,dW(s)$$
$$+ \int_0^t M \frac{\kappa_1\kappa_3 D_2^N(s)}{\kappa_2+\kappa_3+\kappa_1 D_2^N(s)} \begin{pmatrix} -1 \\ 1 \end{pmatrix} ds.$$

We compare simulations for 500 realizations of the original model with 500 realizations of the Gaussian approximation $N_0 Z_2(\cdot) + N_0^{1/2} U_2(\cdot), N_0 Z_3(\cdot) + N_0^{1/2} U_3(\cdot)$ and of the diffusion approximation $N_0 D_2^{N_0}(\cdot), N_0 D_3^{N_0}(\cdot)$. For comparison we also give the deterministic value given by $N_0 Z_2(\cdot), N_0 Z_3(\cdot)$. We use

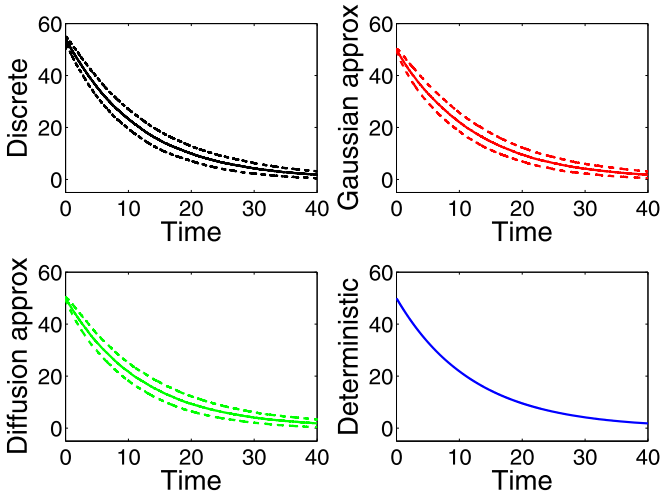

FIG. 3. *Mean and standard deviation of the amount of substrate in the Michaelis–Menten model* [500 *simulations with parameters* $N_0 = 100$, $\gamma = 0$, $X_1(0) = X_3(0) = 0$, $X_2(0) = 50$, $M = 5$, $\kappa_1' = 0.1, \kappa_2' = 500, \kappa_3' = 100$].

$N_0 = 100$ and a time interval on the scale $\gamma = 0$. The initial values are set to $X_1(0) = X_3(0) = 0, X_2(0) = 50$ and $M = 5, \kappa_1' = 0.1$, $\kappa_2' = 500$ and $\kappa_3' = 100$. Figure 3 shows the mean and one standard deviation above and below the mean for each of the three processes, and Figure 4 shows five trajectories for the three processes. In this example, both Gaussian and diffusion approximations give good approximations for the means and the standard deviations of the pair of processes $X_2(\cdot), X_3(\cdot)$.

5.3. *Another enzyme model.* Another model for an enzymatic reaction includes an additional form for the enzyme which cannot bind to the substrate. There are now four species, substrate, active enzyme, enzyme-substrate complex and inactive enzyme, involved in five reactions:

$$\begin{aligned}
&(1) \quad S + E \xrightarrow{\kappa_1'} SE,\\
&(2) \quad SE \xrightarrow{\kappa_2'} S + E,\\
&(3) \quad SE \xrightarrow{\kappa_3'} P + E,\\
&(4) \quad F \xrightarrow{\kappa_4'} E,\\
&(5) \quad E \xrightarrow{\kappa_5'} F,
\end{aligned}$$

with mass-action kinetics and rate constants such that $\kappa_1' = O(1), \kappa_2', \kappa_3' = O(N)$, $\kappa_4', \kappa_5' = O(N^2)$ so that $\kappa_1' = \kappa_1, \kappa_2' = \kappa_2 N, \kappa_3' = \kappa_3 N, \kappa_4' = \kappa_4 N^2, \kappa_5' = \kappa_5 N^2$.

