ABSTRACT

Title of dissertation: State and Parameter Estimation in Stochastic Dynamical Systems

Mingkai Yu

Dissertation directed by: Dr. Muruhan Rathinam
Department of Mathematics and Statistics

Stochastic dynamical systems arise in the modeling of intracellular biological processes driven by diffusion and reaction of molecules. While several fluorescence based techniques exist for partial measurements of these systems, it remains a challenge to estimate the full state and parameters of such systems from the partially observed data. In this thesis, we study two problems: inferring the diffusivity from raw Fluorescence Correlation Spectroscopy (FCS) data, and estimating state and parameters in partially observed chemical reaction networks.

Fluorescence intensity could be modeled as a stochastic process governed by the motion of particles. The autocorrelation of the intensity has an analytical form as a function of time lag and diffusion coefficient, hence the diffusion coefficient could be fitted. We present a mathematical derivation of the autocorrelation function. We also derive a formula for the variance of the time average of the autocorrelation function in integral form, and give a closed-form upper bound. We examine several different approaches to fit the diffusivity via Monte Carlo simulations. Furthermore,

we analyze the sensitivity of the diffusivity D obtained via the process of nonlinear least squares fit.

We also study chemical reaction networks modeled by a discrete-state continuous-time Markov process, where the state vector represents the copy number of the species. We consider two scenarios, one in which exact observations of some species are made continuously in time over a window and the other in which observations are made in snapshots of time.

For the continuous in time observation problem, we derive equations for the conditional probability distribution of the unobserved states. We also provide a novel particle filter to compute the conditional probability distribution of the unobserved species. We also adapt our algorithm for the estimation of parameters and for a past state value based on observations up to a future time.

For the filtering of stochastic reaction networks when some states are observed noiselessly in snapshots of time, we propose a targeting algorithm that ensures the filter process reaches the target state. We present a rigorous proof of our algorithm, discuss the choices we could make within the implementation of the algorithm, and use numerical examples to illustrate our algorithm.

State and Parameter Estimation in Stochastic Dynamical Systems

by

Mingkai Yu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore County in partial fulfillment
of the requirements for the degree of Doctor of Philosophy 2023

Advisory Committee:
Dr. Muruhan Rathinam, Chair/Advisor
Dr. Avanti Athreya
Dr. Animikh Biswas
Dr. Anindya Roy
Dr. Jinglai Shen

Acknowledgments

I would like to thank my advisor Dr. Muruhan Rathinam for his continuous help and guidance throughout my Ph.D. study. I am grateful for him introducing me to this interesting topic and giving me lots of good ideas. He has always been willing to help me with his vast knowledge. His guidance in my research and writing of the thesis have been invaluable to me.

I thank Dr. Avanti Athreya, Dr. Animikh Biswas, Dr. Anindya Roy and Dr. Jinglai Shen for serving in my advisory committee and providing insightful comments.

I thank all professors in the math department at UMBC who taught me courses. Special thanks to Dr. Muruhan Rathinam, Dr. Animikh Biswas, Dr. Justin Webster, and Dr. Hye-Won Kang and for their amazing lectures on analysis, differential equations, probability theory and stochastic processes.

I owe my deepest gratitude to my parents and boyfriend for their love and support. It wouldn't be possible for me to achieve this without their understanding and encouragement.

Table of Contents

Chapter 1

Introduction

Many real-world phenomena could be modeled as stochastic dynamical systems, but measuring parameters or the full state of such systems can be challenging. Sometimes we have to make an indirect measurement of the process to infer the unobservable states and parameters. In this thesis, we study two problems with applications in biochemical field: inferring the diffusivity from Fluorescence Correlation Spectroscopy (FCS), and estimating state and parameters in partially observed chemical reaction networks.

1.1 Literature Review

FCS is a tool to determine the diffusivity of fluorescently labeled molecules in solution by measuring the fluctuation of fluorescence intensity [35, 36] and it's useful in the study of drug delivery, where scientists are interested in the diffusion of drug molecules in macromolecular solutions and polymer gels [41, 42]. This technology has also improved scientists' ability to track biological processes at the intracellular level [33]. Reaction networks are models for systems consisting of chemical species and a set of reaction channels. Fluorescence technologies can track the temporal dynamics of a limited number of species, rasing the a critical question of estimating the full dynamics from the partial time-course observation data.

In an FCS experiment, as particles undergo reflected Brownian motion in a container diffusing into and out of a very small confocal volume illuminated by a laser beam, the fluorescence intensity fluctuates. The fluorescence intensity could be modeled as a stochastic process governed by the motion of particles. The autocorrelation of the intensity has an analytical form as a function of time lag and diffusion coefficient, so the diffusion coefficient could be fitted from the experimentally measured autocorrelation function. In [30], Koppel provided the first statistical analysis of the standard deviations for experimental correlation functions assuming Gaussian intensities. In [40], the authors have computed the standard deviation and bias of the autocorrelation function from an engineer's perspective.

Chemical reaction networks occurring at the intracellular level often have some or all molecular species present in low copy numbers. As such, these are best modeled by a discrete state Markov process in continuous time, where the state $Z(t) \in \mathbb{Z}_+^n$ of the process is the vector copy number of the molecular species at time t. We consider a stochastic reaction network with n molecular species and m reaction channels. The dynamics of the reaction network are characterized by *propensity functions* $a_j(z, c)$ where $z \in \mathbb{Z}_+^n$ is the state and $c \in \mathbb{R}^p$ is a vector of parameters, and also by the stoichiometric vectors $\nu_j \in \mathbb{Z}^n$ for $j = 1, \ldots, m$. The occurrence of a reaction event j at time t leads to a state change $Z(t) = Z(t-) + \nu_j$ and given $Z(t) = z$, the probability that a reaction event j occurs during $(t, t + h]$ is $a_j(z, c)h + o(h)$ as $h \to 0^+$.

This model and the corresponding stochastic simulation algorithm introduced by Gillespie [19, 18] assumed a well-stirred system where spatial variations in molec-

ular concentrations were considered negligible. Later on when the spatial variations were considered important, this model still proved useful in that its framework allowed for a spatial compartmental model where transport of molecules across compartments could be treated mathematically as reactions [6].

Stochastic filtering methods provide a framework for addressing the problem of determining the state of a system from incomplete and potentially noisy observations. The stochastic filtering methods usually involve generating recursive updates in time so that the additional computation required in going from the knowledge of conditional probability at a time instant t to a future time instant $t+h$ only involves the observations made during $(t, t+h]$ and the conditional probability computed by time t.

The widely known instance of stochastic filtering is the Kalman filter [26] which was concerned with a Gaussian process with a linear evolution in discrete time. Later on, the Kalman-Bucy filter [27] was introduced in the continuous time setting of linear stochastic differential equations (SDEs). In these linear and Gaussian settings the conditional probability distribution is also Gaussian, and hence its time evolution can be reduced to studying the time evolution of the mean and the covariance.

In general, one may not expect the conditional probability to be Gaussian and hence the methods are more complex. In the general nonlinear setting for SDEs, there are several methods and one may find [4] as a general reference. In the setting of Markov processes in discrete time, one may find recursive filtering equations which describe the time evolution of the conditional probability distribution [14].

For discrete and finite state Markov processes in continuous time, Confortola

and Fuhrman [10] provided a rigorous derivation of the evolution equation for the conditional probability distribution $\pi(t, x)$ for the case of finite state Markov chain. Our work [39] gave an intuitive derivation for reaction networks with unbounded state space. The authors [12] later gave a rigours derivation for unbounded state space. These evolution equations take the form of differential equations with jumps. These equations, known as the *filtering equations*, are analogous to *Kolmogorov's forward equations* for the time evolution of the (unconditional) probability distribution $P\{Z(t) = z\}$ of a continuous time and discrete state Markov process. In the case of reaction networks, Kolmogorov's forward equations are known as the *Chemical Master Equations* (CME).

It is well known that when the number of states is infinite or very large, the solution of the CME is not practical and Monte Carlo simulation is the more efficient method of choice. The well known Gillespie algorithm [19, 21] and its variants[17, 1] are the methods of choice for exact realizations of sample paths. Likewise, when the state space is very large or infinite, direct numerical solution of the filtering equations is not practical.

We describe some previous results for state and parameter estimation in the context of reaction networks. All these methods are for observations based on discrete time snapshots. In [7], the authors consider Bayesian inference of parameters based on exact partial state observations in discrete time snapshots and proposes a Markov Chain Monte Carlo (MCMC) method. Golightly and collaborators [23, 24] use the *Chemical Langevin Equation* (CLE) as the underlying model and propose MCMC methods for state and parameter estimation based on partial state obser-

vations corrupted by additive Gaussian noise in discrete time snapshots. In [9], the authors consider the case where the reaction network is approximated by the *linear noise approximation*. The observations are assumed to be linear combination of the states corrupted by an additive Gaussian noise with an unknown variance and they propose an extended Kalman-Bucy filter. They also consider pure delays in the system. The recent work in [13] considers the case where the reaction network may be approximated by an ODE model with jumps. A function of the state corrupted by an additive Gaussian noise is observed in discrete time snapshots. Rigorous results are derived to show that by solving the filtering problem for the reduced model based on observations of the original model, one can also obtain a good approximation of the estimation of the original model.

1.2 Contribution of this Thesis

In Chapter 2, we present an alternative mathematical derivation of the autocorrelation function. To quantify the uncertainty, we also derive a formula for the variance of the time average of the autocorrelation function $C(T,\tau) = \frac{1}{T}\int_0^T I(t)I(t+\tau)dt$ in integral form in terms of system specifications, and we find a closed-form upper bound (See equation (2.42)). We show that as observation time window $T \to \infty$, the variance of the time average converges in the order of $O(\frac{1}{T})$. We then examine several different approaches to fit the diffusivity via Monte Carlo simulations. Then in order analyze the sensitivity of the diffusivity D obtained via the procedure of nonlinear least squares fit, we apply the implicit function theorem to

the system established by the first-order optimality condition. As we are interested in the sensitivity of the fitted D in response to the perturbation of the observed data, which is a stochastic process, we first characterize the perturbation down to a single random scalar by keeping the first mode of the Karhunen-Loeve expansion. We find that although the first mode carries a significant amount of the energy, it contributes very little to the inference error. In fact, more modes are needed in the sensitivity analysis.

In Chapter 3 and Chapter 4, we investigate the estimation of unobserved states and parameters of a stochastic reaction network based on exact (noiseless) partial state observations. We consider the situation where observations are made continuously in time in Chapter 3 (which leads to published paper [39]), and then in Chapter 4, we consider the scenario where observations are made in snapshots of time. We write $Z(t) = (X(t), Y(t))$ where $X(t) \in \mathbb{Z}_+^{n_1}$ is the unobserved component of the state and $Y(t) \in \mathbb{Z}_+^{n_2}$ is the observed component of the state.

In Chapter 3 for the continuous-time observation, we are interested in computing the *conditional probability*

$$\pi(t, x) = P\{X(t) = x \,|\, Y(s) = y(s),\, 0 \le s \le t\} \quad \forall x \in \mathbb{Z}_+^{n_1}. \tag{1.1}$$

We show in Section 3.1 that the evolution equation for $\pi(t, x)$ satisfies a system of nonlinear differential equations with jumps and hence it is not possible to regard $\pi(t, x)$ as the (unconditional) probability distribution of a Markov process. We define a nonnegative *unnormalized conditional distribution* $\rho(t, x)$ which satisfies a

linear evolution equation and when normalized yields $\pi(t, x)$.

We provide a Monte Carlo filtering algorithm in Section 3.2 which generates a pair of processes (V, w) where V has the same state space as X such that

$$\pi(t, x) = \frac{\mathbb{E}[1_{\{x\}}(V(t))\, w(t)]}{\mathbb{E}[w(t)]}.$$

To estimate unknown parameters, we adapt our algorithm in a Bayesian framework by absorbing unknown parameters into the augmented state space. Similar strategies could be applied to address the smoothing problem (i.e. to estimate the conditional probability of a past event based on the observations made until a later time). We illustrate our algorithms via numerical examples in Section 3.4.

In Chapter 4, we are concerned with the filtering of stochastic reaction networks when some states are observed noiselessly in snapshots of time. We aim to compute the conditional probabilities

$$\pi(t, z) = P\{Z(t) = z \,|\, Y(t_i) = y_i,\, i = 1, \ldots, T_s\} \tag{1.2}$$

. where $t \geq 0$, $z \in \mathbb{Z}_+^n$ and T_s is the number of snapshots.

We describe a straightforward Monte Carlo algorithm based on a simple acceptance/rejection strategy. This algorithm could be highly inefficient with low effective sample size in situation where the probability for the observed outcome to happen is small. We therefore propose a new algorithm which we call the *targeting algorithm* that ensures the filter process reaches the target state. This is achieved by im-

posing the constraints on the reaction count vector between observation snapshots. Conditioned on that, we simulate the paths in between by interpolating the filter process as a conditional inhomogeneous Poisson process and computing the associated weights due to a change of probability measure. We present a rigorous proof of our algorithm in Section 4.3. We also discuss the choices we could make within the implementation of the algorithm in Section 4.4 and use numerical examples to illustrate our algorithm in Section 4.5.

Chapter 2

Parameter Estimation in Fluorescence Correlation Spectroscopy

2.1 Introduction

Fluorescence Correlation Spectroscopy (FCS) is a tool to determine the diffusivity of fluorescently labeled molecules in solution by measuring the fluctuation of fluorescence intensity. In the FCS experiment, the solution in a container is placed under the laser beam which creates a tiny confocal volume inside the solution. When the fluorescent particles cross the confocal volume, photons are emitted and a detector records the emitted photon intensity.

The fluorescence intensity could be modeled as a stochastic process governed by the motion of particles. Let N be the total number of particles in solution. Let $X_i(t)$ be the location of the ith fluorescently labeled particle at time t, $X_i(t) \in \mathbb{R}^3$, $i = 1, 2, ..., N$. Let $F : \mathbb{R}^3 \to \mathbb{R}$ be the laser intensity function the maps the location of a particle to an intensity it contributes. So the intensity contributed by particle j is given by

$$i_j(t) = F(X_j(t)), \tag{2.1}$$

where

$$F(x) = F_0 \exp\left(\frac{-2x_1^2}{r^2} + \frac{-2x_2^2}{r^2} + \frac{-2x_3^2}{l^2}\right). \tag{2.2}$$

The total intensity contributed by all the fluorescent particles is,

$$I(t) = \sum_{j=1}^{N} i_j(t). \tag{2.3}$$

Define the autocorrelation function $G(\tau)$ by

$$G(\tau) = \frac{\mathrm{Cov}\big(I(t), I(t+\tau)\big)}{\big(\mathbb{E}[I(t)]\big)^2}. \tag{2.4}$$

In the next section, we will show that $G(\tau)$ has an analytical formula as a function of the time lag τ and the diffusion coefficient D, so diffusion coefficient could be fitted from the experimentally measured autocerrelation function.

2.2 Derivation of Autocorrelation Function $G(\tau)$

Assumption 1. Processes $\{X_i(t)\}_{i=1}^{N}$ are i.i.d. (independent identically distributed) and can be represented by a process $X(t)$.

Hence,

$$\mathrm{Cov}\big(I(t), I(t+\tau)\big) = \mathrm{Cov}\big(\sum_{j=1}^{N} i_j(t), \sum_{k=1}^{N} i_k(t+\tau)\big) = \sum_{j=1}^{N}\sum_{k=1}^{N} \mathrm{Cov}\big(i_j(t), i_k(t+\tau)\big)$$

$$= \sum_{j=1}^{N} \mathrm{Cov}\big(i_j(t), i_j(t+\tau)\big) = N\mathrm{Cov}(i(t), i(t+\tau)) = N\Big(\mathbb{E}[i(t+\tau)i(t)] - \mathbb{E}[i(t+\tau)]\mathbb{E}[i(t)]\Big).$$

And,

$$\mathbb{E}[I(t)] = \mathbb{E}\left[\sum_{j=1}^{N} i_j(t)\right] = N\mathbb{E}[i(t)].$$

Thus,

$$G(\tau) = \frac{\mathrm{Cov}\big(I(t+\tau), I(t)\big)}{\big(\mathbb{E}[I(t)]\big)^2} = \frac{N\Big(\mathbb{E}[i(t+\tau)i(t)] - \mathbb{E}[i(t)]\mathbb{E}[i(t+\tau)]\Big)}{N^2(\mathbb{E}[i(t)])^2} = \frac{\mathbb{E}[i(t+\tau)i(t)]}{N(\mathbb{E}[i(t)])^2} - \frac{1}{N} \tag{2.5}$$

Next, we shall compute $\mathbb{E}[i(t)]$ and $\mathbb{E}[i(t)i(t+\tau)]$.

Assumption 2. The law of $X(t)$ is independent of t, and is the uniform measure on $B \subset \mathbb{R}^3$, where B is the box that contains the solution. Thoughout our analysis, we assume $N, |B| \to \infty$ with $\frac{N}{|B|}$ fixed. Thus,

$$\mathbb{E}[i(t)] = \mathbb{E}[F(X(t))] = \frac{1}{|B|}\int_{x\in B} F(x)dx \approx \frac{1}{|B|}\int_{\mathbb{R}^3} F(x)dx = F_0\frac{\pi^{3/2}r^2 l}{\sqrt{8}|B|}. \tag{2.6}$$

Here, we provide some explanation of the last two steps in the above equation.
Note that $F(x) = F_0 \exp\left(\frac{-2x_1^2}{r^2} + \frac{-2x_2^2}{r^2} + \frac{-2x_3^2}{l^2}\right) = F_0\exp(-x^T Ax)$ is a three-dimensional Gaussian function, where $A = \mathrm{diag}(\frac{2}{r^2}, \frac{2}{r^2}, \frac{2}{l^2})$. For $\int_{\mathbb{R}^3} F(x)dx$, we know that 99.7% of the contribution comes from within the ellipsoid $\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = 1$ with $a = \frac{3r}{2}, b = \frac{3r}{2}$ and $c = \frac{3l}{2}$. As long as the length of each side of the solution box B is much larger than r and l, the contribution from outside the box could be safely ignored.
For the integral evaluation step, recall the Gaussian integral

$$\int_{\mathbb{R}} e^{-x^2}dx = \sqrt{\pi}.$$

In general, for a multidimensional Gaussian function $f(z) = \exp(-z^T Q z)$, where $z \in \mathbb{R}^n$ and Q is symmetric positive definite, there exists an orthogonal matrix P such that $P^T Q P = D$, where $D = (d_i)$ is diagonal. Let $w = P^T z$, we have

$$\int_{\mathbb{R}^n} \exp(-z^T Q z) dz = \int_{\mathbb{R}^n} \exp(-w^T D w) dw = \frac{\sqrt{\pi^n}}{\sqrt{d_1 d_2 ... d_n}} = \frac{\sqrt{\pi^n}}{\sqrt{\det(Q)}}.$$

Assumption 3. The Markov process $X(t)$ has Brownian increments, i.e. $X(t+\tau) - X(t) \sim \mathcal{N}(0, \sigma^2 \tau I)$. That is, it has a transition density $k(\tau, x, y)$.

$$k(\tau, x, y) \approx (2\pi\sigma^2\tau)^{-3/2} \exp\Big\{ - \frac{\|y - x\|^2}{2\sigma^2\tau} \Big\}. \tag{2.7}$$

Note that σ is related with the diffusivity D in the sense that $D = \frac{\sigma^2}{2}$. Now we can compute $\mathbb{E}[i(t)i(t+\tau)]$.

$$\begin{aligned}
&\mathbb{E}[i(t)i(t+\tau)] = \mathbb{E}\left[F(X(t))F(X(t+\tau))\right] = \mathbb{E}\Big[\mathbb{E}\big[\ F(X(t))F(X(t+\tau))|X(t)\big]\Big] \\
&= \mathbb{E}\Big[F(X(t))\ \mathbb{E}[F(X(t+\tau))|X(t)]\Big] = \int_{x\in B} F(x)\ \mathbb{E}[F(X(t+\tau))|X(t) = x]\ \frac{1}{|B|} dx \\
&= \frac{1}{|B|} \int_{x\in B} \int_{y\in B} F(x)F(y)k(\tau, x, y) dy dx \approx F_0^2 \frac{\pi^{3/2} r^2 l}{8|B|(1 + \frac{4D\tau}{r^2})\sqrt{1 + \frac{4D\tau}{l^2}}}.
\end{aligned}$$

The detail of the integral evaluation step is provided as follows,

$$\begin{aligned}
&\frac{1}{|B|} \int_{x\in B} \int_{y\in B} F(x)F(y)k(\tau, x, y) dy dx \\
&\approx \frac{F_0^2}{|B|} \frac{1}{(2\pi\sigma^2\tau)^{3/2}} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \exp(-x^T A x) \exp(-y^T A y) \exp\left(- \frac{(y-x)^T (y-x)}{2\sigma^2\tau} \right) dx dy \\
&= \frac{F_0^2}{|B|} \frac{1}{(2\pi\sigma^2\tau)^{3/2}} \int_{\mathbb{R}^6} \exp(-z^T Q z) dz
\end{aligned}$$

where

$$Q = \begin{bmatrix} A + \frac{I}{2\sigma^2\tau} & -\frac{I}{2\sigma^2\tau} \\ \\ -\frac{I}{2\sigma^2\tau} & A + \frac{I}{2\sigma^2\tau} \end{bmatrix}.$$

Evaluating the Guassian integral in $\mathbb{R}^6$, we get

$$\mathbb{E}[i(t)i(t+\tau)] \approx F_0^2 \frac{\pi^{3/2} r^2 l}{8|B|(1+\frac{4D\tau}{r^2})\sqrt{1+\frac{4D\tau}{l^2}}}. \tag{2.8}$$

With the analytical formulas for $\mathbb{E}[i(t)]$ and $\mathbb{E}[i(t)i(t+\tau)]$ available, we can derive $G(\tau)$

$$G(\tau) = \frac{\mathrm{Cov}\big(I(t+\tau), I(t)\big)}{\big(\mathbb{E}[I(t)]\big)^2} = \frac{\mathbb{E}[i(t+\tau)i(t)]}{N(\mathbb{E}[i(t)])^2} - \frac{1}{N} = \frac{1}{N}\frac{|B|}{\pi^{3/2}r^2 l}\frac{1}{(1+\frac{4D\tau}{r^2})}\frac{1}{\sqrt{1+\frac{4D\tau}{l^2}}} - \frac{1}{N}. \tag{2.9}$$

The total number of particles N is very large (say, more than a thousand), while $\frac{N\ \pi^{3/2}r^2 l}{|B|}$ is usually chosen to be less than ten, so we keep the first term and drop the second term of the above equation. Thus

$$G(\tau) \approx \frac{1}{N}\frac{|B|}{\pi^{3/2}r^2 l}\frac{1}{(1+\frac{4D\tau}{r^2})}\frac{1}{\sqrt{1+\frac{4D\tau}{l^2}}} = \frac{1}{N_e}\frac{1}{(1+\frac{4D\tau}{r^2})}\frac{1}{\sqrt{1+\frac{4D\tau}{l^2}}}.$$

where

$$N_e = \frac{N}{|B|}\pi^{3/2} r^2 l \tag{2.10}$$

is called the effective number of particles in the confocal volume. While the function

$F(x)$ decays exponentially from the center of the laser beam, we can still think of an effective confocal volume of $\pi^{3/2}r^2l$. It follows that $\frac{N\pi^{3/2}r^2l}{|B|}$ is the number of particle in that effective confocal volume . Eventually, we reach the analytical formula used by scientists [42],

$$G(\tau) = \frac{1}{N_e}\frac{1}{(1+\frac{4D\tau}{r^2})}\frac{1}{\sqrt{1+\frac{4D\tau}{l^2}}}. \tag{2.11}$$

2.3 Higher Moments of Autocorrelation Function

2.3.1 Moments Contributed by a Single Particle

Here we list the moments of autocorrelation function contributed by a single particle. These formulas are derived by successive conditioning and integrating Gaussian functions.

We define diffusion time τ_r and τ_l by

$$\tau_r = \frac{r^2}{4D}, \qquad \tau_l = \frac{l^2}{4D}. \tag{2.12}$$

And we introduce ϕ_n and ψ_n functions for the n^{th} moment of autocorrelation function, where ϕ takes in absolute time while ψ takes in relative time lag as shortcuts. The list of moments are as follows:

$$\phi_1(t) = \psi_1 = \mathbb{E}[i(t)] = F_0\frac{\pi^{3/2}r^2l}{\sqrt{8}|B|}. \tag{2.13}$$

$$\begin{aligned}\phi_2(t,t+\tau) = \psi_2(\tau) &= \mathbb{E}[i(t)i(t+\tau)] \\ &= F_0^2 \frac{\pi^{3/2} r^2 l}{\sqrt{8}|B|(2+\frac{2\tau}{\tau_r})\sqrt{2+\frac{2\tau}{\tau_l}}}.\end{aligned} \tag{2.14}$$

$$\begin{aligned}\phi_3(t,t+\tau_1,t+\tau_1+\tau_2) = \psi_3(\tau_1,\tau_2) &= \mathbb{E}[i(t)i(t+\tau_1)i(t+\tau_1+\tau_2)] \\ &= F_0^3 \frac{\pi^{3/2} r^2 l}{\sqrt{8}|B|(3+\frac{4(\tau_1+\tau_2)}{\tau_r}+\frac{4\tau_1\tau_2}{\tau_r^2})\sqrt{3+\frac{4(\tau_1+\tau_2)}{\tau_l}+\frac{4\tau_1\tau_2}{\tau_l^2}}}.\end{aligned} \tag{2.15}$$

$$\begin{aligned}&\phi_4(t,t+\tau_1,t+\tau_1+\tau_2,t+\tau_1+\tau_2+\tau_3) = \psi_4(\tau_1,\tau_2,\tau_3) \\ &= \mathbb{E}[i(t)i(t+\tau_1)i(t+\tau_1+\tau_2)i(t+\tau_1+\tau_2+\tau_3)] \\ &= F_0^4 \frac{\pi^{3/2} r^2 l}{\sqrt{8}|B|} \frac{1}{(4+2\frac{3\tau_1+4\tau_2+3\tau_3}{\tau_r}+8\frac{\tau_1\tau_2+\tau_1\tau_3+\tau_2\tau_3}{\tau_r^2}+3\frac{\tau_1\tau_2\tau_3}{\tau_r^3})} \\ &\quad \times \frac{1}{\sqrt{4+2\frac{3\tau_1+4\tau_2+3\tau_3}{\tau_l}+8\frac{\tau_1\tau_2+\tau_1\tau_3+\tau_2\tau_3}{\tau_l^2}+3\frac{\tau_1\tau_2\tau_3}{\tau_l^3}}}.\end{aligned} \tag{2.16}$$

2.3.2 Moments of the Total Intensity

Since $I(t) = \sum_{j=1}^{N} i_j(t)$, we can obtain moments of total intensity by doing combinatorics.

$$\mathbb{E}[I(t)] = N\mathbb{E}[i(t)] = F_0 \frac{N_e}{\sqrt{8}}, \text{ where } N_e = \frac{N\pi^{3/2} r^2 l}{|B|}. \tag{2.17}$$

$$\begin{aligned}\mathrm{Corr}(I(t), I(t+\tau)) :=& \mathbb{E}[I(t)I(t+\tau)] = N\mathbb{E}[i(t)i(t+\tau)] + N(N-1)\big(\mathbb{E}[i(t)]\big)^2 \\ &\approx F_0^2 \left(\frac{N_e}{8(1+\frac{4D\tau}{r^2})\sqrt{1+\frac{4D\tau}{l^2}}} + \left(\frac{N_e}{\sqrt{8}}\right)^2 \right).\end{aligned} \tag{2.18}$$

$$\mathrm{Cov}(I(t), I(t+\tau)) = \mathrm{Corr}(I(t), I(t+\tau)) - (\mathbb{E}[I(t)])^2 = F_0^2 \frac{N_e}{8(1+\frac{4D\tau}{r^2})\sqrt{1+\frac{4D\tau}{l^2}}}. \tag{2.19}$$

$$\begin{aligned}&\mathbb{E}[I(t)I(t+\tau_1)I(t+\tau_1+\tau_2)] = N\mathbb{E}[i(t)i(t+\tau_1)i(t+\tau_1+\tau_2)] + +N(N-1)(N-2)(\mathbb{E}[i])^3 \\ &+ N(N-1)\ \mathbb{E}[i]\ \big(\mathbb{E}[i(t)i(t+\tau_1)] + \mathbb{E}[i(t)i(t+\tau_1+\tau_2)] + \mathbb{E}[i(t+\tau_1)i(t+\tau_1+\tau_2)]\big) \\ &\approx F_0^3 \Bigg(\frac{N_e}{\sqrt{8}} \frac{1}{(3+\frac{4(\tau_1+\tau_2)}{\tau_r}+\frac{4\tau_1\tau_2}{\tau_r^2})\sqrt{3+\frac{4(\tau_1+\tau_2)}{\tau_l}+\frac{4\tau_1\tau_2}{\tau_l^2}}} + \frac{N_e^3}{8\sqrt{8}} + \frac{N_e^2}{8}\frac{1}{(2+\frac{2\tau_1}{\tau_r})\sqrt{2+\frac{2\tau_1}{\tau_l}}} \\ &\quad + \frac{N_e^2}{8}\frac{1}{(2+\frac{2\tau_2}{\tau_r})\sqrt{2+\frac{2\tau_2}{\tau_l}}} + \frac{N_e^2}{8}\frac{1}{(2+\frac{2(\tau_1+\tau_2)}{\tau_r})\sqrt{2+\frac{2(\tau_1+\tau_2)}{\tau_l}}} \Bigg).\end{aligned} \tag{2.20}$$

2.4 Time Average $C(T,\tau) = \frac{1}{T}\int_0^T I(t)I(t+\tau)dt$

Recall that

$$G(\tau) = \frac{\mathrm{Cov}\big(I(t+\tau), I(t)\big)}{\big(\mathbb{E}[I(t)]\big)^2} = \frac{\mathbb{E}[I(t+\tau)I(t)] - \mathbb{E}[I(t)]\mathbb{E}[I(t+\tau)]}{(\mathbb{E}[I(t)])^2} = \frac{\mathbb{E}[I(t+\tau)I(t)]}{\big(\mathbb{E}[I(t)]\big)^2} - 1.$$

In practice, we use time average instead of ensemble average:

$$\hat{G}(\tau) = \frac{\frac{1}{T}\int_0^T I(t)I(t+\tau)dt}{\left(\frac{1}{T}\int_0^T I(t)dt\right)^2} - 1. \tag{2.21}$$

We denote

$$S(T) = \frac{1}{T}\int_0^T I(t)dt, \tag{2.22}$$

and

$$C(T,\tau) = \frac{1}{T}\int_0^T I(t)I(t+\tau)dt. \tag{2.23}$$

Since $X(t)$, $i(t)$ and $I(t)$ are stationary processes, we can see that $S(T)$ and $C(T,\tau)$ are unbiased estimators of $\mathbb{E}[I(t)]$ and $\mathbb{E}[I(t)I(t+\tau)]$, respectively. That is, $\mathbb{E}\left[\frac{1}{T}\int_0^T I(t)dt\right] = \mathbb{E}[I(t)]$ and $\mathbb{E}\left[\frac{1}{T}\int_0^T I(t)I(t+\tau)dt\right] = \mathbb{E}[I(t)I(t+\tau)]$.

However, when it comes to G, it is not clear whether $\hat{G}$ is an unbiased estimator of G.

$$G(\tau) = \frac{\mathbb{E}[I(t+\tau)I(t)]}{\left(\mathbb{E}[I(t)]\right)^2} - 1, \qquad \hat{G}(\tau) = \frac{\frac{1}{T}\int_0^T I(t)I(t+\tau)dt}{\left(\frac{1}{T}\int_0^T I(t)dt\right)^2} - 1.$$

$$\mathbb{E}[\hat{G}(\tau)] \neq \frac{\mathbb{E}\left[\frac{1}{T}\int_0^T I(t)I(t+\tau)dt\right]}{\left(\mathbb{E}\left[\frac{1}{T}\int_0^T I(t)dt\right]\right)^2} - 1 = G(\tau). \qquad \mathbb{E}\left[\frac{C}{S^2}\right] \neq \frac{\mathbb{E}[C]}{(\mathbb{E}[S])^2}.$$

In general, we cannot distribute the expectation over the numerator and denominator. However, once the random quantities C and S have almost no variance, $\mathbb{E}[\hat{G}(\tau)]$ becomes a justifiable approximation of $G(\tau)$. Since $I(t) = F(X(t))$, X is an ergodic Markov process, F is a bounded measurable function, we have that $S(T)$ converges in probability to $\mathbb{E}[I(t)]$ as $T \to \infty$. What remains is to evaluate or estimate the

variance

$$\mathrm{Var}(C(T,\tau)) = \mathbb{E}[C(T,\tau)^2] - (\mathbb{E}[C(T,\tau)])^2. \tag{2.24}$$

2.4.1 $\mathrm{Var}[C(T,\tau)]$

$$\mathbb{E}[C(T,\tau)^2] = \frac{1}{T^2}\int_0^T \int_0^T \mathbb{E}[I(t)I(t+\tau)I(s)I(s+\tau)]dtds \tag{2.25}$$

We consider the general case of $\mathbb{E}[I(t_1)I(t_2)I(t_3)I(t_4)]$

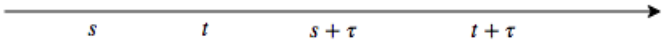
$$\begin{aligned}
&\mathbb{E}[I(t_1)I(t_2)I(t_3)I(t_4)] = \mathbb{E}\Big[\sum_{j=1}^{N} i_j(t_1)\sum_{k=1}^{N} i_k(t_2)\sum_{l=1}^{N} i_l(t_3)\sum_{m=1}^{N} i_m(t_4)\Big] \\
&= N\phi_4(t_1,t_2,t_3,t_4) \\
&+ N(N-1)[\phi_3(t_1,t_2,t_3)\phi_1(t_4) + \phi_3(t_1,t_2,t_4)\phi_1(t_3) + \phi_3(t_1,t_3,t_4)\phi(t_2) + \phi_3(t_2,t_3,t_4)\phi_1(t_1)] \\
&+ N(N-1)[\phi_2(t_1,t_2)\phi_2(t_3,t_4) + \phi_2(t_1,t_3)\phi_2(t_2,t_4) + \phi_2(t_1,t_4)\phi_2(t_2,t_3)] \\
&+ N(N-1)(N-2)[\phi_2(t_1,t_2)\phi_1(t_3)\phi_1(t_4) + \phi_2(t_1,t_3)\phi_1(t_2)\phi_1(t_4) + \phi_2(t_1,t_4)\phi_1(t_2)\phi_1(t_3) \\
&\quad + \phi_2(t_2,t_3)\phi_1(t_1)\phi_1(t_4) + \phi_2(t_2,t_4)\phi_1(t_1)\phi_1(t_3) + \phi_2(t_3,t_4)\phi_1(t_1)\phi_1(t_2)] \\
&+ N(N-1)(N-2)(N-3)\phi_1(t_1)\phi_1(t_2)\phi_1(t_3)\phi_1(t_4)
\end{aligned} \tag{2.26}$$

$$
\begin{aligned}
&\mathbb{E}[I(s)I(s+\tau)I(t)I(t+\tau)] \\
&= N\phi_4(s, s+\tau, t, t+\tau) \\
&+ N(N-1)[\phi_3(s, s+\tau, t)\phi_1(t+\tau) + \phi_3(s, s+\tau, t+\tau)\phi_1(t) \\
&\quad + \phi_3(s, t, t+\tau)\phi_1(s+\tau) + \phi_3(s+\tau, t, t+\tau)\phi_1(s)] \\
&+ N(N-1)[\phi_2(s, s+\tau)\phi_2(t, t+\tau) + \phi_2(s, t)\phi_2(s+\tau, t+\tau) + \phi_2(s, t+\tau)\phi_2(s+\tau, t)] \\
&+ N(N-1)(N-2)[\phi_2(s, s+\tau)\phi_1(t)\phi_1(t+\tau) + \phi_2(s, t)\phi_1(s+\tau)\phi_1(t+\tau) \\
&\quad + \phi_2(s, t+\tau)\phi_1(s+\tau)\phi_1(t) + \phi_2(s+\tau, t)\phi_1(s)\phi_1(t+\tau) \\
&\quad + \phi_2(s+\tau, t+\tau)\phi_1(s)\phi_1(t) + \phi_2(t, t+\tau)\phi_1(s)\phi_1(s+\tau)] \\
&+ N(N-1)(N-2)(N-3)\phi_1(s)\phi_1(s+\tau)\phi_1(t)\phi_1(t+\tau)
\end{aligned}
$$

Now, we study the case when $t - s > \tau$, we have

$$
\begin{aligned}
&\mathbb{E}[I(s)I(s+\tau)I(t)I(t+\tau)] \\
&= N\psi_4(\tau, t-s-\tau, \tau) \\
&+ N(N-1)[\psi_3(\tau, t-s-\tau)\psi_1 + \psi_3(\tau, t-s)\psi_1 + \psi_3(t-s, \tau)\psi_1 + \psi_3(t-s-\tau, \tau)\psi_1] \\
&+ N(N-1)[\psi_2(\tau)\psi_2(\tau) + \psi_2(t-s)\psi_2(t-s) + \psi_2(t+\tau-s)\psi_2(t-s-\tau)] \\
&+ N(N-1)(N-2)[\psi_2(\tau)\psi_1\psi_1 + \psi_2(t-s)\psi_1\psi_1 + \psi_2(t+\tau-s)\psi_1\psi_1 \\
&\quad + \psi_2(t-s-\tau)\psi_1\psi_1 + \psi_2(t-s)\psi_1\psi_1 + \psi_2(\tau)\psi_1\psi_1] \\
&+ N(N-1)(N-2)(N-3)\psi_1\psi_1\psi_1\psi_1.
\end{aligned}
\tag{2.27}
$$

For case 2 when $0 < t - s < \tau$,



$$
\begin{aligned}
&\mathbb{E}[I(s)I(s+\tau)I(t)I(t+\tau)] \\
&= N\psi_4(t-s,\tau-(t-s),t-s) \\
&+ N(N-1)[\psi_3(t-s,\tau-(t-s))\psi_1 + \psi_3(\tau,t-s)\psi_1 \\
&\quad + \psi_3(t-s,\tau)\psi_1 + \psi_3(\tau-(t-s),t-s)\psi_1] \\
&+ N(N-1)[\psi_2(\tau)\psi_2(\tau) + \psi_2(t-s)\psi_2(t-s) + \psi_2(t-s+\tau)\psi_2(\tau-(t-s))] \\
&+ N(N-1)(N-2)[\psi_2(\tau)\psi_1\psi_1 + \psi_2(t-s)\psi_1\psi_1 + \psi_2(t-s+\tau)\psi_1\psi_1 \\
&\quad + \phi_2(s+\tau,t)\phi_1(s)\phi_1(t+\tau) + \psi_2(t-s)\psi_1\psi_1 + \psi_2(\tau)\psi_1\psi_1] \\
&+ N(N-1)(N-2)(N-3)\psi_1\psi_1\psi_1\psi_1.
\end{aligned}
\tag{2.28}
$$

2.4.2 Asymptotic Analysis as $T \to \infty$

Recall that we define the effective number of particles in the focal volume

$$N_e = \frac{N}{|B|}\pi^{3/2}r^2 l$$

We multiply the the moments ψ_i contributed by a single particle by N and get Ψ_i $(i = 1, 2, 3, 4)$:

$$\Psi_1 = N\psi_1 = F_0\frac{N_e}{\sqrt{8}}. \tag{2.29}$$

$$\Psi_2(\tau) = N\psi_2(\tau) = F_0^2\frac{N_e}{\sqrt{8}(2+\frac{2\tau}{\tau_r})\sqrt{2+\frac{2\tau}{\tau_l}}}. \tag{2.30}$$

$$\begin{aligned}\Psi_3(\tau_1,\tau_2) &= N\psi_3(\tau_1,\tau_2)\\ &= F_0^3\frac{N_e}{\sqrt{8}(3+\frac{4(\tau_1+\tau_2)}{\tau_r}+\frac{4\tau_1\tau_2}{\tau_r^2})\sqrt{3+\frac{4(\tau_1+\tau_2)}{\tau_l}+\frac{4\tau_1\tau_2}{\tau_l^2}}}.\end{aligned} \tag{2.31}$$

$$\begin{aligned}\Psi_4(\tau_1,\tau_2,\tau_3) &= N\psi_4(\tau_1,\tau_2,\tau_3)\\ &= F_0^4\frac{N_e}{\sqrt{8}}\frac{1}{(4+2^{\frac{3\tau_1+4\tau_2+3\tau_3}{\tau_r}}+8\frac{\tau_1\tau_2+\tau_1\tau_3+\tau_2\tau_3}{\tau_r^2}+3\frac{\tau_1\tau_2\tau_3}{\tau_r^3})}\\ &\quad\frac{1}{\sqrt{4+2^{\frac{3\tau_1+4\tau_2+3\tau_3}{\tau_l}}+8\frac{\tau_1\tau_2+\tau_1\tau_3+\tau_2\tau_3}{\tau_l^2}+3\frac{\tau_1\tau_2\tau_3}{\tau_l^3}}}.\end{aligned} \tag{2.32}$$

With the new notation, we can rewrite

$$\begin{aligned}\mathbb{E}[C(T,\tau)] &= N\psi_2(\tau)+N(N-1)(\psi_1)^2 = N\psi_2(\tau)+N^2(\psi_1)^2-N(\psi_1)^2\\ &= \Psi_2(\tau)+(\Psi_1)^2+O(\frac{1}{N}).\end{aligned} \tag{2.33}$$

Since the number of particles N is large, the term $N(\psi_1)^2 = \frac{1}{N}\Psi_1^2$ is very small compared to the terms $N\psi_2(\tau)$ and $N^2(\psi_1)^2$, in which cases the order of N matches with the order of $\frac{\pi^{3/2}r^l}{|B|}$ in ψ. Similarly, for $\mathbb{E}[I(s)I(s+\tau)I(t)I(t+\tau)]$, we can see that some terms is small with the order of $O(\frac{1}{N})$.

Now, we try to estimate how fast $\mathrm{Var}[C(T,\tau)]$ decays in terms of the observation time T. That requires a more detailed look at the double time integral.

$$\mathbb{E}[C(T,\tau)^2] = \frac{1}{T^2}\int_0^T\int_0^T \mathbb{E}[I(t)I(t+\tau)I(s)I(s+\tau)]dtds. \tag{2.34}$$

Now we rewrite $\mathbb{E}[I(s)I(s+\tau)I(t)I(t+\tau)]$ for two cases $t-s>\tau$ and $0<t-s<\tau$.

When $t-s>\tau$,

$$
\begin{aligned}
&\mathbb{E}[I(s)I(s+\tau)I(t)I(t+\tau)] \\
&= \Psi_4(\tau, t-s-\tau, \tau) \\
&+ \Psi_1[\Psi_3(\tau, t-s-\tau) + \Psi_3(\tau, t-s) + \Psi_3(t-s, \tau) + \Psi_3(t-s-\tau, \tau)] \\
&+ \Psi_2(\tau)\Psi_2(\tau) + \Psi_2(t-s)\Psi_2(t-s) + \Psi_2(t+\tau-s)\Psi_2(t-s-\tau) \\
&+ \Psi_1\Psi_1[\Psi_2(\tau) + \Psi_2(t-s) + \Psi_2(t+\tau-s) + \Psi_2(t-s-\tau) + \Psi_2(t-s) + \Psi_2(\tau)] \\
&+ \Psi_1\Psi_1\Psi_1\Psi_1 + O(\frac{1}{N}).
\end{aligned}
$$

And when $0<t-s<\tau$, we have

$$
\begin{aligned}
&\mathbb{E}[I(s)I(s+\tau)I(t)I(t+\tau)] \\
&= \Psi_4(t-s, \tau-(t-s), t-s) \\
&+ \Psi_1[\Psi_3(t-s, \tau-(t-s)) + \Psi_3(\tau, t-s) + \Psi_3(t-s, \tau) + \Psi_3(\tau-(t-s), t-s)] \\
&+ \Psi_2(\tau)\Psi_2(\tau) + \Psi_2(t-s)\Psi_2(t-s) + \Psi_2(t-s+\tau)\Psi_2(\tau-(t-s))] \\
&+ \Psi_1\Psi_1[\Psi_2(\tau) + \Psi_2(t-s) + \Psi_2(t-s+\tau) + \Psi_2(\tau-(t-s)) + \Psi_2(t-s) + \Psi_2(\tau)] \\
&+ \Psi_1\Psi_1\Psi_1\Psi_1 + O(\frac{1}{N}).
\end{aligned}
$$

We notice that both cases have the same terms (colored in blue) which don't depend on $t-s$ and we denote them by g_0:

$$
g_0 = (\Psi_2(\tau))^2 + 2\Psi_2(\tau)(\Psi_1)^2 + (\Psi_1)^4 = (\Psi_2(\tau) + \Psi_1^2)^2. \tag{2.35}
$$

By equation (2.33), we see that

$$\frac{1}{T^2}\int_0^T\int_0^T g_0 dsdt = g_0 = (\mathbb{E}[C(T,\tau)])^2 + O(\frac{1}{N}). \tag{2.36}$$

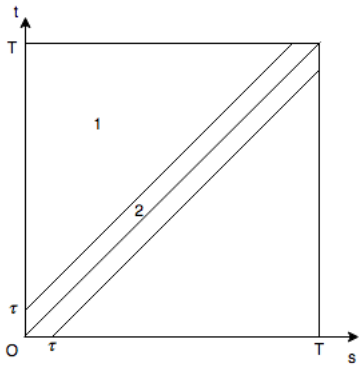
Next, we inspect the rest of the terms in $\mathbb{E}[I(s)I(s+\tau)I(t)I(t+\tau)]$, noting that its value just depend on the time difference $t-s$. So we perform a change of variable that $s=s$, $u=t-s$ and study its behavior on the $s-u$ plane.

