

DISSERTATION

SHORT TIME ANALYSIS OF DETERMINISTIC ODE SOLUTIONS AND THE  
EXPECTED VALUE OF THE CORRESPONDING BIRTH-DEATH PROCESS

Submitted by  
Megan H. Buzby  
Mathematics Department

In partial fulfillment of the requirements  
For the Degree of Doctor of Philosophy  
Colorado State University  
Fort Collins, CO  
Summer 2009

UMI Number: 3385149

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3385149

Copyright 2009 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

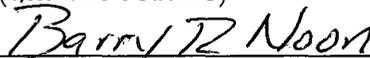
July 13, 2009

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED  
UNDER OUR SUPERVISION BY MEGAN H. BUZBY ENTITLED "SHORT  
TIME ANALYSIS OF DETERMINISTIC ODE SOLUTIONS AND THE  
EXPECTED VALUE OF A CORRESPONDING BIRTH-DEATH PROCESS" BE  
ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE  
OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

  
\_\_\_\_\_  
(Colleen Webb)

  
\_\_\_\_\_  
(Jennifer Mueller)

  
\_\_\_\_\_  
(Barry Noon)

  
\_\_\_\_\_  
Advisor (Donald Estep)

  
\_\_\_\_\_  
Acting Department Head (Donald Estep)

## ABSTRACT OF DISSERTATION

### SHORT TIME ANALYSIS OF DETERMINISTIC ODE SOLUTIONS AND THE EXPECTED VALUE OF A CORRESPONDING BIRTH-DEATH PROCESS

There is a standard way to construct a discrete birth-death probability model for an evolution system, in which an ODE model of the system is used to define the probabilities governing the evolution of the stochastic model. Given the significant differences in the dynamical behavior of ODE solutions which are inherently smooth, and stochastic models which are subject to random variation, the question naturally arises about the connection between the two models. In particular, we investigate the validity of using a continuum model to define the evolution of a stochastic model.

We show a consistent way to define the probabilities for the stochastic model if the ODE has the form  $\dot{y} = f(y) = yg(y)$ . The deterministic model can then be compared to the expected value of the discrete probability model. For an ODE of this form describing population dynamics, we can describe each individual of the population as a categorical random variable. In this formulation, the probability for a birth or death to occur in a population of size  $y$  over a time interval of length  $\Delta t$  is given by  $yg(y)\Delta t + o(\Delta t)$ , depending on the sign of  $g(y)$ .

We derive local and global bounds for the difference between the expected value of the discrete probability model and the solution of the ODE. Locally, the two models behave similarly. Global bounds, however, imply the difference between these two models may be at most exponential in nature. Such a large bound must account for the possibility that the associated probability model may jump across a steady state of the ODE and exhibit divergent behavior while the ODE remains stable.

We explore our results for a number of models. In particular, we provide examples that show that there can be fundamental differences in the dynamical behavior of the stochastic and ODE model solutions, even when they are close over any given step. Our results represent a different approach from another view, which derives some ODE that governs the expected value of the stochastic system. That ODE and the original ODE are *not* the same in general.

Megan H. Buzby  
Mathematics Department  
Colorado State University  
Fort Collins, CO 80523  
Summer 2009

## Acknowledgements

First and foremost, I would like to thank my advisor, Don Estep, for his immense support and for the opportunities he provided throughout my entire time at CSU. He always made time in his very busy schedule to meet with me and never failed to show up with a positive, motivating attitude.

I would like to thank my committee for their time and unique insight.

I would like to thank NSF and the students and faculty of the PRIMES program for the many and varied opportunities I was able to pursue through their support.

I would like to thank the entire math department, faculty, staff, and students, for creating and being a part of a very supportive and enjoyable work environment.

To maintain my sanity during this pursuit, through encouragement as well as distraction, I would sincerely like to thank my family and friends, both near and far.

Without all of these people, I could not have completed this degree.

# Contents

<b>1</b>	<b>Introduction and Background</b>	<b>2</b>
1.1	The Basic Stochastic Model . . . . .	6
1.2	Probability Model for a General Non-linear ODE . . . . .	8
1.2.1	Conditional Expectation of the Probability Model . . . . .	12
1.3	Scalar Examples . . . . .	15
1.3.1	Exponential Growth . . . . .	15
1.3.2	Logistic Growth . . . . .	16
1.3.3	Gompertz growth for tumors with time dependent parameter .	19
1.3.4	Periodic Functions . . . . .	20
1.3.5	Multiple Steady-State Models . . . . .	21
<b>2</b>	<b>Scalar Difference Bounds</b>	<b>32</b>
2.1	Local Bounds . . . . .	32
2.2	Global Difference Bounds . . . . .	50
<b>3</b>	<b>Extension to Higher Dimensions</b>	<b>60</b>
3.0.1	Assumptions for the System Model . . . . .	61
3.1	The System Probability Model . . . . .	62

3.2	Expected Value . . . . .	64
3.3	System Examples . . . . .	66
3.3.1	SI Model of Fox Rabies . . . . .	66
3.3.2	SI Model of pneumonic plague . . . . .	68
<b>4</b>	<b>System Difference Bounds</b>	<b>71</b>
4.1	Local Difference Bounds . . . . .	71
4.2	Global Difference Bounds . . . . .	82
<b>5</b>	<b>Conclusions</b>	<b>90</b>
<b>A</b>	<b>Mathematical Background</b>	<b>92</b>
A.1	Unbiased Estimators and Statistical Properties . . . . .	96
<b>B</b>	<b>Review of Discrete, Continuous, and Stochastic Models</b>	<b>98</b>
B.1	Chemical Kinetics and the Gillespie Algorithm . . . . .	98
B.2	Evolutions and biological applications of Gillespie's algorithm . . . . .	100
B.3	Modeling prior to 1970's in chemical kinetics and biology . . . . .	102
B.4	Convergence from Probability to Deterministic Model - R. G. Dolgo- arshinnykh . . . . .	104
	<b>Bibliography</b>	<b>107</b>

# Chapter 1

## Introduction and Background

Probability models are used in almost every scientific discipline; including ecological population modeling, weather modeling and prediction, chemical reactions, disease prevalence, and more. Stochastic models are often used specifically to describe such rare events that may not follow a smooth trajectory (extreme weather events, earthquakes, volcanoes, etc.). We consider a common approach in ecological population and disease modeling which defines a birth-death Markov process with probabilities determined by a standard ODE model of the same physical system.