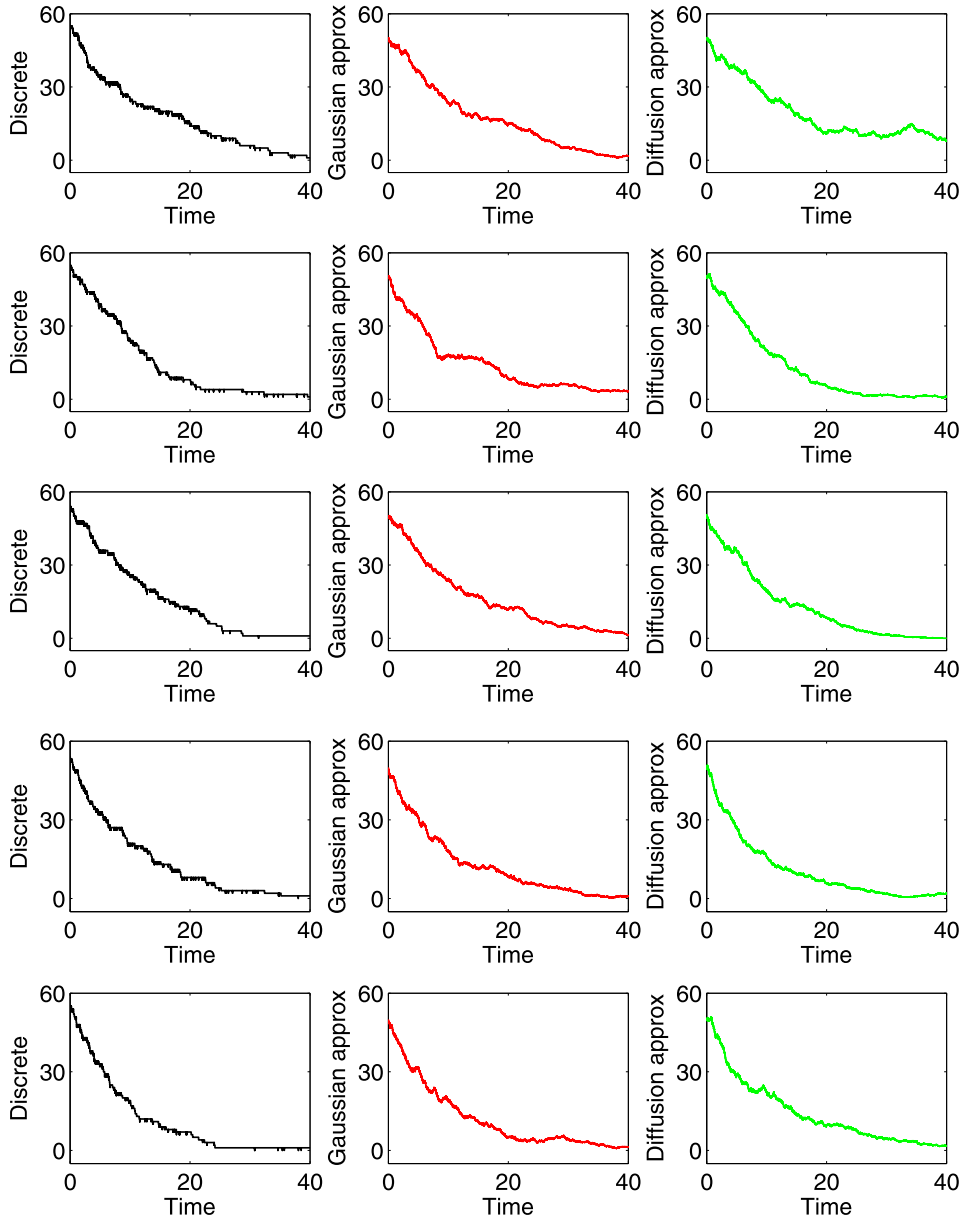

FIG. 4. *Five trajectories of the amount of substrate in the Michaelis–Menten model* (*parameters as in Figure* 3).