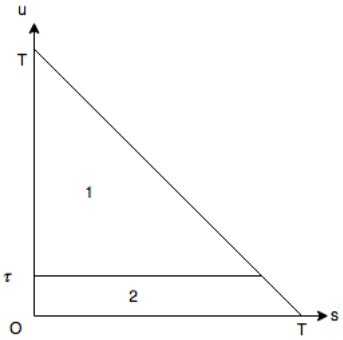
When $u>\tau$,

$$\begin{aligned} g(u) &:= \Psi_4(\tau, u-\tau, \tau) \\ &+ \Psi_1[\Psi_3(\tau,u-\tau)+\Psi_3(\tau,u)+\Psi_3(u,\tau)+\Psi_3(u-\tau,\tau)] \\ &+ \Psi_2(u)\Psi_2(u)+\Psi_2(u+\tau)\Psi_2(u-\tau) \\ &+ \Psi_1\Psi_1[\Psi_2(u)+\Psi_2(u+\tau)+\Psi_2(u-\tau)+\Psi_2(u)], \end{aligned} \tag{2.37}$$

when $0<u<\tau$,

$$\begin{aligned} g(u) &= \Psi_4(u, \tau-u, u) \\ &+ \Psi_1[\Psi_3(u,\tau-u)+\Psi_3(\tau,u)+\Psi_3(u,\tau)+\Psi_3(\tau-u,u)] \\ &+ \Psi_2(u)\Psi_2(u)+\Psi_2(u+\tau)\Psi_2(\tau-u) \\ &+ \Psi_1\Psi_1[\Psi_2(u)+\Psi_2(u+\tau)+\Psi_2(\tau-u)+\Psi_2(u)]. \end{aligned} \tag{2.38}$$





Note that

$$\begin{aligned}\mathbb{E}[C(T,\tau)^2] &= \frac{1}{T^2}\int_0^T\int_0^T \mathbb{E}[I(t)I(t+\tau)I(s)I(s+\tau)]dtds \\ &= g_0 + \frac{2}{T^2}\int_0^T\int_0^{T-u} g(u)dsdu.\end{aligned}$$

Since $g_0 = \big(\mathbb{E}[C(T,\tau)]\big)^2$,

$$\mathrm{Var}[C(T,\tau)] = \mathbb{E}[C(T,\tau)^2] - \big(\mathbb{E}[C(T,\tau)]\big)^2 = \frac{2}{T^2}\int_0^T\int_0^{T-u} g(u)dsdu. \tag{2.39}$$

Lemma 1.

$$2\int_0^T\int_0^{T-u} g(u)dsdu = O(T). \tag{2.40}$$

Proof. Since $g(u)$ has different expressions on region 1 and 2, we split the region.

$$\int_0^T\int_0^{T-u} g(u)dsdu = \int_0^\tau\int_0^{T-u} g(u)dsdu + \int_\tau^T\int_0^{T-u} g(u)dsdu.$$

In region 2, where $0 < u < \tau$

$$J_2 := \int_0^{\tau} \int_0^{T-u} g(u) ds du \leq T\tau \sup_{0 \leq u \leq \tau} g(u) = O(T).$$

In region 1, we inspect term by term and consider the asymptotics when $u \to \infty$

$$\Psi_4(\tau, t-s-\tau, \tau) = \frac{N_e}{\sqrt{8}} \frac{1}{(4 + 2\frac{4u+2\tau}{\tau_r} + 8\frac{2\tau(u-\tau)+\tau^2}{\tau_r^2} + 3\frac{\tau^2(u-\tau)}{\tau_r^3})}$$
$$\frac{1}{\sqrt{4 + 2\frac{4u+2\tau}{\tau_l} + 8\frac{2\tau(u-\tau)+\tau^2}{\tau_l^2} + 3\frac{\tau^2(u-\tau)}{\tau_l^3}}} = O(u^{-3/2}).$$

$$\Psi_1(2\Psi_3(\tau, u-\tau) + 2\Psi(\tau, u))$$
$$= \frac{2N_e}{\sqrt{8}(3 + \frac{4u}{\tau_r} + \frac{4\tau(u-\tau)}{\tau_r^2})\sqrt{3 + \frac{4u}{\tau_l} + \frac{4\tau(u-\tau)}{\tau_l^2}}}$$
$$+ \frac{2N_e}{\sqrt{8}(3 + \frac{4(u+\tau)}{\tau_r} + \frac{4\tau u}{\tau_r^2})\sqrt{3 + \frac{4(u+\tau)}{\tau_l} + \frac{4\tau u}{\tau_l^2}}} = O(u^{-3/2}).$$

$$\Psi_2(t-s)\Psi_2(t-s) + \Psi_2(t-s+\tau)\Psi_2(\tau-(t-s))$$
$$= \frac{N_e}{\sqrt{8}|B|(2 + \frac{2u}{\tau_r})\sqrt{2 + \frac{2(u+\tau)}{\tau_l}}} \frac{N_e}{\sqrt{8}|B|(2 + \frac{2(u-\tau)}{\tau_r})\sqrt{2 + \frac{2(u-\tau)}{\tau_l}}}$$
$$+ \left(\frac{N_e}{\sqrt{8}|B|(2 + \frac{2u}{\tau_r})\sqrt{2 + \frac{2u}{\tau_l}}}\right)^2 = O(u^{-3}).$$

$$
\begin{aligned}
&(\Psi_1)^2\big(2\Psi_2(u)+\Psi_2(u+\tau)+\Psi_2(u-\tau)\big)\\
&=\frac{N_e^2}{8}\Big(\frac{2N_e}{\sqrt{8}|B|(2+\frac{2u}{\tau_r})\sqrt{2+\frac{2u}{\tau_l}}}+\frac{N_e}{\sqrt{8}|B|(2+\frac{2(u+\tau)}{\tau_r})\sqrt{2+\frac{2(u+\tau)}{\tau_l}}}\\
&\quad+\frac{N_e}{\sqrt{8}|B|(2+\frac{2(u-\tau)}{\tau_r})\sqrt{2+\frac{2(u-\tau)}{\tau_l}}}\Big)=O(u^{-3/2}).
\end{aligned}
$$

Since $g(u)=O(u^{-3/2})$ as $u\to\infty$, that is, there exists $u_0>0$, $M>0$ such that $g(u)\le Mu^{-3/2}$ for all $u\ge u_0$. (Also note that u_0 and M are T independent). Without loss of generality, we choose $u_0>\tau$. Thus, we have

$$
\int_{u_0}^{T} g(u)du\le\int_{u_0}^{T} Mu^{-3/2}du=2M\Big(\frac{1}{\sqrt{u_0}}-\frac{1}{\sqrt{T}}\Big)=O(1).
$$

Hence,

$$
\int_{u_0}^{T} g(u)(T-u)du\le\int_{u_0}^{T} g(u)\ Tdu=O(T).
$$

Similar to the case in region 2, now we consider

$$
\int_{0}^{u_0} g(u)(T-u)du\le Tu_0\sup_{0\le u\le\ u_0} g(u)=O(T).
$$

Therefore,

$$
\int_0^T\int_0^{T-u} g(u)dsdu=O(T).
$$

□

It follows immediately from lemma 1 that

$$\mathrm{Var}[C(T,\tau)] = \frac{2}{T^2}\int_0^T \int_0^{T-u} g(u)dsdu = O\left(\frac{1}{T}\right). \tag{2.41}$$

2.4.3 An Upper Bound for $\mathrm{Var}[C(T,\tau)]$

In the previous section, we derive the exact expression of $\mathrm{Var}[C(T,\tau)]$ as a time integral in equation (2.39) and show that it decays in order $O(\frac{1}{T})$. In this section, we would like to look into the constant in front of $\frac{1}{T}$ and find an upper bound for $\mathrm{Var}[C(T,\tau)]$ in closed-form. Recall that

$$\mathrm{Var}[C(T,\tau)] = \frac{2}{T^2}\int_0^T \int_0^{T-u} g(u)dsdu \leq \frac{2}{T}\int_0^T g(u)du.$$

We shall find an upper bound for each of the terms in $g(u)$

$\boldsymbol{\Psi_4}$ Terms $\int_0^\tau \Psi_4(u, T-u, u)du + \int_\tau^T \Psi_4(\tau, u-\tau, \tau)du$

Recall that

$$\Psi_4(u,\tau-u,u)$$
$$= \Psi_1 \frac{1}{(4+2\frac{6u+4(\tau-u)+}{\tau_r}+8\frac{2u(\tau-u)+u^2}{\tau_r^2}+3\frac{u^2(\tau-u)}{\tau_r^3})} \frac{1}{\sqrt{4+2\frac{6u+4(\tau-u)}{\tau_l}+8\frac{2u(\tau-u)+u^2}{\tau_l^2}+3\frac{u^2(\tau-u)}{\tau_l^3}}},$$

$$\Psi_4(\tau,u-\tau,\tau)$$
$$= \Psi_1 \frac{1}{(4+2\frac{6\tau+4(u-\tau)+}{\tau_r}+8\frac{2\tau(u-\tau)+\tau^2}{\tau_r^2}+3\frac{\tau^2(u-\tau)}{\tau_r^3})} \frac{1}{(4+2\frac{6\tau+4(u-\tau)+}{\tau_l}+8\frac{2\tau(u-\tau)+\tau^2}{\tau_l^2}+3\frac{\tau^2(u-\tau)}{\tau_l^3})}.$$

In both cases, denominator $\geq \left(\frac{4}{\tau_l}u+4+\frac{4\tau}{\tau_l}\right)^{3/2}$.

Since

$$\int_{t_1}^{t_2} \frac{1}{(Au+B)^{3/2}} du = \frac{2}{A\sqrt{At_1+B}}\Big(1 - \sqrt{\frac{At_1+B}{At_2+B}}\Big).$$

Hence,

$$\begin{aligned}&\int_0^{\tau} \Psi_4(u, T-u, u)du + \int_{\tau}^{T} \Psi_4(\tau, u-\tau, \tau)\\ &\leq \Psi_1 \frac{2}{\frac{4}{\tau_l}\sqrt{4+\frac{4\tau}{\tau_l}}} = \Psi_1 \frac{\tau_l}{4\sqrt{1+\frac{\tau}{\tau_l}}}.\end{aligned}$$

$\Psi_1\Psi_3$ Terms

$$2\Psi_1\Big(\int_0^{T} \Psi_3(\tau, u)du + \int_0^{\tau} \Psi_3(u, \tau-u)du + \int_{\tau}^{T} \Psi_3(\tau, u-\tau)du\Big)$$

Recall that

$$\Psi_3(\tau, u) = \Psi_1 \frac{1}{(3+\frac{4(u+\tau)}{\tau_r}+\frac{4u\tau}{\tau_r^2})\sqrt{3+\frac{4(u+\tau)}{\tau_l}+\frac{4u\tau}{\tau_l^2}}}.$$

Let $A = \frac{4}{\tau_l} + \frac{4\tau}{\tau_l^2}$, and $B = 3 + \frac{4\tau}{\tau_l}$, we have

$$\begin{aligned}&2\Psi_1 \int_0^{T} \Psi_3(\tau, u) \leq 2(\Psi_1)^2 \int_0^{T} \frac{1}{(Au+B)^{3/2}} du\\ &\leq 2(\Psi_1)^2 \frac{2}{(\frac{4}{\tau_l}+\frac{4\tau}{\tau_l^2})\sqrt{3+\frac{4\tau}{\tau_l}}} = (\Psi_1)^2 \frac{\tau_l}{(1+\frac{\tau}{\tau_l})\sqrt{3+\frac{4\tau}{\tau_l}}}.\end{aligned}$$

Recall that

$$\Psi_3(u,\tau-u)=\Psi_1\frac{1}{(3+\frac{4\tau}{\tau_r}+\frac{4u(\tau-u)}{\tau_r^2})\sqrt{3+\frac{4\tau}{\tau_l}+\frac{4u(\tau-u)}{\tau_l^2}}},$$

$$2\Psi_1\int_0^\tau \Psi_3(u,\tau-u)\le 2\Psi_1\tau\sup_{u\in[0,\tau]}\Psi_3(u,\tau-u)=2(\Psi_1)^2\frac{\tau}{(3+\frac{4\tau}{\tau_r})\sqrt{3+\frac{4\tau}{\tau_l}}}.$$

Recall that

$$\Psi_3(\tau,u-\tau)=\Psi_1\frac{1}{(3+\frac{4u}{\tau_r}+\frac{4\tau(u-\tau)}{\tau_r^2})\sqrt{3+\frac{4u}{\tau_l}+\frac{4\tau(u-\tau)}{\tau_l^2}}}.$$

$$2\Psi_1\int_\tau^T \Psi_3(\tau,u-\tau)=2\Psi_1\int_0^{T-\tau}\Psi_3(\tau,v)dv\le(\Psi_1)^2\frac{\tau_l}{(1+\frac{\tau}{\tau_l})\sqrt{3+\frac{4\tau}{\tau_l}}}.$$

$\Psi_2\Psi_2$ Terms

$$\int_0^T\Psi_2(u)\Psi_2(u)du+\int_0^\tau\Psi_2(u+\tau)\Psi_2(\tau-u)du+\int_\tau^T\Psi_2(u+\tau)\Psi_2(u-\tau)du$$

$$\Psi_2(u)\Psi_2(u)=(\Psi_1)^2\frac{1}{(2+\frac{2u}{\tau_r})^2(2+\frac{2u}{\tau_l})}.$$

Since

$$\int_0^T\frac{1}{(Au+B)^3}du=\frac{1}{2AB^2}-\frac{1}{2A(AT+B)^2},$$

let $A = \frac{2}{\tau_l}$ and $B = 2$,

$$\int_0^T \Psi_2(u)\Psi_2(u)du \leq (\Psi_1)^2 \int_0^T \frac{1}{(Au+B)^3}du \leq \frac{(\Psi_1)^2}{2AB^2} = (\Psi_1)^2\frac{\tau_l}{16}.$$

$$\int_0^\tau \Psi_2(u+\tau)\Psi_2(\tau-u) \leq \tau \sup_{u\in[0,\tau]} \big\{\Psi_2(u+\tau)\Psi_2(\tau-u)\big\}$$
$$\leq \tau\Psi_2(\tau)\Psi_2(0) = (\Psi_1)^2\frac{\tau}{(2+\frac{2\tau}{\tau_r})\sqrt{2+\frac{2\tau}{\tau_l}}}.$$

$$\int_\tau^T \Psi_2(u+\tau)\Psi_2(u-\tau)du = \int_0^{T-\tau} \Psi_2(v+2\tau)\Psi_2(v)dv$$
$$\leq \int_0^{T-\tau} \Psi_2(v)\Psi_2(v)dv \leq (\Psi_1)^2\frac{\tau_l}{16}.$$

$\Psi_1\Psi_1\Psi_2$ Terms

$$\Psi_1\Psi_1\Big(\int_0^T 2\Psi_2(u)du + \int_0^T \Psi_2(u+\tau)du + \int_0^\tau \Psi_2(\tau-u)du + \int_\tau^T \Psi_2(u-\tau)du\Big)$$

Let $A = \frac{2}{\tau_l}$ and $B = 2$,

$$\Psi_1\Psi_1\int_0^T 2\Psi_2(u)du \leq (\Psi_1)^3\int_0^T \frac{1}{(Au+B)^{3/2}}du \leq (\Psi_1)^3\frac{2\tau_l}{\sqrt{2}}.$$

$$\Psi_1\Psi_1\int_0^T 2\Psi_2(u+\tau)du = \Psi_1\Psi_1\int_\tau^{T+\tau}\Psi_2(v)dv$$
$$\leq (\Psi_1)^3\int_\tau^{T+\tau}\frac{1}{(Au+B)^{3/2}}du \leq (\Psi_1)^3\frac{\tau_l}{\sqrt{2+\frac{2\tau}{\tau_l}}}.$$

$$\Psi_1\Psi_1\int_0^\tau 2\Psi_2(\tau-u)du = \Psi_1\Psi_1\int_0^\tau\Psi_2(v)dv$$
$$\leq (\Psi_1)^3\int_0^\tau\frac{1}{(Au+B)^{3/2}}du = (\Psi_1)^3\tau_l\Big(\frac{1}{\sqrt{2}}-\frac{1}{\sqrt{2+\frac{2\tau}{\tau_l}}}\Big).$$

$$\Psi_1\Psi_1\int_\tau^T 2\Psi_2(u-\tau)du = \Psi_1\Psi_1\int_0^{T-\tau}\Psi_2(v)dv \leq (\Psi_1)^3\frac{\tau_l}{\sqrt{2}}.$$

Putting them all together,

$$\begin{aligned}\mathrm{Var}[C(T,\tau)] &\leq \frac{2}{T}\int_0^T g(u)du \\ &\leq \frac{2}{T}\Bigg(\Psi_1\frac{\tau_l}{4\sqrt{1+\frac{\tau}{\tau_l}}}+(\Psi_1)^2\frac{2\tau_l}{(1+\frac{\tau}{\tau_l})\sqrt{3+\frac{4\tau}{\tau_l}}}+2(\Psi_1)^2\frac{\tau}{(3+\frac{4\tau}{\tau_r})\sqrt{3+\frac{4\tau}{\tau_l}}} \\ &\quad +(\Psi_1)^2\frac{\tau_l}{8}+(\Psi_1)^2\frac{\tau}{(2+\frac{2\tau}{\tau_r})\sqrt{2+\frac{2\tau}{\tau_l}}}+(\Psi_1)^3\frac{4\tau_l}{\sqrt{2}}\Bigg).\end{aligned} \tag{2.42}$$

2.4.4 Numerical Results on $\mathrm{Var}[C(T,\tau)]$

In section 2.4.2, we manage to quantify $\mathrm{Var}[C(T,\tau)]$ analytically by a double integral (2.39) as $N\to\infty$. In this section we would like to integrate it numerically

and compare with the variance estimated by Monte Carlo simulation.

We show the plots of **relative error** against T for different τ values.

$$RE(T,\tau) := \frac{\sqrt{\mathrm{Var}(C(T,\tau))}}{\mathbb{E}[C(T,\tau)]} = \sqrt{\frac{\mathbb{E}[C(T,\tau)^2]}{(\mathbb{E}[C(T,\tau)])^2} - 1}\ . \tag{2.43}$$

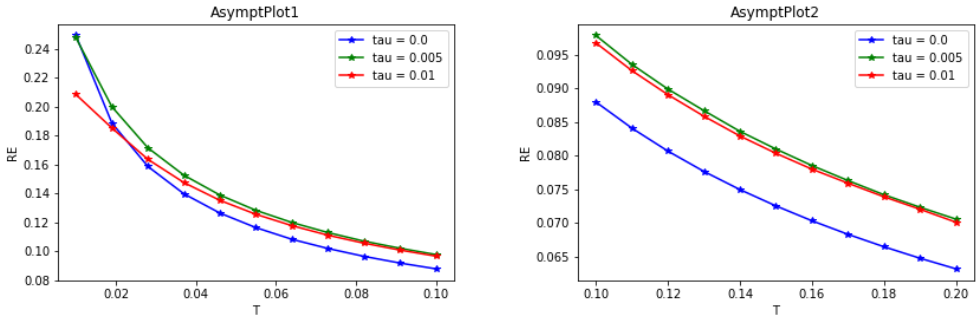


Figure 2.1: $RE(T,\tau)$ against T for three fixed values of $\tau = 0,\ 0.005,\ 0.1$. $RE(T,\tau)$ is computed by numerical integration.

In order to study the rate of decay α, we assume

$$RE(T,\tau) = b\ T^{\alpha}.$$

That is,

$$\log(RE) = \alpha \log(T) + \log(b).$$

The slope of the loglog plot (not shown) for adjacent data point is near -0.5, that is, the relative error decays at order $O(T^{-1/2})$, which agrees with the theoretical result that variance decays at order $O(\frac{1}{T})$.

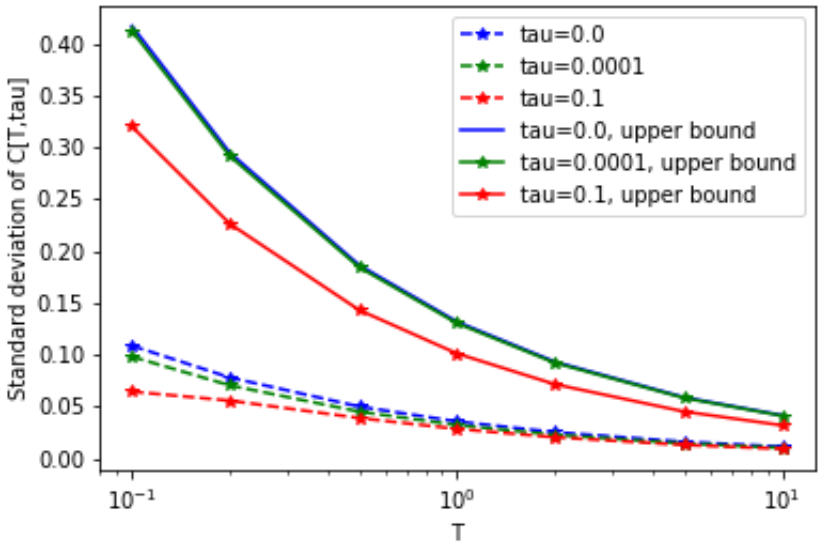


Figure 2.2: Standard deviation of $C(T, \tau)$ for fixed τ values $\tau = 0,\ 0.005,\ 0.1$. Upper bound and numerical estimate.

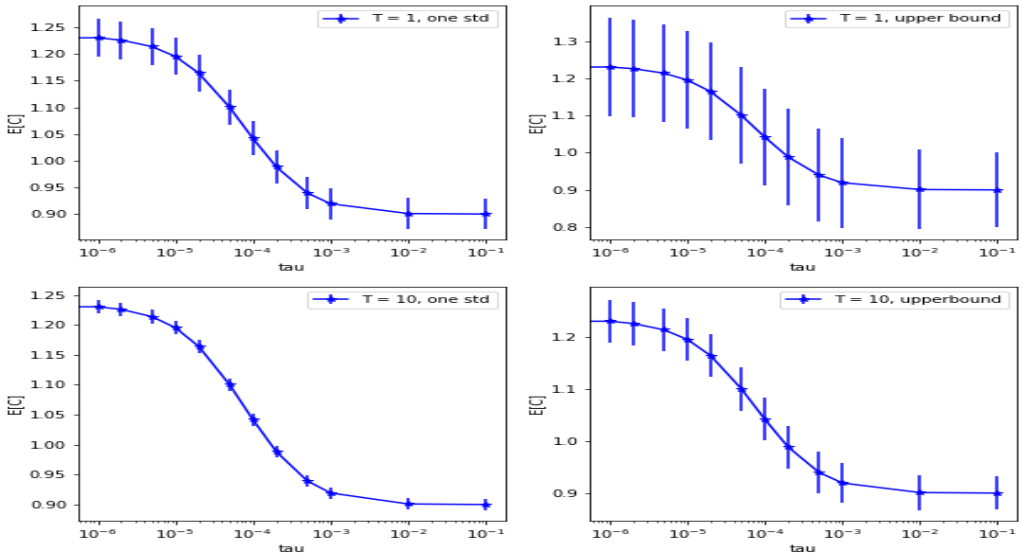


Figure 2.3: $\mathbb{E}[C(T, \tau)]$ as a function of τ with error bars showing the standard deviation. Left column shows the actual standard deviation computed by numerical integration, while right column shows the upper bound of standard deviation. The first row is for $T = 1$ and the second row for $T = 10$.

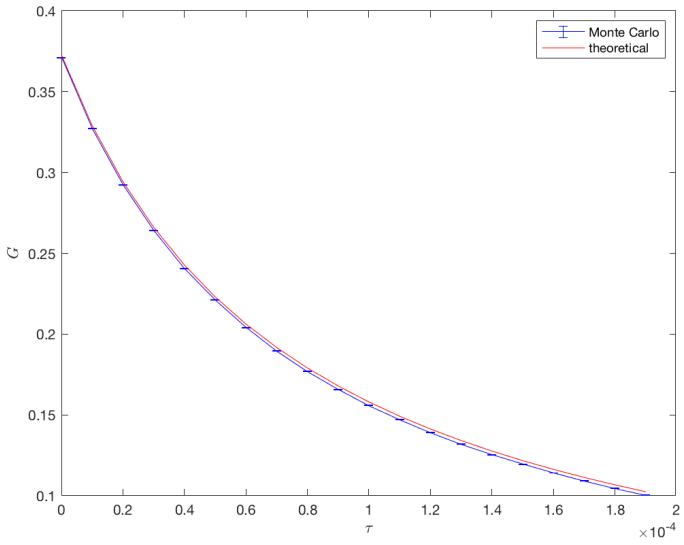


Figure 2.4: $G(\tau)$ simulated and theoretical. $\mathbb{E}[\hat{G}(\tau)]$ estimated from 10000 trajectories. $T = 0.1$.

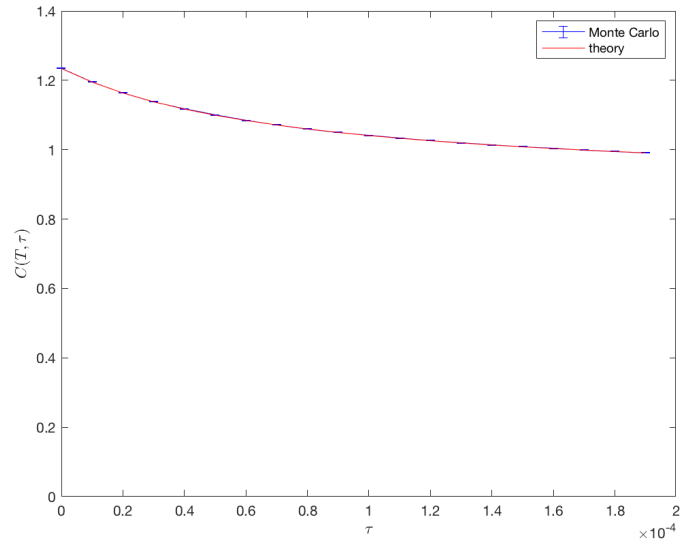


Figure 2.5: $C(\tau)$ simulated and theoretical. Theory: $\mathbb{E}[I(t)I(t+\tau)]$. Monte Carlo: $\mathbb{E}[C(\tau)] = \mathbb{E}\left[\frac{1}{T}\int_0^T I(t)I(t+\tau)dt\right]$ estimated from 10000 trajectories. $T = 0.1$.

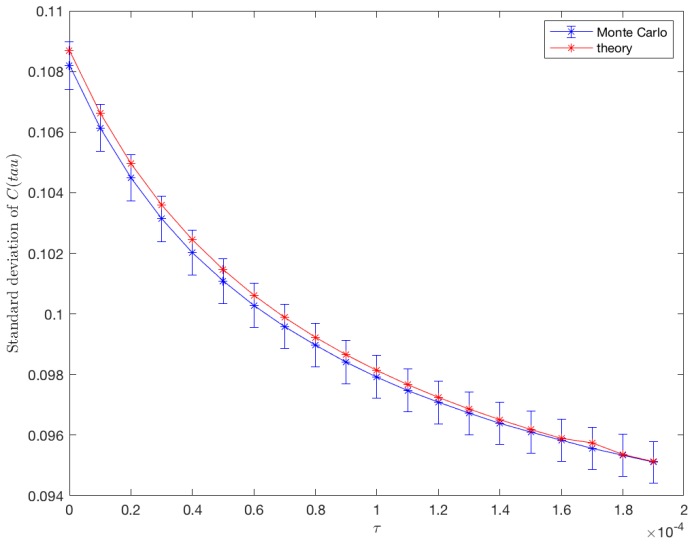


Figure 2.6: Standard deviation of $C(\tau)$. $T = 0.1$.

2.5 Fitting D and N_e

When we estimate diffusivity D by a nonlinear least squares fit, there are several slightly different options. Traditionally, people fit the time-averaged observation $\hat{G}(\tau)$ to the model $G(\tau)$ using equation (2.11)

$$G(\tau) = \frac{1}{N_e} \frac{1}{(1 + \frac{4D\tau}{r^2})} \frac{1}{\sqrt{1 + \frac{4D\tau}{l^2}}}.$$

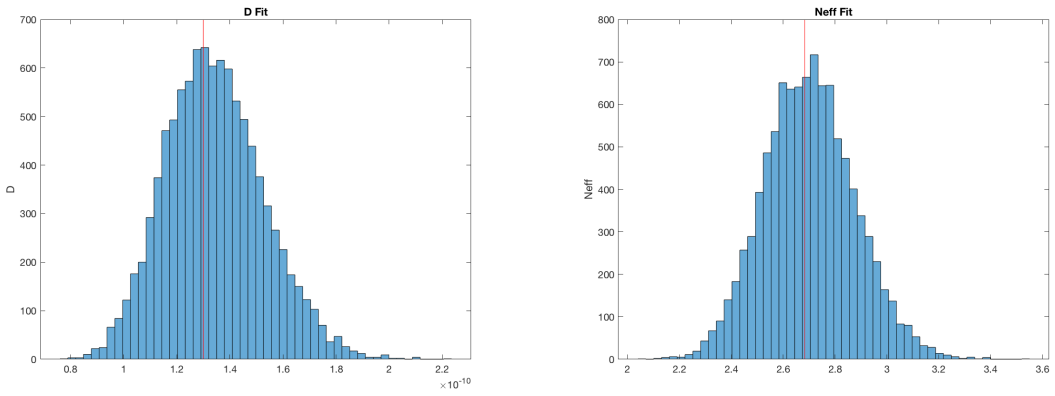


Figure 2.7: Histograms of fitted D and N_e. Single trajectory fit. $T = 0.1$. 10000 independent simulations.

Scientist usually do the repeated experiments and get several trajectories of intensity data. One idea is to simply average the $\hat{G}(\tau)$ obtained from each trajectory and then fit with the analytical formula. In FCS experiment, intensity data is recorded in discrete time, so the time integral is further approximated by the Riemann sum. That is,

$$\hat{G}^{(k)}(\tau) = \frac{\frac{1}{n}\sum_{j=1}^{n} I^{(k)}(t_j)I^{(k)}(t_j+\tau)}{\left(\frac{1}{n}\sum_{j=1}^{n} I^{(k)}(t_j)\right)^2} - 1. \qquad \hat{G}_{avg}(\tau) = \frac{1}{m}\sum_{k=1}^{m} \hat{G}^{(k)}(\tau).$$

Another idea is to average the numerator $\frac{1}{T}\int_0^T I(t)I(t+\tau)dt$ and the denominator

$\frac{1}{T}\int_0^T I(t)dt$ separately. That is,

$$\frac{1}{T}\int_0^T I(t)I(t+\tau)dt \approx \frac{1}{m}\sum_{k=1}^{m}\frac{1}{n}\sum_{j=1}^{n} I^{(k)}(t_j)I^{(k)}(t_j+\tau) =: \hat{\mathrm{C}}\mathrm{orr}(\tau).$$

$$\frac{1}{T}\int_0^T I(t)dt \approx \frac{1}{m}\sum_{k=1}^{m}\frac{1}{n}\sum_{j=1}^{n} I^{(k)}(t_j) =: \hat{I}.$$

$$\hat{G}_{avg}(\tau) = \frac{\frac{1}{m}\sum_{k=1}^{m}\frac{1}{n}\sum_{j=1}^{n} I^{(k)}(t_j)I^{(k)}(t_j+\tau)}{\left(\frac{1}{m}\sum_{k=1}^{m}\frac{1}{n}\sum_{j=1}^{n} I^{(k)}(t_j)\right)^2} - 1.$$

Alternatively, it is also possible to fit the observed $\frac{1}{T}\int_0^T I(t)I(t+\tau)$ to the analytical expression $\mathrm{Corr}(I(t), I(t+\tau))$ given by equation (2.18).

$$\hat{\mathrm{C}}\mathrm{orr}(\tau) = \frac{1}{m}\sum_{k=1}^{m}\frac{1}{n}\sum_{j=1}^{n} I^{(k)}(t_j)I^{(k)}(t_j+\tau).$$

Or fit the observed $\frac{1}{T}\int_0^T I(t)I(t+\tau) - (\frac{1}{T}\int_0^T I(t)dt)^2$ to the analytical formula $\mathrm{Cov}(I(t), I(t+\tau))$ given by equation (2.19), which actually yields the same result as average the numerator and denominator of $\hat{G}(\tau)$ separately.

$$\hat{\mathrm{C}}\mathrm{ov}(\tau) = \frac{1}{m}\sum_{k=1}^{m}\frac{1}{n}\sum_{j=1}^{n} I^{(k)}(t_j)I^{(k)}(t_j+\tau) - \left(\frac{1}{m}\sum_{k=1}^{m}\frac{1}{n}\sum_{j=1}^{n} I^{(k)}(t_j)\right)^2.$$

Since the process $I(t)$ is not well-approximate by a Gaussian process (to be explained in Remark 1), its first two moments don't capture all the information, so it makes sense to use its higher moments. We choose to look at the $\mathbb{E}[I(t)^2 I(t+\tau)]$,

from equation (2.20), we have

$$\begin{aligned}&\text{Corr}\,3(\tau) := \mathbb{E}[I(t)^2 I(t+\tau)] \\ &\approx F_0^3 \left(\frac{N_e}{\sqrt{8}} \frac{1}{(3+\frac{16D\tau}{r^2})\sqrt{3+\frac{16D\tau}{l^2}}} + \frac{N_e^3}{8\sqrt{8}} + \frac{N_e^2}{8\sqrt{8}} \frac{2}{1+\frac{4D\tau}{r^2})\sqrt{1+\frac{4D\tau}{l^2}}} \right).\end{aligned} \tag{2.44}$$

$$\hat{\text{C}}\text{orr}3^{(k)}(\tau) = \frac{1}{n}\sum_{j=1}^{n} I(t_j)^2 I(t_j+\tau) \qquad \hat{\text{C}}\text{orr}3_{avg}(\tau) = \frac{1}{m}\sum_{k=1}^{m} \hat{C}\text{orr}3^{(k)}(\tau)(\tau).$$

We ran numerical experiments to fit diffusivity D with various options. For all these methods, we used the same set of $I(t)$ trajectories generated by Monte Carlo simulation. Parameters are chosen from a real scientific example in [42], $D = 1.3 \times 10^{-10}$, $r = 2 \times 10^{-7}$, $l = 1 \times 10^{-6}$, $F_0 = 1$, $\Delta t = 1 \times 10^{-6}$,$T = 0.1$. We performed nonlinear least squre fit using Matlab program `lsqnonlin` and record the error $|\hat{D} - D|$. Results are shown in Table 2.5.

All method shows decent accuracy in terms of the relative error. As expected, using multiple trajectories to compute $\hat{G}_{avg}$, $\hat{\text{C}}\text{orr}_{avg}$, $\hat{\text{C}}\text{ov}_{avg}$, $\hat{\text{C}}\text{orr}3_{avg}$ yields better estimation than using only a single trajectory. $\hat{\text{C}}\text{orr}_{avg}$ has the best performance in our experiments.

Remark 1. The process $I(t)$ is not well-approximate by a Gaussian process.

Recall the central limit theorem that let X_i, $i = 1, 2, ..., n$ be identically distributed random variables with $\mathbb{E}[X] = \mu$ and $\text{Var}(X) = \sigma^2$. Then $Z_n := \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{n}\sigma} \to \mathcal{N}(0,1)$ in distribution. Berry-Essen Theorem further gives a quanti-

T	Method	Mean Error	RE (%)	Mean std
0.1	single trajectory fit G	4.23523e-12	3.26658	1.85854e-11
0.1	single trajectory fit Cov	4.23529e-12	3.26658	1.85854e-11
0.1	single trajectory fit Corr	1.12131e-12	0.862544	1.57308e-11
0.1	single trajectory fit three-point Corr	3.04654e-12	2.3435	1.86964e-11
0.1	5 trajectories average G directly	2.81276e-12	2.16393	8.20962e-12
0.1	5 trajectories average num and deno of G	8.07951e-13	0.621192	8.13551e-12
0.1	5 trajectories fit Cov	8.07951e-13	0.621192	8.13551e-12
0.1	5 trajectories fit Corr	4.18006e-13	0.321543	7.17572e-12
0.1	5 trajectories fit three-point Corr	2.02914e-12	1.56088	8.54895e-12

Table 2.1: Different methods to to fit D. Actual $D = 1.3 \times 10^{-10}$. All methods use the same set of $10,000$ independent trajectories. $T = 0.1$.

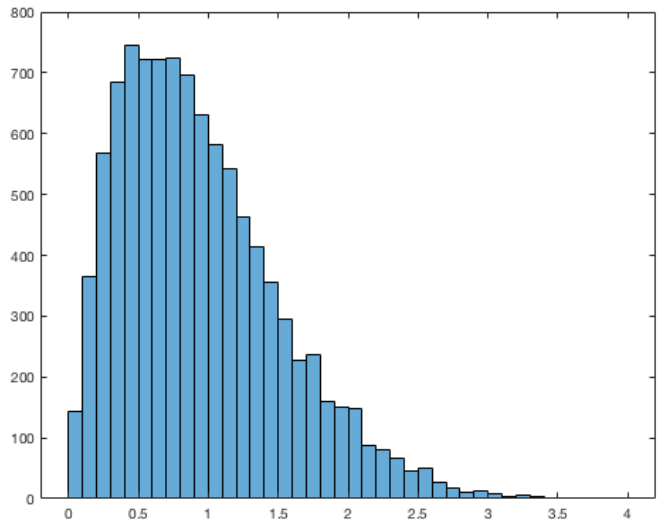


Figure 2.8: Histogram of $I(t_0)$ draw from 10000 independent trajectories

tative estimation of the convergence rate:

$$|F_n^X(z) - \Psi(z)| \leq \frac{C\rho}{\sigma^3\sqrt{n}},$$

where $F_n^X(z)$ is the cumulative density function (c.d.f) of Z_n, $\Psi(z)$ is the c.d.f of standard normal distribution $\mathcal{N}(0,1)$, ρ is the third moment $\mathbb{E}[X^3]$.

For a fixed n, we define $Y_n = \sum_{i=1}^n X_i$, and define the transformation $g_n(z) = \sqrt{n}\sigma z + n\mu$, let $Y \sim \mathcal{N}(n\mu, n\sigma^2)$, then we have

$$|F_n^Y(y) - F^Y(y)| = |F_n^X(z) - \Psi(z)| \leq \frac{C\rho}{\sigma^3\sqrt{n}}.$$

The constant C is shown to be $0.40975 \leq C \leq 0.4748$. For FCS setup, we have

$$\mathbb{E}[i(t)] = F_0\frac{\pi^{3/2}r^2l}{\sqrt{8}|B|}, \qquad \mathbb{E}[i(t)^2] = F_0^2\ \frac{\pi^{3/2}r^2l}{8|B|}, \qquad \rho = \mathbb{E}[i(t)^3] = F_0^3\ \frac{\pi^{3/2}r^2l}{\sqrt{8}|B|3\sqrt{3}}.$$

Since $\pi^{3/2}r^2l \ll |B|$, hence $(\mathbb{E}[i(t)])^2 \ll \mathbb{E}[i(t)^2]$, $\sigma = \sqrt{\mathrm{Var}(i(t)} = \sqrt{\mathbb{E}[i(t)^2] - (\mathbb{E}[i(t)])^2} \approx \sqrt{\mathbb{E}[i(t)^2]}$. Thus,

$$\frac{C\rho}{\sigma^3\sqrt{N}} \approx \frac{C}{\sqrt{N}}\frac{\frac{\pi^{3/2}r^2l}{\sqrt{8}|B|3\sqrt{3}}}{\left(\frac{\pi^{3/2}r^2l}{8|B|}\right)^{3/2}}\frac{N}{N} = C\frac{\frac{N_e}{\sqrt{8}\ 3\sqrt{3}}}{\left(\frac{N_e}{8}\right)^{3/2}} = \frac{8C}{3\sqrt{3}\sqrt{N_e}}.$$

That is, for any time t, the deviation of the distribution of $I(t)$ from Gaussian is of the order $O(\frac{1}{N_e})$ rather than $O(\frac{1}{N})$. Therefore, it is not appropriate to approximate $I(t)$ by a Gaussian process.

Remark 2. Fitting covariance gives the same $\hat{D}$ as fitting G when averaging $\frac{1}{T}\int_0^T I(t)I(t+\tau)dt$ and $\frac{1}{T}\int_0^T I(t)dt$ separately.

Proof. We have seen that

$$G(\tau) = \frac{1}{N_e}\frac{1}{(1+\frac{4D\tau}{r^2})}\frac{1}{\sqrt{1+\frac{4D\tau}{l^2}}}, \qquad \mathrm{Cov}(\tau) = \frac{F_0^2 N_e}{8}\frac{1}{(1+\frac{4D\tau}{r^2})\sqrt{1+\frac{4D\tau}{l^2}}}.$$

In Cov function, we replace $\frac{F_0^2 N_e}{8}$ by C and formulate the models as

$$G(\tau_i; D, N_e) = \frac{1}{N_e}\frac{1}{(1+\frac{4D\tau}{r^2})}\frac{1}{\sqrt{1+\frac{4D\tau}{l^2}}}, \qquad \mathrm{Cov}(\tau_i; D, C) = \frac{C}{(1+\frac{4D\tau}{r^2})\sqrt{1+\frac{4D\tau}{l^2}}}.$$

To Perform nonlinear least squares fit for G and Cov, we solve minimization problems respectively

$$\begin{aligned}
(\hat{D}_G, \hat{N}_{e,G}) &= \underset{(D,N_e)}{\arg\min}\, J_G(D, N_e), \quad \text{where } J_G(D, N_e) = \sum_{i=1}^{n_\tau}\left|\hat{G}(\tau_i) - G(\tau_i; D, N_e)\right|^2 \\
(\hat{D}_{cov}, \hat{C}_{cov}) &= \underset{(D,C)}{\arg\min}\, J_C(D, C), \quad \text{where } J_C(D, C) = \sum_{i=1}^{n_\tau}\left|\hat{C}\mathrm{ov}(\tau_i) - \mathrm{Cov}(\tau_i; D, C)\right|^2
\end{aligned} \tag{2.45}$$

For data, recall that

$$\hat{G}_{avg}(\tau) = \frac{\frac{1}{m}\sum_{k=1}^{n}\frac{1}{n}\sum_{j=1}^{n} I^{(k)}(t_j)I^{(k)}(t_j+\tau)}{\left(\frac{1}{m}\sum_{k=1}^{n}\frac{1}{n}\sum_{j=1}^{n} I^{(k)}(t_j)\right)^2} - 1 = \frac{\hat{\mathrm{C}}\mathrm{ov}(\tau)}{(\hat{I})^2}.$$

That is, we have $\hat{G}(\tau_i) = (\hat{I})^2\ \hat{\mathrm{C}}\mathrm{ov}(\tau_i)$ for all τ_i, $i = 1, ..., n_\tau$ with $\hat{I}$ not depending on τ.

Also, let $C = \frac{(\hat{I})^2}{N_e}$, then we have $G(\tau_i; D, N_e) = (\hat{I})^2 \mathrm{Cov}(\tau_i; D, C)$.

Thus we have, $\hat{D}_G = \hat{D}_{cov}$, $\hat{C}\mathrm{ov} = \frac{(\hat{I})^2}{\hat{N}_e}$. □

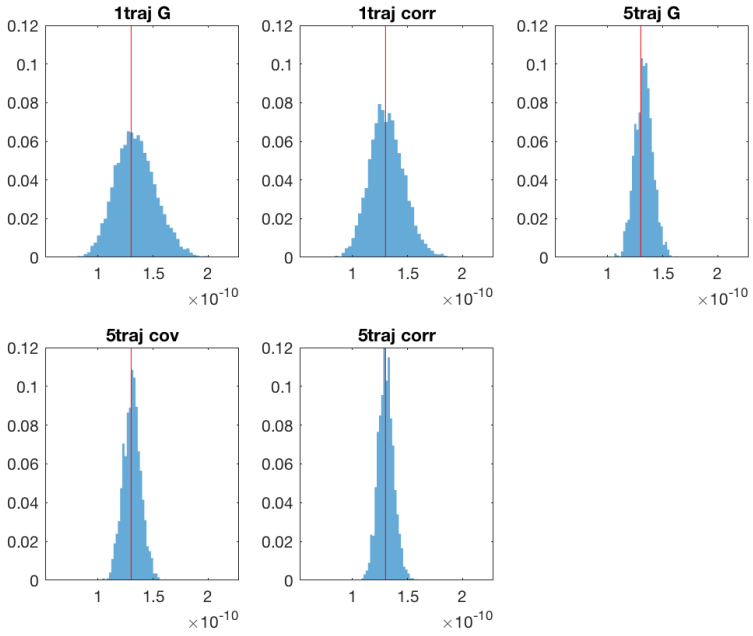


Figure 2.9: Histograms of fitted D obtained by different methods. Actual $D = 1.3 \times 10^{-10}$. All methods use the same set of 10000 independent simulated trajectories. $T = 0.1$

2.6 Sensitivity of diffusivity D

We would like to address the nonlinear least squares fit problem that arises in FCS setup. We denote the parameters to be fitted by $\theta = (D, N_e, F_0)$, and we solve the minimization problem

$$\hat{\theta} = \arg\min_{\theta > 0} J(\theta), \qquad J(\theta) = \int_0^{\tau_{max}} |\mathrm{Corr}(\tau; \theta) - \hat{\mathrm{Corr}}(\tau)|^2 d\tau, \tag{2.46}$$

where the model $\mathrm{Corr}(\tau;\theta)$ is given by

$$\mathrm{Corr}(\tau; D, N_e, F_0) = \frac{F_0^2 N_e}{8} \frac{1}{(1+\frac{4D\tau}{r^2})\sqrt{1+\frac{4D\tau}{l^2}}} + \frac{F_0^2 N_e^2}{8}, \tag{2.47}$$

and $\hat{\mathrm{Corr}}(\tau)$ is an observed trajectory of the correlation of the intensity data

$$\hat{\mathrm{Corr}}(\tau) = \frac{1}{T}\int_0^T I(t)I(t+\tau)\,dt. \tag{2.48}$$

Let $\bar{\theta} = (\bar{D}, \bar{N}_e, \bar{F}_0)$ be the true parameter that governs the system, we have $\mathbb{E}[\hat{\mathrm{Corr}}(\tau)] = \mathrm{Corr}(\tau, \bar{\theta})$. Further, each realization of $\hat{\mathrm{Corr}}(\tau)$ could be written as $\mathrm{Corr}(\tau;\bar{\theta})$ plus a random perturbation $\eta(\tau,\omega)$

$$\hat{\mathrm{Corr}}(\tau,\omega) = \mathrm{Corr}(\tau;\bar{\theta}) + \eta(\tau,\omega). \tag{2.49}$$

We shall justify that the minimizer $\hat{\theta}$ is well-defined in a local neighborhood of the true parameter $\bar{\theta}$ when the perturbation $\eta(t,\omega)$ is small.