Our main purpose in this dissertation is to explore the mathematical consistency between the deterministic behavior of ordinary differential equations and statistics of the discrete events of the associated probability model. Depending on the form of  $\dot{y} = f(y)$ , the mathematical connection between the rates of the ODE and the probabilities of a birth-death process may not be entirely clear.

The ODE has the property that solutions that start near each other, remain close for at least some time, and solutions diverge and converge in a smooth fashion. The stochastic birth-death process we are interested in uses probabilities determined by  $f(y)$ , but the discrete changes vary significantly, so that two real solutions starting

near each other may not be close at all. This means that there can be significant differences in the behavior of the two models over any sizeable time interval. For example, consider an ODE model with a saddle point equilibrium which is convergent for solutions less than the steady state and divergent behavior for solutions greater than the steady state. In the equivalent probability model we describe, a particular simulation may approach the steady state for some time and then jump to the divergent family of solutions with some positive probability, which depends on the size of the allowable discrete jump. The ODE from which the probabilities are taken has no solution with this type of behavior. Fig. 1.1 shows 20 simulations of such a model, which we consider in more detail in Section 1.3. This significantly different dynamical behavior motivates studying the difference between continuous rates of a differential equation and the discrete probabilities of the related birth-death process.

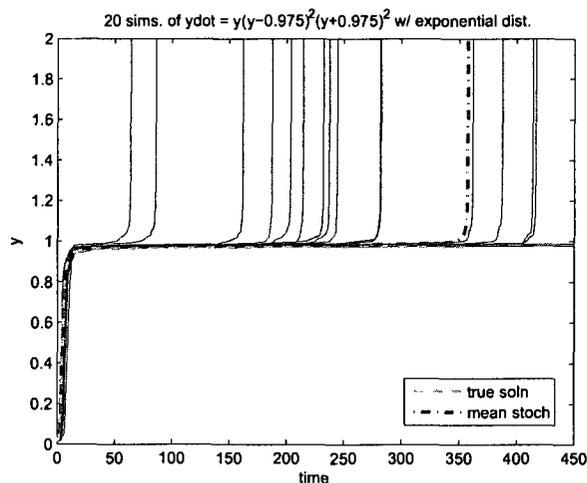


Figure 1.1: Simulations for  $\dot{y} = y(y - 0.975)^2(y + 0.975)^2$  using the 'basic model' and the Gillespie algorithm described in Appendix B.

To explore this connection, however, we cannot compare particular simulations

of a probability model to solutions of an ODE. Rather, we consider some sort of statistical quantity, e.g. the expected value over many simulations.

The discrete probability model we consider in this dissertation begins with a differential equation of a particular form,  $\dot{y} = yg(y)$ , that provides a natural way to define probabilities and a discrete birth-death probability model. We then seek to estimate the difference between the expected value of the discrete probability model and the ODE solution. If we assume the particular form of the ODE  $f(y) = yg(y)$ , it is clear that the expected value is the correct statistic to consider. We show that it remains close to the solution of the differential equation for short time. In particular, we determine local and global bounds for the difference between the expectation of the probability model and the solution of the differential equation, then consider how the properties of  $g(y)$  affect this difference.

Here, we note that another commonly used technique in population models [28], [34], chemical kinetics [19], and disease modeling [13], [24] considers another approach. In this approach, a discrete probability model is assumed and it is shown that its expected value converges to a solution of some differential equation, in the limit of large populations. Note the difference between this method and the one we discuss - we approach the problem *beginning* with the ODE and defining a ‘close’ probability model as opposed to beginning with a *probability* model and showing convergence to a differential equation. It is not clear that the expectation of this stochastic model should converge to the same differential equation as we might use to define the probability model in our method. We will show some examples, such as the one shown in Fig. 1.1 for which the expected value of the probability model does not

converge to a trajectory of the original ODE.

## 1.1 The Basic Stochastic Model

We now review the set-up from Kot [28] and Renshaw [34] of a simple linear birth process, also called the Yule-Furry process [28]. We use this basic model as justification for a more general non-linear model.

The pure-birth (or general birth-death) process is a continuous-time Markov process, wherein the probability of future events occurring only depends on the current state. For a population model, we assume the state space is any potential number of individuals in the population ( $n = 0, 1, 2, \dots$ ) at any time  $t \geq 0$ . Note that in expanding this to non-population models, the state space may also be negative. We assume that the population is well-mixed and each individual acts independently of the others. The assumption of independence is necessary for this formulation, but may not hold for some populations. Animals with complex social structures, for example killer whales [9] and primates, may deviate far from this assumption.

We let  $X(t)$  be a random variable representing the population at time  $t$ . Thus,

$$P(X(t+1) = n | X(t) = \hat{X}), \quad n = 0, 1, 2, \dots$$

gives the probability that the random population variable  $X$  is of size  $n$  at time  $t$ , given the knowledge that the population at the previous time step was  $\hat{X}$ .

For a single individual in a pure birth process, the following probabilities are assumed

$$P(1 \text{ birth in } (t, t + \Delta t] | X(t) = 1) = \lambda \Delta t + o(\Delta t)$$

$$P(> 1 \text{ birth in } (t, t + \Delta t] | X(t) = 1) = o(\Delta t)$$

$$P(0 \text{ births in } (t, t + \Delta t] | X(t) = 1) = 1 - \lambda\Delta t + o(\Delta t),$$

for some constant  $\lambda$ , where  $\Delta t$  is taken small enough so that the probability of more than one birth occurring in a given time interval is negligible.

Now if the population is of size  $n$  and a single birth occurs for one individual, then  $n - 1$  individuals did *not* give birth. There are  $\binom{n}{1} = n$  ways for this to happen, which means the probability for a single birth to occur in the entire population over an interval of length  $\Delta t$  is given by

$$\begin{aligned} P(1 \text{ event in } (t, t + \Delta t] | X(t) = n) &= \binom{n}{1}(\lambda\Delta t + o(\Delta t))(1 - \lambda\Delta t + o(\Delta t))^{n-1} \\ &= n\lambda\Delta t + o(\Delta t). \end{aligned} \tag{1.1}$$

Similarly,

$$P(0 \text{ events in } (t, t + \Delta t] | X(t) = n) = 1 - n\lambda\Delta t + o(\Delta t) \tag{1.2}$$

and

$$P(> 1 \text{ event in } (t, t + \Delta t] | X(t) = n) = o(\Delta t). \tag{1.3}$$

At this point, we diverge from the standard model described by Kot and Renshaw. Kot and Renshaw proceed to long time results by using Chapman-Kolmogorov equations [10] to find a closed form of the probability a population is size  $n$  at time  $t$ . This leads to a probability model where event times are the random variable, drawn from an exponential distribution with parameter  $n\lambda$ . Simulations of this model use Gillespie's algorithm, described in the Appendix B.