We denote E, S, F as species 1, 2 and 3, respectively, and let $X_i(t)$ be the number of molecules of species i in the system at time t. The total number M of active, inactive and substrate-bound enzyme molecules is conserved. The stochas-

tic model is

$$\begin{aligned}
X_1(t) &= X_1(0) - Y_1\Big(\int_0^t \kappa_1' X_1(s)X_2(s)\,ds\Big) \\
&\quad + Y_2\Big(\int_0^t \kappa_2'(M - X_1(s) - X_3(s))\,ds\Big) \\
&\quad + Y_3\Big(\int_0^t \kappa_3'(M - X_1(s) - X_3(s))\,ds\Big) + Y_4\Big(\int_0^t \kappa_4' X_3(s)\,ds\Big) \\
&\quad - Y_5\Big(\int_0^t \kappa_5' X_1(s)\,ds\Big), \\
X_2(t) &= X_2(0) - Y_1\Big(\int_0^t \kappa_1' X_1(s)X_2(s)\,ds\Big) \\
&\quad + Y_2\Big(\int_0^t \kappa_2'(M - X_1(s) - X_3(s))\,ds\Big), \\
X_3(t) &= X_3(0) - Y_4\Big(\int_0^t \kappa_4' X_3(s)\,ds\Big) + Y_5\Big(\int_0^t \kappa_5' X_1(s)\,ds\Big).
\end{aligned}$$

If the initial amount of substrate is $O(N) \gg M$, then the scaling exponents for the species abundances are

$$\alpha_1 = 0, \qquad \alpha_2 = 1, \qquad \alpha_3 = 0,$$

and the scaling exponents for the rate constants are

$$\beta_1 = 0, \qquad \beta_2 = 1, \qquad \beta_3 = 1, \qquad \beta_4 = 2, \qquad \beta_5 = 2.$$

The normalized system becomes

$$\begin{aligned}
Z_1^N(t) &= Z_1^N(0) - Y_1\Big(\int_0^t N\kappa_1 Z_1^N(s)Z_2^N(s)\,ds\Big) \\
&\quad + Y_2\Big(\int_0^t N\kappa_2(M - Z_1^N(s) - Z_3^N(s))\,ds\Big) \\
&\quad + Y_3\Big(\int_0^t N\kappa_3(M - Z_1^N(s) - Z_3^N(s))\,ds\Big) + Y_4\Big(\int_0^t N^2\kappa_4 Z_3^N(s)\,ds\Big) \\
&\quad - Y_5\Big(\int_0^t N^2\kappa_5 Z_1^N(s)\,ds\Big), \\
Z_2^N(t) &= Z_2^N(0) - N^{-1}Y_1\Big(\int_0^t N\kappa_1 Z_1^N(s)Z_2^N(s)\,ds\Big) \\
&\quad + N^{-1}Y_2\Big(\int_0^t N\kappa_2(M - Z_1^N(s) - Z_3^N(s))\,ds\Big), \\
Z_3^N(t) &= Z_3^N(0) - Y_4\Big(\int_0^t N^2\kappa_4 Z_3^N(s)\,ds\Big) + Y_5\Big(\int_0^t N^2\kappa_5 Z_1^N(s)\,ds\Big).
\end{aligned}$$

The fastest time-scale has $m_2 = 2$ and $r_{2,N} = N^2$, with $\zeta_{2,4} = e_1 - e_3$, $\zeta_{2,5} = -e_1 + e_3$. The operator L_2 is

$$L_2 h(z) = \kappa_4 z_3\big(h(z + e_1 - e_3) - h(z)\big) + \kappa_5 z_1\big(h(z - e_1 + e_3) - h(z)\big),$$

with $\ker(L_2)$ consisting of functions of coordinates z_2 and $z_1 + z_3$ only. To simplify our calculations we make a change of variables to $(v_0, v_1, v_2) = (z_2, z_1 + z_3, z_3)$, so in this system of variables $\zeta_{2,4} = \widetilde{e}_2$, $\zeta_{2,5} = -\widetilde{e}_2$ with the operator L_2

$$L_2 h(v) = \kappa_4 v_2\big(h(v - \widetilde{e}_2) - h(z)\big) + \kappa_5 (v_1 - v_2)\big(h(v + \widetilde{e}_2) - h(v)\big).$$

Functions $h(v) \in \ker(L_2)$ are now functions of v_0, v_1 only. Thus $\mathbb{E}_2 = \mathcal{R}(S_2) = \operatorname{span}\{\widetilde{e}_2\}$ and $\mathbb{E}_1 \times \mathbb{E}_0 = \mathcal{N}(S_2^T) = \operatorname{span}\{\widetilde{e}_1, \widetilde{e}_0\}$.