Proposition 1. When there is no perturbation (i.e. $\hat{\mathrm{Corr}}(\tau) = \mathrm{Corr}(\tau)$), then $\hat{\theta} = \bar{\theta}$ is the unique minimizer of nonlinear least squares fit problem (2.46) with $J(\hat{\theta}) = 0$.

Proof. Suppose there is $\tilde{\theta} = (\tilde{D}, \tilde{N}_e, \tilde{F}_0)$ so that $J(\tilde{\theta}) = 0$, we shall show that $\tilde{\theta} = \bar{\theta}$. Since $J(\theta) = 0$ if and only if $\mathrm{Corr}(\tau;\theta) = \hat{\mathrm{Corr}}(\tau)$ for all $\tau \in [0, \tau_{max}]$, all we need to show is that $\mathrm{Corr}(\tau;\theta) = \mathrm{Corr}(\tau;\tilde{\theta})$ for all τ implies $\theta = \tilde{\theta}$.

Let $g(t) := \frac{1}{(1+\frac{4t}{r^2})\sqrt{1+\frac{4t}{l^2}}}$, we could rewrite

$$\mathrm{Corr}(\tau; D, N_e, F_0) = \frac{F_0^2 N_e}{8} g(D\tau) + \frac{F_0^2 N_e^2}{8}.$$

When we take the first and second order partial derivatives with respect to τ for $\mathrm{Corr}(\tau;\theta) = \mathrm{Corr}(\tau;\tilde{\theta})$, we have

$$\frac{F_0^2 N_e}{8} g'(D\tau)D = \frac{\tilde{F}_0^2 \tilde{N}_e}{8} g'(\tilde{D}\tau)\tilde{D}, \qquad \frac{F_0^2 N_e}{8} g''(D\tau)D^2 = \frac{\tilde{F}_0^2 \tilde{N}_e}{8} g''(\tilde{D}\tau)\tilde{D}^2.$$

In particular, when $\tau = 0$

$$\frac{F_0^2 N_e}{8} g'(0)D = \frac{\tilde{F}_0^2 \tilde{N}_e}{8} g'(0)\tilde{D}, \qquad \frac{F_0^2 N_e}{8} g''(0)D^2 = \frac{\tilde{F}_0^2 \tilde{N}_e}{8} g''(0)\tilde{D}^2.$$

Since $g'(0) \neq 0$ and $g''(0) \neq 0$, we have that $D = \tilde{D}$, $F_0^2 N_e = \tilde{F}_0^2 \tilde{N}_e$.

If we evaluate $\mathrm{Corr}(\tau;\theta) = \mathrm{Corr}(\tau;\tilde{\theta})$ at $\tau = 0$, we get

$$\frac{F_0^2 N_e}{8} + \frac{F_0^2 N_e^2}{8} = \frac{\tilde{F}_0^2 \tilde{N}_e}{8} + \frac{\tilde{F}_0^2 \tilde{N}_e^2}{8}, \text{ that is, } \frac{F_0^2 N_e}{8}(1 + N_e) = \frac{\tilde{F}_0^2 \tilde{N}_e}{8}(1 + \tilde{N}_e).$$

Hence, $N_e = \tilde{N}_e$, $F_0 = \tilde{F}_0$. □

Now that we've shown $\hat{\theta} = \bar{\theta}$ is the unique minimizer when there is no perturbation, we shall use the implicit function theorem to show that $\hat{\theta}$ is well-defined locally near $\bar{\theta}$ when a small perturbation η is in place. Note that the first order optimality condition $DJ(\hat{\theta}) = 0$ establishes a system of equations of the minimizer $\hat{\theta}$ and perturbation η:

$$\frac{\partial}{\partial\theta_1}J(\theta)=0,\quad \frac{\partial}{\partial\theta_2}J(\theta)=0,\quad \frac{\partial}{\partial\theta_3}J(\theta)=0. \tag{2.50}$$

Note that the cost function $J=\int_0^{\tau_{max}}|\mathrm{Corr}(\tau;\theta)-\hat{\mathrm{Corr}}(\tau)|^2d\tau$ involves perturbation $\eta(\tau,\omega)=\hat{\mathrm{Corr}}(\tau)-\mathrm{Corr}(\tau;\bar{\theta})$, which is a random process of τ. Although there are versions of the implicit function theorem in functional space, we conduct our sensitivity analysis where η is reduced to a random vector. Karhunen-Loeve expansion decomposes a stochastic process into a collection of deterministic orthogonal functions with the random coefficients. Using the first few terms of the truncated Karhunen-Loeve expansion, we characterize the random process η by its coefficients, which are time independent. This enables us to use the classical implicit function theorem in a finite dimensional space.

We state two important tools we use: the implicit function theorem and the Karhunen-Loeve expansion.

Theorem 2. *(Implicit Function Theorem) Let $F:\mathbb{R}^{n+m}\to\mathbb{R}^m$ be a continuously differentiable function. We write points in $\mathbb{R}^{m+n}$ in the form (x,y), where $x=(x_1,...,x_n)\in\mathbb{R}^n$ and $y=(y_1,...,y_m)\in\mathbb{R}^m$ and denote*

$$D_xF=\begin{bmatrix}\frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n}\end{bmatrix},\quad D_yF=\begin{bmatrix}\frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m}\\ \vdots & \ddots & \vdots\\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m}\end{bmatrix}.$$

Fix a point (a,b) with $F(a,b)=0$. If $D_yF(a,b)$ is invertible, then there exists an open set $U\subset\mathbb{R}^n$ containing point a such that there exists a unique continuous

differentiable function $g : U \to \mathbb{R}^m$ *such that* $g(a) = b$ *and* $F(x, g(x)) = 0$ *for all* $x \in U$*. Moreover, the partial derivative of* g *in* U *are given by*

$$D_y F|_{(x,g(x))}\ Dg(x) = -D_x F|_{(x,g(x))}.$$

$Dg(a)$ is particularly helpful to approximate the change in y under a small perturbation in x:

$$g(a + h) \approx g(a) + Dg(a)h \quad \text{for } h \text{ small.}$$

Theorem 3. *(Karhunen-Loeve Theorem [29, 34]) Let* $X : [0, t_{max}] \times \Omega \to \mathbb{R}$ *be a centered mean-square continuous stochastic process with* $X \in L^2([0, t_{max}] \times \Omega)$*. There exists an orthonormal basis* $\{e_i\}$ *of* $L^2([0, t_{max}])$ *such that* $\forall t \in [0, t_{max}]$*,*

$$X(t, \omega) = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\ \xi_i(\omega) e_i(t) \qquad \text{in } L^2(\Omega), \tag{2.51}$$

where ξ_i *are centered mutually uncorrelated random variables with unit variance and are given by*

$$\xi_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_0^{t_{max}} X(t, \omega) e_i(t) dt.$$

Hence, given a realization of the perturbation η, we could truncate it by its the first term of the Karhunen-Loeve expansion and characterize it by a single parameter x_1, that is,

$$\eta(\tau, \omega) = \tilde{\xi}_1(\omega, \bar{N}_e, \bar{F}_0)\ e_1(\bar{D}\tau) = x_1 e_1(\bar{D}\tau). \tag{2.52}$$

Recall the optimality condition (2.50),

$$\frac{\partial}{\partial\theta_1}J(\theta)=0,\quad \frac{\partial}{\partial\theta_2}J(\theta)=0,\quad \frac{\partial}{\partial\theta_3}J(\theta)=0.$$

where $J=\int_0^{\tau_{max}}|\mathrm{Corr}(\tau;\theta)-\hat{\mathrm{Corr}}(\tau)|^2d\tau=\int_0^{\tau_{max}}|\mathrm{Corr}(\tau;\theta)-\mathrm{Corr}(\tau;\bar{\theta})-\eta(\tau,\omega)|^2d\tau=\int_0^{\tau_{max}}|\mathrm{Corr}(\tau;\theta)-\mathrm{Corr}(\tau;\bar{\theta})-x_1e_1(\bar{D}\tau)|^2d\tau$.

The optimality condition as in terms of θ and x_1 could be viewed as $F(x_1,\theta)=0$ with $F(0,\bar{\theta})=0$, so locally there exists an implicit function $\theta=\hat{\theta}(x_1)$ and we could use that to approximate how the minimizer deviates under a small perturbation, namely, $\hat{\theta}(x_1)=\hat{\theta}(0)+D\hat{\theta}(0)x_1$.

We will then solve the linear system to solve for the sensitivity $D\hat{\theta}(0)$

$$D_\theta F(0,\bar{\theta})\ D\hat{\theta}(0)=-D_xF(0,\bar{\theta}). \tag{2.53}$$

2.6.1 Characterizing the Perturbation

We perform Karhunen-Loeve expansion for the process $\hat{\mathrm{Corr}}(\tau)=\frac{1}{T}\int_0^T I(t)I(t+\tau)dt=\hat{\mathbb{E}}_T[I(t)I(t+\tau)]$. We estimate the covariance of $\hat{\mathrm{Corr}}(\tau)$ via Monte Carlo simulation. In Monte Carlo simulations, we take diffusion time $\tau_r=7.6923\times10^{-5}$, time mesh for observed trajectories $\Delta t=1\times10^{-5}$, $T=0.09$, and we try different τ_{max} values.

We can see from Figure 2.10 that the largest eigenvalue captures overwhelming majority (over 95%) of the energy of $\hat{\mathrm{Corr}}(\tau)$. Thus, $\hat{\mathrm{Corr}}(\tau)$ could be approximated

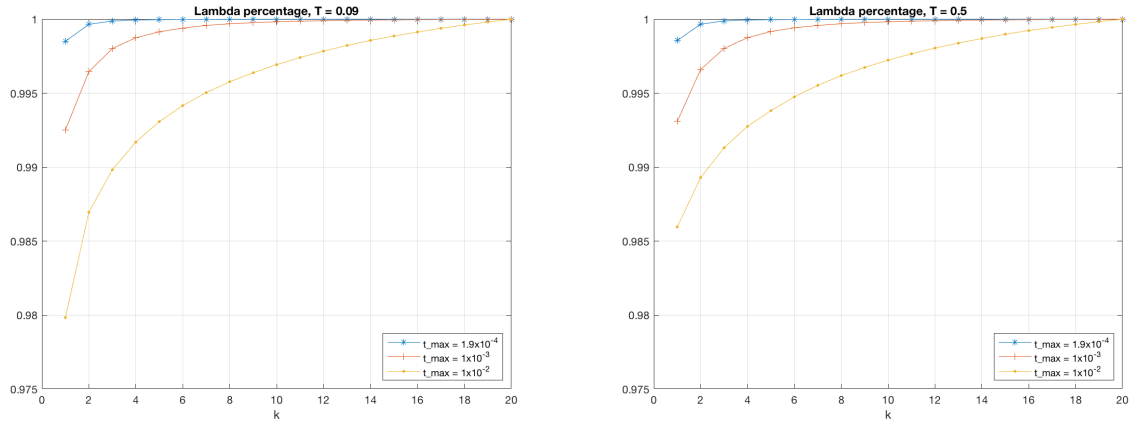


Figure 2.10: The first k eigenvalues over the sum of all eigenvalues. $\frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{n}\lambda_i}$, $T = 0.09, T = 0.5$.

by

$$\hat{\text{Corr}}(\tau) = \mathbb{E}[\hat{\text{Corr}}(\tau)] + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\ \xi_i e_i(\tau) \approx \mathbb{E}[\hat{\text{Corr}}(\tau)] + \sqrt{\lambda_1}\ \xi_1 e_1(\tau). \tag{2.54}$$

Further, we notice that the eigenfunction corresponding to the leading eigenvalue is practically constant. Hence, we assume that the process $\hat{\text{Corr}}(\tau)$ deviates from $\mathbb{E}[\hat{\text{Corr}}(\tau)]$ simply up to a random perturbation η that doesn't depend on τ:

$$\hat{\text{Corr}}(\tau, \omega) = \mathbb{E}[\hat{\text{Corr}}(\tau)] + \eta(\omega). \tag{2.55}$$

Now we would like to address the nonlinear least squares fit problem. Letting $\theta = (D, N_e, F_0)$,

$$\min_{\theta>0} J(\theta), \qquad J(\theta) = \int_0^{\tau_{max}} |\text{Corr}(\tau;\theta) - \hat{Corr}(\tau)|^2 d\tau, \tag{2.56}$$

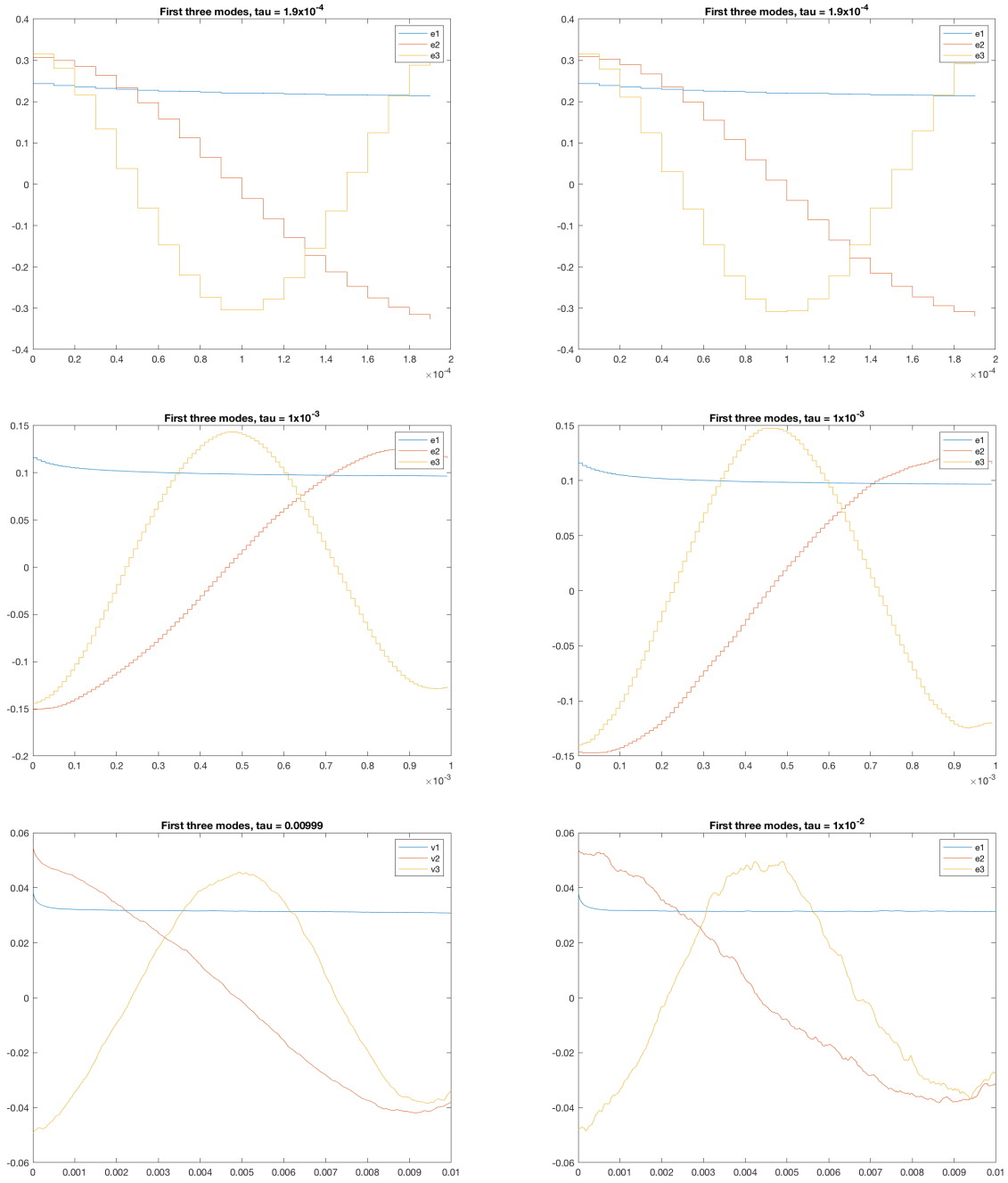


Figure 2.11: First three modes.

where

$$\text{Corr}(\tau; D, N_e, F_0) = F_0^2 \frac{N_e}{8(1+\frac{4D\tau}{r^l 2})\sqrt{1+\frac{4D\tau}{l^2}}} + \frac{F_0^2 N_e^2}{8}. \tag{2.57}$$

Proposition 2. Let $\bar{\theta} = (D, N_e, F_0)$ be the true parameters that drive the stochastic dynamical system. Suppose $\hat{\text{Corr}}(\tau, \omega) = \mathbb{E}[\hat{\text{Corr}}(\tau)] + \eta(\omega)$. For each given value of $\eta(\omega)$, the minimization problem has a unique global minimizer $\hat{\theta}$ with $J(\hat{\theta}) = 0$. The minimizer $\hat{\theta} = (\hat{D}, \hat{N}_e, \hat{F}_0)$ is given by $\hat{D} = D, \hat{N}_e = N_e + \frac{8\eta}{F_0^2 N_e}$, $\hat{F}_0 = F_0 \frac{F_0 N_e}{\sqrt{F_0^2 N_e^2 + 8\eta}}$.

Proof. Substituting $\hat{D} = D, \hat{N}_e = N_e + \frac{8\eta}{F_0^2 N_e}$, $\hat{F}_0 = F_0 \frac{F_0 N_e}{\sqrt{F_0^2 N_e^2 + 8\eta}}$ into (2.57), we can verify that,

$$\text{Corr}(\tau; \hat{D}, \hat{N}_e, \hat{F}_0) = \text{Corr}(\tau; D, N_e, F_0) + \eta = \hat{\text{Corr}}(\tau), \quad \forall \tau > 0. \tag{2.58}$$

To see this,

$$\begin{aligned}
&\text{Corr}(\tau; \hat{D}, \hat{N}_e, \hat{F}_0) = \hat{F}_0{}^2 \frac{\hat{N}_e}{8(1+\frac{4D\tau}{r^l 2})\sqrt{1+\frac{4D\tau}{l^2}}} + \frac{\hat{F}_0{}^2 \hat{N}_e{}^2}{8} \\
&= \left(F_0 \frac{F_0 N_e}{\sqrt{F_0^2 N_e^2 + 8\eta}}\right)^2 \frac{N_e + \frac{8\eta}{F_0^2 N_e}}{8(1+\frac{4D\tau}{r^l 2})\sqrt{1+\frac{4D\tau}{l^2}}} + \frac{1}{8}\left(F_0 \frac{F_0 N_e}{\sqrt{F_0^2 N_e^2 + 8\eta}}\right)^2 \left(N_e + \frac{8\eta}{F_0^2 N_e}\right)^2 \\
&= F_0^2 \frac{N_e}{8(1+\frac{4D\tau}{r^l 2})\sqrt{1+\frac{4D\tau}{l^2}}} + \frac{F_0^2 N_e^2}{8} + \eta = \text{Corr}(\tau; D, N_e, F_0) + \eta = \hat{\text{Corr}}(\tau).
\end{aligned}$$

Since $J(\theta) = \int_0^{\tau_{max}} |\text{Corr}(\tau; \theta) - \hat{\text{Corr}}(\tau)|^2 d\tau \geq 0$, $J(\hat{\theta}) = 0$, hence $\hat{\theta}$ is a global minimizer. For any $\theta \neq \hat{\theta}$, the equation (2.58) fails to hold (due to the injective property shown in Proposition 1), thus $J(\theta) > 0$, and $\hat{\theta}$ is the unique global minimizer. □

Note that this nice result that $\hat{\theta}$ is the unique global minimizer with $J(\hat{\theta}) = 0$

is achieved under the assumption that $\hat{\text{Corr}}(\tau,\omega) = \mathbb{E}[\hat{\text{Corr}}(\tau)] + \eta(\omega)$. However, we shouldn't forget that this assumption is an approximation of the actual case that $\hat{\text{Corr}}(\tau) = \mathbb{E}[\hat{\text{Corr}}(\tau)] + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\, \xi_i e_i(\tau)$. So now we would like to take a closer look at the KL expansion of the perturbation $\hat{\text{Corr}}(\tau) - \mathbb{E}[\hat{\text{Corr}}(\tau)]$.

Let us denote the truncated KL expansion of $\hat{\text{Corr}} - \mathbb{E}[\hat{\text{Corr}}(\tau)]$ by S_k and the remainder R_k,

$$S_k(\tau,\omega) = \sum_{i=1}^{k} \sqrt{\lambda_i}\, \xi_i(\omega) e_i(\tau), \tag{2.59}$$

$$R_k(\tau,\omega) = \hat{\text{Corr}}(\tau,\omega) - \mathbb{E}[\hat{\text{Corr}}(\tau)] - S_k(\tau,\omega). \tag{2.60}$$

In Figure 2.12, we show the first three modes of the KL expansion for three independent realizations of $\hat{\text{Corr}}(\tau)$. In the left column, we plot the truncated KL expansion (namely, the first mode $S_1(\tau)$, the first two modes $S_2(\tau)$, the first three modes $S_3(\tau)$) in comparison with the process $\hat{\text{Corr}}(\tau,\omega) - \mathbb{E}[\hat{\text{Corr}}(\tau)]$. We also plot the remainder $R_1(\tau), R_2(\tau), R_3(\tau)$ compared with the perturbation $R_0(\tau) = \hat{\text{Corr}}(\tau,\omega) - \mathbb{E}[\hat{\text{Corr}}(\tau)]$ in the right column.

Recall the KL expansion,

$$X(t,\omega) = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\, \xi_i(\omega) e_i(t) = \sum_{i=1}^{\infty} z_i(\omega) e_i(t) \qquad \text{in } L^2(\Omega), \tag{2.61}$$

where

$$\xi_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_0^{t_{max}} X(t,\omega) e_i(t) dt$$

Now we look at the distribution of random variables z_1, z_2 and z_3. The standard deviations of z_1, z_2 and z_3 are $\sigma_1 = 0.4679$, and $\sigma_2 = 0.0163$, $\sigma_3 = 0.0071$, respec-

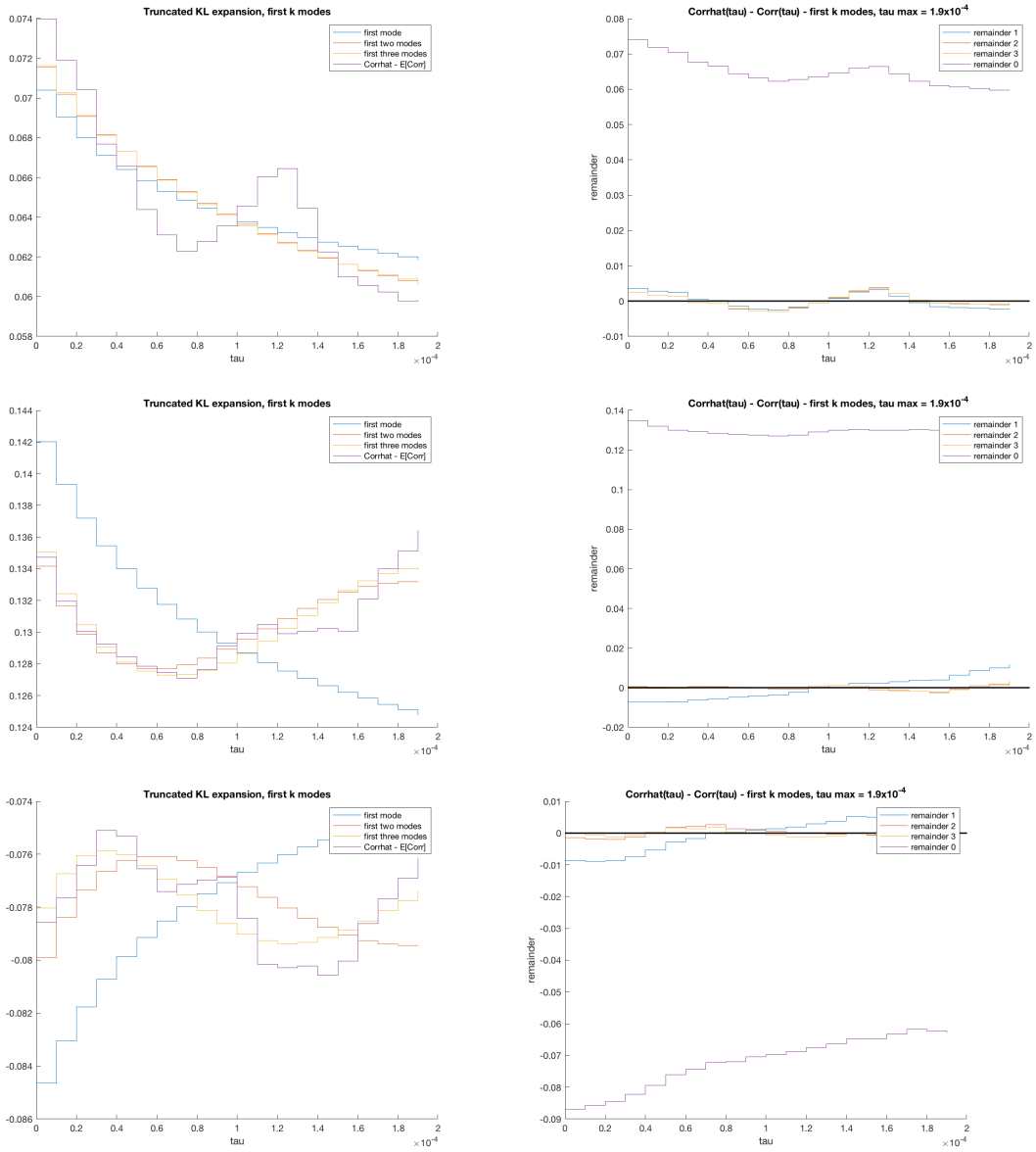


Figure 2.12: Truncated KL expansion of trajectories of $\hat{\text{Corr}}(\tau, \omega) - \mathbb{E}[\hat{\text{Corr}}(\tau)]$ and the remainders for three different sample trajectories.

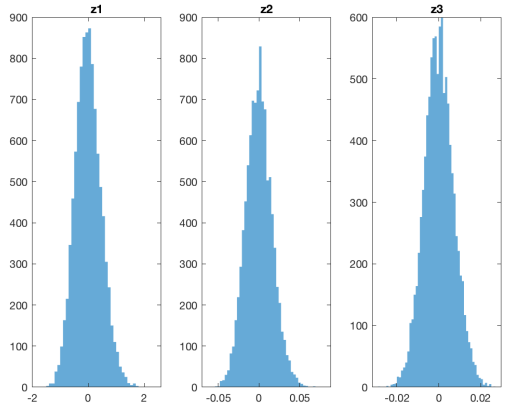


Figure 2.13: Distribution of random variables z_1, z_2 and z_3

tively.

2.6.2 Solving for the Sensitivity

Having justified the characterization of the perturbation, we now go back to the sensitivity problem where we need to solve the linear system

$$D_\theta F(0, \bar{\theta})\ D\hat{\theta}(0) = -D_x F(0, \bar{\theta}).$$

Here the function $F(x_1, \theta) = 0$ comes from the first-order optimality condition (2.50)

$$\frac{\partial}{\partial \theta_1} J(\theta) = 0, \quad \frac{\partial}{\partial \theta_2} J(\theta) = 0, \quad \frac{\partial}{\partial \theta_3} J(\theta) = 0.$$

Then we use implicit function theorem to obtain a linear approximation $\hat{\theta}(x_1) = \hat{\theta}(0) + D\hat{\theta}(0)x_1$.

To be concrete, equation (2.50) is now reformulated as

$$F_1(x_1,\theta) = \frac{\partial}{\partial\theta_1}J(\theta) = \frac{\partial}{\partial D}\int_0^{\tau_{max}} |\mathrm{Corr}(\tau;\theta) - \mathrm{Corr}(\tau;\bar{\theta}) - x_1 e_1(\bar{D}\tau)|^2 d\tau = 0. \tag{2.62}$$

$$F_2(x_1,\theta) = \frac{\partial}{\partial\theta_1}J(\theta) = \frac{\partial}{\partial N_e}\int_0^{\tau_{max}} |\mathrm{Corr}(\tau;\theta) - \mathrm{Corr}(\tau;\bar{\theta}) - x_1 e_1(\bar{D}\tau)|^2 d\tau = 0. \tag{2.63}$$

$$F_3(x_1,\theta) = \frac{\partial}{\partial\theta_1}J(\theta) = \frac{\partial}{\partial F_0}\int_0^{\tau_{max}} |\mathrm{Corr}(\tau;\theta) - \mathrm{Corr}(\tau;\bar{\theta}) - x_1 e_1(\bar{D}\tau)|^2 d\tau = 0. \tag{2.64}$$

Recall that,

$$\begin{aligned}\mathrm{Corr}(\tau; D, N_e, F_0) &= \frac{F_0^2 N_e}{8}\frac{1}{(1+\frac{4D\tau}{r^2})\sqrt{1+\frac{4D\tau}{l^2}}} + \frac{F_0^2 N_e^2}{8} \\ &= \frac{F_0^2 N_e}{8} g(D\tau) + \frac{F_0^2 N_e^2}{8}, \quad \text{with } g(t) = \frac{1}{(1+\frac{4t}{r^2})\sqrt{1+\frac{4t}{l^2}}}\end{aligned}$$

So

$$\partial_{\theta_1}\mathrm{Corr}(\tau,\theta) = \frac{F_0^2 N_e}{8} g'(D\tau)\tau = \frac{F_0^2 N_e \tau}{8} g(D\tau)\left(-\frac{2}{l^2}\cdot\frac{1}{1+\frac{4D\tau}{l^2}} - \frac{4}{r^2}\cdot\frac{1}{1+\frac{4D\tau}{r^2}}\right),$$

$$\partial_{\theta_2}\mathrm{Corr}(\tau,\theta) = \frac{F_0^2}{8} g(D\tau) + \frac{2F_0^2 N_e}{8},$$

$$\partial_{\theta_3}\mathrm{Corr}(\tau,\theta) = \frac{2F_0 N_e}{8} g(D\tau) + \frac{2F_0 N_e^2}{8}.$$

We could obtain that

$$F_1(x_1,\theta) = 2\int_0^{\tau_{max}} \Big(\operatorname{Corr}(\tau;\theta) - \operatorname{Corr}(\tau;\bar{\theta}) - x_1 e_1(\bar{D}\tau)\Big)\left(\frac{F_0^2 N_e}{8} g'(D\tau)\tau\right) d\tau = 0.$$
$$F_2(x_1,\theta) = 2\int_0^{\tau_{max}} \Big(\operatorname{Corr}(\tau;\theta) - \operatorname{Corr}(\tau;\bar{\theta}) - x_1 e_1(\bar{D}\tau)\Big)\left(\frac{F_0^2}{8} g(D\tau) + \frac{2F_0^2 N_e}{8}\right) d\tau = 0.$$
$$F_3(x_1,\theta) = 2\int_0^{\tau_{max}} \Big(\operatorname{Corr}(\tau;\theta) - \operatorname{Corr}(\tau;\bar{\theta}) - x_1 e_1(\bar{D}\tau)\Big)\left(\frac{2F_0 N_e}{8} g(D\tau) + \frac{2F_0 N_e^2}{8}\right) d\tau = 0.$$

$$\frac{\partial}{\partial\theta_1}F_1(x_1,\theta) = 2\int_0^{\tau_{max}} \Big(\operatorname{Corr}(\tau;\theta) - \operatorname{Corr}(\tau;\bar{\theta}) - x_1 e_1(\bar{D}\tau)\Big)\frac{\partial}{\partial D}\left(\frac{F_0^2 N_e}{8} g'(D\tau)\tau\right) d\tau$$
$$+ 2\int_0^{\tau_{max}} \frac{\partial}{\partial D}\left[\operatorname{Corr}(\tau;\theta) - \operatorname{Corr}(\tau;\bar{\theta}) - x_1 e_1(\bar{D}\tau)\right]\left(\frac{F_0^2 N_e}{8} g'(D\tau)\tau\right) d\tau.$$

$$\frac{\partial}{\partial\theta_2}F_1(x_1,\theta) = 2\int_0^{\tau_{max}} \Big(\operatorname{Corr}(\tau;\theta) - \operatorname{Corr}(\tau;\bar{\theta}) - x_1 e_1(\bar{D}\tau)\Big)\frac{\partial}{\partial N_e}\left(\frac{F_0^2 N_e}{8} g'(D\tau)\tau\right) d\tau$$
$$+ 2\int_0^{\tau_{max}} \frac{\partial}{\partial N_e}\left[\operatorname{Corr}(\tau;\theta) - \operatorname{Corr}(\tau;\bar{\theta}) - x_1 e_1(\bar{D}\tau)\right]\left(\frac{F_0^2 N_e}{8} g'(D\tau)\tau\right) d\tau.$$

$$\frac{\partial}{\partial\theta_3}F_1(x_1,\theta) = 2\int_0^{\tau_{max}} \Big(\operatorname{Corr}(\tau;\theta) - \operatorname{Corr}(\tau;\bar{\theta}) - x_1 e_1(\bar{D}\tau)\Big)\frac{\partial}{\partial F_0}\left(\frac{F_0^2 N_e}{8} g'(D\tau)\tau\right) d\tau$$
$$+ 2\int_0^{\tau_{max}} \frac{\partial}{\partial F_0}\left[\operatorname{Corr}(\tau;\theta) - \operatorname{Corr}(\tau;\bar{\theta}) - x_1 e_1(\bar{D}\tau)\right]\left(\frac{F_0^2 N_e}{8} g'(D\tau)\tau\right) d\tau.$$

Thus,

$$\frac{\partial}{\theta_1}F_1(0,\bar{\theta}) = 2\int_0^{\tau_{max}} \left(\frac{\bar{F_0}^2 \bar{N_e}}{8} g'(\bar{D}\tau)\tau\right)^2 d\tau = 2\int_0^{\tau_{max}} \Big(\frac{\partial \operatorname{Corr}}{\partial\theta_1}(\tau,\bar{\theta})\Big)^2 d\tau.$$

$$\frac{\partial}{\theta_2}F_1(0,\bar{\theta}) = 2\int_0^{\tau_{max}} \left(\frac{\bar{F_0}^2}{8} g(\bar{D}\tau) + \frac{2\bar{F_0}^2 \bar{N_e}}{8}\right)\left(\frac{\bar{F_0}^2 \bar{N_e}}{8} g'(\bar{D}\tau)\tau\right) d\tau.$$
$$= 2\int_0^{\tau_{max}} \Big(\frac{\partial \operatorname{Corr}}{\partial\theta_2}(\tau,\bar{\theta})\Big)\Big(\frac{\partial \operatorname{Corr}}{\partial\theta_1}(\tau,\bar{\theta})\Big) d\tau.$$

$$\frac{\partial}{\theta_3}F_1(0,\bar{\theta}) = 2\int_0^{\tau_{max}} \left(\frac{2\bar{F}_0\bar{N}_e}{8}g(\bar{D}\tau) + \frac{2\bar{F}_0{\bar{N}_e}^2}{8}\right)\left(\frac{{\bar{F}_0}^2\bar{N}_e}{8}g'(\bar{D}\tau)\tau\right)d\tau$$
$$= 2\int_0^{\tau_{max}} \Big(\frac{\partial\,\mathrm{Corr}}{\partial\theta_3}(\tau,\bar{\theta})\Big)\Big(\frac{\partial\,\mathrm{Corr}}{\partial\theta_1}(\tau,\bar{\theta})\Big)d\tau$$

We could compute $D_\theta F_2(0,\bar{\theta})$ and $D_\theta F_3(0,\bar{\theta})$ in a similar manner, hence we can write the Jacobian matrix $D_\theta F(0,\bar{\theta})$ as

$$D_\theta F\theta(0,\bar{\theta}) =$$
$$\begin{bmatrix} 2\int_0^{\tau_{max}} \left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_1}(\tau,\bar{\theta})\right)^2 d\tau, & 2\int_0^{\tau_{max}} \left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_1}(\tau,\bar{\theta})\right)\left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_2}(\tau,\bar{\theta})\right)d\tau, & 2\int_0^{\tau_{max}} \left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_1}(\tau,\bar{\theta})\right)\left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_3}(\tau,\bar{\theta})\right)d\tau \\ 2\int_0^{\tau_{max}} \left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_1}(\tau,\bar{\theta})\right)\left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_2}(\tau,\bar{\theta})\right)d\tau, & 2\int_0^{\tau_{max}} \left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_2}(\tau,\bar{\theta})\right)^2 d\tau, & 2\int_0^{\tau_{max}} \left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_2}(\tau,\bar{\theta})\right)\left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_3}(\tau,\bar{\theta})\right)d\tau \\ 2\int_0^{\tau_{max}} \left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_1}(\tau,\bar{\theta})\right)\left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_3}(\tau,\bar{\theta})\right)d\tau, & 2\int_0^{\tau_{max}} \left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_2}(\tau,\bar{\theta})\right)\left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_3}(\tau,\bar{\theta})\right)d\tau, & 2\int_0^{\tau_{max}} \left(\frac{\partial\,\mathrm{Corr}}{\partial\theta_3}(\tau,\bar{\theta})\right)^2 d\tau. \end{bmatrix} \tag{2.65}$$

We shall show that $D_\theta F\theta(0,\bar{\theta})$ is positive definite from its own structure. We introduce $L^2([0,\tau_{max}])$ functions f_1, f_2, f_3 to be that $f_i(\tau) = \frac{\partial\,\mathrm{Corr}}{\partial\theta_i}(\tau,\bar{\theta})$ for $i = 1,2,3$. Then the matrix $M := D_\theta F\theta(0,\bar{\theta})$ has the form $M_{ij} = (f_i, f_j)$. To see M is positive semidefinite, let $z \in \mathbb{R}^3$ be arbitrary, we have

$$z^T M z = \sum_{i,j=1,2,3} M_{ij} z_i z_j = \sum_{i,j=1,2,3} (f_i, f_j) z_i z_j = \sum_{i,j=1,2,3} (z_i f_i, z_j f_j)$$
$$= (\sum_{i=1}^{3} z_i f_i, \sum_{j=1}^{3} z_j f_j) \geq 0.$$

Moreover, note that $z^T M z = 0$ if and only if $\sum_{i=1}^3 z_i f_i = 0$. Since f_1, f_2, f_3 are linearly independent functions in $L^2([0,\tau_{max}])$, we know M is positive definite and hence nonsingular. Hence the linear system $D_\theta F(0,\bar{\theta})\ D\hat{\theta}(0) = -D_x F(0,\bar{\theta})$ has a unique solution.

We compute the right hand side

$$D_x F(0,\bar{\theta}) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1}(0,\bar{\theta}) \\ \frac{\partial F_2}{\partial x_1}(0,\bar{\theta}) \\ \frac{\partial F_3}{\partial x_1}(0,\bar{\theta}) \end{bmatrix} = \begin{bmatrix} 2\int_0^{\tau_{max}} \Big(- e_1(\bar{D}\tau)\Big)\Big(\frac{\partial \,\mathrm{Corr}}{\partial \theta_1}(\tau,\bar{\theta})\Big)d\tau \\ 2\int_0^{\tau_{max}} \Big(- e_1(\bar{D}\tau)\Big)\Big(\frac{\partial \,\mathrm{Corr}}{\partial \theta_2}(\tau,\bar{\theta})\Big)d\tau \\ 2\int_0^{\tau_{max}} \Big(- e_1(\bar{D}\tau)\Big)\Big(\frac{\partial \,\mathrm{Corr}}{\partial \theta_3}(\tau,\bar{\theta})\Big)d\tau \end{bmatrix} \tag{2.66}$$

For Jacobian matrix $D_\theta F(0,\bar{\theta})$, we introduce a few L^2 inner product notations for integrals. Let $f, h, 1, e : [0, \tau_{max}] \to \mathbb{R}$ be given by

$$f(\tau) = g(D\tau) = \frac{1}{(1+\frac{4\bar{D}\tau}{r^2})\sqrt{1+\frac{4\bar{D}\tau}{l^2}}},$$

$$h(\tau) = g'(\bar{D}\tau)\tau = \frac{1}{\bar{D}} f'(\tau)\tau,$$

$$1(\tau) = 1,$$

$$e(\tau) = e_1(\bar{D}\tau).$$

Thus, we could rewrite

$$\begin{aligned} \partial_{\theta_1} \mathrm{Corr}(\tau,\theta) &= \frac{F_0^2 N_e}{8} g'(D\tau)\tau = \frac{F_0^2 N_e}{8} h(\tau), \\ \partial_{\theta_2} \mathrm{Corr}(\tau,\theta) &= \frac{F_0^2}{8} g(D\tau) + \frac{2F_0^2 N_e}{8} = \frac{F_0^2}{8} f(\tau) + \frac{2F_0^2 N_e}{8} 1(\tau), \\ \partial_{\theta_3} \mathrm{Corr}(\tau,\theta) &= \frac{2F_0 N_e}{8} g(D\tau) + \frac{2F_0 N_e^2}{8} = \frac{2F_0 N_e}{8} f(\tau) + \frac{2F_0 N_e^2}{8} 1(\tau). \end{aligned}$$

We have previously shown that $\hat{D} = D, \hat{N}_e = N_e + \frac{8\eta}{F_0^2 N_e}$, $\hat{F}_0 = F_0 \frac{F_0 N_e}{\sqrt{F_0^2 N_e^2 + 8\eta}}$ is the unique minimizer to the minimization problem (2.46) provided that $\eta(\tau,\omega) = \eta$ independent of τ. Now in this sensitivity analysis setup, this case could be translated

as $\eta = x_1 e_1(\bar{D}\tau)$ with $e_1(\bar{D}\tau) = 1$. Solving the linear system with this particular right hand side,

$$D_\theta F(0, \bar{\theta})\ D\hat{\theta}(0) = -D_x F(0, \bar{\theta}). \tag{2.67}$$

we obtain the solution $D\hat{\theta}(0) = (0, \frac{8}{F_0^2 N_e}, -\frac{4}{F_0^2 N_e^2})$. The $D\hat{\theta}(0)$ solution says, $\hat{D} = \bar{D} + o(x_1)$, $\hat{N}_e = \bar{N}_e + \frac{8x_1}{F_0^2 N_e^2} + o(x_1)$, $\hat{F}_0 = \bar{F}_0 - \frac{4x_1}{F_0^2 N_e^2} + o(x_1)$. The results obtained here from perturbation analysis agrees with the proposition we obtained earlier.

However, solving equation $D_\theta F(0, \bar{\theta})\ D\hat{\theta}(0) = -D_x F(0, \bar{\theta})$ with a general right hand side gives a complicated expression of $D\hat{\theta}(0)$.

$$\begin{aligned} z_1 = &- \frac{8}{F_0^2 N_e}\big[- (e,h)(f,1)^2 + (e,f)(f,1)(h,1) - (e,1)(f,f)(h,1) + (e,1)(f,1)(h,f) \\ &+ (e,h)(f,f)(1,1) - (e,f)(h,f)(1,1)\big] / \\ &[(f,f)(h,1)^2 - 2(f,1)(h,1)(h,f) + (f,1)^2(h,h) + (h,f)^2(1,1) - (f,f)(h,h)(1,1)]. \end{aligned} \tag{2.68}$$

When we substitute a constant-valued function $e = 1$ into equation (2.68), we can see that z_1 simplifies to 0. That is, under a constant perturbation η that is independent of τ, the solution $z_1 = 0$.

We have obtained

$$\hat{D} = \bar{D} + z_1 x_1 + o(x_1), \tag{2.69}$$

where z_1 is the sensitivity coefficient given by (2.68), and x_1 is a random scalar that

describes the magnitude of the perturbation

$$\eta(\tau,\omega) = \tilde{\xi_1}(\omega, \bar{N}_e, \bar{F}_0)\ e_1(\bar{D}\tau) = x_1 e_1(\bar{D}\tau). \tag{2.70}$$

From the Karhuen-Loeve expansion of η, we have $\forall\tau \in [0, \tau_{max}]$,

$$\eta(\tau,\omega) = \hat{\text{Corr}}(\tau,\omega) - \text{Corr}(\tau;\bar{\theta}) = \sum_{i=1}^{\infty} x_i(\omega) e_i(\tau) \qquad \text{in } L^2(\Omega), \tag{2.71}$$

with $\mathbb{E}[x_i] = 0$, $\mathbb{E}[x_i x_j] = \delta_{ij}\lambda_i$.

Thus, we get

$$\mathbb{E}[\hat{D} - \bar{D}] = \mathbb{E}[z_1 x_1 + o(x_1)] = z_1\mathbb{E}[x_1] + o(x_1) = o(x_1). \tag{2.72}$$

That is, if we only consider the first mode and neglect the error caused by linearization, $\hat{D}$ should be an unbiased estimator of $\bar{D}$. In the numerical experiments of fitting 10^4 trajectories, we saw a small relative bias of 0.87% when $T = 0.1$, while the bias gets larger for shorter T, namely, 1.64% for $T = 0.05$, 3.95% for $T = 0.02$, and 7.37% for $T = 0.01$.

$$\mathbb{E}[|\hat{D}-\bar{D}|^2] = \text{Var}[\hat{D}-\bar{D}]+(\mathbb{E}[\hat{D}-\bar{D}])^2 = \text{Var}[z_1x_1+o(x_1)]+o(x_1^2) = (z_1)^2\lambda_1+o(x_1^2). \tag{2.73}$$

First mode only accounts for 0.593% of the relative error. However, in the nonlinear least squares fit experiments, we saw a relative error of 12.1% for $T = 0.1$. $z_1 =$

-1.6476×10^{-12}, $\sqrt{\lambda_1} = 0.468$.

We check how well the linear approximation $\hat{D} = \bar{D} + z_1 x_1 + o(x_1)$ from implicit function theorem works on a local neighborhood. Here we use the perturbation of the form

$$\eta_1(\tau, \omega) = x_1(\omega) e_1(\tau) \approx \hat{\text{Corr}}(\tau, \omega) - \text{Corr}(\tau; \bar{\theta}) \tag{2.74}$$

where $e_1(\tau)$ is the leading eigenfunction, and we take a mesh of x_1 within the three standard deviations. We compare the prediction $\hat{D} = \bar{D} + z_1 x_1$ derived from (2.74) with the fitted $\hat{D}$ obtained from minimizing square errors. As Figure 2.14 shows, this linear approximation is good in a local neighborhood when we only use the first mode of the perturbation, so we might have to consider the influence of higher modes.