Alternatively, we may directly use Eqns (1.1), (1.2), and (1.3) to arrive at the basic *discrete short time probability model* which we simulate for a given initial population.

Beginning at some initial time, we progress an interval of length  $\Delta t$ , check a random number against the event probability, and update the population size accordingly. We then proceed another interval of length  $\Delta t$  and repeat the process. In this model, the time interval is fixed and the population is the random variable.

We now use the form of this basic model to motivate probability definitions for a birth-death process of a general non-linear model.

## 1.2 Probability Model for a General Non-linear ODE

Consider a general non-linear ODE

$$\begin{cases} \dot{y} = f(y) \\ y(0) = y_0 \end{cases} \quad (1.4)$$

Now, we aim to use  $f(y)$  to define the probabilities of a birth-death process. In this general form, it is not clear that there is a connection to any probabilities. However, if  $f(y) = \lambda y$ , then we can use the rate  $\lambda$  to define the probability  $\lambda \Delta t$  for a single individual to give birth (since  $f(y) > 0$ ) over the time interval of length  $\Delta t$ . Assuming independence, individuals in the population can be defined as categorical variables, each with identical probabilities for events occurring. (A categorical random variable may have more than two possible outcomes and is described as a generalization of a Bernoulli random variable.) The expression  $y\lambda\Delta t$  then represents the probability for an entire population of size  $y$  to increase by one.

We want to use a similar motivation to define a birth-death process for a general non-linear ODE. To justify such a model, we consider *only* differential equations of the

form  $f(y) = yg(y)$ . This structure implies a population of size  $y$  with each individual a categorical random variable i.i.d. with probability  $g(y)\Delta t$  of an event occurring. As in the linear case, this gives  $yg(y)\Delta t$  as the probability a single event occurs in the entire population over  $[t, t + \Delta t]$ , with a birth or death occurring depending on the sign of  $g$ .

To make this more clear, we are now restricting ourselves to any ODE of the form

$$\begin{cases} \dot{y} = f(y) = yg(y) \\ y(0) = y_0 \end{cases}, \quad (1.5)$$

where

$$g(y) = g_1(y) - g_2(y)$$

and  $g_1(y)$  and  $g_2(y)$  are nonnegative.

With these assumptions, we use the rates of the ODE to define a discrete birth-death process. The probability an event occurs to a *single individual* in a population of size  $Y_{n-1}$  is defined

$$P(1 \text{ birth in } (t, t + \Delta t] | Y(t) = Y_{n-1}) = g_1(Y_{n-1})\Delta t$$

$$P(1 \text{ death in } (t, t + \Delta t] | Y(t) = Y_{n-1}) = g_2(Y_{n-1})\Delta t$$

$$P(0 \text{ events in } (t, t + \Delta t] | Y(t) = Y_{n-1}) = 1 - (g_1(Y_{n-1}) + g_2(Y_{n-1}))\Delta t.$$

Furthermore, we define the probability of more than one event occurring in the interval  $[t_{n-1}, t_n]$  to be zero.

Given an ODE of the form (1.5) and event probabilities for individuals defined above, we now consider event probabilities for an entire population of size  $Y_{n-1}$ .

Again, independence implies the entire population is a sum of  $Y_{n-1}$  categorical random variables with three 'success' probabilities occurring in a time interval of length  $[t_{n-1}, t_n]$ . Within this framework, we can define the probability for the entire population  $Y(t + \Delta t)$  to have 1 birth, 0 deaths, and no events occurring for the remaining  $Y_{n-1} - 1$  individuals in the following way:

$$\begin{aligned} P(Y(t + \Delta t) - Y(t) = 1 | Y(t) = Y_{n-1}) \\ &= \frac{Y_{n-1}!}{1!0!(Y_{n-1}-1)!} (g_1(Y_{n-1})\Delta t)^1 \\ &\quad \cdot (g_2(Y_{n-1})\Delta t)^0 \\ &\quad \cdot (1 - (g_1(Y_{n-1}) + g_2(Y_{n-1}))\Delta t)^{Y_{n-1}-1-0}, \end{aligned}$$

or

$$P(Y(t + \Delta t) - Y(t) = 1 | Y(t) = Y_{n-1}) = Y_{n-1}g_1(Y_{n-1})\Delta t + o(\Delta t).$$

Similarly,

$$P(Y(t + \Delta t) - Y(t) = -1 | Y(t) = Y_{n-1}) = Y_{n-1}g_2(Y_{n-1})\Delta t + o(\Delta t)$$

and

$$P(Y(t + \Delta t) - Y(t) = 0 | Y(t) = Y_{n-1}) = 1 - Y_{n-1}(g_1(Y_{n-1}) + g_2(Y_{n-1}))\Delta t + o(\Delta t).$$

In this model, each individual is a categorical random variable with three possible outcomes - birth, death, and no event - with probabilities listed above. Note that without this particular structure of  $f(y)$ , the connection between the expected value of the discrete stochastic model with categorical random variables and the solution of the ODE is not apparent. This becomes more clear in the description of conditional expectation below.

However, when the ODE is in this form, we can derive a direct connection between continuous rates from an ODE and discrete probabilities of a birth-death process. A limitation of this model is that we must have  $y(g_1(y) + g_2(y))\Delta t \leq 1$  so the equations above do, in fact, describe probabilities. This implies that the probability model we describe is only valid for small  $\Delta t$  over a time interval where this inequality holds. If this is not the case, the assumptions and the probability model are not valid beyond this point. For several models, decreasing the size of  $\Delta t$  extends the validity of the probability model at least for some time. However, this does increase computational cost and may not completely resolve the issue depending on the behavior of the ODE, e.g. exponential growth or some non-linear behaviors. Note in the case of exponential growth, we may also place an artificial bound on  $g(y)$  for a fixed  $\Delta t$  to ensure the probability assumption. Unless otherwise stated, from here on we assume  $\Delta t$  is small enough to ensure the probability assumption of our model over the time interval of interest.

### 1.2.0.1 Scaling

We note that one apparent issue is that the discrete probability model involves changes of 1, while an ODE describes infinitely small changes. However, if we scale  $Y_n = Y_n/M$ ,  $M$  a large maximum population, then we can convert the discrete probability model to a model for density that increments by  $1/M$ . We can compute the appropriately scaled ODE and obtain the same results.