The next time-scale has $m_1 = 1$, $r_{1,N} = N$ and $\zeta_{1,1} = (0, -1)$, $\zeta_{1,2} = \zeta_{1,3} = (0, 1)$. Also

$$\begin{aligned} L_1 h(v) = {} & \kappa_1 v_0 (v_1 - v_2)\big(h\big((v_0, v_1 - 1) - h(v_0, v_1)\big)\big) \\ & + (\kappa_2 + \kappa_3)(M - v_1)\big(h\big((v_0, v_1 + 1) - h(v_0, v_1)\big)\big) \end{aligned}$$

with $\ker(L_1)$ consisting of functions of v_0 only. Thus $\mathbb{E}_1 = \mathcal{R}(S_1) = \operatorname{span}\{\widetilde{e}_1\}$ and $\mathbb{E}_1 = \mathcal{N}(S_1^T) = \operatorname{span}\{\widetilde{e}_0\}$.

Finally, L_0 is

$$L_0 h(v) = -\kappa_1 v_0 (v_1 - v_2)\partial_{v_0} h(v_0) + \kappa_2 (M - v_1)\partial_{v_0} h(v_0).$$

The conditional stationary distribution $\mu_{v_0, v_1}(dv_2)$ of Markov chain with generator L_2 is such that $\rho_0(v_0, v_1) = \int v_2 \mu_{v_0, v_1}(dv_2) = \frac{v_1 \kappa_5}{\kappa_4 + \kappa_5}$, thus

$$\begin{aligned} \overline{L}_1 h(v) = {} & \kappa_1 v_0 \frac{v_1 \kappa_4}{\kappa_4 + \kappa_5}\big(h\big((v_0, v_1 - 1) - h(v_0, v_1)\big)\big) \\ & + (\kappa_2 + \kappa_3)(M - v_1)\big(h\big((v_0, v_1 + 1) - h(v_0, v_1)\big), \end{aligned}$$

which has conditional stationary distribution $\mu_{v_0}(dv_1)$ such that

$$\begin{aligned} \rho_1(v_0) &= \int v_1 \mu_{v_0}(dv_1) = \frac{M(\kappa_4 + \kappa_5)(\kappa_2 + \kappa_3)}{\kappa_1 \kappa_4 v_0 + (\kappa_4 + \kappa_5)(\kappa_2 + \kappa_3)}, \\ \rho_2(v_0) &= \int v_2 \mu_{v_0, v_1}(dv_2)\mu_{v_0}(dv_1) = \frac{M \kappa_5 (\kappa_2 + \kappa_3)}{\kappa_1 \kappa_4 v_0 + (\kappa_4 + \kappa_5)(\kappa_2 + \kappa_3)}. \end{aligned}$$

The compensator for the process V_0^N is $F^N(v) = \kappa_2(M - v_1) - \kappa_1 v_0 (v_1 - v_2) = F(v)$, and averaging F gives $\overline{F}_1(v_0, v_1) = \kappa_2(M - v_1) - \kappa_1 v_0 (v_1 - \rho_0(v_0, v_1))$, and

$$\begin{aligned} \overline{F}(v_0) &= \kappa_2\big(M - \rho_1(v_0)\big) - \kappa_1 v_0\big(\rho_1(v_0) - \rho_2(v_0)\big) \\ &= -\frac{M \kappa_1 \kappa_3 \kappa_4 v_0}{\kappa_1 \kappa_4 v_0 + (\kappa_4 + \kappa_5)(\kappa_2 + \kappa_3)}, \end{aligned}$$

so

$$\nabla\overline{F}(v_0) = -\frac{M\kappa_1\kappa_3\kappa_4(\kappa_4+\kappa_5)(\kappa_2+\kappa_3)}{(\kappa_1\kappa_4 v_0 + (\kappa_4+\kappa_5)(\kappa_2+\kappa_3))^2}.$$