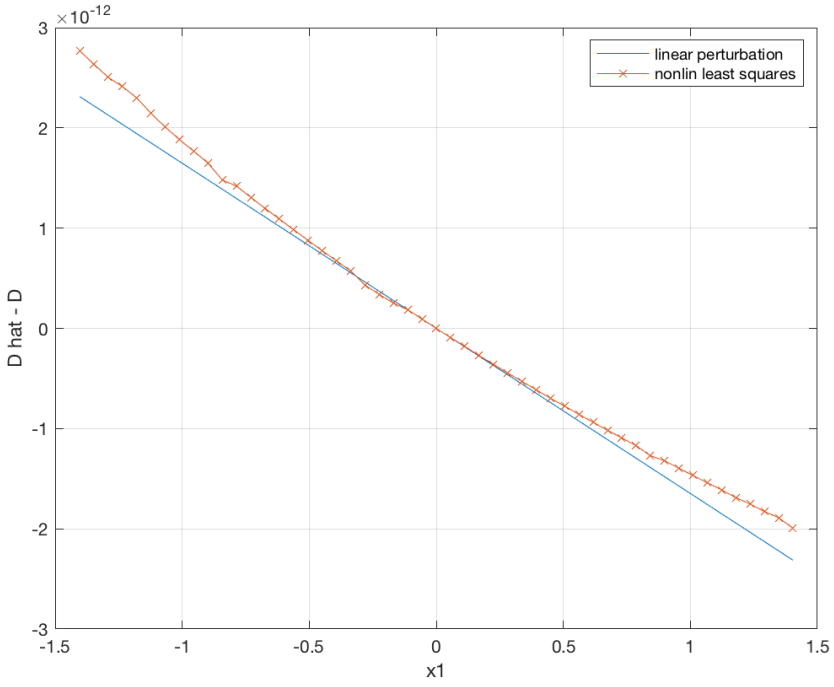


Figure 2.14: The error $\hat{D} - D$ against x_1 when $\hat{\text{Corr}}(\tau) = \text{Corr}(\tau; \bar{\theta}) + x_1 e_1(\tau)$, with $x_1 \in [-3\sigma_1, 3\sigma_1]$ where $\sigma_1 = \sqrt{\text{Var}[x_1]} = \sqrt{\lambda_1} = 0.468$

Recall the (2.74) is an approximation of the Karhunen-Loeve expansion truncated with only the first mode,

$$\eta(\tau,\omega)=\sum_{i=1}^{\infty}x_i(\omega)e_i(t) \tag{2.75}$$

Previously, we have obtained $\lambda_1 \gg \sum_{i=2}^{n_\tau}\lambda_i$, the first mode approximates the perturbation very well. So it is intuitive to think that $\eta_1 \approx \eta$ should yield very close solutions to the minimization problem (2.46). That is, $\hat{D}_1 \approx \hat{D}$,

$$\begin{aligned}
&(\hat{D}_1,\hat{N}_e,\hat{F}_0)=\underset{(D,N_e,F_0)}{\arg\min}\, J_1(D,N_e,F_0),\\
&\text{where } J_1(D,N_e,F_0)=\int_0^{\tau_{max}}|\operatorname{Corr}(\tau,\theta)-\operatorname{Corr}(\tau,\bar{\theta})-x_1e_1|^2d\tau,\\
&(\hat{D},\hat{N}_e,\hat{F}_0)=\underset{(D,N_e,F_0)}{\arg\min}\, J(D,N_e,F_0),\\
&\text{where } J(D,N_e,F_0)=\int_0^{\tau_{max}}|\operatorname{Corr}(\tau,\theta)-\hat{\operatorname{Corr}}(\tau,\theta)|^2d\tau.
\end{aligned}$$

However, Figure 2.15 suggests that it doesn't appear this way. It seems that although the rest of the modes are small compared to the first mode, their influence on fitted $\hat{D}$ seems nonnegligible.

Hence, besides the truncated $\eta(\tau,\omega) = x_1(\omega)e_1(\tau)$ that only has the first mode of the KL expansion, we could also look at a general perturbation $\eta(\tau,\omega) = \hat{\operatorname{Corr}}(\tau,\omega) - \operatorname{Corr}(\tau;\bar{\theta})$ coming from a simulated $\hat{\operatorname{Corr}}$ trajectory.

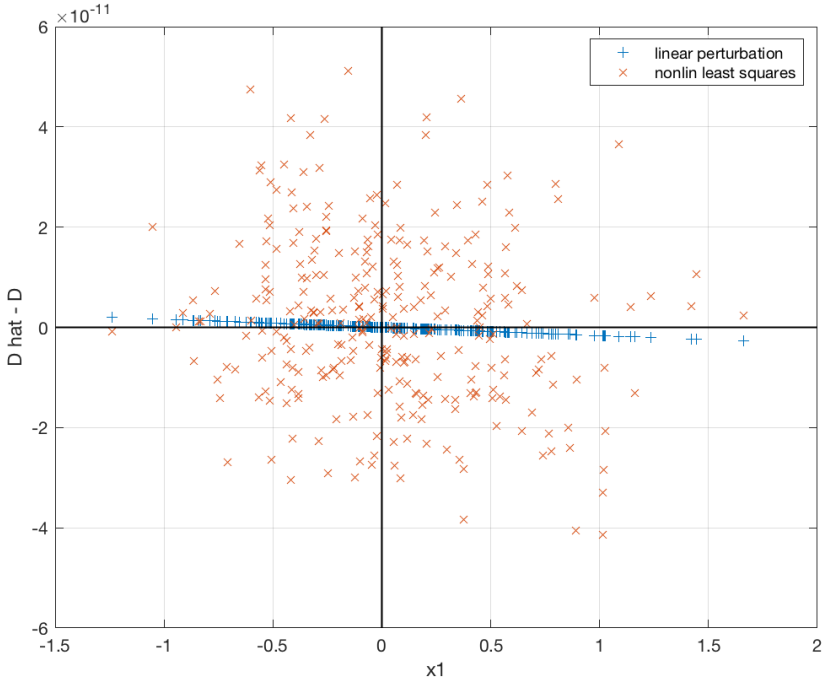


Figure 2.15: The error $\hat{D} - D$ against KL coefficient x_1 for general $\hat{\text{Corr}}$ data. Perturbation analysis that only considers the first mode e_1 fails to capture a large portion of the actual error.

2.6.3 More Modes in Perturbation

When we consider the perturbation of the form

$$\eta(\tau,\omega)=\sum_{i=1}^{n}x_i(\omega)e_i(\tau), \tag{2.76}$$

with $\mathbb{E}[x_i]=0$, $\mathbb{E}[x_ix_j]=\delta_{ij}\lambda_i$.

Now the relation of the best fit $\theta=(D,N_e,F_0)$ and the perturbation data $x=(x_1,...,x_n)$ is given by the system of equations:

$$F_1(x,\theta)=\frac{\partial}{\partial\theta_1}J(\theta)=\frac{\partial}{\partial D}\int_0^{\tau_{max}}|\mathrm{Corr}(\tau;\theta)-\mathrm{Corr}(\tau;\bar{\theta})-\sum_{i=1}^{n}x_ie_i(\bar{D}\tau)|^2d\tau=0. \tag{2.77}$$

$$F_2(x,\theta)=\frac{\partial}{\partial\theta_1}J(\theta)=\frac{\partial}{\partial N_e}\int_0^{\tau_{max}}|\mathrm{Corr}(\tau;\theta)-\mathrm{Corr}(\tau;\bar{\theta})-\sum_{i=1}^{n}x_ie_i(\bar{D}\tau)|^2d\tau=0. \tag{2.78}$$

$$F_3(x,\theta)=\frac{\partial}{\partial\theta_1}J(\theta)=\frac{\partial}{\partial F_0}\int_0^{\tau_{max}}|\mathrm{Corr}(\tau;\theta)-\mathrm{Corr}(\tau;\bar{\theta})-\sum_{i=1}^{n}x_ie_i(\bar{D}\tau)|^2d\tau=0. \tag{2.79}$$

Applying the implicit function theorem near the point $(0,\bar{\theta})$, locally there exists $g:\mathbb{R}^n\to\mathbb{R}^3$, so that $F(x,g(x))=0$, and we could obtain the sensitivity of $\hat{\theta}$ with respect to x by solving the linear system

$$D_\theta F(0,\bar{\theta})\ Dg(0)=-D_xF(0,\bar{\theta}). \tag{2.80}$$

Note that $D_\theta F(0,\bar{\theta})$ is the same invertible matrix we computed before for a single mode perturbation. The sensitivity $Z := Dg(0)$ could be viewed as a block matrix $[Z_{:,1}, Z_{:,2}, ..., Z_{:,n}]$, and the right hand side $B := -D_x F(0,\bar{\theta}) = [B_1, ..., B_n]$ with

$$B_i = -D_{x_i} F(0,\bar{\theta}) = \begin{bmatrix} \frac{\partial F_1}{\partial x_i}(0,\bar{\theta}) \\ \frac{\partial F_2}{\partial x_i}(0,\bar{\theta}) \\ \frac{\partial F_3}{\partial x_i}(0,\bar{\theta}) \end{bmatrix} = \begin{bmatrix} 2\int_0^{\tau_{max}} \Big(e_i(\bar{D}\tau)\Big)\Big(\frac{\partial \text{Corr}}{\partial \theta_1}(\tau,\bar{\theta})\Big)d\tau \\ 2\int_0^{\tau_{max}} \Big(e_i(\bar{D}\tau)\Big)\Big(\frac{\partial \text{Corr}}{\partial \theta_2}(\tau,\bar{\theta})\Big)d\tau \\ 2\int_0^{\tau_{max}} \Big(e_i(\bar{D}\tau)\Big)\Big(\frac{\partial \text{Corr}}{\partial \theta_3}(\tau,\bar{\theta})\Big)d\tau \end{bmatrix}. \quad (2.81)$$

We are particularly interested in predicting the error in D,

$$\hat{D} = \bar{D} + \sum_{i=1}^{n} \frac{\partial g_1}{\partial x_i}(0) x_i + o(\|x\|) = \bar{D} + \sum_{i=1}^{n} z_{1,i} x_i + o(\|x\|). \quad (2.82)$$

where $z_{1,i}$ is obtained from solving linear system $D_\theta F(0,\bar{\theta})\ Z_{:,i} = B_i$

$$z_{1,i} = -\frac{8}{F_0^2 N_e}[-(e_i,h)(f,1)^2 + (e_i,f)(f,1)(h,1) - (e_i,1)(f,f)(h,1)$$
$$+ (e_i,1)(f,1)(h,f) + (e_i,h)(f,f)(1,1) - (e_i,f)(h,f)(1,1)]$$
$$/[(f,f)(h,1)^2 - 2(f,1)(h,1)(h,f) + (f,1)^2(h,h) + (h,f)^2(1,1) - (f,f)(h,h)(1,1)]. \quad (2.83)$$

Thus,

$$\mathbb{E}[\hat{D} - \bar{D}] = \mathbb{E}\left[\sum_{i=1}^{n} z_{1,i} x_i + o(\|x\|)\right] = o(\|x\|). \quad (2.84)$$

$$\mathbb{E}[|\hat{D} - \bar{D}|^2] = \text{Var}[\hat{D} - \bar{D}] + (\mathbb{E}[\hat{D} - \bar{D}])^2 = \sum_{i=1}^{n} \lambda_i z_{1,i}^2 + o(\|x\|^2). \quad (2.85)$$

We could compute the relative L_2 error in D caused by the ith mode alone, and by

the first i's modes together. That is, under the perturbation $\eta = x_i e_i$, the relative error $RE_i = \frac{\sqrt{\mathbb{E}[|\hat{D}-\bar{D}|^2]}}{\overline{D}} = \frac{\sqrt{\lambda_i}\cdot|z_{1,i}|}{\overline{D}}$, and under the perturbation of $\eta = \sum_{j=1}^{i} x_j e_j$, the cumulative relative error $RE_i = \frac{\sqrt{\mathbb{E}[|\hat{D}-\bar{D}|^2]}}{\overline{D}} = \frac{\sqrt{\sum_{j=1}^{i}\lambda_j (z_{1,j})^2}}{\overline{D}}$. We summarize the predicted relative error using first six modes of KL expansion in Table 2.2. The relative error in D obtained from fitting $10,000$ simulated trajectories is 12.1248%. We need to consider second and third modes to capture the error.

modes	1	2	3	4	5	6
$z_{1,i}$	-1.6476×10^{-12}	-7.88178×10^{-10}	1.41486×10^{-09}	-8.3639×10^{-10}	7.25645×10^{-10}	5.00278×10^{-10}
RE_i (%) (ith mode alone)	0.593026	9.88288	7.70081	2.4555	1.33673	0.618167
RE_1^i (%) (first i modes)	0.593026	9.90066	12.5429	12.781	12.8507	12.8656

Table 2.2: Predicted relative error using first six modes of KL expansion.

Chapter 3

Continuous-time Filtering in Stochastic Reaction Networks

Chemical reaction networks occurring at the intracellular level often have some or all molecular species present in low copy numbers. As such, these are best modeled by a discrete state Markov process in continuous time, where the state $Z(t) \in \mathbb{Z}_+^n$ of the process is the vector copy number of the molecular species at time t. Fluorescence technologies can track the temporal dynamics of a limited number of species, leaving a critical question of estimating the full dynamics from the partial time-course observation data.

In this chapter, we consider the problem of estimating the state $Z(t)$ from exact and continuous observations of some of the states in time, and later Chapter 4, we will investigate the case when observations are made in discrete time. Suppose that the reaction network consists of n molecular species where the vector copy number of n_2 of the molecular species may be observed exactly as a function of time over the interval $[0, t]$. Based on this observation, we are interested in estimating the vector copy number of the remaining $n_1 = n - n_2$ species at time t. If we denote the vector copy number of the observable species by $Y(t) \in \mathbb{Z}_+^{n_2}$ and that of the rest of the (unobservable) species by $X(t) \in \mathbb{Z}_+^{n_1}$, then our goal is to compute the conditional probability distribution

$$\pi(t, x) = P\{X(t) = x \,|\, Y(s) = y(s),\ 0 \le s \le t\}$$

for $x \in \mathbb{Z}_+^{n_1}$.

In Section 3.1 we briefly review the basics of stochastic reaction networks and provide the filtering equations. We also introduce the useful *unnormalized filtering equations*. In Section 3.2 we provide a Monte Carlo filtering algorithm which generates a pair of processes (V, w) where V has the same state space as X and w is an associated weight process such that

$$\pi(t, x) = \frac{\mathbb{E}[1_{\{x\}}(V(t))\, w(t)]}{\mathbb{E}[w(t)]}.$$

We also show how our algorithm can be easily adapted to do parameter estimation based on exact partial state observation in a Bayesian framework and also how to estimate the conditional probability of a past event based on the observations made until a later time. That is, we show how to compute

$$P\{X(t) = x \,|\, Y(s) = y(s),\, 0 \le s \le T\},$$

where $x \in \mathbb{Z}_+^{n_1}$ and $0 < t < T$, via an easy modification of our algorithm. In Section 3.4 we illustrate our algorithms via numerical examples and Section 3.5 makes some concluding remarks and discusses future work.

As we shall describe in Section 3.1, the evolution equation for $\pi(t, x)$ is nonlinear, and hence it is not possible to regard $\pi(t, x)$ as the (unconditional) probability distribution of a Markov process. We shall define a nonnegative function $\rho(t, x)$

which satisfies a linear evolution equation and when normalized yields $\pi(t,x)$:

$$\pi(t,x) = \frac{\rho(t,x)}{\sum_{\tilde{x}} \rho(t,\tilde{x})}.$$

We refer to $\rho(t,x)$ as the *unnormalized conditional distribution* and refer to the evolution equations (3.8) and (3.9) for $\rho(t,x)$, that are defined in Section 3.1, as the *unnormalized filtering equations*.

We shall show that $\rho(t,x)$ equals the expected value of a function of a suitably defined Markov process $(V(t), w(t))$:

$$\rho(t,x) = \mathbb{E}\left(1_{\{x\}}(V(t))\, w(t)\right),$$

and thereby enabling an unbiased Monte Carlo estimate of $\rho(t,x)$. Here, V has $\mathbb{Z}_+^{n_1}$ as its state space (thus V has the same state space as the unobserved species X) and w is a nonnegative weight.

We provide a weighted Monte Carlo algorithm, also referred to as a *particle filter*, that is tailored to chemical reaction networks. Our algorithm provides an unbiased estimate of $\rho(t,x)$, the unnormalized conditional distribution. The algorithm simulates N_s identically distributed copies of (V,w) where V has $\mathbb{Z}_+^{n_1}$ as its state space and w is a weight such that $\rho(t,x)$ is estimated by

$$\hat{\rho}(t,x) = \frac{1}{N_s} \sum_{i=1}^{N_s} 1_{\{x\}}(V^{(i)}(t))\, w^{(i)}(t).$$

Thus, $\pi(t,x)$ is estimated by

$$\hat{\pi}(t,x) = \frac{\sum_{i=1}^{N_s} 1_{\{x\}}(V^{(i)}(t))\, w^{(i)}(t)}{\sum_{i=1}^{N_s} w^{(i)}(t)}.$$

Due to the division, $\hat{\pi}(t,x)$ is not an unbiased estimator of $\pi(t,x)$. However, by virtue of law of large numbers, as $N_s \to \infty$ we expect $\hat{\pi}(t,x)$ to converge to $\pi(t,x)$. In fact, Monte Carlo methods of filtering in most other contexts (discrete time Markov chains, nonlinear SDEs etc.) also obtain the conditional probabilities by estimating an unnormalized version and then normalizing.

Our weighted Monte Carlo method is concerned with the computation of the conditional distribution of the unobserved species based on exact observation of some species in continuous time. In reality, perfect observations are not possible. However, it is often reasonable to assume that the observation noise is discrete in nature and hence the noise could be incorporated into the model via the introduction of additional species and reactions. For instance, during fluorescence intensity observations, photons produced constitute a species and their production by a fluorescently tagged molecule constitutes a reaction channel. The assumption of continuous in time observation is less realistic. Nevertheless, if the sampling rate is high compared to reaction rates, then we expect this to provide a good approximation. Moreover, having an algorithm for the idealised baseline case of continuous in time observations is expected to provide valuable insights into how to analyse and develop techniques for the case of discrete time observations of a continuous time process.

3.1 Stochastic Reaction Networks and Filtering Equations

We consider a stochastic reaction network with n molecular species and m reaction channels. The process $Z(t) \in \mathbb{Z}_+^n$ stands for the copy number vector of the species at time t. The dynamics of the reaction network are characterized by *propensity functions* $a_j(z, c)$ where $z \in \mathbb{Z}_+^n$ is the state and $c \in \mathbb{R}^p$ is a vector of parameters, and also by the stoichiometric vectors $\nu_j \in \mathbb{Z}^n$ for $j = 1, \ldots, m$. The occurrence of a reaction event j at time t leads to a state change $Z(t) = Z(t-) + \nu_j$ and given $Z(t) = z$, the probability that a reaction event j occurs during $(t, t+h]$ is $a_j(z, c)h + o(h)$ as $h \to 0^+$. We shall adopt the convention that the process $Z(t)$ is right continuous and assume $Z(t)$ to be non-explosive, meaning that there are only finitely many reaction events during any finite time interval. For brevity, we suppress the dependence on parameters except when we are concerned with parameter estimation. It is well known[19, 21, 18, 21] that the time evolution of the (unconditional) probability mass function

$$p(t, z) = P\{Z(t) = z\}$$

is given by the *Kolmogorov's forward equations* also known as the *chemical master equations* (CME):

$$p'(t, z) = \sum_{j=1}^{m} a_j(z - \nu_j)\, p(t, z - \nu_j) - \sum_{j=1}^{m} a_j(z)\, p(t, z). \tag{3.1}$$

We consider the situation where we make exact (noiseless) observations of the

copy number of the last n_2 species in continuous time. We write $Z(t) = (X(t), Y(t))$ where $X(t) \in \mathbb{Z}_+^{n_1}$ is the unobserved component of the state and $Y(t) \in \mathbb{Z}_+^{n_2}$ is the observed component of the state. Suppose we make a particular observation $Y(s) = y(s)$ for $0 \leq s \leq t$ for $t \geq 0$. We are interested in computing the *conditional probability*

$$\pi(t, x) = P\{X(t) = x \,|\, Y(s) = y(s),\, 0 \leq s \leq t\} \quad \forall x \in \mathbb{Z}_+^{n_1}. \tag{3.2}$$

Let us denote by ν_j' the first n_1 components and by ν_j'' the last n_2 components of ν_j for $j = 1, \ldots, m$ and define the subset $\mathcal{O} \subset \{1, \ldots, m\}$ consisting of *observable reaction channels*, that is those that alter $Y(t)$. Thus $j \in \mathcal{O}$ if and only if $\nu_j'' \neq 0$. We denote by $\mathcal{U}$, the complement of $\mathcal{O}$. Thus $\mathcal{U}$ consists of the *unobservable reaction channels*.

We shall denote by $t_k, k = 1, 2, \ldots$ the successive jump times of $y(t)$ and let $t_0 = 0$. Let's define $a^{\mathcal{O}}(x, y)$ and $a^{\mathcal{U}}(x, y)$ by

$$a^{\mathcal{O}}(x, y) = \sum_{j \in \mathcal{O}} a_j(x, y), \quad a^{\mathcal{U}}(x, y) = \sum_{j \in \mathcal{U}} a_j(x, y), \tag{3.3}$$

which are respectively the total propensity of the observable and the unobservable reactions. For each $k \in \mathbb{N}$ define $\mathcal{O}_k$ by

$$\mathcal{O}_k = \{j \in \mathcal{O} \,|\, \nu_j'' = y(t_k) - y(t_k-)\}.$$

Thus $\mathcal{O}_k$ is the subset of observable reactions that are consistent with the observed

jump $y(t_k) - y(t_k-) = y(t_k) - y(t_{k-1})$ at time t_k.

The time evolution of $\pi(t,x)$ follows a system of differential equations in between jump times t_k and at the jump times $\pi(t,x)$ undergoes jumps. A rigorous derivation of the evolution equations for $\pi(t,x)$ for finite state Markov processes is given in [10], and [12] provides a proof for potentially infinite state space assuming non-explosive condition. Here we provide a more intuitive derivation in the Appendix. We summarize these equations which we call the *filtering equations* here.

For $k = 0, 1, \ldots,$ and for $t_k \leq t < t_{k+1}$, π satisfies the following system of differential equations:

$$\begin{aligned} \pi'(t,x) = &\sum_{j\in\mathcal{U}} \pi(t, x-\nu_j')\, a_j(x-\nu_j', y(t_k)) - \sum_{j\in\mathcal{U}} \pi(t,x)\, a_j(x, y(t_k)) \\ &- \pi(t,x)\left(a^{\mathcal{O}}(x, y(t_k)) - \sum_{\tilde{x}} a^{\mathcal{O}}(\tilde{x}, y(t_k))\pi(t,\tilde{x}) \right) \quad \forall x \in \mathbb{Z}_+^{n_1} \end{aligned} \tag{3.4}$$

and for $k = 1, 2, \ldots,$ and at times t_k, $\pi(t,x)$ jumps according to:

$$\pi(t_k, x) = \frac{\sum_{l\in\mathcal{O}_k} a_l(x-\nu_l', y(t_{k-1}))\, \pi(t_k-, x-\nu_l')}{\sum_{\tilde{x}} \sum_{l\in\mathcal{O}_k} a_l(\tilde{x}, y(t_{k-1}))\, \pi(t_k-, \tilde{x})} \quad \forall x \in \mathbb{Z}_+^{n_1}. \tag{3.5}$$

We note that $\pi(t,x)$ is right continuous since by our convention the process $Z(t)$ is right continuous and as a consequence the observed trajectory $y(t)$ is right continuous as well.

Since we are interested in the case where the state space of the system of ODEs (3.4) is either an infinite or a very large finite subset of $\mathbb{Z}_+^{n_1}$, a direct solution of this equation is not practical. Instead our goal is a Monte Carlo simulation. Our

Monte Carlo algorithm involves generating a trajectory $V(t)$ with same state space $\mathbb{Z}_+^{n_1}$ as $X(t)$, along with a weight trajectory $w(t)$, such that at any time $t \geq 0$ and for $x \in \mathbb{Z}_+^{n_1}$

$$\pi(t,x) = \frac{\mathbb{E}[w(t)1_{\{x\}}(V(t))]}{\mathbb{E}[w(t)]}. \tag{3.6}$$

We note that, given a set A, 1_A is the indicator function of the set. In practice, from a sample of N_s identically distributed trajectories $V^{(i)}$ along with weights $w^{(i)}$ for $i = 1, \ldots, N_s$, $\pi(t,x)$ is estimated by

$$\hat{\pi}(t,x) = \frac{\sum_{i=1}^{N_s} 1_{\{x\}}(V^{(i)}(t))\, w^{(i)}(t)}{\sum_{i=1}^{N_s} w^{(i)}(t)} \tag{3.7}$$

Since (3.4) is not linear, π cannot be interpreted as the probability mass function of some Markov process. However, we can define a related quantity $\rho(t,x)$ which evolves according to a linear equation and can be related to a Markov process. We define $\rho(t,x)$ to be a nonnegative function that satisfies the *unnormalized filtering equations* defined as below.

For $k = 0, 1, \ldots$ on the interval $t_k \leq t < t_{k+1}$, $\rho(x,t)$ satisfies

$$\begin{aligned} \rho'(t,x) = &\sum_{j \in \mathcal{U}} \rho(t, x-\nu_j')\, a_j(x-\nu_j', y(t_k)) - \sum_{j \in \mathcal{U}} \rho(t,x)\, a_j(x, y(t_k)) \\ &- \rho(t,x)\, a^{\mathcal{O}}(x, y(t_k)) \quad \forall x \in \mathbb{Z}_+^{n_1}. \end{aligned} \tag{3.8}$$

For $k = 1, 2, \ldots$ at jump times t_k, $\rho(t,x)$ jumps according to

$$\rho(t_k, x) = \frac{1}{|\mathcal{O}_k|} \sum_{j \in \mathcal{O}_k} a_j(x-\nu_j', y(t_{k-1}))\, \rho(t_k-, x-\nu_j') \quad x \in \mathbb{Z}_+^{n_1}. \tag{3.9}$$

We note that $|\mathcal{O}_k|$ is the number of reaction channels in the set $\mathcal{O}_k$ and this is non-zero. Let $\bar{\rho}(t) = \sum_{\tilde{x}} \rho(t, \tilde{x})$. It is shown in Appendix A.2 that π is given by $\pi(t, x) = \rho(t, x)/\bar{\rho}(t)$. In the next section we shall see that an appropriate stochastic interpretation of $\rho(t, x)$ leads to the stochastic simulation algorithm to generate V and w such that

$$\mathbb{E}\left[1_{\{x\}}(V(t))\, w(t)\right] = \rho(t, x). \tag{3.10}$$

3.2 Monte Carlo Algorithms

3.2.1 Weighted Monte Carlo and Resampling

In Section 3.2.2 we shall define a pair of processes (V, w) with state space $\mathbb{Z}_+^{n_1} \times [0, \infty)$ such that (3.10) holds where ρ satisfies the unnormalized filtering equations (3.8) and (3.9). The estimation of $\pi(t, x)$ is accomplished by simulating N_s identically distributed copies of the pair of processes (V, w) and estimating $\pi(t, x)$ via (3.7).

We note that this type of procedure differs from the most common form of simulating a reaction network via N_s i.i.d. trajectories, say $Z^{(i)}$ and then estimating $\mathbb{E}[\varphi(Z(t))]$, the expectation of a function φ of the state via the sample average

$$\frac{1}{N_s} \sum_{i=1}^{N_s} \varphi(Z^{(i)}(t)).$$

In this most familiar form of Monte Carlo simulation, all trajectories are weighted equally. In our situation, the Monte Carlo algorithm involves a weighted average.

The weights carry information about the relative importance of trajectories of V.

Corresponding to the sample $(V^{(i)}, w^{(i)})$ where $i = 1, \ldots, N_s$, we may associate the (random) empirical measure (mass) $M(t)$ at time t given by

$$M(t) = \frac{1}{N_s} \sum_{i=1}^{N_s} w^{(i)}(t)\, \delta_{V^{(i)}(t)}, \tag{3.11}$$

where δ_x stands for the Dirac mass or unit point mass concentrated at $x \in \mathbb{Z}_+^{n_1}$. If we choose (V, w) to satisfy (3.10), then for each $x \in \mathbb{Z}_+^{n_1}$ the expected value of the empirical measure of the singleton $\{x\}$ is $\rho(t, x)$:

$$\mathbb{E}[M(t)(\{x\})] = \rho(t, x). \tag{3.12}$$

The fact that the weights may grow (or decay) unevenly as time progresses leads to two different issues. The first issue is that the simulation may run into numerical problems where some weights become very large while others become very small, exceeding the finite precision of the computer. Especially, very large weights will become `Inf` in the finite precision representation of the computer, rendering computations impractical. The second issue is more fundamental and is unrelated to the finite precision nature of computations. The accuracy of the Monte Carlo estimate of $\pi(t, x)$ depends on the variance of $w(t)$ as well as the variance of $1_{\{x\}}(V(t))\, w(t)$ and also the covariance between these two terms. If the algorithm allows for a large variance in $w(t)$, naturally this would have negative implications for the accuracy of the estimate.

In order to avoid these issues, one may *resample* the empirical measure at certain time points to produce a new empirical measure that is still the sum of N_s Dirac masses but with equal weights. By resampling, we mean the creation a new weighted sample of size N_s from an existing weighted sample of size N_s that captures the information in the original sample. In particular, if an empirical measure $M(t)$ at time t is resampled to produce another empirical measure $\tilde{M}(t)$, then we want the resampling procedure to satisfy the basic condition that

$$\mathbb{E}[\tilde{M}(t)(\{x\})] = \mathbb{E}[M(t)(\{x\})] \quad \forall x \in \mathbb{Z}_+^{n_1}. \tag{3.13}$$

In order to maintain above requirement, the resampling procedure itself introduces extra randomness which in turn can result in loss of accuracy. Thus the frequency with which resampling is performed will be an important topic of investigation. We also note that the resampling procedure introduces dependence among the sample trajectories. Thus, while $(V^{(i)}, w^{(i)})$ for $i = 1, \ldots, N_s$ are identically distributed, the collection is not necessarily independent.

Suppose that at time $t_0 > 0$, we have simulated pairs $(V^{(i)}(t_0), w^{(i)}(t_0))$ for $i = 1, \ldots, N_s$, and we resample to produce a new set of pairs $(\tilde{V}^{(i)}(t_0), \tilde{w}^{(i)}(t_0))$ for $i = 1, \ldots, N_s$. Then we shall have that for all i and j,

$$\mathbb{E}[1_{\{x\}}(\tilde{V}^{(j)}(t_0))\, \tilde{w}^{(j)}(t_0)] = \mathbb{E}[1_{\{x\}}(V^{(i)}(t_0))\, w^{(i)}(t_0)],$$

and $\tilde{w}^{(j)}(t_0) = 1$ for $j = 1, \ldots, N_s$. The consequent evolution of $(\tilde{V}^{(j)}(t), \tilde{w}^{(j)}(t))$ for

$j = 1, \ldots, N_s$ (which we shall rename $(V^{(j)}, w^{(j)})$), will proceed in a conditionally independent fashion until the next resampling.

We shall employ a particular resampling algorithm described in Refs. [4] and [11]. This is given in Algorithm 4 and involves regarding each sample point/particle as giving birth to a number (possibly zero) of offsprings. It is most convenient to think of this algorithm as propagating N_s particles so that at time t, the ith particle is situated at location $V^{(i)}(t) \in \mathbb{Z}_+^{n_1}$ and has weight $w^{(i)}(t) \geq 0$. When resampling occurs at say, time t_0, Algorithm 4 considers the weight of each particle i and assigns a random number $o^{(i)} \in \mathbb{Z}_+$ of offsprings to the particle. The number $o^{(i)}$ is chosen such that $\mathbb{E}[o^{(i)} \,|\, w^{(i)}(t_0)] = w^{(i)}(t_0)$ and subject to the condition that the total number of offsprings of all particles is N_s. Each particle i is considered "dead" after it gives "birth" to $o^{(i)}$ offsprings each of which start at the same location $V^{(i)}(t_0)$ as its parent and with weight $w^{(i)}(t_0) = 1$. Subsequently all the offsprings evolve independently. We refer the reader to [4] (Section 9.2) and [11] regarding further details of a specific version of this algorithm that is also used by us and described in Algorithm 4.

3.2.2 Processes V and w

In this section we define a pair of processes (V, w) where $V(t) \in \mathbb{Z}_+^{n_1}$ and $w(t) \geq 0$ as follows. We initialize $V(0)$ with the same distribution as that of the initial state $X(0) = x_0$ and set $w(0) = 1$. In between jump times, that is, for $t_k \leq t < t_{k+1}$ where $k = 0, 1, \ldots,$ the process V is evolved according to the underlying

reaction network with *all observable reactions removed*. Thus only the copy number of the first n_1 species can change during (t_k, t_{k+1}). Moreover, for $t_k \leq t < t_{k+1}$ where $k = 0, 1, \ldots$, the weight process $w(t)$ is evolved according to

$$w(t) = w(t_k) \exp\left\{ - \int_{t_k}^{t} a^{\mathcal{O}}(V(s), y(t_k))\, ds \right\}. \tag{3.14}$$

Equivalently, we note that $w(t)$ satisfies the ODE

$$w'(t) = w(t)\, a^{\mathcal{O}}(V(t), y(t_k))$$

for $t_k \leq t < t_{k+1}$ (at $t = t_k$ the right-hand side derivative is considered). At jump times $t = t_k$ for $k = 1, 2, \ldots$ the process $w(t)$ jumps as follows. For $j \in \mathcal{O}_k$, set

$$w(t_k) = w(t_k-)\, a_j(V(t_k-), y(t_{k-1})) \quad \text{with probability } \frac{1}{|\mathcal{O}_k|}. \tag{3.15}$$

Moreover, at jump times $t = t_k$ for $k = 1, 2, \ldots$ the process $V(t)$ jumps as follows. For $j \in \mathcal{O}_k$, set

$$V(t_k) = V(t_k-) + \nu'_j \quad \text{with probability } \frac{1}{|\mathcal{O}_k|}. \tag{3.16}$$

We note that if $a_j(V(t_k-), y(t_{k-1})) = 0$ for the chosen $j \in \mathcal{O}_k$, it may be possible that $V(t_k)$ is assigned an infeasible state, that is a state that has negative components. However, at the same time $w(t_k)$ would be zero and hence the value of $V(t_k)$ would not matter.

Our goal is to show that

$$\mathbb{E}[1_{\{x\}}(V(t))\ w(t)] = \rho(t,x)$$

for all t and x, where ρ is defined by (3.8) and (3.9). The following two lemmas basically establish that.

Lemma 4. *Let V and w be defined as above, and let ρ defined by*

$$\rho(t,x) = \mathbb{E}[1_{\{x\}}(V(t))\, w(t)]. \tag{3.17}$$

If V is non-explosive then $\rho(t,x)$ satisfies (3.8) on $t_k \le t < t_{k+1}$.

Proof. As a function of t, $1_{\{x\}}(V(t))$ is of bounded variation and piecewise constant on every bounded interval of time. Moreover, $w(t)$ is absolutely continuous in t. Thus we may write for $t_k \le t < t_{k+1}$

$$\begin{aligned}
1_{\{x\}}(V(t))\, w(t) &= 1_{\{x\}}(V(t_k))\, w(t_k) + \int_{t_k}^{t} 1_{\{x\}}(V(s))\, w'(s) ds \\
&\quad + \sum_{t_k < s \le t} w(s)\ \left(1_{\{x\}}(V(s)) - 1_{\{x\}}(V(s-))\right) \\
&= 1_{\{x\}}(V(t_k))\, w(t_k) - \int_{t_k}^{t} 1_{\{x\}}(V(s))\, w(s)\, a^{\mathcal{O}}(V(s), y(t_k))\, ds \\
&\quad + \sum_{j \in \mathcal{U}} \sum_{t_k < s \le t} w(s)\ \left(1_{\{x\}}(V(s-) + \nu'_j) - 1_{\{x\}}(V(s-))\right)\ (R_j(s) - R_j(s-))\,,
\end{aligned}$$

where R_j is the process that counts the number of firings of reaction channel j during $(0,t]$. Since the stochastic intensity [8, 3] of R_j is given by $a_j(V(t-))$, taking

expectations we obtain

$$\rho(t,x) = \rho(t_k,x) + \sum_{j\in\mathcal{U}} \mathbb{E}\left[\int_0^t \left(1_{\{x\}}(V(s)+\nu_j') - 1_{\{x\}}(V(s))\right)\, w(s)\, a_j(V(s),y(t_k))\, ds\right]$$
$$- \mathbb{E}\left[\int_0^t 1_{\{x\}}(V(s))\, w(s)\, a^{\mathcal{O}}(V(s),y(t_k))\, ds\right].$$

Using Fubini and differentiating with respect to t, We obtain that

$$\begin{aligned}
\rho'(t,x) &= \sum_{j\in\mathcal{U}} \mathbb{E}[(1_{\{x\}}(V(t)+\nu_j') - 1_{\{x\}}(V(t)))\, a_j(V(t),y(t_k))\, w(t)] \\
&\quad - \mathbb{E}[1_{\{x\}}(V(t))\, w(t)\, a^{\mathcal{O}}(V(t),y(t_k))] \\
&= \sum_{j\in\mathcal{U}} \mathbb{E}[1_{\{x-\nu_j'\}}(V(t))\, a_j(x-\nu_j',y(t_k))\, w(t)] - \sum_{j\in\mathcal{U}} \mathbb{E}[1_{\{x\}}(V(t)))\, a_j(x,\nu_j')\, w(t)] \\
&\quad - \mathbb{E}[1_{\{x\}}(V(t))\, w(t)\, a^{\mathcal{O}}(x,y(t_k))] \\
&= \sum_{j\in\mathcal{U}} \rho(t,x-\nu_j')\, a_j(x-\nu_j',y(t_k)) - \sum_{j\in\mathcal{U}} \rho(t,x)\, a_j(x,y(t_k)) - \rho(t,x)\, a^{\mathcal{O}}(t,x),
\end{aligned}$$

which agrees with (3.8). □

Lemma 5. *Let $k \in \{1,2,\dots\}$ and suppose*

$$\mathbb{E}[1_{\{x\}}(V(t_k-))\, w(t_k-)] = \rho(t_k-,x)$$

for all x. Then

$$\mathbb{E}[1_{\{x\}}(V(t_k))\, w(t_k)] = \rho(t_k,x)$$

Proof. It follows that the conditional expectation

$$\begin{aligned}&\mathbb{E}[w(t_k)\,1_{\{x\}}(V(t_k))\,|\,w(t_k-),V(t_k-)]\\&\quad=\frac{1}{|\mathcal{O}_k|}\sum_{j\in\mathcal{O}_k} w(t_k-)\,a_j(x-\nu_j',y(t_{k-1}))\,1_{\{x-\nu_j'\}}(V(t_k-)).\end{aligned}$$

Taking expectation, we obtain

$$\mathbb{E}[w(t_k)\,1_{\{x\}}(V(t_k))]=\frac{1}{|\mathcal{O}_k|}\sum_{j\in\mathcal{O}_k}\rho(t_k-,x-\nu_j')\,a_j(x-\nu_j',y(t_{k-1})).$$

The result follows from (3.9). □

From the above lemmas and mathematical induction, it follows that

$$\mathbb{E}[1_{\{x\}}(V(t))\,w(t)]=\rho(t,x),$$

for all $t\geq 0$ and x.

Algorithm 1 describes the overall filtering algorithm whereas Algorithm 2 describes the continuous evolution and Algorithm 3 describes the jumps at observed jump times t_k. Algorithm 4 describes the resampling via offsprings [4]. We observe that in Algorithm 1, resampling is only performed at the observed jump times t_k and not necessarily always. In Section 3.4.5 we numerically explore different resampling strategies. One extreme option is always to resample at t_k, the other extreme is never to resample. The third option is to resample at t_k if either the number of zero weights among the N_s particles exceeds a predetermined number or if the ratio

of the largest weight to the smallest non-zero weight among the particles exceeds a predetermined value. If resampling is not performed at t_k, then the weights are rescaled so that the average weight is 1. This latter step prevents all weights from growing to be too large or too small. We also note that it is possible to consider resampling at time points which may not coincide with the observed jump times t_k. However, since the only point in time where some weights may become zero is at the jump times t_k, it makes sense to consider the jump times as the points in time to determine if a resampling step is required.

Algorithm 1 Overall scheme

Input: Jump times of Y $(t_1, \ldots, t_{N_k})$ and observed Y at jump times $(y_1, \ldots, y_{N_k})$. Initial distribution μ_0 for $X(0) = x_0$, parameter value c, final time T, filter sample size N_s.
Generate N_s i.i.d sample from μ_0 and assign to $V^{(1)}, \ldots, V^{(N_s)}$. Set $w^{(i)} = 1$ for $i = 1, \ldots, N_s$.
$k = 1$, $t = 0$, $y = y_0$
for $k = 1$ to N_k **do**
    for $i = 1$ to N_s **do**
        $(V_-^{(i)}, w_-^{(i)}) = \text{Continuous-evolution}(V^{(i)}, w^{(i)}, t, t_k, y, c)$
        $(V^{(i)}, w^{(i)}) = \text{Jump}(V_-^{(i)}, w_-^{(i)}, y, y_k, c)$
    end for
    if resampling **then** $(V, w) = \text{Offsprings}(V, w)$
    else rescale w so that $\sum_{i=1}^{N_s} w^{(i)} = N_s$
    end if
    Set $t = t_k$, $y = y_k$
end for
for $i = 1$ to N_s **do**
    $(V^{(i)}, w^{(i)}) = \text{Continuous-evolution}(V^{(i)}, w^{(i)}, t_{N_k}, T, y_{N_k}, c)$
end for

3.2.3 Treating Parameters as Random Variables

In general, the propensity functions $a_j(x, y)$ depend on a vector c of parameters, thus $a_j = a_j(x, y, c)$ where $c \in \mathbb{R}^p$. Here we consider a Bayesian framework

Algorithm 2 Continuous evolution

function CONTINUOUS-EVOLUTION(V_0, w_0, t_0, t_f, y, c)
  Assume unobservable reactions are numbered $1, 2, \ldots, m_u$
  Assume observable reactions are numbered $m_u + 1, \ldots, m$
  Set $V = V_0$, $w = w_0$, $t = t_0$
  while $t < t_f$ **do**
    for $j = 1$ to m_u **do**
      Set $\lambda_j = a_j(V, y, c)$
    end for
    Set $\lambda^{\mathcal{O}} = \sum_{j=m_u+1}^{m} \lambda_j$
    Set $\lambda^{\mathcal{U}} = \sum_{j=1}^{m_u} \lambda_j$
    Generate $\tau \sim$ Exponential($\lambda^{\mathcal{U}}$)
    if $t + \tau < t_f$ **then**
      Generate $u \sim$ Uniform$[0, 1]$
      Find $j \in \{1, \ldots, m_u\}$ such that $\sum_{l=1}^{j-1} \lambda_l < u\lambda^{\mathcal{U}} \leq \sum_{l=1}^{j} \lambda_l$
      Set $V = V + \nu'_j$
    end if
    Set $t_n = \min\{t + \tau, t_f\}$
    Set $w = w \times \exp\{-(t_n - t)\, \lambda^{\mathcal{O}}\}$
    Set $t = t_n$
  end while
  return (V, w)
end function

for inferring the parameters c from the observation $Y(s) = y(s)$ for $0 \leq s \leq t$. This involves treating c as a random variable, or rather a stochastic process $C(t)$ which remains constant in time $t \geq 0$. Thus, we may "absorb" $C(t)$ into $X(t)$, thus expanding the unobserved components of the state by p extra dimensions. The Bayesian framework involves starting with a prior distribution $\bar{\mu}$ on the parameter space $\mathbb{R}^p$. Then we compute the posterior distribution probability mass/density function $\pi(t, x, c)$ which is characterized by

$$P\{X(t) = x, C(t) \in A \,|\, Y(s) = y(s),\, 0 \leq s \leq t\} = \int_{c \in A} \pi(t, x, c)\, dc.$$

Algorithm 3 Jump

function JUMP(V_-, w_-, y_-, y, c)
  Assume unobservable reactions are numbered $1, 2, \ldots, m_u$
  Assume observable reactions are numbered $m_u + 1, \ldots, m$
  Find indices $j \in \{m_u + 1, \ldots, m\}$ that satisfy $\nu''_j = y - y_-$.
  Let above indices be stored in array J of length L.
  for $i = 1$ to L **do**
    Set $\lambda_i = a_{J(i)}(V_-, y_-, c)$
  end for
  if $L > 1$ **then**
    Generate $u \sim$ Uniform$[0, 1]$
    Let $i \in \{1, \ldots, L\}$ be such that $i - 1 < u\, L \leq i$
  else
    Set $i = 1$
  end if
  Set $V = V_- + \nu'_{J(i)}$
  Set $w = w_- \, \lambda_{J(i)}$
  return (V, w)
end function

Since the parameters are distributed continuously, the filtering equations (3.4) and (3.5) need to be modified as follows. For $t_k \leq t < t_{k+1}$, π satisfies the following system of differential equations:

$$\begin{aligned}\pi'(t,x,c) = &\sum_{j\in\mathcal{U}} \pi(t, x-\nu'_j, c)\, a_j(x-\nu'_j, y(t_k), c) - \sum_{j\in\mathcal{U}} \pi(t,x,c)\, a_j(x, y(t_k), c) \\ &- \pi(t,x,c)\left(a^{\mathcal{O}}(x, y(t_k), c) - \int_{\tilde{c}} \sum_{\tilde{x}} a^{\mathcal{O}}(\tilde{x}, y(t_k), \tilde{c}) \pi(t, \tilde{x}, \tilde{c}) d\tilde{c}\right),\end{aligned} \tag{3.18}$$

which holds for all $x \in \mathbb{Z}_+^{n_1}$ and $c \in \mathbb{R}^p$. At times t_k, $\pi(t, x, c)$ jumps according to:

$$\pi(t_k, x, c) = \frac{\sum_{l\in\mathcal{O}_k} a_l(x - \nu'_l, y(t_{k-1}), c)\, \pi(t_k-, x-\nu'_l, c)}{\int_{\tilde{c}} \sum_{\tilde{x}} \sum_{l\in\mathcal{O}_k} a_l(\tilde{x}, y(t_{k-1}), \tilde{c})\, \pi(t_k-, \tilde{x}, \tilde{c}) d\tilde{c}}, \tag{3.19}$$

which holds for all $x \in \mathbb{Z}_+^{n_1}$ and $c \in \mathbb{R}^p$.