For the linear case where  $f(y) = \lambda y$  or for functions which describe total numbers of individuals (not densities), scaling is not necessary.

### 1.2.0.2 Lipschitz continuity

We must also assume the functions  $f$  and  $g$  are Lipschitz continuous. This continuity has a direct effect on the validity of the difference bounds in Chapters 2 and 4. Recall that  $f$  is Lipschitz continuous with constant  $L_f$  if the following holds:

$$|f(x_1) - f(x_2)| \leq L_f |x_1 - x_2| \quad \text{for all } x_1, x_2 \text{ in a specified compact set.} \quad (1.6)$$

We denote the Lipschitz constant of  $g$  by  $L_g$ .

In practice,  $f$  and  $g$  are differentiable and we compute the Lipschitz constant using the maximum of the function's derivative,  $f'(y)$  or  $g'(y)$ , over its domain.

### 1.2.1 Conditional Expectation of the Probability Model

Assuming that the individual probabilities are between 0 and 1 and the assumptions of the ODE hold, we consider the expectation of the discrete probability model. As the birth-death process is a Markov process, we define the conditional expectation using the dependence on the population at the previous time step.

Similar to what is used for an explicit Euler method, the model indicates

$$Y_n - Y_{n-1} = f(Y_{n-1})\Delta t = Y_{n-1}g(Y_{n-1})\Delta t$$

or

$$Y_n - Y_{n-1} = Y_{n-1}g_1(Y_{n-1})\Delta t - Y_{n-1}g_2(Y_{n-1})\Delta t.$$

If we think of the right hand side as a conditional expected change in the population size over the time interval  $[t_{n-1}, t_n]$ , this supports the assumption that  $P(\text{birth}) = g_1(y)\Delta t$  and  $P(\text{death}) = g_2(y)\Delta t$  give the probabilities a birth and death occur,

respectively, and the probability of an event (whether it is a birth or death) is  $p = P(\text{birth}) + P(\text{death})$ .

We compute the expected population *change* conditioned on the previous population at time  $t_{n-1}$  and use it to determine the total expectation of the population size at time  $t_n$ . We use the notation  $\text{Cat}(p_k)$  for the categorical variable with probability  $p_k$ ,  $k = 1, 2, 3$ .

$$\begin{aligned} E(Y_n - Y_{n-1}|Y_{n-1}) &= (-1)E\left(\sum_{i=1}^{Y_{n-1}} \text{Cat}(g_2(Y_{n-1})\Delta t)|Y_{n-1}\right) \\ &\quad + (1)E\left(\sum_{i=1}^{Y_{n-1}} \text{Cat}(g_1(Y_{n-1})\Delta t)|Y_{n-1}\right) \\ &\quad + (0)E\left(\sum_{i=1}^{Y_{n-1}} \text{Cat}(1 - g_1(Y_{n-1})\Delta t - g_2(Y_{n-1})\Delta t)|Y_{n-1}\right) \end{aligned}$$

Using the linearity of expectation, we have

$$\begin{aligned} E(Y_n - Y_{n-1}|Y_{n-1}) &= (-1) \sum_{i=1}^{Y_{n-1}} E(\text{Cat}(g_2(Y_{n-1})\Delta t)|Y_{n-1}) \\ &\quad + (1) \sum_{i=1}^{Y_{n-1}} E(\text{Cat}(g_1(Y_{n-1})\Delta t)|Y_{n-1}). \end{aligned}$$

Since the probabilities do not depend on  $i$  and the expectation of a categorical random variable is its success probability  $p_k$ , this expression becomes

$$E(Y_n - Y_{n-1}|Y_{n-1}) = -Y_{n-1}g_2(Y_{n-1})\Delta t + Y_{n-1}g_1(Y_{n-1})\Delta t$$

or

$$E(Y_n - Y_{n-1}|Y_{n-1}) = Y_{n-1}g(Y_{n-1})\Delta t. \quad (1.7)$$

Again, using the linearity of expectation, this gives

$$E(Y_n|Y_{n-1}) = E(Y_{n-1}) + Y_{n-1}g(Y_{n-1})\Delta t.$$

Taking the expected value of both sides one more time then gives the total expectation,

$$E(Y_n) = E(Y_{n-1}) + E(Y_{n-1}g(Y_{n-1}))\Delta t. \quad (1.8)$$

## 1.3 Scalar Examples

In this section, we consider several examples of ODEs. We form the corresponding birth-death process and observe the simulations and expected value as they compare to the solution of the ODE. We illustrate several issues that may require adjustments to the model: probability assumptions for large populations or time steps, single or randomly chosen initial conditions, and stability of steady states.

### 1.3.1 Exponential Growth

We first consider the basic ODE of exponential growth,

$$\begin{cases} \dot{y} = \lambda y \\ y(0) = y_0, \end{cases} \quad (1.9)$$

with  $\lambda = 3$ ,  $y_0 = 6$ .

Note that in this simple case, the ODE function is linear and analysis is greatly simplified. We discuss implications of this linearity for each of the difference bounds in the following sections. Since  $Y_{n-1}$  is increasing, the probability condition of  $\lambda Y_{n-1} \Delta t \leq 1$  requires  $\Delta t$  to be very small for any significant time interval. Fig. 1.2 shows the stochastic simulations, expected value, and ODE solution for  $\Delta t = 0.001$ ,  $y_0 \in U([1, 11])$ .

In this figure, note that the expected value of these simulations follows very closely to the solution of the ODE initially. At  $t = 0.971$ , however, the probability of an event occurring becomes greater than 1 for simulations with  $Y > 333$ . The discrete probability model is not valid after this time.

In this case, it is illustrative to consider what happens if we use the model past

this time interval. Once the probability threshold is reached, we see linear behavior in each of the simulations. This occurs because we test whether an event occurs by comparing a random number on  $U([0, 1])$  to the birth probability. If the random number is less than the probability, we say an event occurred and we update the population by 1. This means that we accept every random number compared to the event probability and a birth is guaranteed to occur at every time step. This results in an underestimation of the number of births that occurred, thus maintaining linear, instead of exponential, growth.