Setting

$$u_1(v_0) = \frac{\kappa_1\kappa_4 v_0 + \kappa_2(\kappa_4+\kappa_5)}{\kappa_1\kappa_4 v_0 + (\kappa_2+\kappa_3)(\kappa_4+\kappa_5)}, \qquad u_2(v_0) = \frac{\kappa_1 v_0}{\kappa_4+\kappa_5},$$

and

$$u_3(v_0) = -\frac{\kappa_1 v_0}{\kappa_4+\kappa_5}u_1(v_0) = -\frac{(\kappa_1\kappa_4 v_0 + \kappa_2(\kappa_4+\kappa_5))\kappa_1 v_0}{(\kappa_1\kappa_4 v_0 + (\kappa_2+\kappa_3)(\kappa_4+\kappa_5))(\kappa_4+\kappa_5)},$$

the solutions to the Poisson equations are given by functions

$$h_1(v) = v_1 u_1(v_0), \qquad h_2(v) = -v_2 u_2(v_0), \qquad h_3(v) = -v_2 u_3(v_0),$$

and $H^N = \frac{1}{N}h_1 + \frac{1}{N^2}(h_2+h_3)$.

Let $r_N = N^{1/2}$ and observe that $\frac{1}{N^2}(h_2+h_3)$ makes a negligible contribution to the quadratic variation. Consequently,

$$\begin{aligned}
&N\big[V_0^N - H^N\circ V^N\big]_t \\
&\quad\approx \sum_{k=1}^{5} N^{-1}\int_0^t \big(\zeta_{k2} + h_1\big(V^N(s-)\big) - h_1\big(V^N(s-) + T\Lambda_N\zeta_k\big)\big)^2\, dR_k^N(s) \\
&\quad\approx \int_0^t \big(-1 + u_1\big(V_0^N\big)\big)^2\kappa_1 V_0^N\big(V_1^N - V_2^N\big)\, ds \\
&\qquad + \int_0^t \big(1 - u_1\big(V_0^N\big)\big)^2\kappa_2\big(M - V_1^N\big)\, ds + \int_0^t u_1\big(V_0^N\big)^2\kappa_3\big(M - V_1^N\big)\, ds.
\end{aligned}$$

Hence

$$\begin{aligned}
G(v) = \big(&\big(\kappa_3(\kappa_4+\kappa_5)\big)^2\big(\kappa_1 v_0(v_1-v_2) + \kappa_2(M-v_1)\big) \\
&+ \big(\kappa_1\kappa_4 v_0 + \kappa_2(\kappa_4+\kappa_5)\big)^2\big(\kappa_3(M-v_1)\big)\big) \\
&/\big(\big(\kappa_1\kappa_4 v_0 + (\kappa_2+\kappa_3)(\kappa_4+\kappa_5)\big)^2\big)
\end{aligned}$$

and

$$\overline{G}(v_0) = \frac{M\kappa_1\kappa_3\kappa_4 v_0(\kappa_3(\kappa_4+\kappa_5)^2(2\kappa_2+\kappa_3) + (\kappa_1\kappa_4 v_0 + \kappa_2(\kappa_4+\kappa_5))^2)}{(\kappa_1\kappa_4 v_0 + (\kappa_2+\kappa_3)(\kappa_4+\kappa_5))^3}.$$

If V_0 is the solution of

$$V_0(t) = V_0(0) - \int_0^t \frac{M\kappa_1\kappa_3\kappa_4 V_0(s)}{\kappa_1\kappa_4 V_0(s) + (\kappa_4+\kappa_5)(\kappa_2+\kappa_3)}\, ds,$$

then, since $G_0 = G_1 \equiv 0$, $U^N = N^{1/2}(V_0^N - V_0) \Rightarrow U$ where

$$U(t) = U(0) + \int_0^t \sqrt{\overline{G}(V_0(s))}\, dW_s - \int_0^t \frac{M\kappa_1\kappa_3\kappa_4(\kappa_4+\kappa_5)(\kappa_2+\kappa_3)}{(\kappa_1\kappa_4 V_0(s) + (\kappa_4+\kappa_5)(\kappa_2+\kappa_3))^2} U(s)\, ds.$$