Algorithm 5 describes the Monte Carlo algorithm to generate a weighted sam-

Algorithm 4 Offsprings

function OFFSPRINGS(V, w)
    We note that $V = (V_1, \ldots, V_{N_s})$ and $w = (w_1, \ldots, w_{N_s})$. We denote $[x]$ by the integer part of a number x, while $\{x\} = x - [x]$ the fractional part of x.
    Normalize $\bar{w}_i = w_i / \sum_{i=1}^{N} w_i$.
    Simulate i.i.d random variables $u_i \sim$ Unif[0,1], for $i = 1, ..., N-1$
    Initialize $g = N, h = N$
    for $i = 1$ to $N - 1$ **do**
        if $\{N\bar{w}_i\} + \{g - N\bar{w}_i\} < 1$ **then**
            if $u_i < 1 - (\{N\bar{w}_o\}/\{g\})$ **then**
                $o_i = [N\bar{w}_i]$
            else
                $o_i = [N\bar{w}_i] + (h - [g])$
            end if
        else
            if $u_i < 1 - \frac{1-\{N\bar{w}_i\}}{1-\{g\}}$ **then**
                $o_i = [N\bar{w}_i] + 1$
            else
                $o_i = [N\bar{w}_i] + (h - [g])$
            end if
        end if
    end for
    We now have o_i copies of particle i at location $V^{(i)}$. Reindex the positions of the particles as $v_1(t), v_2(t), ..., v_N(t)$
    return $(\tilde{V}, w)$
end function

ple from the posterior distribution for the parameter. Having computed $C^{(i)}$ and $w^{(i)}$ for $= 1, \ldots, N_s$, we may estimate the posterior mean $\bar{C}$ and standard deviation σ_C by

$$\bar{C} = \frac{\sum_{i=1}^{N_s} w^{(i)} C^{(i)}}{\sum_{i=1}^{N_s} w^{(i)}}, \tag{3.20}$$

and

$$\sigma_C^2 = \frac{\sum_{i=1}^{N_s} w^{(i)} (C^{(i)} - \bar{C})^2}{\sum_{i=1}^{N_s} w^{(i)}}. \tag{3.21}$$

Algorithm 5 Overall scheme for Bayesian inference

Input: Jump times of Y $(t_1, \ldots, t_{N_k})$ and observed Y at jump times $(y_1, \ldots, y_{N_k})$. Initial distribution μ_0 for $X(0) = x_0$, prior distribution $\bar{\mu}$ for parameter c, final time T, filter sample size N_s.
Generate a sample x_0 from μ_0 and a sample c from $\bar{\mu}$. Set $V^{(i)} = x_0$, $w^{(i)} = 1$, $C^{(i)} = c$ for $i = 1$ to N_s
Set $k = 1$, $t = 0$, $y = y_0$
for $k = 1$ to N_k **do**
    for $i = 1$ to N_s **do**
        $(V_-^{(i)}, w_-^{(i)}) = \text{Continuous-evolution}(V^{(i)}, w^{(i)}, t, t_k, y, C^{(i)})$
        $(V^{(i)}, w^{(i)}) = \text{Jump}(V_-^{(i)}, w_-^{(i)}, y, y_k, C^{(i)})$
    end for
    Set $U^{(i)} = (V^{(i)}, C^{(i)})$ for $i = 1$ to N_s
    if resampling **then** $(U, w) = \text{Offsprings}(U, w)$
    else rescale w so that $\sum_{i=1}^{N_s} w^{(i)} = N_s$
    end if
    Set $(V^{(i)}, C^{(i)}) = U^{(i)}$ for $i = 1$ to N_s
    Set $t = t_k$, $y = y_k$
end for
for $i = 1$ to N_s **do**
    $(V^{(i)}, w^{(i)}) = \text{Continuous-evolution}(V^{(i)}, w^{(i)}, t_{N_k}, T, y_{N_k}, C^{(i)})$
end for

3.2.4 Estimating Past State

Besides the real-time update of the conditional probability mass function $\pi(t, x) = P\{X(t) = x \,|\, Y(s) = y(s), 0 \le s \le t\}$, we would like to look back and consider estimation of the state at time t with the observation made until a later time T. A special case is when the initial state of the system is not known precisely, but rather a prior distribution is the best of our knowledge.

To tackle the problem of finding

$$P\{X(t_0) = x \,|\, Y(s) = y(s),\, 0 \le s \le T\}$$

with $0 \le t_0 \le T$, we expand the state of the underlying process as $(X, \tilde{X}, Y)$ where

$\tilde{X}$ is simply a copy of X until time t_0 and there after $\tilde{X}(t)$ remains fixed at the value $\tilde{X}(t_0)$.

The process $(X, \tilde{X}, Y)$ with state space $\mathbb{Z}_+^{n_1} \times \mathbb{Z}_+^{n_1} \times \mathbb{Z}_+^{n_2}$ will be a piecewise time homogeneous Markov process in the sense that on the interval $[0, t_0]$ it will evolve according to a certain reaction network and during $(t_0, T]$ it will evolve according to a different reaction network. During $[0, t_0]$ whenever X jumps by ν'_j, $\tilde{X}$ also jumps by $\tilde{\nu}'_j = \nu'_j$. However, during $(t_0, T]$ the stoichiometric vectors $\tilde{\nu}'_j$ corresponding to $\tilde{X}$ are all zero indicating no change in state. Based on this, the original filtering equations (3.4), (3.5), (3.8) and (3.9) are still valid for this extended process. We are simply estimating

$$P\{\tilde{X}(T) = x \,|\, Y(s) = y(s),\ 0 \le s \le T\}$$

which equals

$$P\{X(t_0) = x \,|\, Y(s) = y(s),\ 0 \le s \le T\}$$

since $\tilde{X}(T) = X(t_0)$. Algorithm 6 describes this.

3.3 Estimation Error

In this section, we discuss two notions of error in state or parameter estimation. The first is a measure of the error in the estimated conditional distribution and the true conditional distribution while the second is the L^2 error in estimating the state or a parameter.

Algorithm 6 Past state estimation

Input: Jump times of Y $(t_1, \ldots, t_{N_k})$ and observed Y at jump times $(y_1, \ldots, y_{N_k})$. Initial distribution μ_0 for $X(0) = x_0$, parameter value c, intermediate time t_0, final time T, filter sample size N_s.
Set $k = 1$, $t = 0$, $y = y_0$, $k_1 = \max\{k : t_k \leq t_0\}$.
for $k = 1$ to k_1 **do**
    for $i = 1$ to N_s **do**
        $(V_-^{(i)}, w_-^{(i)}) = \text{Continuous-evolution}(V^{(i)}, w^{(i)}, t, t_k, y, c)$
        $(V^{(i)}, w^{(i)}) = \text{Jump}(V_-^{(i)}, w_-^{(i)}, y, y_k, c)$
    end for
    if resampling **then** $(V, w) = \text{Offsprings}(V, w)$
    else rescale w so that $\sum_{i=1}^{N_s} w^{(i)} = N_s$
    end if
    Set $t = t_k$, $y = y_k$
end for
$(V^{(i)}, w^{(i)}) = \text{Continuous-evolution}(V^{(i)}, w^{(i)}, t, t_0, y_{N_k}, c)$ for $i = 1, \cdots, N_s$
Set $\tilde{V}^{(i)} = V^{(i)}$ for $i = 1, \cdots, N_s$
for $k = k_1 + 1$ to N_k **do**
    for $i = 1$ to N_s **do**
        $(V_-^{(i)}, w_-^{(i)}) = \text{Continuous-evolution}(V^{(i)}, w^{(i)}, t, t_k, y, c)$
        $(V^{(i)}, w^{(i)}) = \text{Jump}(V_-^{(i)}, w_-^{(i)}, y, y_k, c)$
    end for
    Set $U^{(i)} = (V^{(i)}, \tilde{V}^{(i)})$ for $i = 1, \ldots, N_s$
    if resampling **then** $(U, w) = \text{Offsprings}(U, w)$
    else rescale w so that $\sum_{i=1}^{N_s} w^{(i)} = N_s$
    end if
    Set $(V^{(i)}, \tilde{V}^{(i)}) = U^{(i)}$ for $i = 1, \ldots, N_s$
    Set $t = t_k$, $y = y_k$
end for
$(V^{(i)}, w^{(i)}) = \text{Continuous-evolution}(V^{(i)}, w^{(i)}, t, T, y_{N_k}, c)$ for $i = 1, \cdots, N_s$

We first note that the conditional distribution $\pi(t, x)$ is a function of the observed trajectory of Y from 0 to t. For clarity, we write $\pi(t, x, Y_{[0,t]})$. Likewise, we write $\hat{\pi}(t, x, Y_{[0,t]})$ for the estimator.

In order to measure the error between the estimated conditional distribution and the true conditional distribution we introduce the *mean total variation error*

(MTVE):

$$\mathrm{MTVE} = \mathbb{E}\left(\sum_{x\in\mathbb{Z}_+^{n_1}} |\hat{\pi}(t,x,Y_{[0,t]}) - \pi(t,x,Y_{[0,t]})|\right). \tag{3.22}$$

We also introduce the *conditional mean total variation error* (CMTVE):

$$\mathrm{CMTVE} = \mathbb{E}\left(\sum_{x\in\mathbb{Z}_+^{n_1}} |\hat{\pi}(t,x,Y_{[0,t]}) - \pi(t,x,Y_{[0,t]})| \,\Big|\, \mathcal{Y}_t\right), \tag{3.23}$$

where $\mathcal{Y}_t$ is the σ-algebra generated by the observed process Y up to time t. The reader unfamiliar with σ-algebras may regard this as condition on the trajectory of Y up to time t. Thus, CMTVE is a function of the observed trajectory. If we are estimating conditional density of a parameter as opposed to state, the summations in the above definitions need to be replaced by an integrals.

On the other hand, instead of attempting to describe the entire conditional distribution for the (unobserved) state or a parameter, one is frequently interested in obtaining a point estimate. The ideal point estimate of the unobserved state $X(t)$ is the conditional expectation $\mathbb{E}[X(t)\,|\,\mathcal{Y}_t]$. In practice one estimates the latter via the estimator $\mathcal{E}(t)$ defined by

$$\mathcal{E}(t) = \frac{\sum_{i=1}^{N_s} V^{(i)}(t)\, w^{(i)}(t)}{\sum_{i=1}^{N_s} w^{(i)}(t)}. \tag{3.24}$$

Thus the estimation error is given by $e(t) = \mathcal{E}(t) - X(t)$. The quantities of interest are, the conditional bias $\mathbb{E}[e(t)\,|\,\mathcal{Y}_t]$, the bias $\mathbb{E}[e(t)]$, the conditional L^2 error $(\mathbb{E}[e^2(t)\,|\,\mathcal{Y}_t])^{1/2}$ and the L^2 error $(\mathbb{E}[e^2(t)])^{1/2}$.

It is instructive to split $e(t)$ as

$$e(t) = (\mathcal{E}(t) - \mathbb{E}[X(t) \,|\, \mathcal{Y}_t]) + (\mathbb{E}[X(t) \,|\, \mathcal{Y}_t] - X(t)) \,.$$

We make the *important observation that* $\mathcal{E}(t)$ *and* $X(t)$ *are independent when conditioned on* $\mathcal{Y}_t$. Hence

$$\begin{aligned}
&\mathbb{E}\Big\{ \Big(\mathcal{E}(t) - \mathbb{E}[X(t)|\mathcal{Y}_t]\Big) \Big(\mathbb{E}[X(t)|\mathcal{Y}_t] - X(t)\Big) \,\Big|\, \mathcal{Y}_t \Big\} \\
&= \mathbb{E}\Big\{ \Big(\mathcal{E}(t) - \mathbb{E}[X(t)|\mathcal{Y}_t]\Big) \,\Big|\, \mathcal{Y}_t \Big\} \mathbb{E}\Big\{ \Big(\mathbb{E}[X(t)|\mathcal{Y}_t] - X(t)\Big) \,\Big|\, \mathcal{Y}_t \Big\} = 0.
\end{aligned}$$

Hence we may expand

$$\begin{aligned}
\mathbb{E}[e^2(t) \,|\, \mathcal{Y}_t] &= \mathbb{E}\left[(\mathcal{E}(t) - \mathbb{E}[X(t)|\mathcal{Y}_t])^2 \,|\mathcal{Y}_t \right] + \mathbb{E}\left[(\mathbb{E}[X(t)|\mathcal{Y}_t] - X(t))^2 \,|\mathcal{Y}_t \right] \\
&= \mathbb{E}\left[(\mathcal{E}(t) - \mathbb{E}[X(t)|\mathcal{Y}_t])^2 \,|\, \mathcal{Y}_t \right] + \mathrm{Var}\left[X(t) \,|\, \mathcal{Y}_t \right]
\end{aligned} \tag{3.25}$$

We observe that the second term depends only on the filtering problem and not on the filtering algorithm, while the first term depends on the filtering algorithm. We expect as the filter sample size N_s approaches infinity the first term to approach zero in some sense. Taking expectation on (3.25) we obtain that

$$\mathbb{E}[e^2(t)] = \mathbb{E}\left[(\mathcal{E}(t) - \mathbb{E}[X(t)|\mathcal{Y}_t])^2 \right] + \mathbb{E}\left[\mathrm{Var}[X(t) \,|\, \mathcal{Y}_t] \right], \tag{3.26}$$

where as observed earlier, the second term depends only on the filtering problem and provides a lower bound on the L^2 error.

In situations where observations are obtained via simulations as is the case in

this paper, we do know the true value $X(t)$ of the unobserved state and thus $\mathbb{E}[e^2(t)]$ may be estimated as follows. We simulate the original system via the Gillespie algorithm N_r independent times to obtain $(X^{(1)}, Y^{(1)}), \ldots, (X^{(N_r)}, Y^{(N_r)})$. For each such simulation j, we run the overall filter algorithm (with some sample size N_s) for the observed trajectory $Y^{(j)}$ to record $\mathcal{E}_j(t)$, the filter estimate of the conditional expectation. Then we may estimate $\mathbb{E}[e^2(t)]$ as

$$\mathbb{E}[e^2(t)] \approx \frac{1}{N_r} \sum_{j=1}^{N_r} |\mathcal{E}_j(t) - X^{(j)}(t)|^2.$$

On the other hand, for any given observation trajectory, estimation of the conditional error $\mathbb{E}[e^2(t) \,|\mathcal{Y}_t]$ is harder even in situations as in this paper where observations are generated via Monte Carlo simulations. If the filter sample size N_s is very large, we may be justified in approximating $\mathbb{E}[e^2(t) \,|\mathcal{Y}_t]$ by $\mathrm{Var}[X(t) \,|\mathcal{Y}_t]$ according to (3.25). This suggests the following estimation of $\mathbb{E}[e^2(t) \,|\, \mathcal{Y}_t]$. Generate an observation Y of one trajectory and apply the filter with sample size N_s once to obtain an estimate $\mathcal{V}(t)$ of $\mathrm{Var}[X(t) \,|\mathcal{Y}_t]$. Here

$$\mathcal{V}(t) = \frac{\sum_{i=1}^{N_s} (V^{(i)}(t))^2 \, w^{(i)}(t)}{\sum_{i=1}^{N_s} w^{(i)}(t)} - \left(\frac{\sum_{i=1}^{N_s} V^{(i)}(t) \, w^{(i)}(t)}{\sum_{i=1}^{N_s} w^{(i)}(t)} \right)^2 . \tag{3.27}$$

3.4 Numerical Examples

In this section, we illustrate the filtering algorithms via examples. In all examples, observations were made by simulation of the underlying reaction network using the Gillespie algorithm [19] to obtain one or more independent samples of

(X, Y), the unobserved and observed trajectories. Then the filtering algorithms were applied to the observed trajectories.

3.4.1 A Linear Propensity Example

We consider the simple reaction network

$$\begin{aligned} S &\xrightarrow{c_1} S + A, \\ \varnothing &\xrightarrow{c_2} S, \\ S &\xrightarrow{c_3} \varnothing, \end{aligned} \tag{3.28}$$

consisting of two species and three reaction channels and assume mass action form of propensities. Thus the propensities are given by $a_1(z) = c_1 z_2$, $a_2(z) = c_2$, $a_3(z) = c_3 z_2$.

First, consider the case where the copy number of species S is observed exactly while the species A is unobserved. Thus, $Z(t) = (X(t), Y(t)) = (\#A(t), \#S(t))$. In this situation, since the second two reaction channels alone determine the copy number of S, the conditional probability density function $\pi(t, x)$ could be computed exactly as

$$\pi(t, x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where $\lambda = \int_0^t c_1 y(t) dt$. With initial state $\#A = 0$ and $\#S = 5$, parameter values $c = (1, 5, 1)$, and filter sample size $N_s = 10,000$, we ran Algorithm 1, and Figure 3.1 shows the comparison between the conditional distribution computed by the filter with the exact conditional distribution. In order to estimate the CMTVE defined

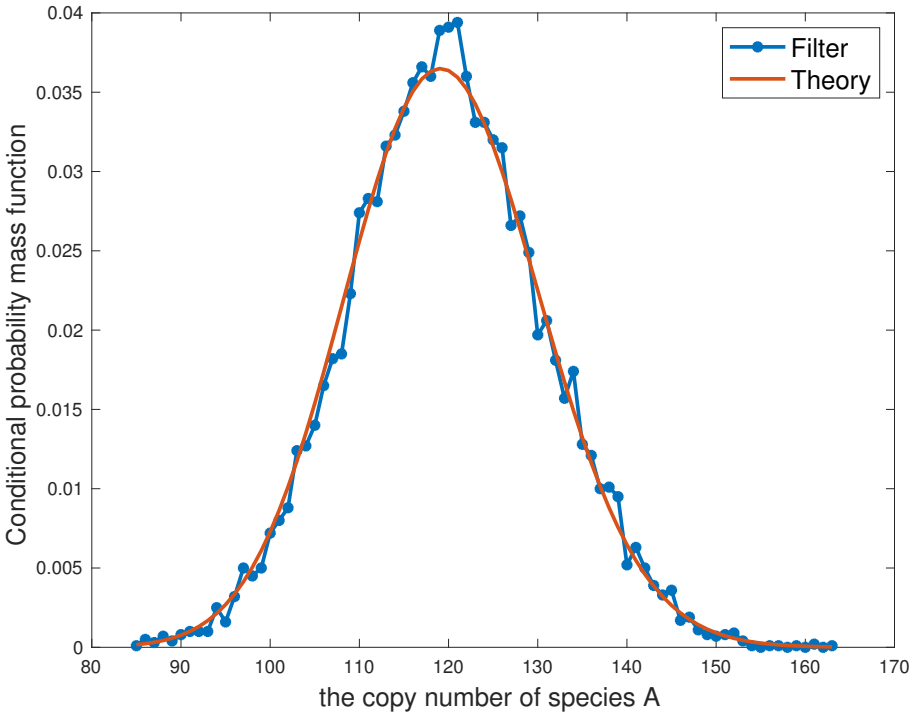


Figure 3.1: The conditional distribution of the number of A at time $T = 20$ conditioned on one observation of the entire trajectory of the copy number of S on the interval $[0, T]$, in the linear propensity example. The estimated conditional distribution $\hat{\pi}(T, x)$ and the exact theoretical conditional distribution $\pi(T, x)$ are shown.

by (3.23), with same initial state, parameter values, final time $T = 20$, filter sample size $N_s = 10,000$, and the same observation trajectory, we ran 100 simulations of algorithm 1. From this we estimated CMTVE for the algorithm to be 0.0475 with a 95%confidence interval of $[0.0461, 0.0488]$.

We also tested the filter's performance on point estimation. To compare estimated state $\mathcal{E}(t)$ with the actual state $X(t)$, as described in Section 3.3, we generated $N_r = 500$ independent realizations of the system (X, Y), and applied the filtering algorithm with filter sample size $N_s = 1000$ to get an estimate $\mathcal{E}(t)$ for each of the realizations. Figure 3.2 shows a scatter plot of the values of $\mathcal{E}(T)$ against $X(T)$. We note that $\mathcal{E}(T)$ is a biased estimator whose bias is expected to approach zero as N_s tends to infinity. The fact that the regression line does not have slope 1 is due to this as well as due to finite sample size of N_r.

Next, with the same system, we considered observing species A rather than species S. So $Z(t) = (X(t), Y(t)) = (\#S(t), \#A(t))$. We kept the initial state and parameter values of c_1 and c_3 to be the same as before and performed a Bayesian estimation of parameter c_2. Thus we considered c_2 as a random variable C_2 with a uniform prior distribution on $[4, 6]$. We randomly generated a sample of $N_r = 1000$ values of C_2 following uniform distribution $[4, 6]$ and generated one observation trajectory Y for each value of C_2. Then we applied Algorithm 5 to compute $\bar{C}_2(t)$, the filter estimate of $\mathbb{E}(C_2 \,|\mathcal{Y}_t)$ (see (3.20). The result is shown in Figure 3.3.

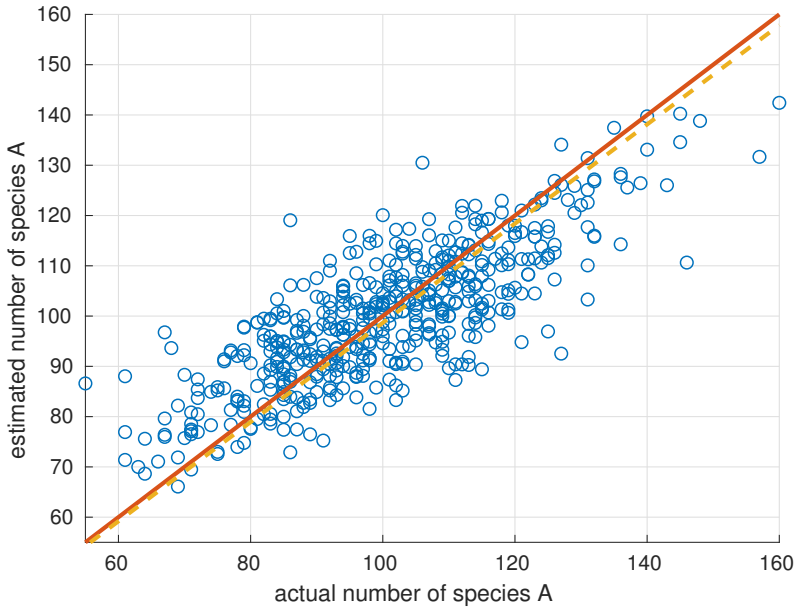


Figure 3.2: Scatter plot of the point estimates $\mathcal{E}(T)$ of the copy number of species A at $T = 20$ (based on the observation of species S on the interval $[0, T]$) against the actual copy numbers $X(T)$ in the linear propensity example. In each trial, a realization of the system (X, Y) was generated, and we computed the point estimates $\mathcal{E}(T)$ of the copy number of species A at time T based on the trajectory of species S. 500 independent trials were performed. The dashed yellow line is the regression line and solid red line is the line with slope 1. The bias is -0.2810 with a 95% confidence interval of $[-1.20427, 0.642256]$ the estimated L^2 error is 10.3159 with a 95% confidence interval of $[9.58143, 11.0014]$.

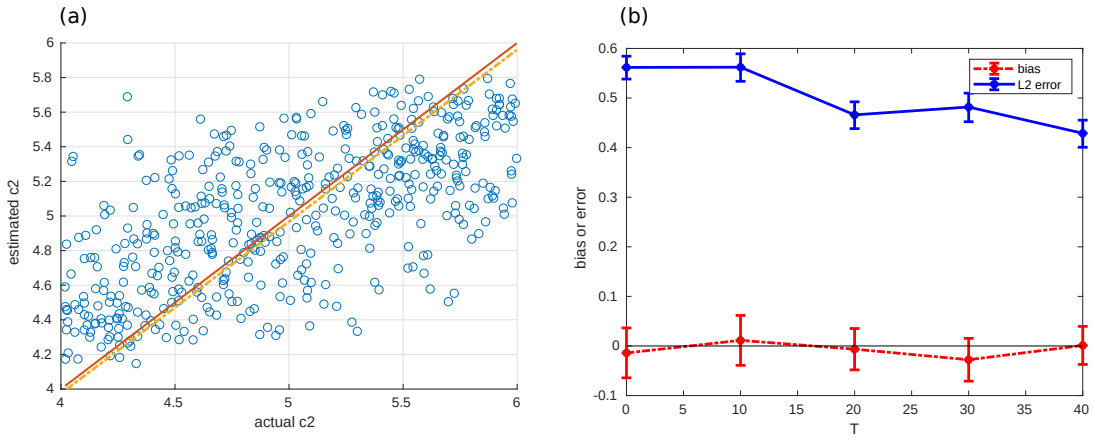


Figure 3.3: (a) A scatter plot of $\bar{C}_2(40)$, the estimated value of C_2 (based on the observation of the copy number trajectory of species A on the interval $[0, 40]$), against the actual value of C_2 in the linear propensity example. In each trial, we randomly generated a value of C_2 following a uniform distribution on $[4, 6]$ and simulated the full system to generate an observation trajectory of species A. Then we applied Algorithm 5 to compute $\bar{C}_2(40)$, where a uniform (prior) distribution on the interval $[4, 6]$ was endowed. Filter sample size was $N_s = 1000$ and trial size $N_r = 500$. (b) The bias $\mathbb{E}[\bar{C}_2(T) - C_2]$ as well as the L^2 error $\mathbb{E}[(\bar{C}_2(T) - C_2)^2]^{1/2}$ (along with 95% confidence intervals) are plotted against T. Decreasing L^2 error with T suggests that longer observations lead to better estimations.

3.4.2 Genetic Circuit Example

We consider a genetic transcription example [38] where a protein A encoded by gene D_A could bind with its own gene promoter to enhance or suppress its transcription:

$$\begin{aligned} D_A + A &\xrightarrow{c_1} D'_A, \\ D'_A &\xrightarrow{c_2} D_A + A, \\ D_A &\xrightarrow{c_3} D_A + A, \\ D'_A &\xrightarrow{c_4} D'_A + A, \\ A &\xrightarrow{c_5} \varnothing. \end{aligned} \tag{3.29}$$

Let $Z(t) = (\#D_A(t), \#D'_A(t), \#A(t))$ and suppose the propensity function has the mass action form $a_1(z) = c_1 z_1 z_3, a_2(z) = c_2 z_2, a_3(z) = c_3 z_1, a_4(z) = c_4 z_2, a_5(z) = c_5 z_3$. Suppose we only observe the molecule counts of protein A and would like to estimate the copy number of the naked form of gene promoter D_A and the bounded form of the gene promoter D'_A. We ran all numerical experiments with initial condition $Z(0) = (3, 0, 15)$ and nominal parameter values $c_1 = 0.3, c_2 = 3, c_3 = 0.5, c_4 = 0.2, c_5 = 0.06$ and filter sample size $N_s = 10,000$.

The estimation of the copy number of D_A at time t based on observation of A over $[0, t]$ is shown in Figure 3.4.

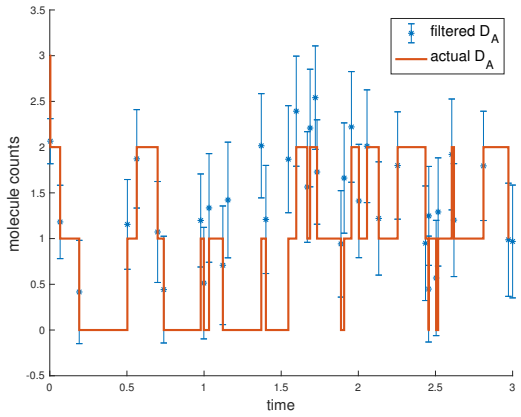


Figure 3.4: Estimation of copy number of D_A at time t based on the observation of species A over $[0, t]$ (genetic circuit example). Estimates are shown at jump times of the observed species A along with their 68% confidence intervals. Filter sample size used was $N_s = 10,000$.

We estimated one parameter at a time while fixing the other parameters at their nominal values mentioned before. We chose a uniform distribution as the prior and the posterior conditional distributions corresponding to the same single observed trajectory at several snapshots in time are shown in Figure 3.5.

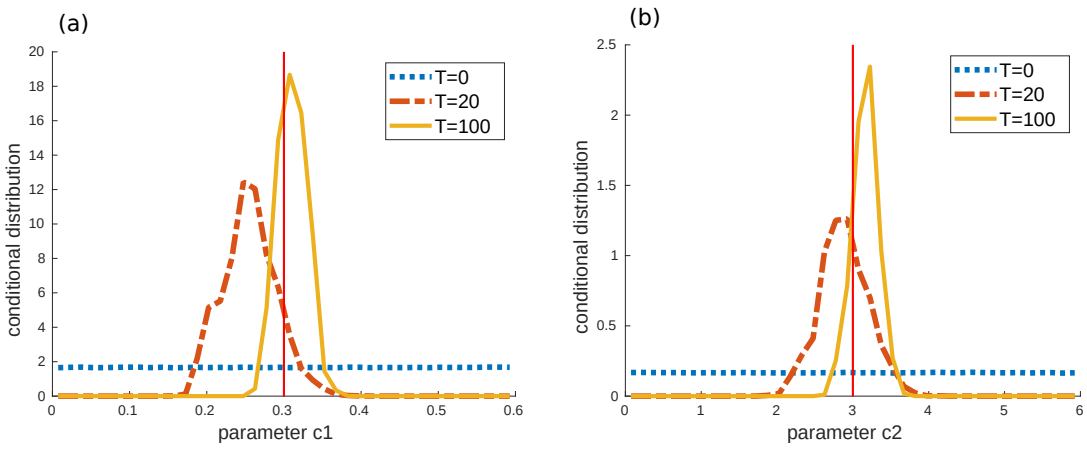


Figure 3.5: The posterior conditional probability density function of parameters C_1 (a) and C_2 (b), based on one observed trajectory of A on interval $[0, T]$ in the genetic circuit example for $T = 0, 20$ and 100. The filter sample size was $N_s = 1,000,000$. $T = 0$ is simply the uniform priors on C_1 (a) and C_2 (b). The vertical lines show the true parameter values.

3.4.3 Genetic Toggle Switch

Consider the system of genetic toggle switch[16]

$$\begin{aligned} \varnothing &\xrightarrow{a_1} S_1, \\ S_1 &\xrightarrow{a_2} \varnothing, \\ \varnothing &\xrightarrow{a_3} S_2, \\ S_2 &\xrightarrow{a_4} \varnothing. \end{aligned} \tag{3.30}$$

Let $Z(t) = (X(t), Y(t)) = (\#S_1(t), \#S_2(t))$ and suppose the propensity functions have the form $a_1(z) = \frac{\alpha_1}{1+y^\beta}$, $a_2(z) = x$, $a_3(z) = \frac{\alpha_2}{1+x^\gamma}$, $a_4(z) = y$. Suppose we only observe species S_2 and we estimate the molecule counts of species S_1. We chose the initial condition $Z(0) = (0, 0)$ and the nominal parameter values $\alpha_1 = 50, \alpha_2 = 16, \beta = 2.5, \gamma = 1$. The estimation of the molecular counts of species

S_1 right after each jump of $\#S_2(t)$ for a particular observed trajectory is shown in Figure 3.6 where a filter sample size of $N_s = 10,000$ was used.

As in the genetic circuit example, we estimated one parameter at a time while fixing the other parameter at its nominal value mentioned before. We chose a uniform distribution as the prior and the posterior conditional distributions computed based on the same single observed trajectory at several snapshots in time are shown in Figure 3.7.

Figure 3.7 suggests that the inference of α_2 is much better than the inference of α_1. To understand why, we note that in the toggle switch example, the trajectories of $\#S_1$ and $\#S_2$ exhibit a switching pattern that switches between two modes. In one mode, $\#S_1$ fluctuates around α_1 while $\#S_2$ is nearly zero, and in the other, $\#S_2$ fluctuates around α_2 while $\#S_1$ is nearly zero. The influence of α_1 on the behavior of S_2 only enters indirectly, and we conjecture by affecting the switching frequency. If this were to be the case, one may need a long observation trajectory, perhaps beyond $T = 10,000$ to make a good estimate of α_1. Longer time duration T may require larger sample size N_s making the computations more tedious.

In order to test our hypothesis regarding the switching frequency, after some trial and error, we chose the parameter values $\alpha_1 = 20$ and $\alpha = 9$ and kept the other two parameters the same. Figure 3.8 shows a comparison of trajectories of the system with the two different sets of parameters. We repeated the parameter inference experiment with this new choice of α_1 and α_2 and the results are shown in Figure 3.9 and show that α_1 is predicted better.

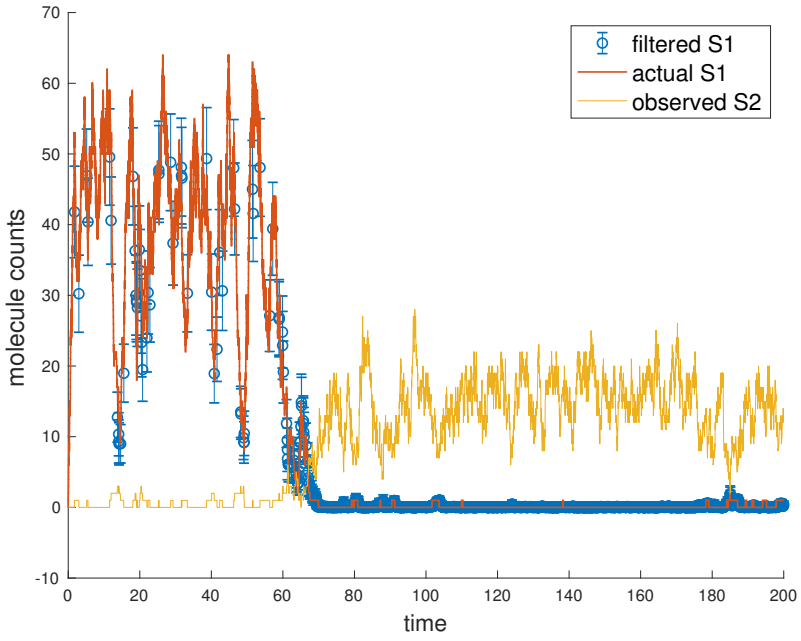


Figure 3.6: Estimation of copy number of S_1 at time t in the genetic toggle switch example based on one observed trajectory of S_2 over $[0, t]$. Estimates are shown at jump times of the observed species S_2 along with their 68% confidence intervals. Filter sample size used was $N_s = 10,000$.

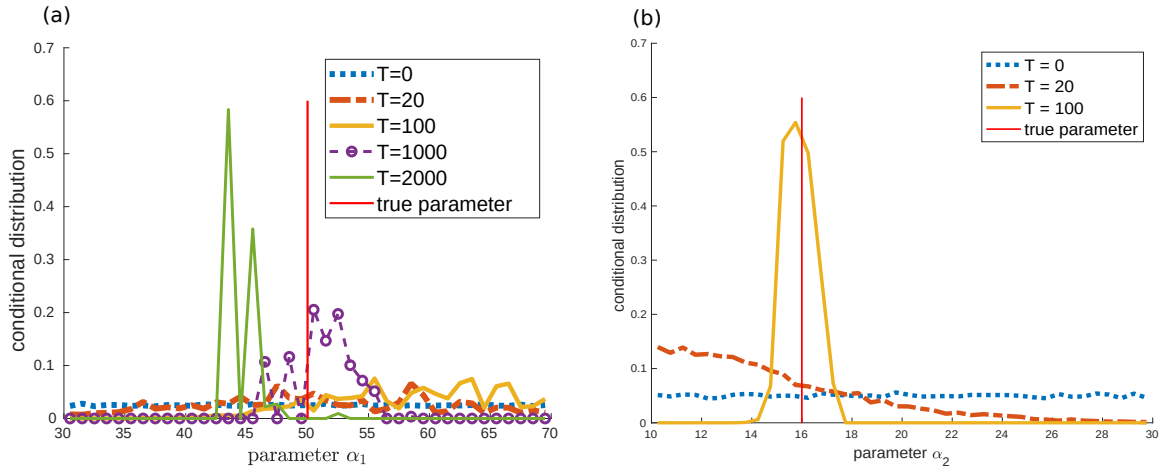


Figure 3.7: The posterior conditional probability density function of parameters α_1 (a) and α_2 (b) in the genetic toggle example based on the observation of the trajectory of $\#S_2$ over $[0, T]$. The nominal parameter values were $\alpha_1 = 50, \alpha_2 = 16, \beta = 2.5, \gamma = 1$. Estimates are shown at times $T = 0, 20, 100, 1000$ and 2000 in (a) and $T = 0, 20$ and 100 in (b). The filter sample size was $N_s = 10,000$. Note that $T = 0$ corresponds to the uniform priors. Two different observation trajectories, one for the estimation of α_1 (a) and the other for α_2 (b) were used.

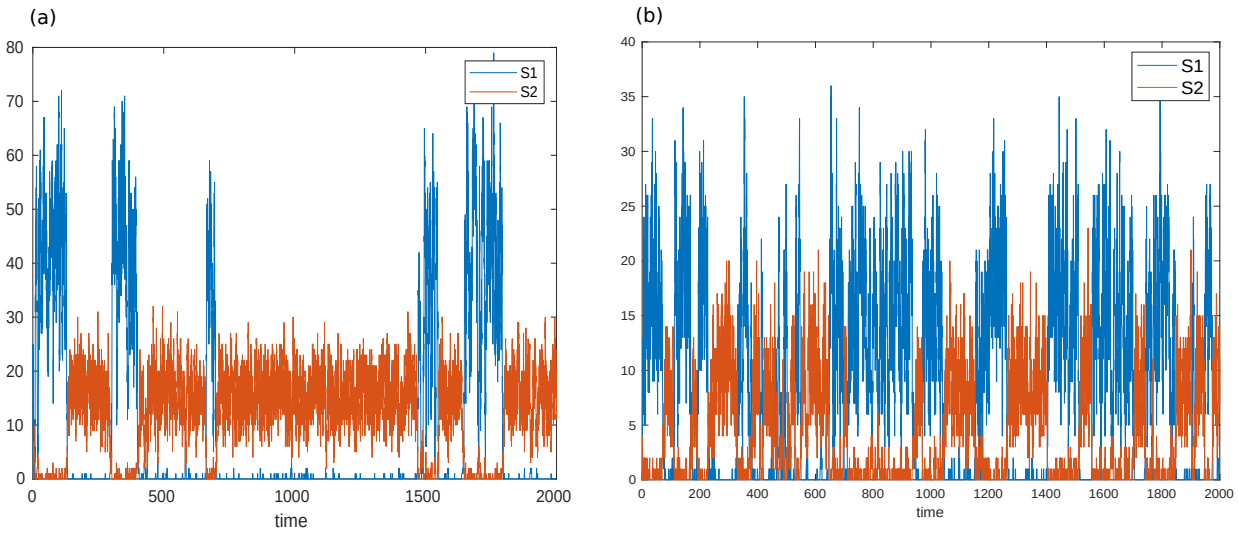


Figure 3.8: Trajectories of the genetic toggle switch system under different parameters. Part (a) has less frequent switching and corresponds to $\alpha_1 = 50, \alpha_2 = 16, \beta = 2.5, \gamma = 1$. Part (b) corresponds to $\alpha_1 = 20, \alpha_2 = 9, \beta = 2.5, \gamma = 1$ and switches more frequently.

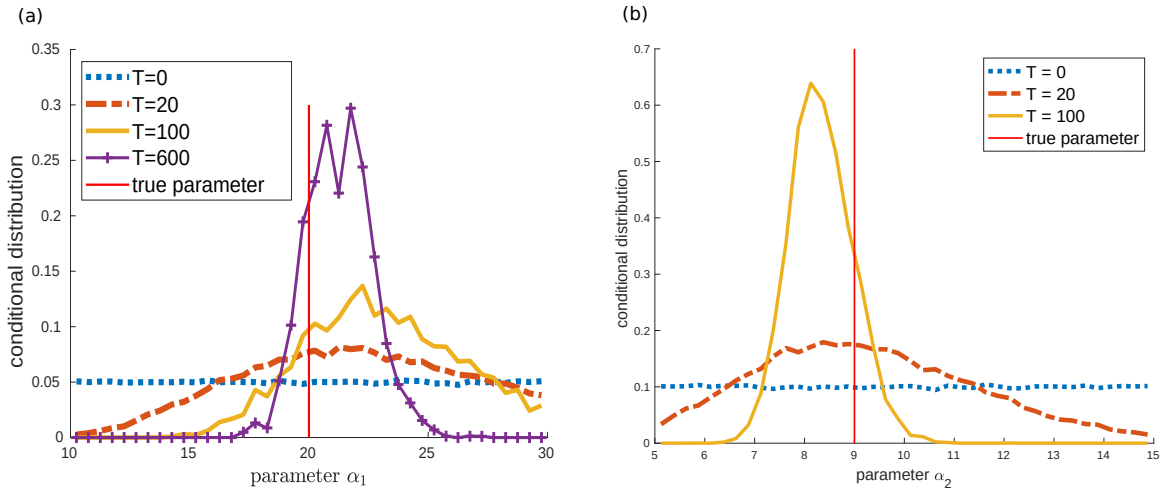


Figure 3.9: The posterior conditional probability density function of parameters α_1 (a) and α_2 (b) in the genetic toggle example based on the observation of the trajectory of $\#S_2$ over $[0, T]$. The nominal parameter values were $\alpha_1 = 20, \alpha_2 = 9, \beta = 2.5, \gamma = 1$. Estimates are shown at times $T = 0, 20, 100$ and 600 in (a) and $T = 0, 20$ and 100 in (b). The filter sample size was $N_s = 100,0000$. Note that $T = 0$ corresponds to the uniform priors.

3.4.4 SEIR Model

Consider the SEIR model for the spread of infectious diseases

$$\begin{aligned} S + I &\xrightarrow{\beta} E + I, \\ E &\xrightarrow{\kappa} I, \\ I &\xrightarrow{\gamma} R. \end{aligned} \tag{3.31}$$

Let $Z(t) = (\#S(t), \#E(t), \#R(t), \#I(t))$ where S, E, I and R stand for susceptible, exposed, infected and recovered respectively. We assume that we could observe the infectious population I exactly. The propensity functions $a_j(\cdot)$ have the form $a_1(z) = \beta z_1 z_4/N$, $a_2(z) = \kappa z_2$, $a_3(z) = \gamma z_4$, where $N = z_1 + z_2 + z_3 + z_4$ is the total population. We ran all numerical experiments with initial condition $Z(0) = (s_0, e_0, r_0, i_0) = (500, 20, 0, 5)$ and parameter values $\beta = 0.05, \kappa = 0.2, \gamma = 0.05$.

The estimation of the susceptible and exposed population after each change of $Z_4(t) = \#I(t)$ for a particular observed trajectory of I is shown in Figure 3.10 where a filter sample size of $N_s = 10,000$ was used. Figure 3.11 shows scatter plots corresponding to state estimation of susceptible and exposed number of individuals.

Keeping other parameters at their nominal values, we explored Bayesian inference of parameter κ (the reciprocal of the incubation period) based on a uniform prior. The scatter plot of the parameter estimation of κ based on $N_r = 1000$ trials each with a filter sample size $N_s = 1000$ is in shown in Figure 3.12.

Next we explore estimation of the state at a past time t_0 based on observation up to a future time T. That is $0 \leq t_0 \leq T$. The case of $t_0 = 0$ corresponds to

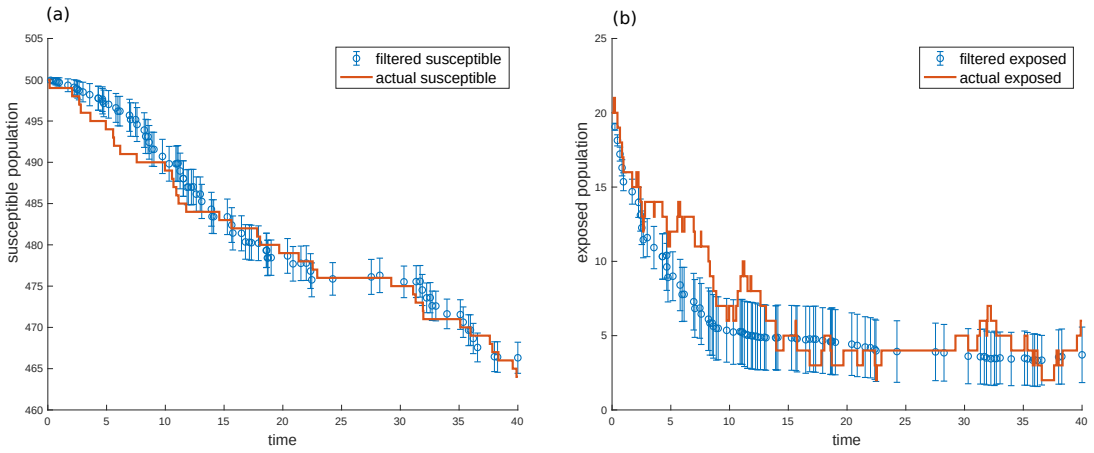


Figure 3.10: Estimation of the susceptible population (a) and the exposed population (b) at time t in the SEIR example given one trajectory of the infected population over $[0, t]$. Estimates are shown at times of observed new infections or recovery along with the 68% confidence intervals. Filter sample size used was $N_s = 10,000$.