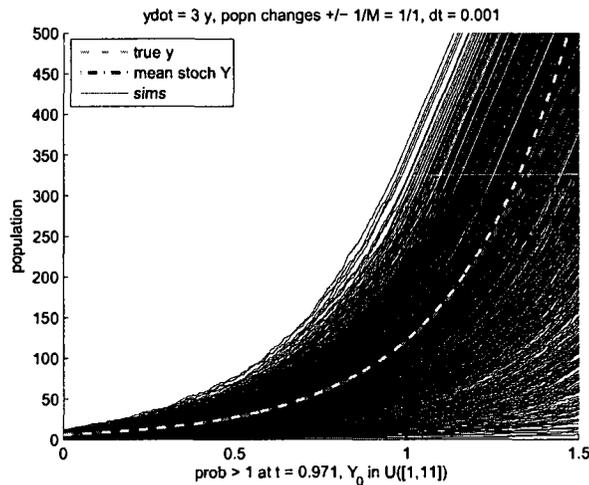


Figure 1.2: Simulations for exponential growth for  $\Delta t = 0.001$  and  $Y_0 \in U([1, 11])$ . The probability model is only valid on the time interval  $[0, 0.971]$  for  $Y \leq 333$ .

### 1.3.2 Logistic Growth

We next consider the ODE for the logistic model with carrying capacity,  $K$ , where  $y(t)$  describes the population density at time  $t$ . For more on the impact of scaling, refer to the discussion at the end of this section.

$$\begin{cases} \dot{y} = ry(1 - y/K) \\ y(0) = y_0. \end{cases} \quad (1.10)$$

For this example and in the following figures, we assume  $r = 1$  and  $K = 0.5$ . Assuming population density, we scaled the discrete probability model such that  $y = y * M$  and simulations describe discrete changes of size  $\pm 1/M$ .

In this example, the probability remains less than one until carrying capacity is reached, but the initial nonlinearity has a significant impact on the expected value of the simulations. If we choose  $y_0$  to be randomly chosen from some interval (for example, uniformly on  $[0, 0.1]$ ), the expected value averages these nonlinearities from simulations that fall on different trajectories of the ODE solution. We show this in Fig. 1.3. Compare this expected value to the expected value in Fig. 1.4, where each simulation has the same initial value, 0.05. See Chapter 2 for more on how this nonlinearity effects the potential difference between discrete probability and deterministic models.

Note that for  $M = 1000$  and  $\Delta t = 0.001$ , the probability model is only valid up to a population equivalent to the carrying capacity. This occurs for at least one simulation at time  $t = 4.047$  and  $t = 3.363$  in Fig. 1.4 and Fig. 1.3, respectively. The shorter end time shown in Fig. 1.3 makes sense as the variability increases with a range of initial conditions. See the following sections for more on the impact of  $M$  and  $\Delta t$  on the discrete probability model.

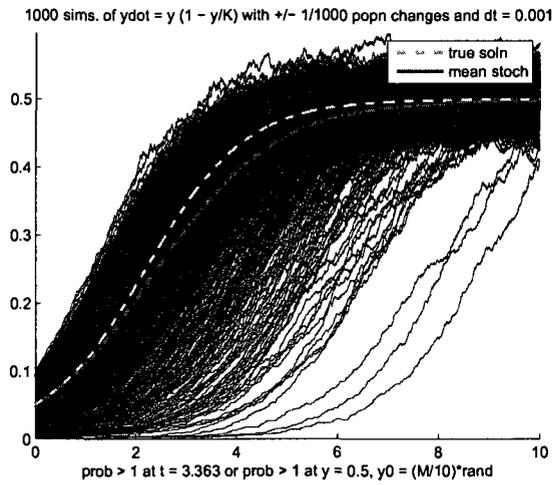


Figure 1.3: Simulations for logistic growth with  $\Delta t = 0.001$ , and random initial conditions on  $U([0, 0.1])$ .

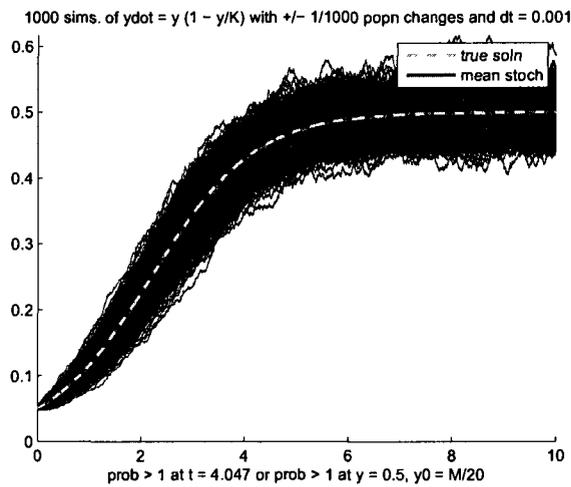


Figure 1.4: Simulations for logistic growth with all simulations having the same initial condition,  $Y_0 = 0.05$ ,  $\Delta t = 0.001$ .

### 1.3.3 Gompertz growth for tumors with time dependent parameter

Cancerous tumors were initially modeled with simple exponential growth. In 1964, A.K. Laird [30] successfully modeled tumor growth data with the Gompertz curve given by the solution of the ODE (1.3.3), which was a significant improvement over purely exponential growth. The ODE is described as

$$\begin{cases} \frac{dV}{dt} = \alpha \ln\left(\frac{K}{V(t)}\right)V(t) \\ V(0) = V_0, \end{cases}$$

with solution

$$V(t) = K \exp(\ln(V_0/K) * \exp(-\alpha t))$$

where  $V$  is the volume of the tumor,  $\alpha$  is a constant related to the potential for rapid growth of the cells, and  $K$  is the maximum possible size of the tumor. Since tumors are just cellular populations growing in a confined space with limited nutrients, the use of a ‘carrying capacity’ in the model makes sense.

As the ODE has a stable behavior at  $K$ , the short time step model does a very good job matching the expected value to the solution of the ODE, regardless of initial conditions less than or greater than  $K$ . See Fig. 1.5. Biologically speaking,  $V_0 < K$  is the scenario where the tumor was found relatively early and  $V_0 > K$  is possible if the tumor is found late and treatment is being pursued.

Note that the issue of time dependent parameters is already covered in the formulation of the general non-linear model. These probabilities were dependent on the population size, which varies with time and is therefore naturally computed as a time-dependent parameter.

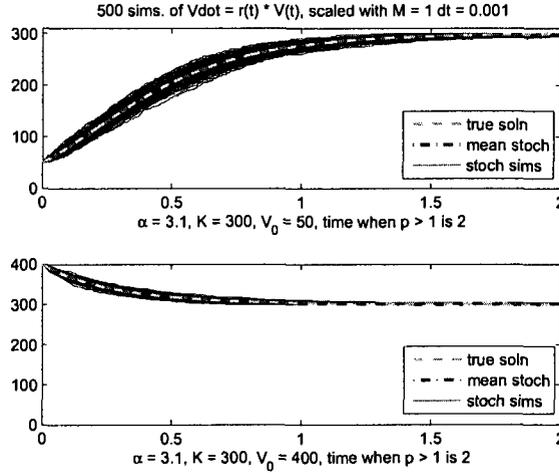


Figure 1.5: Simulations for the birth-death process associated with (1.3.3) with  $\alpha = 3.1$  and  $K = 300$ .