The corresponding diffusion approximation is

$$D^N(t) = Z_2^N(0) + N^{-1/2}\int_0^t \sqrt{\overline{G}(D^N(s))}\, dW(s) - \int_0^t \frac{M\kappa_1\kappa_3\kappa_4 D^N(s)}{\kappa_1\kappa_4 D^N(s) + (\kappa_4+\kappa_5)(\kappa_2+\kappa_3)}\, ds.$$

Finally, we compare simulations for 500 realizations of the original model X_2 with 500 realizations of the Gaussian approximation $N_0 V_0(\cdot) + N_0^{1/2} U(\cdot)$ and the diffusion approximation $N_0 D_2^{N_0}(\cdot)$. For comparison we also give the deterministic value given by $N_0 V_0(\cdot)$. We use $N_0 = 100$, a time interval on the scale $\gamma = 0$, and initial values are set to $X_1(0) = X_3(0) = 0$, $X_2(0) = 50$ as in the previous example. Here the additional parameters are set to $M = 5$, $\kappa_1' = 0.5$, $\kappa_2' = 500$, $\kappa_3' = 100$ and $\kappa_4' = \kappa_5' = 5000$. Figure 5 shows the mean and one standard deviation above and below the mean for each of the three processes, and Figure 6 five trajectories for the three processes. Again, both Gaussian and diffusion approximations give a good approximation for the mean and the standard deviation from the mean of $X_2(\cdot)$.

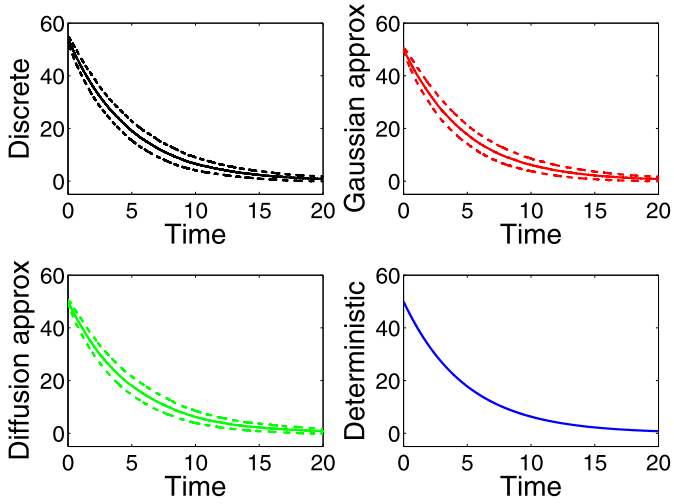

FIG. 5. *Mean and standard deviations of the amount of substrate in the three time-scale enzyme model* [500 *simulations with* $N_0 = 100$, $M = 5$, $\gamma = 0$, $X_1(0) = 0$, $X_2(0) = 50$, $X_3(0) = 0$, $\kappa_1' = 0.5$, $\kappa_2' = 500$, $\kappa_3' = 100$, $\kappa_4' = \kappa_5' = 5000$].