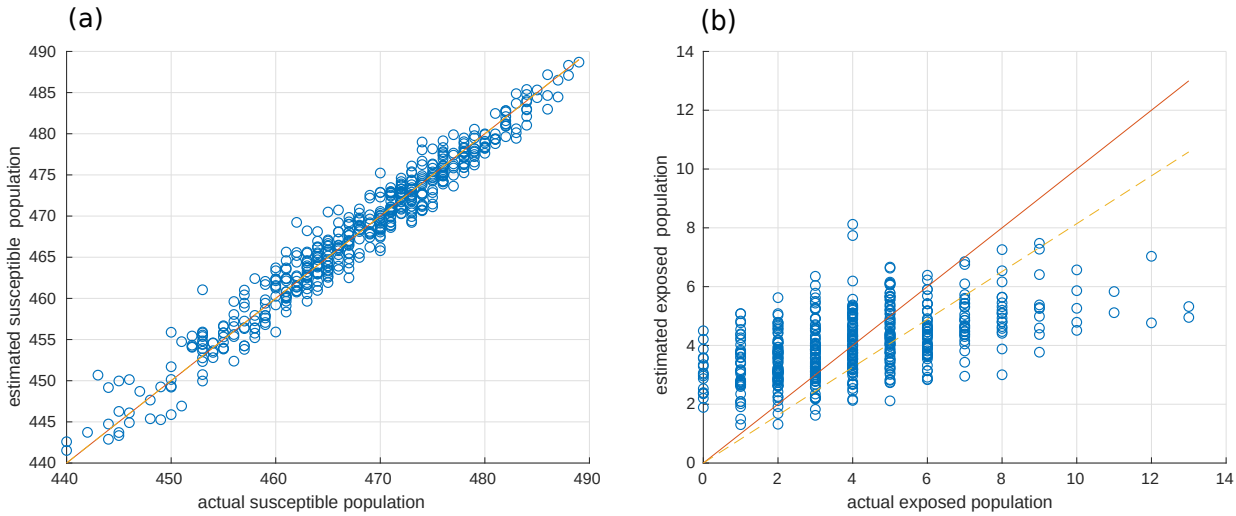


Figure 3.11: Estimate of the susceptible (a) and exposed (b) population in SEIR example at time $T = 40$ given observations of the infected population over $[0, T]$. The filter sample size was $N_s = 1000$, and trial size $N_r = 500$. For the susceptible population, the bias is 0.0156 within a 95% confidence interval $[-0.160955, 0.192202]$ and L_2 error is 1.9723 within $[1.81684, 2.11637]$. For the exposed population, the bias is -0.0156 within a 95% confidence interval $[-0.192202, 0.160955]$ and L_2 error is 1.9723 within $[1.81684, 2.11637]$.

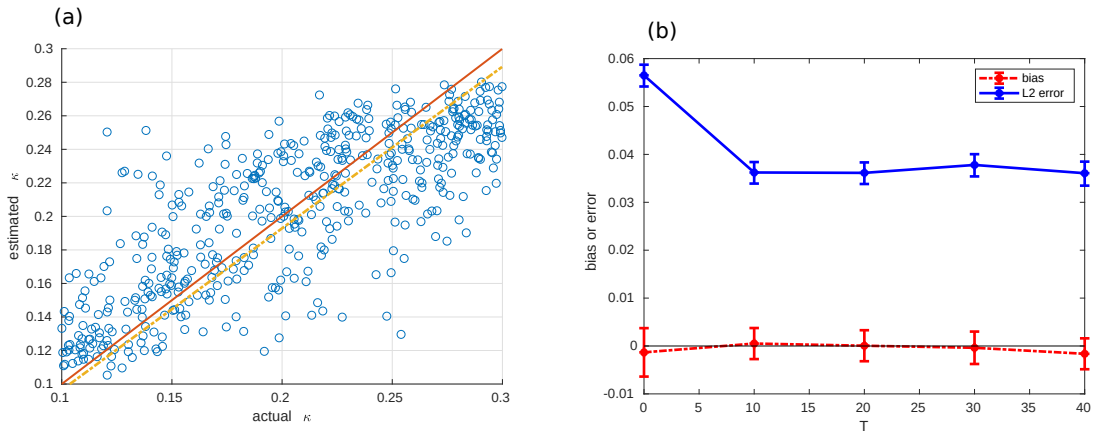


Figure 3.12: (a) A scatter plot of $\hat{\kappa}(T)$, the estimated value of κ against the actual value of κ. Here $T = 40$. Filter sample size was $N_s = 1000$ and trial size $N_r = 500$. (b) The the bias $\mathbb{E}[\hat{\kappa}(T) - \kappa]$ as well as the L^2 error $\mathbb{E}[(\hat{\kappa}(T) - \kappa)^2]^{1/2}$ (along with 95% confidence intervals) are plotted against T.

the situation where the initial state itself is not known, but we only have a prior distribution for it. We assumed that the exposed population at initial time $t = 0$ followed a binomial distribution, with parameters $N = 520$ and $p = 0.04$. Since $\#I(0) = 5$ and the total population is 525, this left us with 520 individuals. We assumed a probability $p = 0.04$ of being exposed, which lead us to choose the binomial distribution. Algorithm 6 was used to estimate the conditional probability mass function

$$P\{X_2(t_0) = x \,|\, Y(s) = y(s),\, 0 \le s \le T\}$$

for various t_0 and T. The results are shown in Figure 3.13.

3.4.5 Investigation of Resampling

Here we investigate the strategies of resampling in our filter. In the numerical examples presented so far, we applied the offspring algorithm to resample right after

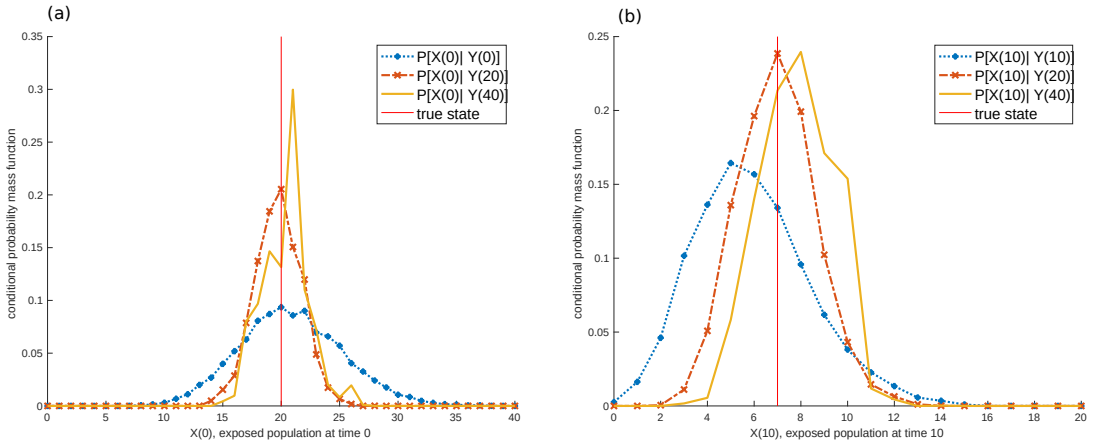


Figure 3.13: (a) Conditional distribution of the exposed population at the initial time $X_2(0)|\mathcal{Y}_T$ for $T = 0, 20$ and 40. (b) Conditional distribution of the exposed population at an intermediate time $X_2(10)|\mathcal{Y}_T$ for $T = 10, 20$ and 40. Filter sample size was $N_s = 10,000$. The prior of the initial distribution of the exposed population follows a binomial distribution $B(520, 0.04)$.

each jump t_k. Here we present the results of three different resampling strategies. One was to resample at each jump time t_k as we did earlier. As an alternative, we applied an adaptive resampling method where resampling is applied at t_k if more than 10 particles had zero weights or if the ratio of the maximum weight to the minimum nonzero weight was greater than 1000. The third alternative was not to resample at all. We note that, whenever resampling was not used, the weights were normalized instead so that the average weight was 1 right after any jump time t_k.

Table 3.1 shows the L^2 error as well as the bias in estimating the conditional mean of the state or a parameter along with the 95% confidence intervals, for the various examples considered earlier. The results are for state estimation unless a parameter is mentioned in the first column. The entrees `NaN` indicate numerical issues in `MATLAB` mainly due to the fact that all weights became either zero or infinity within numerical precision. These results show that the three approaches yield more

Table 3.1: The bias and error of estimation of the states or parameters when different resampling schemes were applied.

System	Method	L^2 error	Confidence interval	bias	Condidence interval
linear propensity	each jump	9.7470	$[9.0677, 10.3819]$	-0.3346	$[-1.2068, 0.5375]$
linear propensity	adaptive (0%)	10.0216	$[9.3373, 10.6621]$	-0.0699	$[-0.9671, 0.8273]$
linear propensity	never	10.0783	$[9.4464, 10.6728]$	-0.4910	$[-1.3922, 0.4103]$
genetic circuit	each jump	0.5638	$[0.5241, 0.6009]$	0.0100	$[-0.0405, 0.0604]$
genetic circuit	adaptive (75.85%)	0.5456	$[0.5066, 0.5820]$	-0.0085	$[-0.0574, 0.0403]$
genetic circuit	never	NaN	NaN	NaN	NaN
genetic toggle	each jump	5.5083	$[5.0655, 5.9181]$	0.3701	$[-0.1219, 0.8622]$
genetic toggle	adaptive (2.20%)	5.5307	$[5.0344, 5.9860]$	0.2406	$[-0.2541, 0.7353]$
genetic toggle	never	5.4175	$[4.9015, 5.8886]$	-0.8134	$[-1.2930, -0.3339]$
SEIR	each jump	2.0853	$[1.9323, 2.2279]$	-0.0176	$[-0.2043, 0.1691]$
SEIR	adaptive (9.91%)	2.3726	$[2.2102, 2.5245]$	0.1238	$[-0.0884, 0.3359]$
SEIR	never	2.0823	$[1.9572, 2.2003]$	-0.0029	$[-0.1893, 0.1835]$
linear propensity c_2	each jump	0.4099	$[0.3850, 0.4334]$	0.0233	$[-0.0134, 0.0599]$
linear propensity c_2	adaptive (23.00%)	0.4237	$[0.3977, 0.4481]$	-0.0295	$[-0.0674, 0.0083]$
linear propensity c_2	never	0.5380	$[0.5022, 0.5716]$	0.0628	$[0.0149, 0.1106]$
SEIR κ	each jump	0.0377	$[0.0354, 0.0398]$	0.0002	$[-0.003, 0.004]$
SEIR κ	adaptive (20.61%)	0.0351	$[0.0329, 0.0372]$	-0.003	$[-0.004, 0.0028]$
SEIR κ	never	0.0371	$[0.0343, 0.0396]$	-0.0064	$[-0.0097, -0.003]$

or less the same accuracy (the estimated errors are within the confidence intervals of each other) except in the case of the genetic circuit example where serious numerical issues manifest without resampling. We suspect that in the other examples, if final time T is increased, the need for resampling will become evident.

3.5 Conclusions and Future Work

We presented a novel filtering algorithm to compute the conditional probability mass function

$$\pi(t, x) = P\{X(t) = x \,|\, Y(s) = y(s), 0 \leq s \leq t\},$$

in the context of reaction networks. Here $Y(t) = y(t)$ is the vector copy number of a subset of species that are assumed to be observed exactly in continuous time and $X(t)$ is the vector copy number of the remaining unobserved species. We also showed how this algorithm can be adapted for the purposes of Bayesian parameter estimation based on exact partial state observation. Furthermore, we also showed how the state $X(t_0)$ at time t_0 can be estimated based on observations up to a later time T where $0 \leq t_0 \leq T$.

The filtering algorithm involves a weighted Monte Carlo method and a resampling strategy needs to be employed. We explored some possibilities for resampling at the observed jump times t_k. Our investigations in this regards were numerical and based on relatively simple examples, and hence not exhaustive. An investigation of adaptive resampling based on some theoretical analysis is the subject of future work.

While we presented an intuitive derivation of the filtering equations, our derivation is not mathematically rigorous. [10] provides a rigorous derivation in the context of finite state Markov processes which in the case of reaction networks correspond to systems where species conservation relations limit the species counts to be bounded.

As is well known, many intracellular reaction networks may have some species in greater abundance and a discrete state model can be tedious to simulate one event at a time. Tau-leap methods [20] as well as model reduction approaches have been proposed for efficient simulation of such systems. See [21] and references therein for several methods as well as Refs. [5, 28, 25, 15, 31, 2, 37] for rigorous mathematical analysis of methods. These same considerations could be applied to the filtering

method proposed here to develop reduced order models and tau-leap simulations.

Our assumption of exact (noiseless) observation of some species may appear unrealistic. However, as mentioned in the introduction, most observation noise may be modeled via extra reactions and extra species such as photons. If the photon counts are very large, the same considerations of reduced order models or tau-leaping mentioned above apply. Less realistic is the assumption of continuous in time observations. In reality, observations are recorded in discrete time snapshots. If the frequency of the snapshots is very high, then the theory of continuous in time observations provides a good approximation. If the frequency is low, then this is not the case.

Chapter 4

Discrete-time Filtering in Stochastic Reaction Networks

4.1 Reaction Networks and The Filtering Problem

We remind readers of the stochastic model of a *reaction network*, which was described in Chapter 3. A reaction network consists of $n \in \mathbb{N}$ *species* and $m \in \mathbb{N}$ *reaction channels*. The copy number of the species i $(1 \leq i \leq n)$ at time $t \geq 0$ is denoted by $Z_i(t) \in \mathbb{Z}_+$ and the number of firings of reaction channel j $(1 \leq j \leq m)$ during $(0, t]$ is denoted by $R_j(t) \in \mathbb{Z}_+$. The vector valued processes X and R are defined by $Z(t) = (Z_1(t), \ldots, Z_n(t))$ and $R(t) = (R_1(t), \ldots, R_m(t))$ and Z is assumed to be a continuous time Markov chain (CTMC) with state space $\mathbb{Z}_+^n$. The characteristics of a reaction network are uniquely defined by the *stoichiometric vectors* $\nu_j \in \mathbb{Z}^n$ and the *propensity functions* $a_j : \mathbb{Z}_+^n \to [0, \infty)$ for $j = 1, \ldots, m$. If the reaction channel j fires at time t, the state Z of the process is changed by ν_j. We assume that the process Z is *cadlag*, that is, right continuous with left hand limits. We shall define the matrix $\nu \in \mathbb{Z}^{n \times m}$ so that its jth column equals ν_j. The (i, j)th entry of ν is denoted ν_{ij} which is the ith entry of the column vector ν_j. Thus we may write

$$Z(t) = Z(0) + \nu\, R(t). \tag{4.1}$$

The propensity function a_j has the interpretation that given $Z(t) = z$ (for $z \in \mathbb{Z}_+^n$), the probability that the reaction channel j fires during $(t, t+h]$ is given by $a_j(z)h + o(h)$ as $h \to 0+$. It follows that R is an m-variate *counting process* with *stochastic intensity* $a(Z(t-))$ [8]. We assume that the processes Z and R are carried by a probability space $(\Omega, \mathcal{F}, P)$.

We shall be concerned with the situation where a subset of the species are observed at time snapshots $0 < t_1 < t_2 \ldots$. We shall take the vector of observed species to consist of the last n_2 components of the state $Z(t)$ and denote it by $Y(t)$. The vector of the rest of the unobserved species shall be denoted by $X(t)$. Thus we write $Z(t) = (X(t), Y(t))$ where $X(t) \in \mathbb{Z}_+^{n_1}$ and $Y(t) \in \mathbb{Z}_+^{n_2}$ with $n = n_1 + n_2$. We shall write $\nu_j = (\nu_j', \nu_j'')$ where $\nu_j' \in \mathbb{Z}^{n_1}$ and $\nu_j'' \in \mathbb{Z}^{n_2}$. We shall define $\nu' \in \mathbb{Z}^{n_1 \times m}$ and $\nu'' \in \mathbb{Z}^{n_2 \times m}$ to consist of the first n_1 rows of ν and the last n_2 rows of ν respectively. Thus we may write

$$\begin{aligned} X(t) &= X_0 + \nu' R(t), \\ Y(t) &= Y_0 + \nu'' R(t), \end{aligned} \tag{4.2}$$

where $Z_0 = (X_0, Y_0)$ is the random initial state.

The filtering problem that we are concerned with is the numerical computation of conditional probabilities of the form:

$$\pi(t, z) = P\{Z(t) = z \,|\, Y(t_i) = y_i,\, i = 1, \ldots, T_s\} \tag{4.3}$$

where $t \geq 0$, $z \in \mathbb{Z}_+^n$ and T_s is the number of snapshots. Due to the fact that the

number of states is either infinity or exceedingly large, Monte Carlo methods are the most feasible way to compute $\pi(t, z)$. In addition to the observations, we assume that we know the distribution of $Z_0 = (X_0, Y_0)$. We note that $\pi(t, z)$ depends also on $\bar{y} = (y_1, \ldots, y_{T_s})$ so that sometimes we write $\pi(t, z, \bar{y})$ to emphasize this dependence.

4.2 The Filtering Algorithm

In this section we present a mathematical setup to describe filtering methods and introduce what we call the Naive Algorithm which motivates the concept of *effective sample size*. We also discuss a framework that underlies few different filtering algorithms in literature as well as present our proposed filtering method.

4.2.1 Mathematical Setup

We shall be concerned with weighted Monte Carlo methods which are also known as *importance sampling* or *particle filters*. We present a mathematical description of a general setup as follows.

We let $(\Omega, \mathcal{F})$ be a measurable space consisting of the sample space and a collection of events. It is useful to think of two sets of processes: one on which observations are made and the other consists of the Monte Carlo simulations performed by the filtering method. We refer to the former as the *lab processes* and the latter as the *method processes* or *filter processes*. Both sets of processes are carried by $(\Omega, \mathcal{F})$. Conditioned on the observations, the lab processes and the method processes are independent.

The lab processes consist of $Z = (X, Y)$ and R, where $Z : [0, T_0] \times \Omega \to \mathbb{Z}_+^n$ and $R : [0, T_0] \times \Omega \to \mathbb{Z}_+^m$ are measurable where $T_0 > 0$ is large enough so that $[0, T_0]$ contains all observation time points t_l.

The method processes consist of $\tilde{Z}^{(i)} = (U^{(i)}, V^{(i)}) : [0, T_0] \times \Omega \to \mathbb{Z}_+^n$, $I^{(i)} : [0, T_0] \times \Omega \to \mathbb{Z}_+^m$ and $W^{(i)} : [0, T_0] \times \Omega \to [0, \infty)$ for $i = 1, \ldots, N_s$. Here N_s is the *filter sample size* (number of particles) and the method processes for different i are all identically distributed and for convenience we drop the superscript i when describing them, or one may regard $\tilde{Z} = (U, V)$, I and W as the template method processes. The method processes U, V and I are proxies for X, Y and R respectively. The process W is the weight process that assigns a degree of importance to the paths of U, V and I.

The method processes are convenient to simulate under a probability measure P_0 on $(\Omega, \mathcal{F})$. The lab processes have the given propensities under a possibly different probability measure P on $(\Omega, \mathcal{F})$. In general, the quantity of interest $\pi(t, z, \bar{y})$ is given as a function of expectations of path functionals of the filter processes. The precise nature of this relationship shall depend on the filtering method.

4.2.2 The Naive Algorithm

It might be clear from the outset that the most straightforward Monte Carlo algorithm is simply based on the following acceptance/rejection strategy. One simulates N_s sample trajectories $\tilde{Z}^{(i)}(t) = (U^{(i)}(t), V^{(i)}(t))$ for $i = 1, \ldots, N_s$ according to the given reaction network characteristics via the Gillespie algorithm or a variant

[19, 21, 17, 1] and assigns a weight $W^i = 1$ (accept) if and only if $V^{(i)}(t_l) = y_l$ for $l = 1, \ldots, T_s$ and $W^i = 0$ (reject) otherwise. We note that $\pi(t, z)$ is given by

$$\pi(t, z) = \frac{\mathbb{E}[1_{\{z\}}(\tilde{Z}(t)\, W]}{\mathbb{E}[W]}$$

and one simply estimates π via the filter sample as

$$\hat{\pi}(t, z) = \frac{\sum_{i=1}^{N_s} 1_z(\tilde{Z}^{(i)}(t))\, W^i}{\sum_{i=1}^{N_s} W^i}.$$

We note that here $P_0 = P$ in relation to Section 4.2.1. We shall refer to this method as the Naive Algorithm. While it is conceptually simple and easy to implement, in many situations the probability of any given sequence of observation snapshots is very low and hence the "effective sample size" is much smaller than N_s. Hence in order to obtain reliable estimates one needs to simulate impractically large numbers N_s of trajectories.

Theorem 6. *Given $\sum_{i=1}^{N_s} W^i > 0$, the naive method is well defined and unbiased under that condition.*

$$\mathbb{E}\left[\frac{\sum_{i=1}^{N_s} 1_{(x,y)}(U^{(i)}(t), V^{(i)}(t))W^i}{\sum_{i=1}^{N_s} W^i} \;\middle|\; \sum_{i=1}^{N_s} W^i > 0\right] = P\{X(t) = x, Y(t) = y \,|\, Y(T) = y_T\}. \tag{4.4}$$

Proof. First, we condition on $\sum_{i=1}^{N_s} W^i = n_s$, we want to show that

$$\mathbb{E}\left[\frac{\sum_{i=1}^{N_s} 1_{(x,y)}(U^{(i)}(t), V^{(i)}(t))W^i}{\sum_{i=1}^{N_s} W^i} \,\middle|\, \sum_{i=1}^{N_s} W^i = n_s\right] = P\{X(t) = x, Y(t) = y \,|\, Y(T) = y_T\}. \tag{4.5}$$

for each value of $n_s > 0$.

Without loss of generality, we may assume $W^i = 1$ for $i = 1, ..., n_s$ while $W^i = 0$ otherwise. Now we have

$$\begin{aligned}
&\mathbb{E}\left[\frac{\sum_{i=1}^{N_s} 1_{(x,y)}(U^{(i)}(t), V^{(i)}(t))W^i}{\sum_{i=1}^{N_s} W^i} \,\middle|\, W^i = 1 \text{ for } i = 1, ..., n_s \text{ and } W^i = 0 \text{ otherwise}\right] \\
&= \frac{1}{n_s}\sum_{i=1}^{N_s} \mathbb{E}\left[1_{(x,y)}(U^{(i)}(t), V^{(i)}(t))W^i \,\middle|\, W^i = 1 \text{ for } i = 1, ..., n_s \text{ and } W^i = 0 \text{ otherwise}\right] \\
&= \frac{1}{n_s}\sum_{i=1}^{n_s} \mathbb{E}\left[1_{(x,y)}(U^{(i)}(t), V^{(i)}(t))W^i \,\middle|\, W^i = 1\right] \\
&\quad + \frac{1}{n_s}\sum_{i=n_s+1}^{N_s} \mathbb{E}\left[1_{(x,y)}(U^{(i)}(t), V^{(i)}(t))W^i \,\middle|\, W^i = 0\right] \\
&= P\{X(t) = x, Y(t) = y \,|\, Y(T) = y_T\}
\end{aligned} \tag{4.6}$$

The last step follows because when $W^i = 1$, that is, when $V^{(i)}(T) = y_T$,

$$\mathbb{E}\left[1_{(x,y)}(U^{(i)}(t), V^{(i)}(t)), W^i \,|\, W^i = 1\right] = P\{X(t) = x, Y(t) = y \,|\, Y(T) = y_T\}$$

and when $W^i = 0$, $\mathbb{E}\left[1_{(x,y)}(U^{(i)}(t), V^{(i)}(t))W^i \,|\, W^i = 0\right] = 0$. □

4.2.3 Effective Sample Size

Often the filtering method involves estimating a quantity c given by

$$c = \frac{\mathbb{E}[C\,W]}{\mathbb{E}[W]}$$

where C is a path functional and W is an associated weight. The quantity c is estimated via the empirical mean

$$\hat{c} = \frac{\sum_{i=1}^{N_s} C^i W^i}{\sum_{i=1}^{N_s} W^i}$$

from a weighted sample (C^i, W^i) for $i = 1, \ldots, N_s$, identically distributed like (C, W). The fidelity of this estimate depends on the joint distribution of (C^i, W^i) for $i = 1, \ldots, N_s$. Law of large numbers states that for sufficiently large N_s, $\hat{c}$ is a good approximation of c. If we assume independence (for different i), when N_s is sufficiently large, under independence assumption we may use the so-called delta method and obtain the approximation that $\hat{c}$ is (approximately) normally distributed with $\mathbb{E}[\hat{c}] \approx c$ and

$$\mathrm{Var}(\hat{c}) \approx \frac{\mathbb{E}^2[CW]}{N_s\,\mathbb{E}^2[W]}\,\mathrm{Var}\left(\frac{CW}{\mathbb{E}[CW]} - \frac{W}{\mathbb{E}[W]}\right).$$

The square of the *relative error* $\mathrm{RE} = \frac{\sqrt{\mathbb{E}[(\hat{c}-c)^2]}}{c}$ is then approximated by

$$\mathrm{RE}^2 \approx \frac{1}{N_s}\mathrm{Var}\left(\frac{CW}{\mathbb{E}[CW]} - \frac{W}{\mathbb{E}[W]}\right).$$

Hence, the fidelity of the estimate clearly depends on the path functional C and its covariance with W among other factors. For instance, if one is interested in estimating $\pi(t,z)$, then $C = 1_{\{z\}}(\tilde{Z}(t))$, and the fidelity of the estimate varies with z. In order to have a measure of fidelity that is independent of the path functional C and solely dependent on W, the quantity

$$N_e = \frac{\left(\sum_{i=1}^{N_s} W^i\right)^2}{\sum_{i=1}^{N_s} (W^i)^2} \tag{4.7}$$

known as *effective sample size* has been proposed in the literature [32]. When N_s is very large, by the law of large numbers

$$N_e \approx N_s \frac{\mathbb{E}^2[W]}{\mathbb{E}[W^2]}. \tag{4.8}$$

Thus, for the naive method, $N_e \approx N_s\, p$ where p is the probability of the sequence of observations. When the molecular copy numbers of the species are in the order of hundreds and if multiple species are observed this probability is very small.

4.2.4 Conditional Propensity Function

Recall that the propensity $a_j(z)$ is defined by the prescription that the conditional probability that one reaction event j occurs during $(t, t+h]$ conditioned on $Z(t) = z$ is given by $a_j(z)h + o(h)$ as $h \to 0+$. We may define the *conditional propensity* $b_j(z, t, t_1, y_1 \dots)$ of the reaction channel j in the above definition by replacing

the conditioned event $\{Z(t) = z\}$ by the event

$$\{Z(t) = z, Y(t_l) = y_l \;\; l = 1, \ldots, T_s\}$$

that includes the observations. Naturally, the conditional propensity will depend not only on the state z at time t, but also the relationship between t and the observation time points t_l as well as the values y_l. See [22] for instance where this is discussed. For a given observation sequence, we may suppress the dependence on the observations and denote b_j by $b_j(z, t)$. We note that due to the explicit dependence on time t, the (conditional) wait time between reaction events is no longer exponential. If an explicit formula for $b_j(z, t)$ is available and the wait times have an analytically tractable formula (as we shall see in a specific example in section 4.5.1) it is indeed possible to simulate the reaction network according to these conditional propensities. In that event, all trajectories will satisfy the observations.

In general, an explicit formula for b_j is not available and hence it is impossible to draw samples from the conditional distribution. However, if a good approximation of b_j is proposed, which we denote by $\hat{a}_j(z, t)$, then one may simulate the reaction network according to these "proposal propensities". Since the proposal propensities may not be exact, a weight $\tilde{L}(T)$ given by the Girsanov change of measure that corresponds to the change in propensity from a_j to $\hat{a}_j$ (for the method process I) is computed and applied. Additionally, the indicator weight W as in the naive method is applied depending on whether the the observations are satisfied, so that the overall weight is $W\tilde{L}(T)$.

The simulation is carried out under a probability measure P_0 and under P_0, I_j have propensities $\hat{a}_j(z,t)$. Under P, given by $d\tilde{P} = \tilde{L}(T)dP_0$, the I_j have the desired propensities $a_j(z)$. Then

$$\pi(t,z) = \tilde{P}\{\tilde{Z}(t) = z \,|\, Y(t_l) = y_l, 1 \le l \le T_s\} = \frac{\tilde{P}\{\tilde{Z}(t) = z, Y(t_l) = y_l, 1 \le l \le T_s\}}{\tilde{P}\{Y(t_l) = y_l, 1 \le l \le T_d\}}$$
$$= \frac{\tilde{\mathbb{E}}[1_{\{z\}}(Z(t))W]}{\tilde{\mathbb{E}}[W]} = \frac{\mathbb{E}_0[1_{\{z\}}(Z(t))\tilde{L}(T)W]}{\mathbb{E}_0[\tilde{L}(T)W]}.$$

If the proposal propensities $\hat{a}_j$ are exactly equal to the conditional propensities b_j, then with probability one, all trajectories $\tilde{Z}$ will satisfy the observations and $W\tilde{L}(T)$ will be a constant making the effective sample size N_s. If $\hat{a}_j$ close enough to b_j, then one expects the effective sample size to be close to N_s. This is the strategy that underlies the methods in [22].

4.2.5 The Targeting Problem

In order to better motivate our new method, we focus on the problem of a single observation snapshot which we take to be $Y(T) = y$ at $t = T > 0$. This becomes a "targeting problem" as we have a random initial state Z_0 with a known distribution and we are given the partial final state condition $Y(T) = y$.

After reordering and splitting the reactions $R(t) = (R'(t), R''(t))$, and possible removal of dependent observable species (corresponding to dependent rows of ν'') we may write the modified ν'' as

$$\nu'' = [A\ B] \tag{4.9}$$

where B is $n_2 \times n_2$ and invertible. This yields

$$y = Y_0 + AR'(T) + BR''(T)$$

and hence

$$R''(T) = B^{-1}(y - Y_0 - AR'(T)). \tag{4.10}$$

Thus, we see that $R''(T)$ is determined by Y_0, y and $R'(T)$.

Here we give an example to illustrate the idea above. We consider a three species and four reactions system

$$S_1 \xrightarrow{c_1} S_2, \quad S_2 \xrightarrow{c_2} S_1, \quad S_1 + S_2 \xrightarrow{c_3} S_3, \quad S_3 \xrightarrow{c_4} S_1 + S_2. \tag{4.11}$$

Let $Z(t) = (\#S_1(t), \#S_2(t), \#S_3(t))$. Suppose species S_3 is observed, so $X(t) = (\#S_1(t), \#S_2(t))$ and $Y(t) = \#S_3(t)$. Since only one species is observed, in the decomposition $R(T) = (R'(T), R''(T))$, $R''(T)$ has dimension one. We can take $R'(T) = (R_1(T), R_2(T), R_3(T)) \in \mathbb{Z}_+^3$ and $R''(T) = R_4(T) \in \mathbb{Z}_+$ and partition the matrix ν as

$$\nu = \left[\begin{array}{c} \nu' \\ \hline \nu'' \end{array}\right] = \left[\begin{array}{cccc} -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ \hline 0 & 0 & 1 & -1 \end{array}\right], \text{ where } \nu'' = \left[\begin{array}{c|c} A & B \end{array}\right] = \left[\begin{array}{ccc|c} 0 & 0 & 1 & -1 \end{array}\right].$$

Since $y = Y_0 + R_3(T) - R_4(T)$, we obtain $R''(T) = R_4(T) = R_3(T) - y + Y_0$, determined by Y_0, y and $R'(T)$.

We shall make our proxy processes $\tilde{Z}$ and I satisfy the condition (4.10). Since we want $V(T) = Y(T) = y$, we shall require

$$y = V_0 + AI'(T) + BI''(T),$$

so

$$I''(T) = B^{-1}(y - V_0 - AI'(T)). \tag{4.12}$$

If we choose $I''(T)$ according to (4.12) we automatically satisfy the target observation. This motivates our key strategy as outlined below. In what follows, we use $K = I(T)$.

1. Given $Y(T) = y$ generate a sample for $\tilde{Z}_0 = (U_0, V_0)$ and $K = I(T)$, and an associated weight $W_p > 0$ so that

   $$P_0\{R(T) = k, Z_0 = z_0 | \, Y(T) = y\} = \mathbb{E}_0[1_{(z_0,k)}(\tilde{Z}_0, K)\, W_p | \, Y(T) = y]/\mathbb{E}_0[W_p],$$

   where P_0 is the simulation probability measure (to be decided). This is accomplished by the following procedure which can be thought of as consisting of prediction and correction steps.

   (a) Generate a sample for $\tilde{Z}_0$ and $K' = I'(T)$ (prediction).

   (b) With the knowledge of $V_0, V(T) = y$ and $K' = I'(T)$ compute $K'' = I''(T)$ via (4.12). If $I''(T) \in \mathbb{Z}_+^{m_2}$ accept, otherwise reject and go back to previous step.

(c) Set the weight W_p to reflect the probability of the assigned value for $I''(T)$. (Correction).

2. Having obtained a value for $\tilde{Z}_0$ and $I(T)$ along with a weight $W_p > 0$, obtain $I(t)$ for $0 \leq t \leq T$ by "interpolation".

Generating a sample from $I'(T)$ can in principle be done by Gillespie method simulation. However, computing the weight W_p as the probability of the value of $I''(T)$ given by (4.12) in closed form is not simple. Furthermore, the interpolation of $I(T)$ for t values in $[0, T]$ is not easy either. However, these steps can be solved with relative ease if the m-variate counting process R is assumed to be an inhomogeneous Poisson process. Hence our strategy is to take R as an m-variate inhomogeneous Poisson process (when conditioned on Z_0) with a time varying intensity that possibly depends on Z_0 and the observed value y of $Y(T)$ under a y-dependent simulation probability measure P_0^y. We evenly divide $[0, T]$ into N pieces so the time grid $(0, t_1, t_2, ..., T)$ satisfies $t_{j+1} - t_j = \Delta t = \frac{T}{N}$. We use a matrix representation for the piecewise constant intensity $\overline{\lambda} \in \mathbb{R}_+^{m \times N}$, where $\overline{\lambda}_{i,j} = \lambda_i(t_j)$. We present the interpolation algorithm in Algorithm 8.

Then a y dependent Girsanov transformation is performed via an associated weight $L_y(T)$ so that under the resulting probability measure P_y given by $dP_y = L_y(T)dP_0^y$, the process Z has the intensity that corresponds to the given reaction network specification. A similar weight $\tilde{L}_y(T)$ is associated with $\tilde{Z}$. The details of the Girsanov transformation are given in section 4.3.4. The weight $L_y(T)$ needs to be multiplied by W_p to obtain the overall weight of a trajectory $\tilde{Z}$.

So far, we have outlined a (weighted) Monte Carlo algorithm for the *targeting problem*. That is, to compute $\pi(t, z)$ for $0 \leq t \leq T$ in the context of a single snapshot observation at time $T > 0$ given the observation $Y(T) = y$ and initial distribution μ for Z_0. We summarize the algorithm in Algorithm 7. This algorithm produces as output an iid weighted sample of size N_s for path functional $C(t) = \phi(\tilde{Z}(t))$ with associated weight $W^{(i)}\tilde{L}^{(i)}(T)$.

Algorithm 7 Targeting Algorithm

1: **Input:** Initial distribution μ, T, filter sample size N_s, observation y.
2: Make a judicious choice of intensity function $\lambda(t, \tilde{z}_0, y)$ that is strictly positive.
3: **for** $i = 1$ to N_s **do**
4: Generate a sample from μ resulting in $\tilde{Z}_0^{(i)} = (U_0^{(i)}, V_0^{(i)})$ and generate an m-variate Poisson random variable $K'^{(i)}$ with mean $\int_0^T \lambda(t, \tilde{Z}_0^{(i)}, y)dt$. Set $K''^{(i)} = G(y - V_0^{(i)}, K'^{(i)})$. If $K''^{(i)} \in \mathbb{Z}_+^{m_2}$ accept and go to next step. Otherwise repeat.
5: Set $W^{(i)} = \rho(K''^{(i)}, \tilde{Z}_0^{(i)}, y)$.
6: Generate interpolated trajectories for $\tilde{Z}^{(i)}(t)$ and $I^{(i)}(t)$ for $t \in [0, T]$ using Algorithm 8.
7: Compute the likelihood ratio $\tilde{L}^{(i)}(T)$ by equation (4.16).
8: Compute the path functional $C^{(i)} = \phi(\tilde{Z}^{(i)})$.
9: **end for**
10: **Output:** Weighted sample $(C^{(i)}, W^{(i)}\tilde{L}^{(i)}(T))$ for $i = 1, \ldots, N_s$.

4.3 Rigorous Treatment of the Targeting Algorithm

The targeting problem for a reaction network is stated as follows. We are given the following information about the lab process Z.

1. An observed set of events O_p from the past (before time $t = 0$ which is referred to as "present") with $P(O_p) > 0$.

2. The conditional probability μ of the lab process Z at time $t = 0$. Thus, for

Algorithm 8 Poisson Interpolation Algorithm

Input: Initial state z_0, final time T, total reaction count $k \in \mathbb{Z}_+^m$, intensity matrix $\overline{\lambda} \in \mathbb{R}_+^{m \times N}$, where N is the number of intervals.
Initialize $t = 0$, $I(t) = 0$, $\tilde{Z}(0) = z_0$, $\Delta t = \frac{T}{N}$
Compute $\mu_i = \sum_{j=1}^{N} \overline{\lambda}_{i,j} \Delta t$ for $i = 1, ..., m$
for $j = 1 : N$ **do**
    Generate reaction count $r_{i,j}$ in $(t_{j-1}, t_j] \sim \text{Bino}(k_i, \frac{\overline{\lambda}_{i,j} \Delta t}{\mu_i})$ for $i = 1, ...m$.
    Compute $r_j = \sum_{i=1}^{m} r_{i,j}$
    Generate t_l^i uniformly distributed in $(t_{j-1}, t_j]$ for $l = 1, ..., r_{i,j}$, $i = 1, ..., m$.
    Sort t_l^i so that $t^{(1)} < t^{(2)} ... < t^{(r_j)}$, $t^{(k)} = t_j^i$, and record $(t^{(k)}, i)$ pair. Let $t^{(0)} = t_{j-1}$
    for $k = 1 : r_j$ **do**
        $I(s) = I(t), \tilde{Z}(s) = \tilde{Z}(t) + \nu_i$ for $s \in [t^{(k-1)}, t^{(k)})$
        $t = t^{(k)} = t_j^i$, $I_i(t) = I_i(t-) + 1$, $\tilde{Z}(t) = \tilde{Z}(t-) + \nu_i$
    end for
    $\mu_i \leftarrow \mu_i - \overline{\lambda}_{i,j} \Delta t, \quad k_i \leftarrow k_i - r_{i,j}$ for $i = 1, .., m$
end for
Output: $I(t)$ and $\tilde{Z}(t)$ for $t \in [0, T]$

$z_0 \in \mathbb{Z}_+^n$

$$\mu(z_0) = P\{Z_0 = z_0 \,|\, O_p\}.$$

3. Future time $T > 0$ and future observation $Y(T) = y$ where $y \in \mathbb{Z}_+^{n_2}$.

The goal is to compute the conditional distribution $\pi(t, z)$ where for $t \in [0, T]$ and $z \in \mathbb{Z}_+^n$

$$\pi(t, z) = P\{Z(T) = z \,|\, O_p, Y(T) = y\}.$$

The targeting algorithm consists of using an alternate probability measure P_0 such that under P_0 and under conditioning on O_p and $\{Z_0 = z_0\}$, the lab process R for reaction counts is an inhomogeneous Poisson with a nonvanishing intensity $\lambda(t, z_0, y) > 0$ on $[0, T]$. Then the filter processes $\tilde{Z}$ and I (proxies for Z and R) are created such that, under P_0 and when conditioned on O_p and $\{\tilde{Z}_0 = z_0\}$, I is an

inhomgeneous Poisson with a nonvanishing intensity $\lambda(t, z_0, y) > 0$ on $[0, T]$.

We remark that it is conceptually easier to think of the role of the future observation y as fixed. Thus, for a given targeting problem, y is fixed and the probability measure P_0 depends on y, thus $P_0 = P_0^y$. We shall suppress the superscript y. The desired probability measure $P = P^y$ under which R has the correct intensity is accomplished by a Girsanov change of measure: $dP = L(T)dP_0$. Since the filter process $\tilde{Z}$ will be used for the estimation, the probability measure to use shall be $d\tilde{P} = \tilde{L}(T)dP_0$.

4.3.1 The Lab Process

We start with $(\Omega, \mathcal{F}, P_0)$, a probability space. Let $T > 0$, $X_0 : \Omega \to \mathbb{Z}_+^{n_1}$ and $Y_0 : \Omega \to \mathbb{Z}_+^{n_2}$ be random variables. Define $Z_0 = (X_0, Y_0)$. Let $\nu \in \mathbb{Z}^{n \times m}$.

Let $R : [0, T] \times \Omega \to \mathbb{Z}_+^m$ be an m-variate counting process which is broken into $R = (R', R'')$ where R' and R'' are m_1-variate and m_2-variate respectively. This means for $i = 1, \ldots, m$, R_i is a *cadlag* process which is non-decreasing, takes values in $\mathbb{Z}_+$ and $R_i(0) = 0$.

Define processes X, Y by

$$X(t) = X_0 + \nu' R(t),$$

$$Y(t) = Y_0 + \nu'' R(t).$$

The $\mathbb{Z}^n$ valued process Z is defined by $Z(t) = (X(t), Y(t))$. Thus we may write

$$Z(t) = Z_0 + \nu R(t).$$

We let $\mathcal{Y}_p \subset \mathcal{F}$ denote a sub σ-algebra of observation events that "happened prior" to time $t = 0$ and we assume $O_p \in \mathcal{Y}_p$ is a set of "past observations" with $P_0(O_p) > 0$. We define

$$\mathcal{F}_t = \mathcal{Y}_p \bigvee \sigma(Z(s), R(s) \,|\, 0 \leq s \leq t).$$

Recall that we may assume that with suitable splitting of the reactions $R(t) = (R'(t), R''(t))$, we may write ν'' as

$$\nu'' = [A \; B] \tag{4.13}$$

where B is invertible. This yields

$$Y(T) = Y_0 + AR'(T) + BR''(T)$$

and hence

$$R''(T) = B^{-1}(Y(T) - Y_0 - AR'(T)).$$

For notational convenience we define $G : \mathbb{Z}^{n_2} \times \mathbb{Z}_+^{m_1} \to \mathbb{Z}_+^{m_2}$ by

$$G(\Delta y, k') = B^{-1}(\Delta y - Ak')$$

for $\Delta y \in \mathbb{Z}^{n_2}$ and $k' \in \mathbb{Z}_+^{m_1}$. For $\Delta y \in \mathbb{Z}^{n_2}$ we also define the subset $S_{\Delta y} \subset \mathbb{Z}_+^{m_1}$ by

$$S_{\Delta y} = \{k' \in \mathbb{Z}_+^{m_1} \mid B^{-1}(\Delta y - Ak') \in \mathbb{Z}_+^{m_2}\}.$$

For $\Delta y \in \mathbb{Z}^{n_2}$ and $k = (k', k'') \in \mathbb{Z}_+^{m_1} \times \mathbb{Z}_+^{m_2}$ the following equivalence holds:

$$\Delta y = \nu'' k \iff k' \in S_{\Delta y} \text{ and } k'' = G(\Delta y, k').$$

Since $Y(T) = Y_0 + \nu'' R(T)$, we note that for $y \in \mathbb{Z}^{n_2}$

$$\{Y(T) - Y_0 = y\} \subset \{R'(T) \in S_y\}.$$

We are given $\lambda : [0, T] \times \mathbb{Z}_+^{n_1} \times \mathbb{Z}_+^{n_2} \times \mathbb{Z}_+^{n_2} \to (0, \infty)^m$, a positive function. We are also given μ, be a probability measure on $\mathbb{Z}_+^n$.

We fix $y \in \mathbb{Z}_+^{n_2}$.

We suppose that the following hold under P_0:

1. Conditioned on O_p, $Z_0 = (X_0, Y_0)$ has the distribution $\mu(x_0, y_0)$.

2. Conditioned on O_p and on $Z_0 = z_0 = (x_0, y_0)$, R is an m-variate inhomogeneous Poisson process with intensity $\lambda(t, x_0, y_0, y)$.

We also split λ into $\lambda = (\lambda', \lambda'')$ where λ' is $\mathbb{R}^{m_1}$ valued and λ'' is $\mathbb{R}^{m_2}$ valued.

Later on we shall consider another probability measure P on $(\Omega, \mathcal{F})$ under which R is an m-variate counting process with $\mathcal{F}_t$-intensity $a(Z(t-))$. Thus, under P, the processes R and Z represent the reaction network. Since λ is strictly positive,

P will be absolutely continuous with respect to P_0.

Remark 3. When we refer to an m-variate Poisson process or an m-variate Poisson random variable with an m-vector valued intensity or mean, we assume that the components are independent and have intensity or mean equal to the corresponding component.

4.3.2 Construction of Filter Random Variables $\tilde{Z}_0 = (U_0, V_0)$, K and W

On $(\Omega, \mathcal{F}, P_0)$ we create a sequence $(U_{0,i}, V_{0,i}, K_i')$ which is $\mathbb{Z}_+^{n_1} \times \mathbb{Z}_+^{n_2} \times \mathbb{Z}_+^{m_1}$ with the following properties (under P_0).

1. Conditioned on O_p, (X_0, Y_0) and $(U_{0,i}, V_{0,i})$ (for $i \in \mathbb{N}$) are all independent, and have the same distribution $\mu(x_0, y_0)$.

2. Conditioned on O_p, K_i' for $i \in \mathbb{N}$ is an iid sequence that is independent of X_0, Y_0 and R.

3. For each i, when conditioned on O_p and on $(U_{0,i}, V_{0,i}) = (x_0, y_0)$, K_i' is an m_1-variate Poisson random variable with mean $\int_0^T \lambda'(t, x_0, y_0, y)dt$. Thus for $k' \in \mathbb{Z}_+^{m_1}$ and $(x_0, y_0) \in \mathbb{Z}_+^n$

$$P_0\{K_i' = k' \,|\, U_{i,0} = x_0, V_{i,0} = y_0, O_p\} = P_0\{R'(T) = k' \,|\, X_0 = x_0, Y_0 = y_0, O_p\}.$$

Define $\rho : \mathbb{Z}_+^{m_2} \times \mathbb{Z}_+^{n_1} \times \mathbb{Z}_+^{n_2} \times \mathbb{Z}_+^{n_2} \to \mathbb{R}$ as follows. For each x_0, y_0, y, $\rho(\cdot, x_0, y_0, y)$ is the p.m.f. of an m_2-variate Poisson random variable with mean $\int_0^T \lambda''(t, x_0, y_0, y)dt \in$

$\mathbb{R}^{m_2}$.