### 1.3.4 Periodic Functions

We also consider the model as it pertains to an ODE with a periodic solution:

$$\begin{cases} \dot{y} = -(0.25 + \sin(\pi t))y^2 \\ y(0) = y_0. \end{cases} \quad (1.11)$$

A time step of  $\Delta t = 0.01$  gives probabilities greater than 1 at  $t = 0.07$ , so we reduce  $\Delta t$  to 0.001 for  $[0, T]$ . Fig. 1.6 shows these simulations as well as the difference,  $|E(Y) - y|$ , at each time step with a single initial condition of  $y_0 = 1.5$ .

In this case, the decreasing trend of the model is sufficient to stabilize the periodic changes in population and results in a fairly accurate probability model with a maximum difference less than 1%. Even for a range of initial conditions on  $[0, 3]$ , the expected value of the probability model captures the periodic behavior of the deterministic solution fairly well.

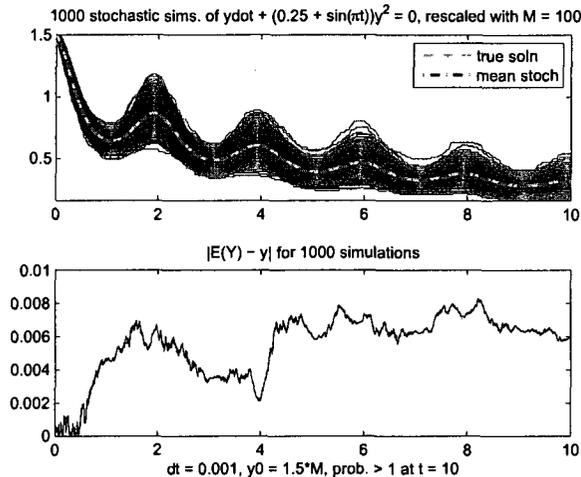


Figure 1.6: Simulations for the birth-death process associated with (1.11) for  $\Delta t = 0.001$ ,  $y_0 = 1.5$ . The second plot shows the difference  $|E(Y) - y|$  for this example.

### 1.3.5 Multiple Steady-State Models

#### 1.3.5.1 Stable Steady States

Next we consider a multiple steady state model with stable and unstable equilibria but no saddle points,

$$\begin{cases} \dot{y} = y(y-1)^3(y+1)^3 \\ y(0) = y_0. \end{cases} \quad (1.12)$$

For the same initial condition on each simulation, the stochastic model matches the solution of the ODE quite well, as shown in Fig. 1.7.

In this case of stable steady states, we do not worry about simulations crossing the steady state or averaging simulations across this value, as seen in Fig. 1.8. There is still a positive probability that some simulations cross the steady state, but the stability properties maintain the correct behavior for the simulations and expected value, even if the steady state is crossed.

Similar to the logistic case in Equation (1.10), using a large range of initial con-

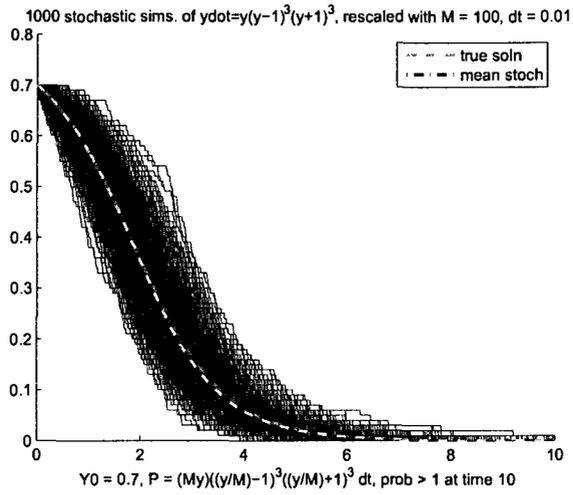


Figure 1.7: Simulations for the birth-death process associated with (1.12).

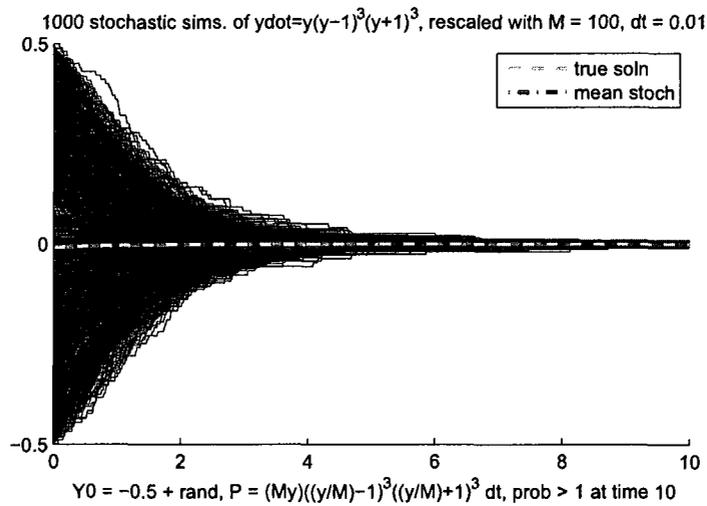


Figure 1.8: Simulations for the birth-death process associated with (1.12).

ditions highly skews the expected value since simulations immediately follow the behavior of many different trajectories, as seen in Fig. 1.9.

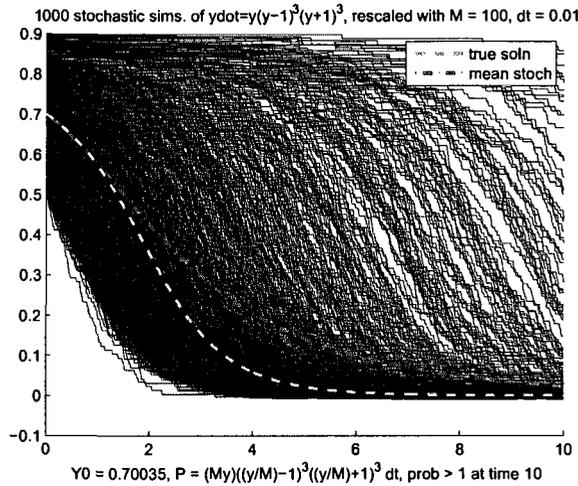


Figure 1.9: Simulations for the birth-death process associated with (1.13).