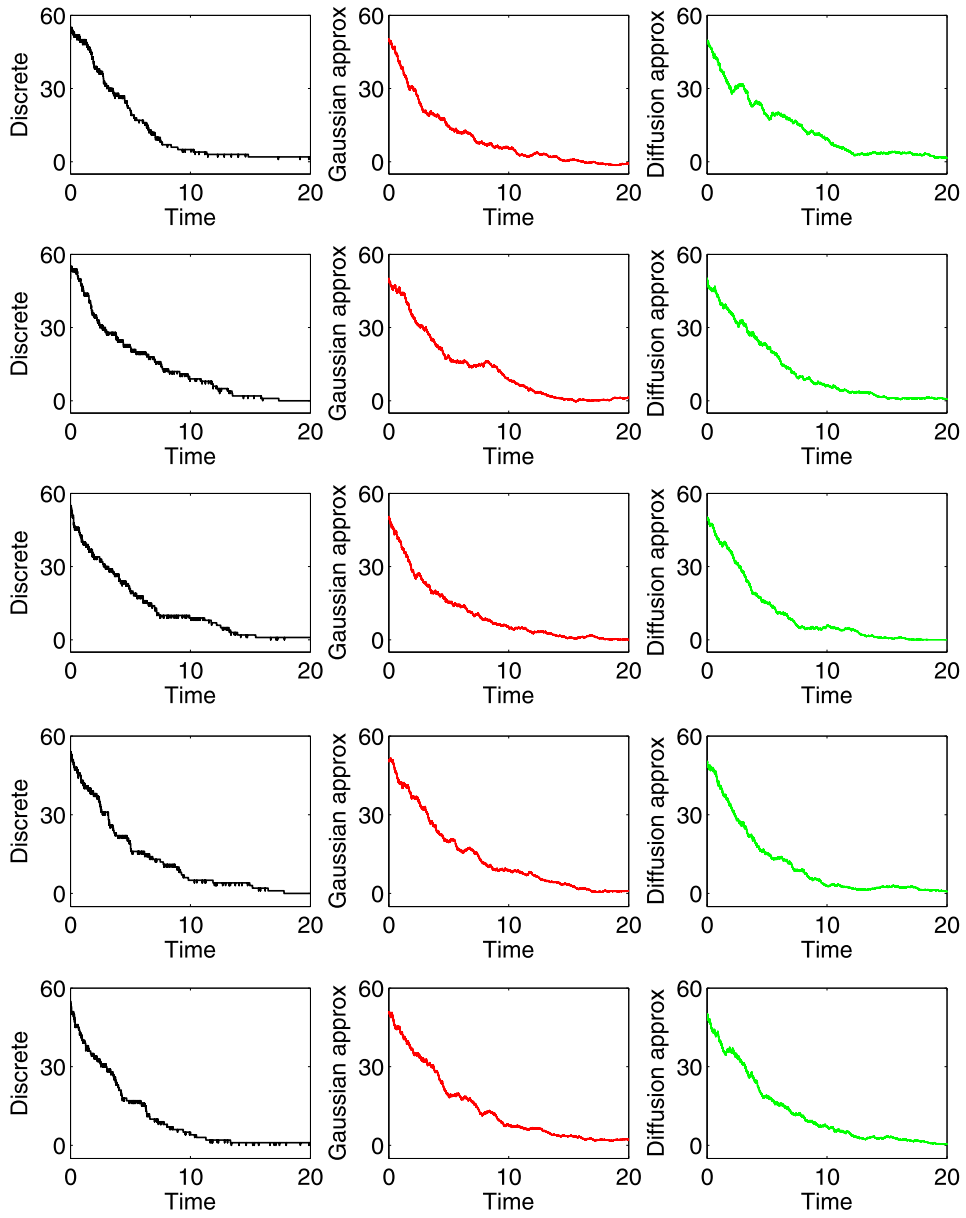

FIG. 6. *Five trajectories for the amount of substrate in the three time-scale enzyme model* (*same parameters as in Figure* 5).

APPENDIX

A.1. Martingale central limit theorem. Various versions of the martingale central limit have been given by McLeish (1974), Rootzén (1977, 1980), Gänssler

and Häusler (1979) and Rebolledo (1980) among others. The following version is from Ethier and Kurtz (1986), Theorem 7.1.4.

THEOREM A.1. *Let $\{M_n\}$ be a sequence of $\mathbb{R}^d$-valued martingales. Suppose*

$$\lim_{n\to\infty} E\Big[\sup_{s\le t}|M_n(s) - M_n(s-)|\Big] = 0 \tag{A.1}$$

and

$$[M_n^i, M_n^j]_t \to c_{i,j}(t)$$

for all $t \ge 0$, where $C = ((c_{i,j}))$ is deterministic and continuous. Then $M_n \Rightarrow M$, where M is Gaussian with independent increments and $E[M(t)M(t)^T] = C(t)$.