Define the random variables N, K', K'', U_0, V_0 and W as follows. Define N by

$$N = \min\{i \in \mathbb{N} \,|\, K'_i \in S_{y-V_{0,i}}\}.$$

Throughout this section we make the following (sensible) assumption.

Assumption 1. $P\{Y(T) = y\} > 0$.

Since P is absolutely continuous with respect to P_0, it follows that $P_0\{Y(T) = y\} > 0$ and hence there exists a y_0 with $P_0\{Y_0 = y_0\} > 0$ such that S_{y-y_0} is nonempty. Therefore

$$P_0\{R'(T) \in S_{y-y_0}\} > 0.$$

Hence, it follows that

$$P_0\{R'_i \in S_{y-V_{0,i}}\} = \sum_{\tilde{y_0}} P_0\{R'_{0,i}(T) \in S_{y-\tilde{y_0}} \,|\, V_0 = \tilde{y_0}\} P_0\{V_0 = \tilde{y_0}\}$$

$$> P_0\{R'(T) \in S_{y-y_0} \,|\, Y_0 = y_0\} P_0\{Y_0 = y_0\} > 0$$

and thus

$$P_0\{N < \infty\} = 1.$$

Thus under Assumption 1, N is finite and K', K'' and W_p below are well defined: $K' = R'_N$, $K'' = G(y - V_{0,N}, K')$, $U_0 = U_{0,N}, V_0 = V_{0,N}$, $W_p = \rho(K'', U_0, V_0, y)$. Let $K = (K', K'')$. We let $\tilde{Z}_0 = (U_0, V_0)$.

We shall denote the observed event

$$O_y = \{Y(T) = y\},$$

and define $O = O_y \cap O_p$ (past and future observations). We note that if there is an observation $Y_0 = y_0$ at $t = 0$ then it is automatically captured in this analysis by taking the $\sum_{x_0} \mu(x_0, y_0) = 1$ and $\mu(x_0, \tilde{y}_0) = 0$ for $\tilde{y}_0 \neq y_0$.

Lemma 7. *For $i \in \mathbb{N}$, $z_0 = (x_0, y_0) \in \mathbb{Z}_+^n$ and $k' \in \mathbb{Z}_+^{m_1}$*

$$P_0\{U_{0,i} = x_0, V_{0,i} = y_0, K_i' = k' \,|\, N = i, O\} = 1_{S_{y-y_0}}(k') \frac{P_0\{X_0 = x_0, Y_0 = y_0, R'(T) = k' \,|\, O_p\}}{P_0\{R'(T) \in S_{y-Y_0} \,|\, O_p\}}.$$

Proof. If $k' \notin S_{y-y_0}$ then both sides of the equation in the lemma are zero. Suppose $k' \in S_{y-y_0}$. Then we have that

$$\begin{aligned}
&P_0\{U_{0,i} = x_0, V_{0,i} = y_0, K_i' = k' \,|\, N = i, O\} \\
&= P_0\{U_{0,i} = x_0, V_{0,i} = y_0, K_i' = k' \,|\, K_i' \in S_{y-V_{0,i}}, K_j' \notin S_{y-V_{0,j}} \text{ for } j < i, O_y, O_p\} \\
&= \frac{P_0\{U_{0,i} = x_0, V_{0,i} = y_0, K_i' = k', K_i' \in S_{y-y_0}, K_j' \notin S_{y-V_{0,j}} \text{ for } j < i, O_y \,|\, O_p\}}{P_0\{K_i' \in S_{y-V_{0,i}}, K_j' \notin S_{y-V_{0,j}} \text{ for } j < i, O_y \,|\, O_p\}} \\
&= \frac{P_0\{U_{0,i} = x_0, V_{0,i} = y_0, K_i' = k' \,|\, O_p\}}{P_0\{K_i' \in S_{y-V_{0,i}} \,|\, O_p\}}.
\end{aligned}$$

We note that the third equality follows since the other events in the numerator are independent of $Y(T)$ (and thus of O_y) when conditioned on O_p. Moreover, when conditioned on O_p, $(U_{0,i}, V_{0,i})$ is distributed like (X_0, Y_0), and conditioned on $(U_{0,i}, V_{0,i}) = (x_0, y_0)$ and O_p, K_i' is distributed like $R'(T)$ when conditioned on

$(X_0, Y_0) = (x_0, y_0)$ and O_p. Thus, we have

$$P_0\{U_{0,i} = x_0, V_{0,i} = y_0, K_i' = k' \,|\, O_p\} = P_0\{X_0 = x_0, Y_0 = y_0, R'(T) = k' \,|\, O_p\}.$$

Also, $P_0\{K_i' \in S_{y-V_{0,i}} \,|\, O_p\} = P_0\{R'(T) \in S_{y-Y_0} \,|\, O_p\}$. These observations clinch the result. □

Proposition 3. For $z_0 = (x_0, y_0) \in \mathbb{Z}^n, k' \in \mathbb{Z}_+^{m_1}$ we have that

$$P_0\{\tilde{Z}_0 = z_0, K' = k' \,|\, O\} = 1_{s_{y-y_0}}(k') \, \frac{P_0\{Z_0 = z_0, R'(T) = k' \,|\, O_p\}}{P_0\{R'(T) \in S_{y-Y_0} \,|\, O_p\}}.$$

Proof. We may write

$$P_0\{U_0 = x_0, V_0 = y_0, K' = k' \,|\, O\} = \sum_{i \in \mathbb{N}} P_0\{U_{0,i} = x_0, V_{0,i} = y_0, K' = k' \,|\, N = i, O\} P_0\{N = i \,|\, O\}.$$

Applying Lemma 7 yields the result. □

Proposition 4. For $z_0 = (x_0, y_0) \in \mathbb{Z}^n, k \in \mathbb{Z}_+^m$ we have that

$$P_0\{Z_0 = z_0, R(T) = k \,|\, O\} = \frac{\mathbb{E}_0[1_{\{(z_0,k)\}}(\tilde{Z}_0, K) \, W \,|\, O]}{\mathbb{E}_0[W \,|\, O]}.$$

Proof. Let $k = (k', k'')$ where $k' \in \mathbb{Z}_+^{m_1}, k'' \in \mathbb{Z}_+^{m_2}$. If $y \neq y_0 + \nu'' k$ then both sides of the equation in Proposition 4 are zero.

Suppose $y = y_0 + \nu'' k$. Then $k'' = G(y - y_0, k')$, $k' \in S_{y-y_0}$ and the event

equality $\{K = k\} = \{K' = k'\}$ holds. Hence

$$\mathbb{E}_0[1_{\{(z_0,k)\}}(\tilde{Z}_0, K)\, W \,|\, O] = \frac{\mathbb{E}_0[W\,;\, \tilde{Z}_0 = z_0, K' = k', O]}{P_0\{O\}}.$$

We note that on $\{K' = k', \tilde{Z}_0 = z_0\}$

$$W = \rho(k'', x_0, y_0, y) = P_0\{R''(T) = k'' \,|\, Z_0 = z_0\} = P_0\{R''(T) = k'' \,|\, Z_0 = z_0, O_p\}.$$

Hence

$$\begin{aligned}
&\mathbb{E}_0[1_{\{(z_0,k)\}}(\tilde{Z}_0, K)\, W \,|\, O] \\
&\quad = P_0\{R''(T) = k'' \,|\, Z_0 = z_0\}\, P_0\{\tilde{Z}_0 = z_0, K' = k' \,|\, O\} \\
&\quad = P_0\{R''(T) = k'' \,|\, Z_0 = z_0\}\, \frac{P_0\{Z_0 = z_0, R'(T) = k' \,|\, O_p\}}{P_0\{R'(T) \in S_{y-Y_0} \,|\, O_p\}} \\
&\quad = P_0\{R''(T) = k'' \,|\, Z_0 = z_0\}\, \frac{P_0\{R'(T) = k' \,|\, Z_0 = z_0\} P_0\{Z_0 = z_0 \,|O_p\}}{P_0\{R'(T) \in S_{y-Y_0} \,|O_p\}} \\
&\quad = \frac{P_0\{R'(T) = k', R''(T) = k'' \,|\, Z_0 = z_0\} P_0\{Z_0 = z_0 \,|\, O_p\}}{P_0\{R'(T) \in S_{y-Y_0} \,|\, O_p\}} \\
&\quad = \frac{P_0\{R(T) = k, Z_0 = z_0 \,|\, O_p\}}{P_0\{R'(T) \in S_{y-Y_0} \,|\, O_p\}}
\end{aligned}$$

where there second equality follows from Proposition 3 (noting that $k' \in S_{y-y_0}$) and the fourth equality follows from the conditional independence of $R'(T)$ and $R''(T)$ given Z_0. We have also used the independence of $(R'(T), R''(T))$ from O_p when conditioned on $Z_0 = z_0$.

Thus for all $k \in \mathbb{Z}_+^m$ we may write

$$\mathbb{E}_0[1_{\{(z_0,k)\}}(\tilde{Z}_0, K)\, W \,|\, O] = 1_{\{y-y_0\}}(\nu'' k)\, \frac{P_0\{Z_0 = z_0, R(T) = k \,|\, O_p\}}{P_0\{R'(T) \in S_{y-Y_0} \,|\, O_p\}}.$$

Since

$$1_{\{y-y_0\}}(\nu'' k) 1_{\{z_0\}}(Z_0) 1_{\{k\}}(K) = 1_{\{y\}}(Y(T) 1_{\{z_0\}}(Z_0) 1_{\{k\}}(K),$$

it follows that

$$\mathbb{E}_0[1_{\{(z_0,k)\}}(\tilde{Z}_0, K)\, W \,|\, O] = \frac{P_0\{Z_0 = z_0, R(T) = k, Y(T) = y \,|\, O_p\}}{P_0\{R'(T) \in S_{y-Y_0} \,|\, O_p\}}. \tag{4.14}$$

Summing over (z_0, k) we obtain

$$\mathbb{E}_0[W \,|\, O] = \frac{P_0\{Y(T) = y \,|\, O_p\}}{P_0\{R'(T) \in S_{y-Y_0} \,|\, O_p\}}. \tag{4.15}$$

Dividing (4.14) by (4.15) we obtain

$$\begin{aligned} \frac{\mathbb{E}_0[1_{\{(z_0,k)\}}(\tilde{Z}_0, K)\, W \,|\, O]}{\mathbb{E}_0[W \,|\, O]} &= \frac{P_0\{Z_0 = z_0, R(T) = k, O_y \,|\, O_p\}}{P_0\{O_y \,|\, O_p\}} \\ &= P_0\{Z_0 = z_0, R(T) = k \,|\, O\}. \end{aligned}$$

□

4.3.3 Construction of the Proxy Processes I, U, V

Having constructed (U_0, V_0, K), we define the process $I : [0,T] \times \Omega \to \mathbb{Z}_+^m$ as follows. Conditioned on $(U_0, V_0) = (x_0, y_0)$, we let $\eta_i(t) = \int_0^t \lambda(s, x_0, y_0, y) ds$ for $t \in [0,T]$. Note that η_i are strictly increasing functions. Let ξ_l^i for $i = 1, 2, \ldots, m$ and $l = 1, 2, \ldots$ be an i.i.d. collection of random variables uniformly distributed on $[0, \eta_i(T)]$ and independent of X_0, Y_0, R and K. Let $t_l^i = \eta_i^{-1}(\xi_l^i)$ Then we set

$$I_i(t) = \sum_{l=1}^{K_i} 1_{[t_l^i, \infty)}(t).$$

The process I is a proxy for R and by construction $I(T) = K$. We define U and V by

$$U(t) = U_0 + \nu' I(t),$$

$$V(t) = V_0 + \nu'' I(t).$$

Let us denote the restriction of R to $[0,T]$ also by R. We note that the conditional law of R given $R(T) = k, X_0 = x_0, Y_0 = y_0, O$ is the same as the conditional law of I given $I(T) = K = k, U_0 = x_0, V_0 = y_0, O$, since both will be "inhomogeneous Poisson bridge" processes with intensity $\lambda(t, x_0, y_0, y)$ under the said conditioning. We summarize this as the following proposition.

Proposition 5. Let $\phi : D_{\mathbb{R}^m}[0,T] \to [0, \infty)$ be a nonnegative measurable map. Then for $k \in \mathbb{Z}_+^m, x_0 \in \mathbb{Z}^{n_1}$, we have that

$$\mathbb{E}_0[\phi(R) \,|\, R(T) = k, X_0 = x_0, Y_0 = y_0, O] = \mathbb{E}_0[\phi(I) \,|\, I(T) = k, U_0 = x_0, V_0 = y_0, O].$$

Proposition 6. Let $\phi : D_{\mathbb{R}^m}[0,T] \times \mathbb{Z}^{n_1} \to [0,\infty)$ be a nonnegative measurable map. Then

$$\mathbb{E}_0[\phi(R, X_0, Y_0) \,|\, O] = \frac{\mathbb{E}_0[\phi(I, U_0, V_0)\, W \,|\, O]}{E_0[W \,|\, O]}.$$

Proof.

$$\begin{aligned}
&\mathbb{E}_0[\phi(R, X_0, Y_0) \,|\, O] \\
&= \sum_{x_0 \in \mathbb{Z}^{n_1}} \sum_{y_0 \in \mathbb{Z}^{n_2}} \sum_{k \in \mathbb{Z}_+^m} \mathbb{E}_0[\phi(R, x_0, y_0) \,|\, R(T) = k, X_0 = x_0, Y_0 = y_0, O] \\
&\qquad \times P_0\{R(T) = k, X_0 = x_0 Y_0 = y_0 \,|\, O\} \\
&= \sum_{k, x_0, y_0} \mathbb{E}_0[\phi(I, x_0, y_0) \,|\, K = k, U_0 = x_0, V_0 = y_0, O]\, \frac{\mathbb{E}_0[1_{\{(x_0,y_0,k)\}}(U_0, V_0, K)\, W \,|\, O]}{\mathbb{E}_0[W \,|\, O]},
\end{aligned}$$

where in the second equality we have used Propositions 4 and 5. Now

$$\begin{aligned}
\frac{\mathbb{E}_0[1_{\{(x_0,y_0,k)\}}(U_0, V_0, K)\, W \,|\, O]}{\mathbb{E}_0[W \,|\, O]} &= \frac{\mathbb{E}_0[W; K = k, U_0 = x_0, V_0 = y_0, O]}{\mathbb{E}_0[W; O]} \\
&= \frac{\mathbb{E}_0[W \,|\, K = k, U_0 = x_0, V_0 = y_0, O]\, P_0\{K = k, U_0 = x_0, V_0 = y_0 \,|\, O\}}{\mathbb{E}_0[W \,|\, O]}.
\end{aligned}$$

Since the process I and W are independent when conditioned on $I(T) = k, U_0 = x_0, V_0 = y_0, O$,

$$\begin{aligned}
&\mathbb{E}_0[\phi(I, x_0, y_0) \,|\, K = k, U_0 = x_0, V_0 = y_0, O]\, \mathbb{E}_0[W \,|\, K = k, U_0 = x_0, V_0 = y_0, O] \\
&= \mathbb{E}_0[\phi(I, x_0, y_0)\, W \,|\, K = k, U_0 = x_0, V_0 = y_0, O].
\end{aligned}$$

Hence

$$\begin{aligned}
&\mathbb{E}_0[\phi(R, X_0, Y_0) \,|\, O] \\
&= \frac{\sum_{k,x_0,y_0} \mathbb{E}_0[\phi(I, x_0, y_0)\, W \,|\, K = k, U_0 = x_0, V_0 = y_0, O]\, P_0\{K = k, U_0 = x_0, V_0 = y_0 \,|\, O\}}{E_0[W \,|\, O]} \\
&= \frac{\mathbb{E}_0[\phi(I, U_0, V_0)\, W \,|\, O]}{E_0[W \,|\, O]}.
\end{aligned}$$

□

4.3.4 Girsanov Change of Measure

We define the processes $L_j, \tilde{L}_j : [0,T] \times \Omega \to \mathbb{R}$ for $j = 1, \ldots, m$ by

$$\begin{aligned}
L_j(t) &= \left(\prod_{i \geq 1} \frac{a_j(X(T_i^j-), Y(T_i^j-))}{\lambda_j(T_i^j-, X_0, Y_0, y)} 1_{\{T_i^j \leq t\}} \right) \exp\left(\int_0^t (\lambda_j(s, X_0, Y_0, y) - a_j(X(s), Y(s)))\, ds \right), \\
\tilde{L}_j(t) &= \left(\prod_{i \geq 1} \frac{a_j(U(\tilde{T}_i^j-), V(\tilde{T}_i^j-))}{\lambda_j(T_i^j-, U_0, V_0, y)} 1_{\{\tilde{T}_i^j \leq t\}} \right) \exp\left(\int_0^t (\lambda_j(s, U_0, V_0, y) - a_j(U(s), V(s)))\, ds \right),
\end{aligned} \tag{4.16}$$

where T_i^j and $\tilde{T}_i^j$ for $i = 1, 2, \ldots$ are the ith jump time of R_j and I_j respectively.

Define the processes $L, \tilde{L}$ by

$$L(t) = L_1(t) \ldots L_m(t), \quad \tilde{L}(t) = \tilde{L}_1(t) \ldots \tilde{L}_m(t). \tag{4.17}$$

We shall show that we can write

$$L(t) = \psi_t(R, X_0, Y_0) \text{ and } \tilde{L}(t) = \psi_t(I, U_0, V_0).$$

for each $t \in [0,T]$ where $\psi_t : D_{\mathbb{R}^m}[0,T] \times \mathbb{Z}^n$ is a measurable map. This is accomplished as follows.

Given $r \in D_{\mathbb{R}^m}[0,T]$, we shall say r is a counting path provided for each $j = 1, \ldots, m$ and for all $t \in (0,T]$

1. $r_j(0) = 0$,
2. $r_j(t) \in \mathbb{Z}_+$,
3. $r_j(t) - r_j(t-) \in \{0,1\}$.

Denote the set of all counting paths by $D_c \subset D_{\mathbb{R}^m}[0,T]$. The set D_c is closed in the Skorohod topology and hence measurable.

Given $r \in D_{\mathbb{R}^m}[0,T]$, $(x_0, y_0) \in \mathbb{Z}^n$ define x, y as follows. For $t \in [0,T]$

$$x(t) = x_0 + \nu' r(t),$$
$$y(t) = y_0 + \nu'' r(t).$$

Define ψ_t^j for $j = 1, \ldots, m$ by

$$\psi_t^j(r, x_0, y_0) = \left(\prod_{i \geq 1} \frac{a_j(x(t_i^j-), y(t_i^j-))}{\lambda_j(t_i^j-, x_0, y_0, y)} 1_{\{t_i^j \leq t\}} \right) \exp\left(\int_0^t (\lambda_j(s, x_0, y_0, y) - a_j(x(s), y(s)))\, ds \right),$$

if $r \in D_c$ and $\psi_t^j(r, x_0, y_0) = 1$ otherwise. Here $t_1^j < t_2^j < \ldots$ are the successive jump times of r_j when $r \in D_c$. Define $\psi_t(r)$ by

$$\psi_t(r, x_0, y_0) = \psi_t^1(r, x_0, y_0)\psi_t^2(r, x_0, y_0) \ldots \psi_t^m(r, x_0, y_0).$$

It can be verified that ψ_t is a measurable map.

Under certain integrability conditions we may assume that $\mathbb{E}_0(L(T)) = 1$. Let $\mathcal{F}_t = \sigma(X_0, Y_0, R(s); 0 \le s \le t)$. It follows that L is a $(\{\mathcal{F}_t\}, P_0)$ martingale and $\mathbb{E}_0(L(t)) = 1$ for all $t \in [0, T]$ [8]. We define the probability measure P on $(\Omega, \mathcal{F})$ by $P = L(T)\, P_0$. It follows that the P intensity of R is $a(X_0 + \nu' R(t-), Y_0 + \nu'' R(t-)) = a(X(t-), Y(t-))$ which is the desired intensity.

Theorem 8. *Let $\phi : D_{\mathbb{R}^m}[0, T] \times \mathbb{Z}_+^n$ be nonnegative measurable. Then*

$$\mathbb{E}[\phi(R, Z_0) \,|\, O] = \frac{\mathbb{E}_0[\phi(I, \tilde{Z}_0)\, \tilde{L}(T)\, W | \, O]}{\mathbb{E}_0[\tilde{L}(T)\, W | \, O]}$$

Proof.

$$\begin{aligned}\mathbb{E}[\phi(R, Z_0) \,|\, O] &= \frac{\mathbb{E}[\phi(R, Z_0);\, O]}{P\{O\}} \\ &= \frac{\mathbb{E}_0[\phi(R, Z_0)\, L(T);\, O]}{\mathbb{E}_0[L(T)\,;\, O]} = \frac{\mathbb{E}_0[\phi(R, Z_0)\, L(T) \,|\, O]}{\mathbb{E}_0[L(T) \,|\, O]}.\end{aligned}$$

Since on O, $L(T) = \psi_T(R, Z_0)$ and $\tilde{L}(T) = \psi_T(I, \tilde{Z}_0)$, by Proposition 6, it follows that

$$\mathbb{E}_0[\phi(R, Z_0)\, L(T) \,|\, O] = \frac{\mathbb{E}_0[\phi(I, \tilde{Z}_0)\, \tilde{L}(T)\, W | \, O]}{\mathbb{E}_0[W | \, O]},$$

and

$$\mathbb{E}_0[L(T) \,|\, O] = \frac{\mathbb{E}_0[\tilde{L}(T)\, W | \, O]}{\mathbb{E}_0[W | \, O]}.$$

From the last three equations the result follows. □

Corollary 9. *Let $\phi : D_{\mathbb{R}^m}[0,T] \times \mathbb{Z}_+^n$ be measurable and $\phi(R, Z_0)$ be integrable. Then*

$$\mathbb{E}[\phi(R, Z_0) \,|\, O] = \frac{\mathbb{E}_0[\phi(I, \tilde{Z}_0)\, \tilde{L}(T)\, W | \, O]}{\mathbb{E}_0[\tilde{L}(T)\, W | \, O]}$$

Corollary 10. *Let $z_0 \in \mathbb{Z}_+^n, k \in \mathbb{Z}_+^m$. Then for $t \in [0,T]$*

$$\mathbb{E}[1_{\{(z_0,k)\}}(Z_0, R(t)) \,|\, O] = \frac{\mathbb{E}_0[1_{\{(z_0,k)\}}(\tilde{Z}_0, I(t))\, \tilde{L}(T)\, W | \, O]}{\mathbb{E}_0[\tilde{L}(T)\, W | \, O]}$$

Corollary 11. *For $z \in \mathbb{Z}_+^n$ and $t \in [0,T]$*

$$\mathbb{E}[1_{\{z\}}(Z(t)) \,|\, O] = \frac{\mathbb{E}_0[1_{\{z\}}(\tilde{Z}(t))\, \tilde{L}(T)\, W | \, O]}{\mathbb{E}_0[\tilde{L}(T)\, W | \, O]}$$

4.4 Tuning the Targeting Algorithm

Within the implementation of our targeting algorithm, there are choices we could make such as the intensity of the inhomogeneous Poisson process $\lambda(t)$, the frequency to resample the filter. In this section, we discuss these factors and also present a "two-stage method" that runs (unconditional) forward evolution in the first portion of the time interval then applies the targeting algorithm in the second portion.

Choosing inhomogeneous Poisson intensities λ: Our targeting method involves choosing inhomogeneous Poisson intensities λ for $R(t)$. It is most convenient to interpolate process I when λ is piecewise constant in time. We evenly divide $[0,T]$ into N pieces with $t_{j+1} - t_j = \Delta t = \frac{T}{N}$ and let $\lambda(t) = \lambda(t_j)$ for $t \in [t_j, t_{j+1})$. We

use a matrix representation for this piecewise constant intensity $\overline{\lambda} \in \mathbb{R}_+^{m\times N}$, where $\overline{\lambda}_{i,j} = \lambda_i(t_j)$. Here we present two strategies we used in our numerical experiments.

One intuitive idea is to only consider the unconditional forward evolution of the system. If we know $\lambda(t) = \mathbb{E}[a(Z(t))]$ analytically, we can construct a left end-point approximation $\overline{\lambda}_1(t) = \lambda(t_j)$ for $t \in [t_j, t_{j+1)})$. Otherwise, we could solve the corresponding reaction rate equations as an approximate or, alternatively, estimate $\mathbb{E}[a(Z(t))]$ via an independent forward simulation.

Our second strategy for choosing λ was based upon the unconditioned $\overline{\lambda}_1(t)$ and involves a constrained optimization scheme that takes the observation into account. We would like our new choice of inhomogeneous Poisson intensity $\overline{\lambda}_2$ to be close to the average intensity $\overline{\lambda}_1(t)$ while also satisfy the condition that the expectation of the process lands precisely at the desired state ($\mathbb{E}_0[\nu'' \int_0^T I(t)dt] = y - y_0$). To be specific, we solve the following constrained optimization problem to get intensity $\overline{\lambda}_2$:

$$\overline{\lambda}_2 = \text{argmin}_{\lambda\in\mathbb{R}_+^{m\times N}} \|\lambda - \overline{\lambda}_1\|_F \quad \text{s.t. } \nu'' r = \Delta y, \text{ where } r_i = \sum_{j=1}^{N} \lambda_{i,j}\Delta t. \tag{4.18}$$

Here $\|\cdot\|_F$ stands for the Frobenius norm. In our numerical examples, we use Matlab program `fmincon` to solve this optimization problem.

Resampling: Since the disparity of the weights will usually grow fast as time progresses, we could again use the resampling technique discussed in Section 3.2.1 to alleviate the issue. Bear in mind that the resampling procedure itself introduces extra randomness which could hurt the accuracy on the other hand. The frequency

to perform resampling is an interesting question to explore.We could resample every Δs, it is easier to implement it numerically if Δs is a multiple of the Δt.

Two-stage method: We introduce the *two-stage method* as another attempt to address the issue of growing variance of the weights. For estimating the system at time t $(0 < t < T)$, especially when t is close to T, we could split the interval $[0, T]$ into two pieces $[0, t_0)$ and $[t_0, T]$ where $0 \leq t_0 \leq t \leq T$ and run the (unconditional) forward simulation in time interval $[0, t_0]$ and then apply the targeting algorithm in the interval $[t_0, T]$. This way, we reduce the length of time interval where the weight grow disparately. We will also explore the choice of t_0 for splitting the interval in Example 3.

4.5 Numerical experiments

In this section, we use numerical examples to demonstrate our targeting algorithms and compare it with other algorithms. We focus on the problem of a single observation snapshot given $Y(0) = y_0$, $Y(T) = y$.

We start with two analytically tractable examples: a pure death process (Example 1) and a reversible isomerization model (Example 2):

$$\text{Example 1:} \qquad S \xrightarrow{c} \varnothing,$$

$$\text{Example 2:} \qquad S_1 \xrightarrow{c_1} S_2, \quad S_2 \xrightarrow{c_2} S_1.$$

In both scenarios, we would like to estimate the conditional distribution at an inter-

mediate time. Since the true conditional distribution is available, we could compare the accuracy of various methods in different scenarios. We choose a common observation as well as a rare one.

Then in Example 3, we look at a reaction network with non-linear propensities. Unlike the first two examples, we could only compute the conditional distribution at the final time by solving the Kolmogorov equations, hence we estimate the distribution at the final time. We also use this example to illustrate the two-stage algorithm.

$$\text{Example 3:} \qquad S_1 \xrightarrow{c_1} S_2, \quad S_2 \xrightarrow{c_2} S_1, \quad S_1 + S_2 \xrightarrow{c_3} S_3, \quad S_3 \xrightarrow{c_4} S_1 + S_2.$$

Lastly, we study an example where we could analytically compute the conditional expectation of the unobserved species A but not the conditional distribution.

$$\text{Example 4:} \qquad S \xrightarrow{c_1} S + A, \qquad \emptyset \xrightarrow{c_2} S.$$

For the first three examples where we could compute the conditional distribution (analytically or by solving Kolmogorov forward equations), we estimate two statistical distances, the mean total variation error (TVE) and Hellinger distance

$$\text{TVE} = \mathbb{E}\left[\sum_z |\hat{\pi}(t,z) - \pi(t,z)| \,\middle|\, Y(T) = y_T\right],$$

$$\text{Hellinger} = \frac{1}{\sqrt{2}}\,\mathbb{E}\left[\sqrt{\sum_z (\sqrt{\hat{\pi}(t,z)} - \sqrt{\pi(t,z)})^2} \,\middle|\, Y(T) = y_T\right].$$

We also record the effective sample fraction (ESF) defined by the ratio of effective sample size N_e over the filter sample size N_s

$$\text{ESF} = \frac{N_e}{N_s} = \frac{1}{N_s} \frac{\left(\sum_{i=1}^{N_s} W^i\right)^2}{\sum_{i=1}^{N_s} (W^i)^2}.$$

4.5.1 Example 1: Pure-death model

We consider an analytically tractable example, a pure death process of a single species

$$S \xrightarrow{c} \varnothing \tag{4.19}$$

with the propensity $a(x) = cx$.

Assume we have noiseless observation of the species at time $t = 0$ and $t = T$, and we are interested in estimating the conditional distribution

$$\pi(x_t|x_0, x_T) := P\{X(t) = x_t \,|\, X(0) = x_0, X(T) = x_T\}. \tag{4.20}$$

The conditional distribution $\pi(x_t|x_0, x_T)$ is given in terms of transition probability

$$\begin{aligned}
\pi(x_t|x_0, x_T) &= \frac{P\{X(t) = x_t, X(0) = x_0, X(T) = x_T\}}{P\{X(0) = x_0, X(T) = x_T\}} \\
&= \frac{P\{X(T) = x_T \,|\, X(0) = x_0, X(t) = x_t\}\, P\{X(t) = x_t \,|\, X(0) = x_0\}\, P\{X(0) = x_0\}}{P\{X(T) = x_T \,|\, X(0) = x_0\} P\{X(0) = x_0\}} \\
&= \frac{P\{X(T) = x_T \,|\, X(t) = x_t\}\, P\{X(t) = x_t \,|\, X(0) = x_0\}}{P\{X(T) = x_T \,|\, X(0) = x_0\}},
\end{aligned} \tag{4.21}$$

where transition probability from t_1 to t_2 ($t_1 \leq t_2$) follows a binomial distribution with $n = x_1$ and $p = e^{-c(t_2-t_1)}$,

$$P\{X(t_2) = x_2 \,|\, X(t_1) = x_1\} = \binom{x_1}{x_2} e^{-c(t_2-t_1)x_2}(1 - e^{-c(t_2-t_1)})^{x_1-x_2}. \tag{4.22}$$

.

Using equations (4.21) and (4.22), we can obtain an analytical formula for the conditional distribution

$$\pi(x_t|x_0, x_T) = \binom{x_0 - x_T}{x_0 - x_t} \left(\frac{e^{-ct} - e^{-cT}}{1 - e^{-cT}}\right)^{x_t - x_T} \left(\frac{1 - e^{-ct}}{1 - e^{-cT}}\right)^{x_0 - x_t}. \tag{4.23}$$

Furthermore, we can also obtain the conditioned propensity

$$\lambda(x_t \,|\, x_T) = a(x_t)\frac{p(x_T|X(t) = x - 1)}{p(x_T|X(t) = x)} = \frac{c(x_t - x_T)}{1 - e^{-c(T-t)}}. \tag{4.24}$$

We also conducted simulations using the conditional propensity function idea discussed in section 4.2.4. The "CP approx" method proposed in [22] uses a piecewise constant approximation of the conditional propensity function that takes constant propensity between two successive events t_1 and t_2, that is, $\lambda(s) = \lambda(x_t \,|\, y_T)$ for $s \in [t_1, t_2)$. This way, the wait time for the next event is exponentially distributed, makes it easier to implement. In what we called the "CP exact" method, we use this time-dependent conditional propensity (4.24) and hence the wait times are non-exponential. Let τ be the waiting time for the first event given $X(0) =$

$x_0, X(T) = x_T$. Then,

$$G(t) := P\{\tau > t\} = \left(\frac{e^{-ct} - e^{-cT}}{1 - e^{-cT}}\right)^{x_0 - x_T}. \tag{4.25}$$

In general, suppose an event happen at time t, given that $X(t) = x_t, X(T) = x_T$, the waiting time τ for the next event has distribution

$$G(s) = P\{\tau > s\} = \left(\frac{e^{-cs} - e^{-c(T-t)}}{1 - e^{-c(T-t)}}\right)^{x_t - x_T}. \tag{4.26}$$

In order to simulate that waiting time τ, we can take an inverse of the cumulative density function $F = 1 - G$

$$\tau = \min\{s : G(s) \leq 1 - u\}, \tag{4.27}$$

where u is a uniform random variable in $[0, 1]$.

Suppose $X(t_1) = x_1, X(T) = x_T$, let t_1 and t_2 denote the times of two successive reaction events. Then the Girsanov transformation likelihood ratio over $[t_1, t_2]$ is given by

$$\begin{aligned} L(t_2) &= L(t_1) \frac{a(x_1)}{\lambda(x(t_2-) \,|\, x_T)} \frac{e^{\int_{t_1}^{t_2} -a(x_1)\, ds}}{e^{\int_{t_1}^{t_2} -\lambda(a(x_s \,|\, x_T)\, ds}} \\ &= L(t_1) \frac{x_1}{x_1 - x_T} \left(1 - e^{-c(T-t_2)}\right) \frac{e^{-cx_1(t_2-t_1)}}{G(\tau)}. \end{aligned}$$

In the following numerical experiments, we took $x_0 = 1000$, $c = 4$, $T = 0.5$, $t = 0.2$, time mesh $\Delta t = 0.02$. In this example, the free dimension $m_1 = 0$, while the determined dimension $m_2 = 1$, so the Poisson weight will always be the same

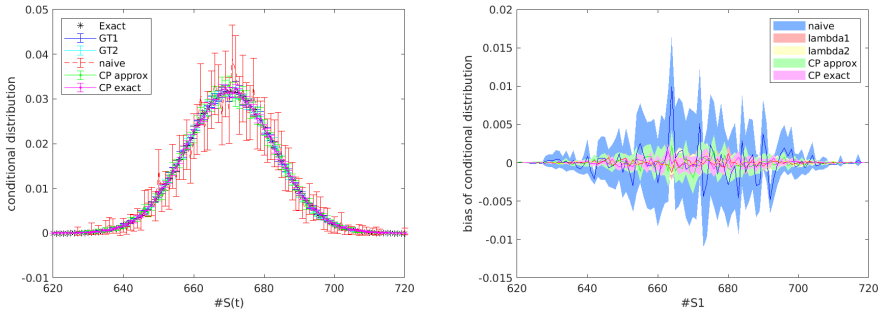


Figure 4.1: Pure death example. Estimated conditional distribution compared with the actual conditional distribution. A common observation. $x_0 = 1000$, $c = 2$, $T = 0.5$, $t = 0.2$. The estimate came with 95% confident interval, which was estimated based on $N_r = 100$ trials, while each trial used filter sample size $N_s = 10,000$.

for different samples. We used inhomogeneous Poisson intensities $\lambda_1(t) = \mathbb{E}[a(Z(t_j)]$ for $t \in [t_j, t_{j+1})$, while $\overline{\lambda}_2$ was obtained via constrained optimization discussed in Section 4.4. We ran each method for $N_r = 100$ independent trials, and for each trial we used the filter sample size $N_s = 1000$. We also recorded the runtime for each filter to compare computational cost. The results are shown in Table 4.1.

Method	Observation	Total Variation Error (95% interval)	Hellinger (95% interval)	Effective Sample Fraction (ESF)	Runtime (seconds)
Naive	$x_T = 368$	1.1353, [1.1077, 1.1630]	0.6057, [0.5968, 0.6147]	0.027	2.3
Targeting ($\bar{\lambda}_1$)	(common observation)	0.2037, [0.1995, 0.2080]	0.1160, [0.1143, 0.1177]	0.919	2.6
Targeting ($\bar{\lambda}_2$)		0.2072, [0.2029, 0.2116]	0.1193, [0.1175, 0.1211]	0.900	2.7
CP (approx)		0.3485, [0.3414, 0.3556]	0.1880, [0.1851, 0.1909]	0.322	2.4
CP (exact)		0.1957, [0.1920, 0.1994]	0.1138, [0.1120, 0.1156]	1.000	2.4
Naive	$x_T = 404$	NaN, [NaN, NaN]	NaN, [NaN, NaN]	NaN	2.4
Targeting ($\bar{\lambda}_1$)	(rare observation)	0.1979, [0.1942, 0.2016]	0.1143, [0.1127, 0.1160]	0.923	2.6
Targeting ($\bar{\lambda}_2$)		0.2297, [0.2248, 0.2347]	0.1412, [0.1386, 0.1438]	0.654	2.4
CP (approx)		0.3351, [0.3275, 0.3427]	0.1834, [0.1804, 0.1864]	0.324	2.1
CP (exact)		0.1909, [0.1872, 0.1947]	0.1117, [0.1100, 0.1135]	1.000	2.3

Table 4.1: Example 1. We took $x_0 = 1000$, $c = 2$, $T = 0.5$, $t = 0.2$, $\Delta t = 0.02$. The estimate came with 95% confident interval, which was estimated based on $N_r = 100$ trials, while each trial has filter sample size $N_s = 1000$.

From Table 4.1, the exact time-dependent conditional propensity method has the best performance, as expected. Our targeting method outperformed the naive

method and the approximate conditional propensity method.

4.5.2 Example 2: Reversible Reaction $S_1 \longleftrightarrow S_2$

We consider the system

$$S_1 \xrightarrow{c_1} S_2, \quad S_2 \xrightarrow{c_2} S_1. \tag{4.28}$$

Let $Z(t) = (\#S_1(t), \#S_2(t))$. Suppose the propensities take mass-action form $a_1(z) = c_1 z_1$ and $a_2(z) = c_2 z_2$. Suppose we know the initial state $z_0 = (x_1, x_2)$ and we have noiseless observation of the copy number of species S_2 at time T.

To find the evolution of the probability distribution, we consider an associated mono-molecular system [38] which is described by a random walk $M(t)$ between two states. This is a continuous-time Markov chain with the rate matrix

$$Q = \begin{pmatrix} -c_1 & c_1 \\ c_2 & -c_2 \end{pmatrix}. \tag{4.29}$$

We can obtain its transition probability matrix $P(t)$ by Kolmogorov's backward equation $P'(t) = Q\,P(t)$, where $P_{i,j}(t) = \text{Prob}\{M(t) = j \,|\, M(0) = i\}$ for $i, j = 1, 2$. For the reversible two species system, suppose the system initially starts with only one species, say $Z_1(0) = x_0$ and $Z_2(0) = 0$, then we have

$$P\{Z_1(t) = x_t, Z_2(t) = x_0 - x_t \,|\, Z_1(0) = x_0, Z_2(0) = 0\} = \text{Binopdf}(x_t; x_0, P_{11}(t)). \tag{4.30}$$

For a general initial condition, we could get the transition probability as well,

$$\begin{aligned} &P\{Z_1(t) = x_t, Z_2(t) = x_1 + x_2 - x_t \,|\, Z_1(0) = x_1, Z_2(0) = x_2\} \\ &= \sum_{k=0}^{x_1+x_2} \text{Binopdf}(k; x_1, P_{11}(t))\text{Binopdf}(x_t - x_1 - k; x_2, P_{21}(t)). \end{aligned} \tag{4.31}$$

Using the transition probability and equation (4.21), we could obtain an analytical formula for the conditional distributions to compare various simulation methods with.

We first took a smaller system with the total copy number of all species being 10. We took $Z(0) = (10, 0)$, $c = (1, 1.5)$, $t = 0.7$, $T = 1$, $\Delta t = 0.1$ and chose a common observation and a rare observation. The results are shown in table 4.5.2.

Method	Observation	TVE (95% interval)	Hellinger (95% interval)	ESF	Runtime
Naive	$y_T = 4$	$0.1188, [0.1104, 0.1272]$	$0.0689, [0.0652, 0.0727]$	0.246	0.05
Targeting ($\overline{\lambda}_1$)	(common observation)	$0.0747, [0.0697, 0.0798]$	$0.0366, [0.0347, 0.0384]$	0.568	0.3
Targeting ($\overline{\lambda}_2$)	$P(Y(T) = 4) = 0.245$	$0.1014, [0.0942, 0.1086]$	$0.0471, [0.0442, 0.0499]$	0.568	0.3
Targeting,resample ($\overline{\lambda}_1$)		$0.0843, [0.0795, 0.0892]$	$0.0543, [0.0515, 0.0571]$	0.955	0.3
Targeting,resample ($\overline{\lambda}_2$)		$0.0772, [0.0725, 0.0820]$	$0.0536, [0.0506, 0.0565]$	0.955	0.3
CP (approx)		$0.0834, [0.0783, 0.0886]$	$0.0459, [0.0437, 0.0480]$	0.518	31.3
Naive	$y_T = 7$	$0.3646, [0.3393, 0.3900]$	$0.2049, [0.1944, 0.2153]$	0.027	0.05
Targeting ($\overline{\lambda}_1$)	(rare observation)	$0.0940, [0.0880, 0.1001]$	$0.0526, [0.0498, 0.0554]$	0.381	0.3
Targeting ($\overline{\lambda}_2$)	$P(Y(T) = 7) = 0.027$	$0.0940, [0.0871, 0.1009]$	$0.0522, [0.0491, 0.0552]$	0.346	0.3
Targeting,resample ($\overline{\lambda}_1$)		$0.0900, [0.0847, 0.0953]$	$0.0597, [0.0575, 0.0619]$	0.834	0.3
Targeting,resample ($\overline{\lambda}_2$)		$0.1136, [0.1071, 0.1202]$	$0.0768, [0.0719, 0.0817]$	0.658	0.3
CP (approx)		$0.0907, [0.0851, 0.0962]$	$0.0526, [0.0498, 0.0554]$	0.429	33.3

Table 4.2: Reversible Reaction $S_1 \longleftrightarrow S_2$. We took $z_0 = (10, 0)$, $c = (1, 1.5)$, $T = 1$, $t = 0.7$, $\Delta t = \Delta s = 0.1$, while taking a common observation as well as a rare observation. The estimate came with 95% confident interval, which was estimated based on $N_r = 100$ trials, while each trial used filter sample size $N_s = 1000$.

We also tried with a larger system with the total copy number of all species being increased to 200. However, the conditional propensity approach takes way too long to compute, so we only compared the naive method and the targeting method

in table 4.5.2.

Method	Observation	TVE (95% interval)	Hellinger (95% interval)	ESF	Runtime (seconds)
Naive	$y_T = 80$	0.1868, [0.1826, 0.1910]	0.1091, [0.1072, 0.1109]	0.0576	19.50
Targeting ($\overline{\lambda}_1$)	(common observation)	0.0665, [0.0646, 0.0684]	0.0311, [0.0303, 0.0318]	0.4309	21.10
Targeting ($\overline{\lambda}_2$)	($p = 0.0575$)	0.0626, [0.0607, 0.0646]	0.0296, [0.0288, 0.0304]	0.4425	21.93
Targeting,resample ($\overline{\lambda}_1$)		0.0732, [0.0710, 0.0754]	0.0478, [0.0468, 0.0488]	0.9489	20.86
Targeting,resample ($\overline{\lambda}_2$)		0.0724, [0.0704, 0.0744]	0.0471, [0.0462, 0.0481]	0.9423,	20.48
Naive	$y_T = 98$	0.6764, [0.6595, 0.6933]	0.3937, [0.3851, 0.4023]	0.0042	19.03
Targeting ($\overline{\lambda}_1$)	(rare observation)	0.0928, [0.0902, 0.0954]	0.0433, [0.0423, 0.0444]	0.2095	21.30
Targeting ($\overline{\lambda}_2$)	($p = 0.0043$)	0.0755, [0.0733, 0.0776]	0.0352, [0.0344, 0.0361]	0.3146	21.95
Targeting,resample ($\overline{\lambda}_1$) $\Delta s = 0.25$		0.0847, [0.0826, 0.0869]	0.0532, [0.0521, 0.0542]	0.8129	21.89
Targeting,resample ($\overline{\lambda}_2$) $\Delta s = 0.25$		0.0894, [0.0871, 0.0917]	0.0549, [0.0538, 0.0560]	0.6194	21.96
Targeting,resample ($\overline{\lambda}_1$), $\Delta s = 0.5$		0.0719, [0.0700, 0.0738]	0.0432, [0.0424, 0.0439]	0.7910	22.26
Targeting,resample ($\overline{\lambda}_2$), $\Delta s = 0.5$		0.0788, [0.0768, 0.0808]	0.0460, [0.0452, 0.0468]	0.6833	22.81

Table 4.3: Example 2 reversible isomerization reaction $S_1 \longleftrightarrow S_2$. We took $z_0 = (100, 100)$, $c = (1, 1.5)$, $T = 2$, $t = 0.7$, $\Delta t = 0.25$, while taking a common observation $y_T = 82$ as well as a rare observation $y_T = 98$. The estimate came with 95% confident interval, which was estimated based on $N_r = 100$ trials, while each trial used filter sample size $N_s = 10,000$.

4.5.3 Example 3: $S_1 \longleftrightarrow S_2, \quad S_1 + S_2 \longleftrightarrow S_3$

We consider

$$S_1 \xrightarrow{c_1} S_2, \quad S_2 \xrightarrow{c_2} S_1, \quad S_1 + S_2 \xrightarrow{c_3} S_3, \quad S_3 \xrightarrow{c_4} S_1 + S_2. \tag{4.32}$$

Let $Z(t) = (\#S_1(t), \#S_2(t), \#S_3(t))$. Suppose we observe species S_3, so $X(t) = (\#S_1(t), \#S_2(t))$ and $Y(t) = \#S_3(t)$. Suppose $a_1(z) = c_1 z_1$, $a_2(z) = c_2 z_2$, $a_3(z) = c_3 z_1 z_2$ and $a_4(z) = c_4 z_3$.

We took $X(0) = (20, 20, 20)$, $c = (0.5, 1, 0.1, 1)$, $T = 1$, and applied the naive method and several variants of targeting algorithm to estimate the conditional distribution of $X(t)$. In this example, we can pick the first three reactions as "free

reactions" and the last one is determined by $R_4 = R_3 - \Delta y$. Since $Z(T) - Z(0) = \nu R(T)$ where

$$\nu = \left[\begin{array}{cccc} -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ \hline 0 & 0 & 1 & -1 \end{array}\right]$$

we have $\Delta Y = Y(T) - Y(0) = R_3(T) - R_4(T)$. In this example, we solved Kolmogorov's forward equation numerically to obtain the conditional distribution to compare our stochastic filter results with.