Intuitively, models with unstable steady states behave similar to the exponential growth example and require a truncation of the model when the probability becomes larger than 1.

### 1.3.5.2 Saddle Point

We next consider the example shown by Fig. 1.1 in the introductory section. This figure shows the result of using sojourn times drawn from an exponential distribution as described by Gillespie’s algorithm in [19]. This example is a pivotal case where the use of the short time step model is vital if we wish to compare the probability model to the differential equation via expected value. For now, we show the results of using short times, holding probability assumptions true and using a single initial condition for each simulation. Then, in the following section, we consider the effect

of using longer time steps.

The deterministic model for this example is given by

$$\begin{cases} \dot{y} = y(y - 0.9755)^2(y + 0.9755)^2 \\ y(0) = y_0, \end{cases} \quad (1.13)$$

with equilibria at  $y = \pm 0.9755$  and  $y = 0$ . We focus on the probability's behavior near the equilibrium  $y^* = 0.9755$ , which is stable for  $y < 0.9755$  and unstable for  $y > 0.9755$ . Fig. 1.10 shows this model using short time steps, which keeps the probabilities less than one and gives the expected value the same behavior as the ODE.

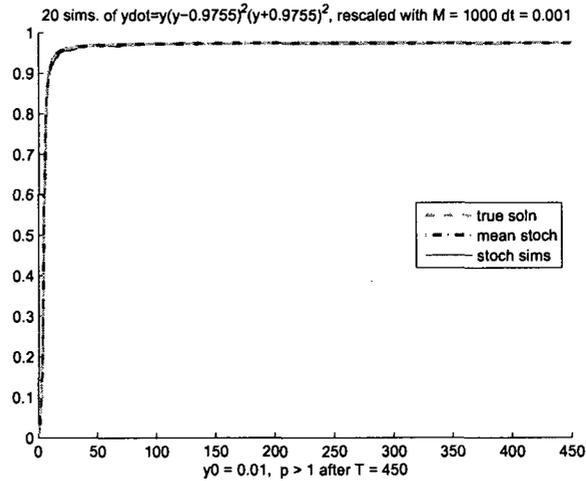


Figure 1.10: Simulations for the birth-death process associated with (1.13) using the short time step model. Notice that by keeping the time step small we avoid jumps across the steady state.

In the next progression of this short time step model, we consider increasing the efficiency by using larger time steps in cases of already stable behavior. We propose such a model in the following section and observe the impact on the above example.

### 1.3.5.3 The Effect of Changing Time Step Size

In trajectories where the behavior of the ODE is stable, we attempt to improve the efficiency of the model by increasing the time step. To implement this, we first choose a larger time step. If it turns out that the time step is too large such that the probability of an event is greater than 1, we decrease the time step until the probability is exactly 1. Updating the population, we repeat the process maintaining the size of  $\Delta t$  for this intermediate behavior.

If the expected value of the stochastic model matches the ODE, we should see similar stable behavior in the expected value. In examples with stable steady states, such as (1.12), this long time model works quite well. If a simulation crosses the steady state (for which there is still a small, but positive probability that this occurs), the behavior of the new trajectory forces it back toward equilibrium.

However, in the case of (1.13), we have a saddle point at  $y^* = 0.9755$  and things do not behave so nicely. Fig. 1.11 shows the result of the moderate time probability model for this ODE. All simulations behave well for small time steps, but if the equilibrium  $y^*$  is not exactly reached (resulting in a zero probability for an event to occur), then there is some positive probability of an event (specifically a birth) occurring at some time. This probability is quite small, so it may seem that we are justified in increasing the time step and therefore decreasing the computation time. However, this also increases the probability of an event occurring, which causes the simulations to jump across the steady state and then diverge from the steady state. We see this transition beginning to occur at about  $t = 400$  in Fig. 1.11. Note that Fig.

1.1 shows the same contradictory behavior, though it was simulated using Gillespie's algorithm for variable time steps. Enough simulations have this significantly different behavior that the expected value also crosses the steady state.

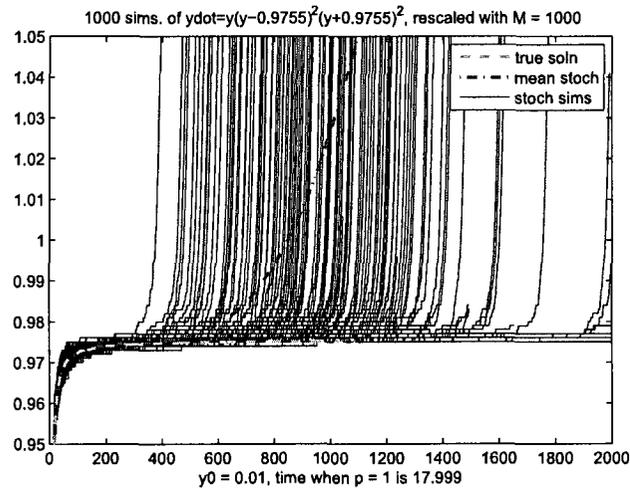


Figure 1.11: Simulations for the birth-death process associated with (1.13) using the adjusted short time step model. Here we zoom in on the transition phase from one side of the steady state to the other.

In the direction of R. Dolgoarshinnykh's paper, beginning with a stochastic model with this behavior, we could likely find an ODE whose solution matches the expected value of the stochastic model. However, such an expected value does not match any trajectory of the *original* ODE, (1.13).

It does not make sense to compare statistics of solutions that cross the steady state to the ODE solution. We can, however, make meaningful comparisons for classes of solutions. For example, in Fig. 1.10, we see that simulations below  $y^* = 0.9755$  converge to the steady state and have an expected value that follows the behavior of the ODE if we keep the time step sufficiently small. This is the first class of solutions that can reasonably be compared to the ODE.

The first class corresponds to those simulations which lie strictly below the steady state and can still be classified as transient behavior. We note here that our difference bounds are based on a Gronwall argument which is only valid for transient behavior. The expected value for these simulations can be compared to the solution of the ODE with initial condition  $y_0 = E(Y_0)$  using our transient analysis. In the third class, we compare those simulations which are strictly above the steady state. These solutions may then be compared to the solution of the ODE with initial condition corresponding to the expected value of these simulations (above the steady state) using our analysis. The middle class describes long term behavior of the ODE for which our analysis does not hold. It is in this intermediate regime that the probability model has the potential to cross the steady state and it does not make sense to compare its expected value to the differential equation in this case. See Fig. 1.12.