REMARK A.2. Note that $C(t) - C(s)$ is nonnegative definite for $t \ge s \ge 0$. If C is absolutely continuous, then the derivative will also be nonnegative definite and will have a nonnegative definite square root. Suppose $\dot{C}(t) = \sigma(t)^2$ where σ is symmetric. Then M can be written as

$$M(t) = \int_0^t \sigma(s)\, dW(s),$$

where W is d-dimensional standard Brownian motion.

REFERENCES

BALL, K., KURTZ, T. G., POPOVIC, L. and REMPALA, G. (2006). Asymptotic analysis of multiscale approximations to reaction networks. *Ann. Appl. Probab.* **16** 1925–1961. MR2288709

BHATTACHARYA, R. N. (1982). On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete* **60** 185–201. MR0663900

COSTA, O. L. V. and DUFOUR, F. (2003). On the Poisson equation for piecewise-deterministic Markov processes. *SIAM J. Control Optim.* **42** 985–1001 (electronic). MR2002143

DAVIS, M. H. A. (1993). *Markov Models and Optimization. Monographs on Statistics and Applied Probability* **49**. Chapman & Hall, London. MR1283589

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York. MR0838085

ETHIER, S. N. and NAGYLAKI, T. (1980). Diffusion approximations of Markov chains with two time scales and applications to population genetics. *Adv. in Appl. Probab.* **12** 14–49. MR0552945

GÄNSSLER, P. and HÄUSLER, E. (1979). Remarks on the functional central limit theorem for martingales. *Z. Wahrsch. Verw. Gebiete* **50** 237–243. MR0554543

GLYNN, P. W. and MEYN, S. P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931. MR1404536

HASELTINE, E. L. and RAWLINGS, J. B. (2002). Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.* **117** 6959–6969.

KANG, H.-W. and KURTZ, T. G. (2013). Separation of time-scales and model reduction for stochastic reaction networks. *Ann. Appl. Probab.* **23** 529–583. MR3059268

KURTZ, T. G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probab.* **8** 344–356. MR0287609

KURTZ, T. G. (1977/78). Strong approximation theorems for density dependent Markov chains. *Stochastic Process. Appl.* **6** 223–240. MR0464414

KURTZ, T. G. (1992). Averaging for martingale problems and stochastic approximation. In *Applied Stochastic Analysis* (*New Brunswick, NJ*, 1991). *Lecture Notes in Control and Inform. Sci.* **177** 186–209. Springer, Berlin. MR1169928

MCLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2** 620–628. MR0358933

REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete* **51** 269–286. MR0566321

ROOTZÉN, H. (1977). On the functional central limit theorem for martingales. *Z. Wahrsch. Verw. Gebiete* **38** 199–210. MR0482956

ROOTZÉN, H. (1980). On the functional central limit theorem for martingales. II. *Z. Wahrsch. Verw. Gebiete* **51** 79–93. MR0566110

SRIVASTAVA, R., YOU, L., SUMMERS, J. and YIN, J. (2002). Stochastic vs. deterministic modeling of intracellular viral kinetics. *J. Theoret. Biol.* **218** 309–321. MR2026325

VAN KAMPEN, N. G. (1961). A power series expansion of the master equation. *Canad. J. Phys.* **39** 551–567. MR0128921

H.-W. KANG
MATHEMATICAL BIOSCIENCES INSTITUTE
OHIO STATE UNIVERSITY
1735 NEIL AVE.
COLUMBUS, OHIO 43210
USA
E-MAIL: kang235@mbi.osu.edu

T. G. KURTZ
DEPARTMENTS OF MATHEMATICS AND STATISTICS
UNIVERSITY OF WISCONSIN
480 LINCOLN DR.
MADISON, WISCONSIN 53706
USA
E-MAIL: kurtz@math.wisc.edu

L. POPOVIC
DEPARTMENT OF MATHEMATICS AND STATISTICS
CONCORDIA UNIVERSITY
1450 DE MAISONNEUVE BLVD. WEST
MONTREAL, QUEBEC H3G1M8
CANADA
E-MAIL: lpopovic@mathstat.concordia.ca