We state our choices of inhomogeneous intensity $\lambda(t)$ as follows. We took time mesh $\Delta t = 0.1$, and used an approximate deterministic system to determine $\lambda(t)$. That is, we solve the reaction rate equations

$$\frac{dz_1}{dt} = -c_1 z_1 + c_2 z_1 - c_3 z_1 z_2 + c_4 z_3, \quad \frac{dz_2}{dt} = c_1 z_1 - c_2 z_1 - c_3 z_1 z_2 + c_4 z_3, \quad \frac{dz_3}{dt} = c_3 z_1 z_2 - c_4 z_3,$$

and let $\lambda_1(t) = a(z(t_i))$ for $t \in [t_i, t_{i+1})$. Then, we got λ_2 via the constrained optimization discussed in Section 4.4.

We also implement the two-stage algorithm where we run our first stage simulation $[0, t_0)$ (where $0 < t_0 < T$) using Gillespie's forward simulation algorithm, then apply our targeting algorithm in the second stage. In the second stage $[t_0, T]$, one option we take is to use the sample mean propensity from the first stage $\overline{\lambda} = \frac{1}{N_s} \sum a(\tilde{Z}(t_0))$. The other option is to assign each individual sample a propensity that takes account of the state $\tilde{Z}(t_0)$ while also steering the trajectory toward the observed state y_T. In this specific example, there are $m = 4$ reactions, with

the number of the number of free reactions $m_1 = 3$ and the number of determined reaction $m_2 = 1$, so we write $\lambda^{(i)} = (\lambda'^{(i)}, \lambda''^{(i)})$ with $\lambda'^{(i)} \in \mathbb{R}^3_+$, $\lambda''^{(i)} \in \mathbb{R}_+$. We choose $\lambda'^{(i)} \in \mathbb{R}^3_+$ to be $\lambda'^{(i)} = \max\{a'(\tilde{Z}^{(i)}(t_0), a'(1,1,1)\}$. The reason we introduce $a'(1,1,1)$ is that $a'(\tilde{Z}^{(i)}(t_1))$ could possibly contain zero components, while our method requires strictly positive intensities, so we enforce a small lower bound corresponding to the situation where there is one molecule for each species present. Our choice $a'(1,1,1)$ is usually a small quantity compared to $a'(\tilde{Z}^{(i)}(t_0)$ while sufficiently safeguarding the strict positivity of intensity $\lambda(t)$. As a targeting algorithm, we obtain $\lambda''^{(i)}$ by solving $\nu''\lambda^{(i)}(T - t_0) = y_T - V^{(i)}(t_0)$, then updated it by $\lambda''^{(i)} = \max\{\lambda''^{(i)}, a''(1,1,1)\}$. This way, we obtain the individual propensity $\lambda^{(i)} = (\lambda'^{(i)}, \lambda''^{(i)})$ for each sample. Numerical results are summarized in Table 4.4 for a common observation and Table 4.5 for a rare observation.

One stage targeting algorithm estimates the conditional distribution more accurately than the naive method. For two stage algorithm, there is an optimal choice of t_0 to split the forward evolution stage and the targeting stage. When the length of the second interval $[t_0, T]$ is large, growing disparity of the Girsanov weights could be a problem. However, if t_0 is very close to the final time T, we may be faced with very disparate Poisson weights W_p, as we're forcing a large number of reactions to occur in a short amount of time.

Method	TVE (95% interval)	ESF	Runtime (seconds)
Naive	$0.2146, [0.2117, 0.2174]$	0.16	0.331
One stage λ_1, $\Delta t = 0.1$	$0.1695, [0.1669, 0.1722]$	0.21	0.773
One stage λ_2, $\Delta t = 0.1$	$0.1699, [0.1673, 0.1726]$	0.21	0.770
Two stage, $\overline{\lambda}$, $t_0 = 0.5$	$0.1331, [0.1310, 0.1352]$	0.35	0.611
Two stage, $\lambda^{(i)}$, $t_0 = 0.5$	$0.1597, [0.1563, 0.1631]$	0.29	0.558
Two stage, $\overline{\lambda}$,$t_0 = 0.8$	$0.1146, [0.1130, 0.1162]$	0.55	0.513
Two stage, $\lambda^{(i)}$, $t_0 = 0.8$	$0.1038, [0.1024, 0.1052]$	0.72	0.520
Two stage, $\overline{\lambda}$,$t_0 = 0.9$	$0.1103, [0.1087, 0.1119]$	0.64	0.500
Two stage, $\lambda^{(i)}$, $t_0 = 0.9$	$0.1024, [0.1010, 0.1038]$	0.78	0.495
Two stage, $\overline{\lambda}$,$t_0 = 0.99$	$0.1857, [0.1832, 0.1882]$	0.39	0.507
Two stage, $\lambda^{(i)}$, $t_0 = 0.99$	$0.1830, [0.1805, 0.1855]$	0.40	0.532
Two stage, $\overline{\lambda}$, $t_0 = 0.999$	$0.2583, [0.2548, 0.2618]$	0.29	0.550
Two stage, $\lambda^{(i)}$, $t_0 = 0.999$	$0.2566, [0.2531, 0.2602]$	0.30	0.590

Table 4.4: Example 3. A common observation. We took $X(0) = (20, 20, 20)$, $c = (0.5, 1, 0.1, 1)$, $t = T = 1$, $y_T = 24$. The estimate came with 95% confident interval, which was estimated based on $N_r = 1,000$ trials, filter sample size $N_s = 1,000$.

Method	TVE (95% interval)	ESF	Runtime (seconds)
Naive	$0.4347, [0.4286, 0.4408]$	0.04	0.326
One stage λ_1, $\Delta t = 0.1$	$0.2184, [0.2152, 0.2217]$	0.15	0.791
One stage λ_2, $\Delta t = 0.1$	$0.2103, [0.2071, 0.2135]$	0.16	0.794
Two stage, $\overline{\lambda}$, $t_0 = 0.5$	$0.1758, [0.1730, 0.1785]$	0.23	0.614
Two stage, $\lambda^{(i)}$, $t_0 = 0.5$	$0.2158, [0.2114, 0.2202]$	0.18	0.590
Two stage, $\overline{\lambda}$, $t_0 = 0.8$	$0.1614, [0.1591, 0.1637]$	0.31	0.527
Two stage, $\lambda^{(i)}$, $t_0 = 0.8$	$0.1639, [0.1617, 0.1660]$	0.32	0.546
Two stage, $\overline{\lambda}$, $t_0 = 0.9$	$0.1971, [0.1942, 0.1999]$	0.26	0.490
Two stage, $\lambda^{(i)}$, $t_0 = 0.9$	$0.1894, [0.1871, 0.1918]$	0.26	0.500
Two stage, $\overline{\lambda}$, $t_0 = 0.99$	$0.3942, [0.3892, 0.3992]$	0.07	0.483
Two stage, $\lambda^{(i)}$,$t_0 = 0.99$	$0.3935, [0.3882, 0.3988]$	0.07	0.547
Two stage, $\overline{\lambda}$,$t_0 = 0.999$	$0.5638, [0.5561, 0.5716]$	0.05	0.479
Two stage, $\lambda^{(i)}$, $t_0 = 0.999$	$0.5601, [0.5527, 0.5674]$	0.05	0.518

Table 4.5: Example 3. A rare observation. We took $X(0) = (20, 20, 20)$, $c = (0.5, 1, 0.1, 1)$, $t = T = 1$, $y_T = 20$. The estimate came with 95% confident interval, which was estimated based on $N_r = 1,000$ trials, filter sample size $N_s = 1,000$.

4.5.4 Example 4: $S \longrightarrow S + A, \quad \emptyset \longrightarrow S$

We consider a two species system

$$S \xrightarrow{c_1} S + A, \qquad \emptyset \xrightarrow{c_2} S.$$

Let $X(t) = \#A(t)$ and $Y(t) = \#S(t)$. The propensity functions are $a_1(z) = c_1 z_2$ and $a_2(z) = c_2$. Suppose we have a deterministic initial condition $Z(0) = z_0 = (x_0, y_0)$. We have

$$\mathbb{E}[Y(t) \,|\, Y(T) = y_T] = y_0 + \frac{t}{T}\Delta y,$$

where $\Delta y = y_T - y_0$. It follows that

$$\mathbb{E}[X(t)|Y(T) = y] = \mathbb{E}\left[x_0 + \int_0^t a_1(Z(s))ds \,|\, Y(T) = y\right] = x_0 + c_1 y_0 t + c_1 \Delta y \frac{t^2}{2T}.$$

We would like to estimate the conditional expectation $\mathbb{E}[X(T)|Y(T) = y]$. We took $Z(0) = (0, 0)$, $t = T = 5$, $c_1 = 0.1$ while c_2 were varied. Naive method has much larger variance in its estimation compared with the targeting algorithm, while targeting method is slightly biased.

Problem	Method	$\hat{\mathbb{E}}[X(T)\|Y(T)=y]$ with 95% interval	ESF	Runtime
$c_2 = 100, T = 5, y = 500$	Naive	$124.899 \pm 0.158, [124.740, 125.057]$	0.0177	27.83
$\mathbb{E}[X(T)\|Y(T)=y] = 125$	Targeting ($\Delta t = 0.1$)	$124.938 \pm 0.040, [124.898, 124.978]$	0.6404	37.44
$c_2 = 100, T = 5, y = 553$	Naive	$138.069 \pm 0.741, [137.328, 138.810]$	0.0011	24.46
$\mathbb{E}[X(T)\|Y(T)=y] = 138.25$	Targeting ($\Delta t = 0.1$)	$138.199 \pm 0.038, [138.162, 138.237]$	0.6150	39.41
$c_2 = 200, T = 5, y = 1000$	Naive	$249.996 \pm 0.254, [249.742, 250.250]$	0.0127	47.03
$\mathbb{E}[X(T)\|Y(T)=y] = 250$	Targeting ($\Delta t = 0.1$)	$249.901 \pm 0.057, [249.844, 249.958]$	0.5326	53.26
$c_2 = 200, T = 5, y = 1074$	Naive	$267.731 \pm 1.246, 266.485, 268.977]$	0.0008	50.04
$\mathbb{E}[X(T)\|Y(T)=y] = 268.50$	Targeting ($\Delta s = 0.1$)	$268.363 \pm 0.056, [268.307, 268.420]$	0.5212	69.68
$c_2 = 400, T = 5, y = 2000$	Naive	$500.462 \pm 0.471, [499.992, 500.933]$	0.0089	94.25
$\mathbb{E}[X(T)\|Y(T)=y] = 500$	Targeting ($\Delta t = 0.1$)	$499.802 \pm 0.105, [499.697, 499.907]$	0.3833	117.29
$c_2 = 400, T = 5, y = 2105$	Naive	$526.271 \pm 1.760, [524.511, 528.030]$	0.0006	91.85
$\mathbb{E}[X(T)\|Y(T)=y] = 526.25$	Targeting ($\Delta t = 0.1$)	$526.013 \pm 0.115, [525.898, 526.127]$	0.3758	114.40

Table 4.6: Example 4. Initial condition $Z(0) = (0, 0)$, $t = T = 5$, parameter $c_1 = 0.1$ c_2 were varied. We took a common observation and a less common observation. Conditional expectation were estimated from $N_r = 100$ independent trials, where each trial we had filter sample size $N_s = 10,000$.

Appendix A

Appendix

A.1 Derivation of the Evolution Equations for $\pi(t, x)$

We note that the authors [10] provide a rigorous derivation of the evolution equation for the conditional probability $\pi(t, z)$ when the state space is finite and the exact observation is of the form $y = h(z)$ where h is a function of the state space. The derivation in textbook [10] may not be easily accessible to applied scientists who may not be familiar with the language of stochastic analysis. Moreover, our filtering equations (while in agreement with [10]) are somewhat simpler in appearance since in our case h corresponds to the projection onto the last n_2 components of the state and also due to the structure of the reaction network. The derivation shown here is more intuitive to follow (at the expense of some rigor) and results in equations consistent with literature [10]. Moreover, we shall not make the assumption that the state space is finite. We believe that the rigorous derivation in [10] can be extended to infinite state space under reasonable assumptions. Such an endeavor is recently done by authors [12].

We start with a discretization of the time interval $[0, \infty)$ by a mesh

$$\{\ell h \,|\ell = 0, 1, \dots\}$$

of spacing h. We consider $(X(\ell h), Y(\ell h))$ as a discrete time Markov chain. If the observed trajectory of $Y(t)$ is given by $y(t)$, then the observations on the mesh points will be given by

$$\bar{y}_\ell = y(\ell h), \quad \ell = 0, 1, \dots$$

We may use the filtering equations for a partially observed discrete time Markov chain derived in [14]. For the discrete time Markov chain $(X(\ell h), Y(\ell h))$ (for $\ell = 0, 1, \dots$) where Y is observed exactly, the conditional probability

$$\pi_\ell(x) = P(X(\ell h) = x \,|\, Y(jh) = \bar{y}_j \; j = 0, \dots, n),$$

is shown in [14] to satisfy

$$\pi_\ell(x') = \frac{\sum_x \kappa((x, \bar{y}_{\ell-1}), (x', \bar{y}_\ell))\pi_{\ell-1}(x)}{\sum_{x_\ell} \sum_{x_{\ell-1}} \kappa((x_{\ell-1}, \bar{y}_{\ell-1}), (x_\ell, \bar{y}_\ell))\pi_{\ell-1}(x_{\ell-1})}, \tag{A.1}$$

where

$$\kappa((x, y), (x', y')) = P(X(\ell h) = x', Y(\ell h) = y' \,|\, X((\ell - 1)h) = x, Y((\ell - 1)h) = y).$$

From the infinitesimal characteristics of continuous time Markov chains, as h ap-

proaches 0+,

$$\begin{aligned}
\kappa((x,y),(x',y')) &= a_j(x,y)h + o(h) \quad (x',y') = (x,y) + \nu_j, \\
\kappa((x,y),(x,y)) &= 1 - h\sum_{j=1}^{m} a_j(x,y) + o(h), \\
\kappa((x,y),(x',y')) &= o(h) \quad \text{otherwise.}
\end{aligned} \tag{A.2}$$

Let's consider two adjacent mesh points t and $t+h$. There are two possibilities; there are no jumps of y on the interval $(t, t+h]$ or there are jumps. Again from the infinitesimal characteristics of continuous time Markov chains, as h approaches 0+, there is either no jump or one jump during $(t, t+h]$. We approximate $\pi(t,x)$ by $\pi_{\ell-1}(x)$ and $\pi(t+h,x)$ by $\pi_\ell(x)$. For the case when there is no jump during $(t, t+h]$, if we suppose $t_k \le t < t+h < t_{k+1}$, then

$$\bar{y}_\ell = \bar{y}_{\ell-1} = y(t_k).$$

Using (A.1) and (A.2), we obtain $\pi(t+h,x)$ as a ratio where the numerator is

$$\sum_{j\in\mathcal{U}} \pi(t, x-\nu_j')a_j(x-\nu_j', y(t_k))h + \pi(t,x) - \pi(t,x)\sum_{j=1}^{m} a_j(x, y(t_k))h + o(h)$$

and the denominator is

$$\sum_{\tilde{x}} \left(1 - \sum_{j\in\mathcal{O}} a_j(\tilde{x}, y(t_k))h\right)\pi(t,\tilde{x}) + o(h).$$

From the above, we may obtain an expression for $(\pi(t+h,x) - \pi(t,x))/h$ which

upon taking limit as $h \to 0+$ yields (3.4). The second case is when $t < t_k < t + h$ and in this case

$$\bar{y}_{\ell-1} = y(t_{k-1}) \neq y(t_k) = \bar{y}_\ell.$$

Using (A.1) and (A.2), we obtain

$$\pi(t+h, x) = \frac{\sum_{l \in \mathcal{O}_k} a_l(x - \nu_l', y(t_{k-1}))\, \pi(t, x - \nu_l')\, h + o(h)}{\sum_{\tilde{x}} \sum_{l \in \mathcal{O}_k} a_l(\tilde{x} - \nu_l', y(t_{k-1}))\, \pi(t, \tilde{x})\, h + o(h)}.$$

Noting that $\pi(t + h, x) \to \pi(t_k, x)$ and $\pi(t, x) \to \pi(t_k-, x)$ as $h \to 0+$, we obtain (3.5).

We note that for a rigorous treatment one needs to use the language of measure theory, since unlike in the discrete time Markov chain case, in the continuous time case we are conditioning on a zero probability event of observing $Y(s) = y(s)$ for $0 \leq s \leq t$. Moreover, a rigorous and mathematically "cleaner" treatment involves working with the integral representation of the differential equation with jumps. The derivation provided here, we hope, provides the "essence" of the idea.

A.2 Unnormalized and Normalized Filtering Equations

Here we present the details of the derivation of the fitering equation (3.4) in section 3.1. That is, we want to show that if $\rho(t, x)$ satisfies the unnormalized filtering equations, then $\pi(t, x) = \rho(t, x) / \sum_{\tilde{x}} \rho(t, \tilde{x})$. We assume the existence and uniqueness of solutions of both the unnormalized and normalized filtering equations.

To that end, suppose ρ solves the unnormalized filtering equations (3.8) and

(3.9), and let $\tilde{\pi}(t,x) = \rho(t,x)/\sum_{\tilde{x}} \rho(t,\tilde{x})$. It is adequate to show that $\tilde{\pi}(t,x)$ satisfies the filtering equations (3.4) and (3.5).

In between jump times, that is, for $t_k \leq t < t_{k+1}$, ρ satisfies

$$\rho'(t,x) = \sum_{j\in\mathcal{U}} \rho(t,x-\nu_j')\, a_j(x-\nu_j', y(t_k)) - \sum_{j\in\mathcal{U}} \rho(t,x)\, a_j(x,y(t_k))$$
$$- \rho(t,x)\, a^{\mathcal{O}}(x,y(t_k)) \quad \forall x \in \mathbb{Z}_+^{n_1}.$$

Then

$$\tilde{\pi}'(t,x) = \frac{\rho'(t,x)}{\sum_{\tilde{x}} \rho(t,\tilde{x})} - \rho(t,x)\frac{\sum_{\tilde{x}} \rho'(t,\tilde{x})}{(\sum_{\tilde{x}} \rho(t,\tilde{x}))^2}.$$

The first term may be written as

$$\frac{\rho'(t,x)}{\sum_{\tilde{x}} \rho(t,\tilde{x})} = \sum_{j\in\mathcal{U}} \tilde{\pi}(t,x-\nu_j')\, a_j(x-\nu_j', y(t_k)) - \sum_{j\in\mathcal{U}} \tilde{\pi}(t,x)\, a_j(x,y(t_k)) - \tilde{\pi}(t,x)\, a^{\mathcal{O}}(x,y(t_k)).$$

The second term can be written as

$$\rho(t,x)\frac{\sum_{\tilde{x}} \rho'(t,\tilde{x})}{(\sum_{\tilde{x}} \rho(t,\tilde{x}))^2} = \pi(t,x)\frac{\sum_{\tilde{x}} \rho'(t,\tilde{x})}{\sum_{\tilde{x}} \rho(t,\tilde{x})}$$

Note that $\sum_{\tilde{x}} \left(\sum_{j\in\mathcal{U}} \rho(t,\tilde{x}-\nu_j')\, a_j(\tilde{x}-\nu_j', y(t_k)) - \sum_{j\in\mathcal{U}} \rho(t,\tilde{x})\, a_j(\tilde{x}, y(t_k)) \right) = 0$, and hence

$$\sum_{\tilde{x}} \rho'(t,\tilde{x}) = \sum_{\tilde{x}} \left(\sum_{j\in\mathcal{U}} \rho(t,\tilde{x}-\nu_j')\, a_j(\tilde{x}-\nu_j', y(t_k)) - \sum_{j\in\mathcal{U}} \rho(t,\tilde{x})\, a_j(\tilde{x}, y(t_k)) - \rho(t,\tilde{x})\, a^{\mathcal{O}}(\tilde{x}, y(t_k)) \right)$$
$$= -\sum_{\tilde{x}} \rho(t,\tilde{x})\, a^{\mathcal{O}}(\tilde{x}, y(t_k))$$

Hence $\tilde{\pi}$ satisfies (3.4).

For $k = 1, 2, \ldots$ at jump times t_k, $\rho(t, x)$ jumps according to

$$\rho(t_k, x) = \frac{1}{|\mathcal{O}_k|} \sum_{j \in \mathcal{O}_k} a_j(x - \nu_j', y(t_{k-1}))\, \rho(t_k-, x - \nu_j') \quad x \in \mathbb{Z}_+^{n_1}.$$

Since $\tilde{\pi}(t, x) = \rho(t, x) / \sum_{\tilde{x}} \rho(t, \tilde{x})$, we have

$$\tilde{\pi}(t_k, x) = \frac{\sum_{l \in \mathcal{O}_k} a_l(x - \nu_l', y(t_{k-1}))\, \tilde{\pi}(t_k-, x - \nu_l')}{\sum_{\tilde{x}} \sum_{l \in \mathcal{O}_k} a_l(\tilde{x}, y(t_{k-1}))\, \tilde{\pi}(t_k-, \tilde{x})} \quad \forall x \in \mathbb{Z}_+^{n_1}$$

which shows that $\tilde{\pi}(t, x)$ satisfies (3.5) at jump times t_k.

A.3 Some Preliminaries on Point Processes

Theorem 12. *(Exponential Formula) Let $a(t)$ be a right-continuous increasing function with $a(0) = 0$, and let u be such that*

$$\int_0^t |u(s)| da(s) < \infty, \qquad t \geq 0.$$

Then the equation

$$x(t) = x(0) + \int_0^t x(s-) u(s) da(s) \tag{A.3}$$

admits a unique locally bounded solution given by

$$x(t) = x(0) \prod_{0 < s \leq t} (1 + u(s) \Delta a(s)) \exp\left\{ \int_0^t u(s) da^c(s) \right\} \tag{A.4}$$

where $\Delta a(t) = a(t) - a(t-)$, $a^c(t) = a(t) - \sum_{0<s\leq t} \Delta a(s)$.

For a single variate point process, it follows immediately from the exponential formula that

$$L(t) = 1 + \int_0^t L(s-)(\mu(s) - 1)dM(s) \tag{A.5}$$

with $M(t) = N(t) - \int_0^t \lambda(s)ds$. admits solution

$$L(t) = L(0) \prod_{0<s\leq t} \mu(s)\Delta N(s) \exp\left\{\int_0^t (\mu(t) - 1)(-\lambda(s))ds\right\}. \tag{A.6}$$

As a corollary, for an $m-$variate point process case, with $L(t) = \prod_{k=1}^m L_k(t)$,

$$L(t) = 1 + \sum_{k=1}^m \int_0^t L(s-)(\mu_k(s) - 1)dM_k(s) \tag{A.7}$$

with $M_k(t) = N_k(t) - \int_0^t \lambda_k(s)ds$. admits solution

$$L_k(t) = L(0) \prod_{0<s\leq t} \mu_k(s)\Delta N_k(s) \exp\left\{\int_0^t \sum_{k=1}^m (\mu_k(t) - 1)(-\lambda_k(s))ds\right\}. \tag{A.8}$$

Theorem 13. *Let* $(N_1(t), ..., N_m(t))$ *be an* $m-$*variate point process adapted to some filtration* $\mathcal{F}_t$*, and let* $\lambda_i(t)$ *be the predictable* $(P, \mathcal{F}_t)$*-intensities of* $N_i(t)$ $(i = 1, ...m)$*. Let* $\mu_i(t)$ $(i = 1, ..., m)$ *be* $\mathcal{F}_t$*-predictable processes, nonnegative, and such that for all* $t \geq 0$ *and all* $1 \leq i \leq m$

$$\int_0^t \mu_i(s)\lambda_i(s)ds < \infty \quad P - a.s.$$

Definte the process L_t by

$$L(t) = \prod_{i=1}^{m} L_i(t),$$

where

$$L_i(t) = \prod_{n\geq 1} \mu_i(T_{i,n})\, 1(T_{i,n} \leq T) \exp\left\{ \int_0^t (1-\mu_i(s))\lambda_i(s) ds \right\}. \tag{A.9}$$

Then $L(t)$ is a $(P, \mathcal{F}_t)$ nonnegative local martingale and a $(P, \mathcal{F}_t)$-suppermartingale.

Theorem 14. *(Direct Randon-Nikodym derivative theorem). Same notion as in the previous theorem. Suppose moreover that*

$$\mathbb{E}[L(1)] = 1.$$

Define the probability measure $\tilde{P}$ by

$$\frac{d\tilde{P}}{dP} = L \tag{A.10}$$

Then, for each $1 \leq i \leq m$, $N_i(t)$ has the $(\tilde{P}, \mathcal{F}_t)$-intensity $\tilde{\lambda}_i(t) = \mu_i(t)\lambda_i(t)$ over $[0, 1]$.

A.4 Speeding up FCS Monte Carlo Simulations

A.4.1 Introduction

Simulating a large sample of Fluorescent Correlation Spectroscopy intensity trajectories could be computationally expensive. For instance, a realistic time in-

crements could be $dt = 1 \times 10^{-6}$ seconds, and a realistic observation time could be a few minutes, say $T = 100$. The *exact simulation* that evolves each particle at each time step could take hours to simulate just one trajectory, so a more efficient Monte Carlo simulation is needed.

Due to the Gaussian nature of the laser intensity function, most of the intensity is contributed by a small number of particles near a tiny confocal volume. Hence in the *efficient simulation* scheme, we would like to update the positions of the particles near the confocal volume over the time step dt, and for the rest of the particles we choose a longer time step Δt to update.

A.4.2 Choosing a Critical Region E_c

We would like to choose a *critical region* E_c so that it captures nearly all of the intensity, and we update the positions of the particles in that region every dt time, the same as the exact simulation.

Computing the portion of the intensity contributed from near the confocal volume is similar to finding the confidence region for a multivariate normal distribution, which involves the density function of χ^2- distribution.

Let $Q = Z_1^2 + Z_2^2 + Z_3^2$, where Z_1, Z_2, Z_3 are independent standard normal random variables, so Q follows χ^2-distribution with three degrees of freedom, and we denote as $Q \sim \chi^2(3)$. Note that the joint probability density function of $Z =$

(Z_1, Z_2, Z_3) is

$$f(z_1, z_2, z_3) = (2\pi)^{-\frac{3}{2}} \exp\left\{-\frac{1}{2}(z_1^2 + z_2^2 + z_3^2)\right\}, \tag{A.11}$$

and we have

$$\mathbb{P}\{Q \leq q\} = \mathbb{P}\{Z_1^2 + Z_2^2 + Z_3^2 \leq q\} = \int_{\{z: z_1^2 + z_2^2 + z_3^2 \leq q\}} f(z)dz.$$

That is, $\mathbb{P}\{Q \leq q\}$ equals the integral of the joint density function coming from inside $B(0, \sqrt{q}) = \{z \mid z_1^2 + z_2^2 + z_3^2 \leq q\}$.

In the FCS scenario, the point spread function is given by

$$F(x, y, z) = F_0 \exp\left\{-\frac{2x^2}{r^2} - \frac{2y^2}{r^2} - \frac{2z^2}{l^2}\right\} = F_0 \exp\left\{-\frac{1}{2}(\frac{4x^2}{r^2} + \frac{4y^2}{r^2} + \frac{4z^2}{l^2})\right\}. \tag{A.12}$$

By a change of variable, $z_1 = \frac{2x}{r}, z_2 = \frac{2y}{r}, z_3 = \frac{2z}{l}$, we could find the probability

$$\mathbb{P}\{Q \leq q\} = \int_E F(x, y, z)dxdydz / \int_{\mathbb{R}^3} F(x, y, z)dxdydz$$

with $E = \{(x, y, z) \mid (\frac{2x}{r})^2 + (\frac{2y}{r})^2 + (\frac{2z}{l})^2 \leq q\}$.

q	Ellipsoid region E	Ellipsoid (a, b, c)	$\mathbb{P}\{Q \le q\}$
1	$(\frac{2x}{r})^2 + (\frac{2y}{r})^2 + (\frac{2z}{l})^2 \le 1$	$(\frac{r}{2}, \frac{r}{2}, \frac{l}{2})$	0.1987
4	$(\frac{2x}{r})^2 + (\frac{2y}{r})^2 + (\frac{2z}{l})^2 \le 2^2$	(r, r, l)	0.7385
9	$(\frac{2x}{r})^2 + (\frac{2y}{r})^2 + (\frac{2z}{l})^2 \le 3^2$	$(\frac{3r}{2}, \frac{3r}{2}, \frac{3l}{2})$	0.9707
16	$(\frac{2x}{r})^2 + (\frac{2y}{r})^2 + (\frac{2z}{l})^2 \le 4^2$	$(2r, 2r, 2l)$	0.9989

If we choose $(\frac{x}{2r})^2 + (\frac{y}{2r})^2 + (\frac{z}{2l})^2 \le 1$, or a even larger rectangular box region $B_{2r} = \{|x| \le 2r, |y| \le 2r, |z| \le 2l\}$, it can cover more than 99.89% of the intensity.

A.4.3 Choosing a Buffer Region E_b

Far away from the critical region E_c, the locations of the particles are updated less frequently, and we take that time increment Δt to be a multiple of the exact time step dt, that is, $\Delta t = m \cdot dt$. We need a *buffer region* to handle the particles that could possibly enter the critical region within Δt time. In that buffer region, the positions of the particles are updated every dt time, and some of the particles might move into the critical region E_c within Δt time. While further outside the buffer region, the probability of entering the critical region within Δt time is nearly zero, namely, we would like to control the conditional probability $\mathbb{P}\{X(t + \Delta t) \in E_c \mid X(t) \in (E_c \cup E_b)^c\}$ to be small. Thus we want to consider the displacement ΔX that a particle could travel over Δt time, and choose a corresponding buffer distance d to let $\mathbb{P}\{\|\Delta X\| > d\}$ be very small, and we can take the buffer region to be $E_b = \{x \in E_c^c \mid \|x - y\| \le d, y \in E_c\}$.

The displacement $\Delta X := X(t+\Delta t) - X(t) \sim \mathcal{N}(0, 2D\Delta t I)$, where I is the 3×3 identity matrix. We could compute the probability of a particle remains in the ball $B(X(t), d)$ after Δt time. Since $\Delta X_i = \sqrt{2D\Delta t} Z_i$, for $i = 1, 2, 3$, we have $\mathbb{P}\{\sum_{i=1}^{3} Z_i^2 \leq q\} = \mathbb{P}\{\sum_{i=1}^{3} \Delta X_i^2 \leq 2qD\Delta t\} = \mathbb{P}\{\|\Delta X\| < d\}$ with $d = \sqrt{2qD\Delta t}$. And we consider one specific example where $dt = 1 \times 10^{-6}$, $\Delta t = 1 \times 10^{-5}$, $D = 1.3 \times 10^{-10}$.

$\sqrt{q}$	Buffer distance d	d example	$\mathbb{P}\{\|\Delta X\| < d\}$
1	$\sqrt{2D\Delta t}$	5.0990×10^{-8}	0.1987
2	$2\sqrt{2D\Delta t}$	1.0198×10^{-7}	0.7385
3	$3\sqrt{2D\Delta t}$	1.5297×10^{-7}	0.9707
4	$4\sqrt{2D\Delta t}$	2.0396×10^{-7}	0.9989

Hence, we might want to choose buffer distance $d = r$ for that example.

A.4.4 Comparison of the Accurate and Efficient Simulations

We choose both E_c and $E_c \cup E_b$ to be cylindrical regions:

$$E_c = \{(x_1, x_2, x_3) \mid x_1^2 + x_2^2 \leq (2r)^2, |x_3| \leq 2l\}, \tag{A.13}$$

$$E_c \cup E_b = \{(x_1, x_2, x_3) \mid x_1^2 + x_2^2 \leq (2r+d)^2, |x_3| \leq 2l + d\} \text{ with } d = r. \tag{A.14}$$

For that specific example, we run Monte Carlo simulation that couples the more accurate simulation with the efficient simulation. 100 pairs of coupled trajectories were simulated.

For each pair of coupled trajectories, we look at sup-norm difference

$$e^{(i)} = \sup_{t\in[0,T]} |I_e^{(i)}(t) - I_a^{(i)}(t)| = \max_{j=1,\dots,nt} |I_e^{(i)}(t_j) - I_a^{(i)}(t_j)| \tag{A.15}$$

We estimate the bias $\mathbb{E}[e]$ and the L_2 error $\sqrt{\mathbb{E}[e^2]}$ from the 100 samples, and we get

$$\mathbb{E}[e] \in [2.49 \times 10^{-4}, 2.80 \times 10^{-4}], \tag{A.16}$$

$$\sqrt{\mathbb{E}[e^2]} \in [2.70 \times 10^{-4}, 3.40 \times 10^{-4}]. \tag{A.17}$$

To provide more detail on computing the confidence interval,

$$\mathbb{E}[e] \approx \overline{e} = \frac{1}{n_s}\sum_{i=1}^{n_s} e^{(i)} = 2.64 \times 10^{-4}, \tag{A.18}$$

$$\sigma_{\overline{e}} = \frac{1}{\sqrt{n_s}}\sigma_e \approx 1.56 \times 10^{-5}, \text{ where } \sigma_e \approx s_e = \sqrt{\frac{1}{n_s - 1}\sum_{i=1}^{n_s}(e^{(i)} - \overline{e})^2}, \tag{A.19}$$

$$\sqrt{\mathbb{E}[e^2]} \approx \sqrt{\overline{e^2}} = \sqrt{\frac{1}{n_s}\sum_{i=1}^{n_s}(e^{(i)})^2} = 3.07 \times 10^{-4}, \tag{A.20}$$

$$\sigma_{\overline{e^2}} = \frac{1}{\sqrt{n_s}}\sigma_{e^2} \approx 2.13 \times 10^{-8}, \tag{A.21}$$

$$\sqrt{\mathbb{E}[e^2]} \in \left[\sqrt{\overline{e^2} - \sigma_{\overline{e^2}}}, \sqrt{\overline{e^2} + \sigma_{\overline{e^2}}}\right]. \tag{A.22}$$

We look at sup-norm difference for $\mathrm{Corr}(t) = \frac{1}{T}\int_0^T I(t)I(t+\tau)dt$,

$$c^{(i)} = \sup_{t\in[0,T]} |\,\mathrm{Corr}_e^{(i)}(t) - \mathrm{Corr}_a^{(i)}(t)| = \max_{j=1,\dots,nt} |\,\mathrm{Corr}_e^{(i)}(t_j) - \mathrm{Corr}_a^{(i)}(t_j)| \tag{A.23}$$

We estimate the bias $\mathbb{E}[e]$ and the L_2 error $\sqrt{\mathbb{E}[e^2]}$ from the 100 samples, and we get

$$\mathbb{E}[c] \in [2.46 \times 10^{-5}, 2.50 \times 10^{-5}], \tag{A.24}$$

$$\sqrt{\mathbb{E}[c^2]} \in [2.47 \times 10^{-5}, 2.51 \times 10^{-5}]. \tag{A.25}$$

Let $d = \hat{D}_e - \hat{D}_a$, the difference of best fit $\hat{D}$ with $I_e(t)$ or $I_a(t)$ data,

$$\mathbb{E}[d] \in [1.52 \times 10^{-16}, 2.82 \times 10^{-16}], \text{ with relative bias } \frac{E[d]}{\overline{D}} \in [1.2 \times 10^{-6}, 2.2 \times 10^{-6}] \tag{A.26}$$

$$\sqrt{\mathbb{E}[d^2]} \in [6.26 \times 10^{-16}, 7.38 \times 10^{-16}] \text{with relative error } \frac{\sqrt{E[d^2]}}{\overline{D}} \in [4.8 \times 10^{-6}, 5.7 \times 10^{-6}]. \tag{A.27}$$

Method	Elapsed time (seconds)
Accurate (C)	130
Accurate (Matlab)	243, 240
Efficient (Matlab)	99, 98
Accurate (Matlab, 6 workers)	104
Efficient (Matlab, 6 workers)	28

Table A.1: Efficiency of the simulations. $n_s = 10$, $T = 0.1s$, $dt = 10^{-6}s$, $\Delta t = 10^5 s$ $N = 1973$. $N_e = 2.68$.

We run the simulation for several different regions E, and record the run time for efficient simulation and the errors obtain by coupling. For each case, we simulate $n_s = 100$ pairs of coupled trajectories using a paralleled Matlab implementation. We are particularly interested in the sup-norm difference $e = \sup_{t\in[0,T]} |I_e(t) - I_a(t)|$ and the relative L_2 error for the best fit D, that is, $\frac{\sqrt{\mathbb{E}[(\hat{D}_e - \hat{D}_a)^2]}}{\overline{D}}$. The run time is for

efficient simulation per trajectory, not for the coupled simulation.

Region E	m	$E/B(\%)$	Time(s)	$\mathbb{E}[\sup_{t\in[0,T]} \lvert I_e(t) - I_a(t)\rvert]$	$\frac{\sqrt{\mathbb{E}[(\hat{D}_e - \hat{D}_a)^2]}}{\bar{D}}$
Ellipsoid $(2r, 2r, 2l)$	10	0.82	2.20	$[8.52 \times 10^{-2}, 9.02 \times 10^{-2}]$	$[4.72 \times 10^{-4}, 5.27 \times 10^{-4}]$
Ellipsoid $(3r, 3r, 2l + r)$	10	2.02	2.42	$[6.33 \times 10^{-4}, 6.75 \times 10^{-4}]$	$[7.51 \times 10^{-6}, 8.70 \times 10^{-6}]$
Ellipsoid $(3r, 3r, 3l)$	10	2.76	2.53	$[2.49 \times 10^{-4}, 2.84 \times 10^{-4}]$	$[2.03 \times 10^{-7}, 2.30 \times 10^{-7}]$
Ellipsoid $(3r, 3r, 3l)$, vectorized	10	2.76	0.72	$[2.49 \times 10^{-4}, 2.84 \times 10^{-4}]$	$[2.03 \times 10^{-7}, 2.30 \times 10^{-7}]$
$x_1^2 + x_2^2 \leq (3r)^2, \lvert x_3\rvert \leq 2l + r$	10	3.04	2.57	$[2.43 \times 10^{-4}, 2.63 \times 10^{-4}]$	$[4.23 \times 10^{-6}, 4.82 \times 10^{-6}]$
Ellipsoid $(2r, 2r, 2l)$	20	0.82	1.16	$[3.31 \times 10^{-1}, 3.39 \times 10^{-1}]$	$[1.36 \times 10^{-3}, 1.60 \times 10^{-3}]$
Ellipsoid $(3r, 3r, 2l + r)$	20	2.02	1.38	$[7.22 \times 10^{-3}, 8.20 \times 10^{-3}]$	$[1.19 \times 10^{-5}, 1.38 \times 10^{-5}]$
Ellipsoid $(3r, 3r, 3l)$	20	2.76	1.53	$[5.70 \times 10^{-3}, 7.20 \times 10^{-3}]$	$[5.53 \times 10^{-6}, 7.89 \times 10^{-6}]$
Ellipsoid $(3r, 3r, 3l)$, vectorized	20	2.76	0.56	$[2.49 \times 10^{-4}, 2.84 \times 10^{-4}]$	$[2.03 \times 10^{-7}, 2.30 \times 10^{-7}]$
$x_1^2 + x_2^2 \leq (3r)^2, \lvert x_3\rvert \leq 2l + r$	20	3.04	1.56	$[7.20 \times 10^{-3}, 8.14 \times 10^{-3}]$	$[1.31 \times 10^{-5}, 1.52 \times 10^{-5}]$

Bibliography

[1] David F Anderson. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *The Journal of chemical physics*, 127(21):214107, 2007.

[2] David F Anderson, Arnab Ganguly, Thomas G Kurtz, et al. Error analysis of tau-leap simulation methods. *The Annals of Applied Probability*, 21(6):2226–2262, 2011.

[3] David F Anderson and Thomas G Kurtz. Continuous time markov chain models for chemical reaction networks. In *Design and analysis of biomolecular circuits*, pages 3–42. Springer, 2011.

[4] Alan Bain and Dan Crisan. *Fundamentals of stochastic filtering*, volume 60. Springer Science & Business Media, 2008.

[5] Karen Ball, Thomas G Kurtz, Lea Popovic, Greg Rempala, et al. Asymptotic analysis of multiscale approximations to reaction networks. *The Annals of Applied Probability*, 16(4):1925–1961, 2006.

[6] David Bernstein. Simulating mesoscopic reaction-diffusion systems using the gillespie algorithm. *Physical Review E*, 71(4):041103, 2005.

[7] Richard J Boys, Darren J Wilkinson, and Thomas BL Kirkwood. Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing*, 18(2):125–135, 2008.

[8] Pierre Brémaud. *Point processes and queues: martingale dynamics*. Springer-Verlag, 1981.

[9] Silvia Calderazzo, Marco Brancaccio, and Bärbel Finkenstädt. Filtering and inference for stochastic oscillators with distributed delays. *Bioinformatics*, 35(8):1380–1387, 2019.

[10] Fulvia Confortola and Marco Fuhrman. Filtering of continuous-time markov chains with noise-free observation and applications. *Stochastics An International Journal of Probability and Stochastic Processes*, 85(2):216–251, 2013.

[11] Dan Crisan and Terry Lyons. Minimal entropy approximations and optimal algorithms. *Monte Carlo methods and applications*, 8(4):343–356, 2002.

[12] Elena Sofia D'Ambrosio, Zhou Fang, Ankit Gupta, and Mustafa Khammash. Filtered finite state projection method for the analysis and estimation of stochastic biochemical reaction networks. *bioRxiv*, pages 2022–10, 2022.

[13] Zhou Fang, Ankit Gupta, and Mustafa Khammash. Stochastic filters based on hybrid approximations of multiscale stochastic reaction networks. *arXiv preprint arXiv:2008.11682*, 2020.

[14] Bert Fristedt, Naresh Jain, N Krylov, and Nikolaĭ Vladimirovich Krylov. *Filtering and Prediction: A Primer*, volume 10. American Mathematical Soc., 2007.

[15] Arnab Ganguly, Derya Altintan, and Heinz Koeppl. Jump-diffusion approximation of stochastic reaction dynamics: error bounds and algorithms. *Multiscale Modeling & Simulation*, 13(4):1390–1419, 2015.

[16] Timothy S Gardner, Charles R Cantor, and James J Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403(6767):339–342, 2000.

[17] Michael A Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The journal of physical chemistry A*, 104(9):1876–1889, 2000.

[18] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.

[19] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[20] Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of chemical physics*, 115(4):1716–1733, 2001.

[21] Daniel T Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.

[22] Andrew Golightly and Chris Sherlock. Efficient sampling of conditioned markov jump processes. *Statistics and Computing*, 29:1149–1163, 2019.

[23] Andrew Golightly and Darren J Wilkinson. Bayesian sequential inference for stochastic kinetic biochemical network models. *Journal of Computational Biology*, 13(3):838–851, 2006.

[24] Andrew Golightly and Darren J Wilkinson. Bayesian parameter inference for stochastic biochemical network models using particle markov chain monte carlo. *Interface focus*, 1(6):807–820, 2011.

[25] Benjamin Hepp, Ankit Gupta, and Mustafa Khammash. Adaptive hybrid simulations for multiscale stochastic reaction networks. *The Journal of chemical physics*, 142(3):034118, 2015.

[26] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. 1960.

[27] Rudolph Emil Kalman and Richard S Bucy. New results in linear filtering and prediction theory. 1961.

[28] Hye-Won Kang, Thomas G Kurtz, et al. Separation of time-scales and model reduction for stochastic reaction networks. *The Annals of Applied Probability*, 23(2):529–583, 2013.

[29] Kari Karhunen. Ueber lineare methoden in der wahrscheinlichkeitsrechnung issue 37 of annales academiae scientiarum fennicae. *Series A*, 1, 1947.

[30] Dennis E Koppel. Statistical accuracy in fluorescence correlation spectroscopy. *Physical Review A*, 10(6):1938, 1974.

[31] Tiejun Li. Analysis of explicit tau-leaping schemes for simulating chemically reacting systems. *Multiscale Modeling & Simulation*, 6(2):417–436, 2007.

[32] Jun S Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and computing*, 6:113–119, 1996.

[33] James CW Locke and Michael B Elowitz. Using movies to analyse gene circuit dynamics in single cells. *Nature Reviews Microbiology*, 7(5):383–392, 2009.

[34] Michel Loeve. Functions aleatoires du second ordre. *Processus stochastique et mouvement Brownien*, pages 366–420, 1948.

[35] Douglas Magde, Elliot Elson, and Watt W Webb. Thermodynamic fluctuations in a reacting systemmeasurement by fluorescence correlation spectroscopy. *Physical Review Letters*, 29(11):705, 1972.

[36] Douglas Magde, Elliot L Elson, and Watt W Webb. Fluorescence correlation spectroscopy. ii. an experimental realization. *Biopolymers: Original Research on Biomolecules*, 13(1):29–61, 1974.

[37] Muruhan Rathinam. Convergence of moments of tau leaping schemes for unbounded markov processes on integer lattices. *SIAM Journal on Numerical Analysis*, 54(1):415–439, 2016.

[38] Muruhan Rathinam and Hana El Samad. Reversible-equivalent-monomolecular tau: A leaping method for small number and stiff stochastic chemical systems. *Journal of Computational Physics*, 224(2):897–923, 2007.

[39] Muruhan Rathinam and Mingkai Yu. State and parameter estimation from exact partial state observation in stochastic reaction networks. *The Journal of Chemical Physics*, 154(3):034103, 2021.

[40] Saveez Saffarian and Elliot L Elson. Statistical analysis of fluorescence correlation spectroscopy: the standard deviation and bias. *Biophysical journal*, 84(3):2030–2042, 2003.

[41] Silviya P Zustiak, Hacene Boukari, and Jennie B Leach. Solute diffusion and interactions in cross-linked poly (ethylene glycol) hydrogels studied by fluorescence correlation spectroscopy. *Soft Matter*, 6(15):3609–3618, 2010.

[42] Silviya P Zustiak, Ralph Nossal, and Dan L Sackett. Hindered diffusion in polymeric solutions studied by fluorescence correlation spectroscopy. *Biophysical journal*, 101(1):255–264, 2011.