#### **1.3.5.4 Discretization Implications**

Intuition suggests that there might be a connection between the evolution of the stochastic model and the numerical advance of the underlying ODE. The stochastic model is strongly reminiscent of a forward Euler approximation, suggesting that it might be worthwhile to consider the analogs of high order integration schemes and schemes that preserve dynamics. We investigated this intuition, but concluded that there is a relatively weak connection between the stochastic model based on categorical random variable and time integration of ODEs.

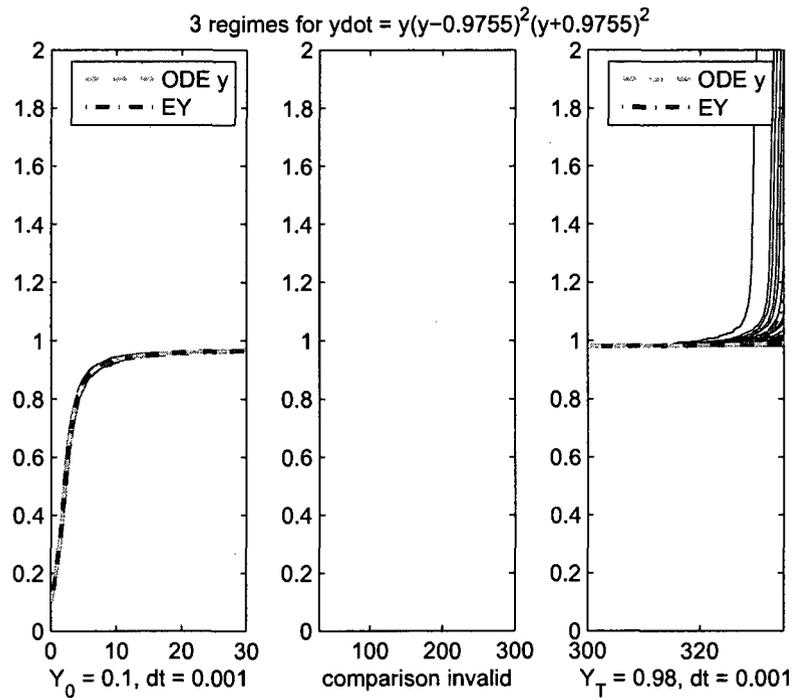


Figure 1.12: We show an example of splitting up the comparison between an ODE solution and the expected value of the discrete stochastic model. The first and third plot show transient behavior for which the two models are comparable. Analysis should not be done on the intermediate time interval as the probability model jumps across the steady state, which is not possible for any trajectory of the ODE.

### 1.3.5.5 Observations on Scaling and Time Step Size

We now take a closer look at how the scaling constant  $M$  and the size of the time step  $\Delta t$  relate to this probability that the stochastic model crosses a steady state. In the probability model, changes in the population occur in discrete increments. The size of  $M$  is related to  $\Delta t$  and the assumption that multiple events occur in the population with probability  $o(\Delta t)$ . Maintaining this assumption requires the population to be relatively small for large  $\Delta t$ . It then follows that the discrete change in the population is relatively large compared to the total population size. Similarly, a small time step implies that one event is likely to occur much more often, meaning that the population must be larger and the population increment relatively smaller.

Reducing  $\Delta t$  and increasing  $M$  increases the time for which this probability model is valid. Making these changes also reduces the variability in the simulations. We show a progression of four plots with increasing  $M$  and decreasing  $\Delta t$  in Fig. 1.13, which illustrate these effects.

Furthermore, in simulating a probability model close to a steady state, larger values of  $M$  result in smaller population increments, which allows the stochastic simulations to get much closer to the steady state. This results in a smaller probability for an event to occur and thus a smaller probability that a simulation steps over the steady state. If the expected value of the probability model becomes closer than one increment of the steady state, we can no longer compare it to the solution of the ODE. This gives a natural time to split the comparison of the two models between transient and long time behavior, as discussed previously. This observation is also

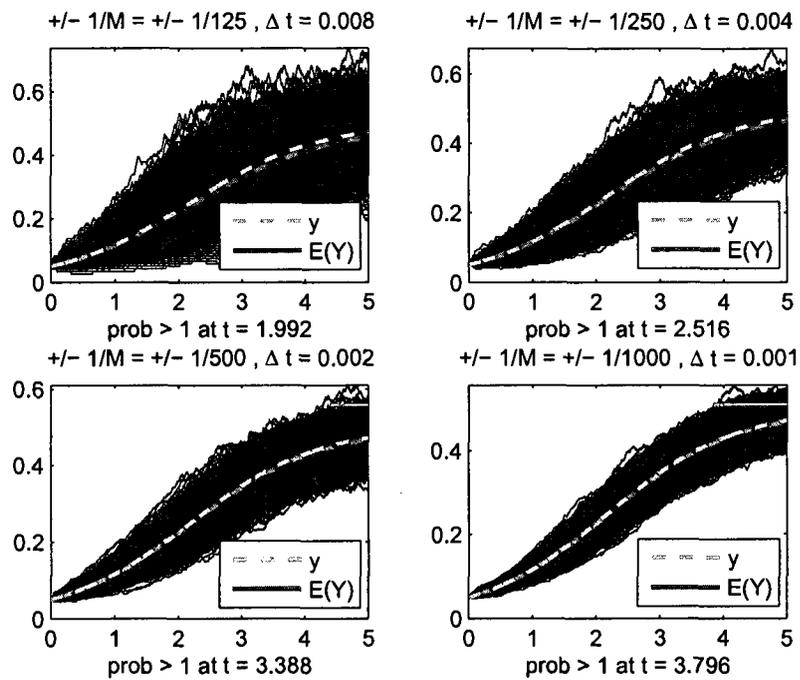


Figure 1.13: From left to right and top to bottom, we show 2000 simulations of the logistic model for increasing  $M$  and decreasing  $\Delta t$ . This progression increases the time for which the model is valid and also decreases the variability in the stochastic simulations.

relevant when we consider system models where one population class may become extinct.

# Chapter 2

## Scalar Difference Bounds

### 2.1 Local Bounds

We now turn to the major aim of this dissertation. We have described a general birth-death process based on an ODE of a particular form, and we wish to determine if its expected value can be reasonably close to the solution of the differential equation.

We first consider a local analysis using the general non-linear ODE with solutions defined on each time interval  $[t_{j-1}, t_j]$ ,  $j = 1, 2, \dots, n$  and initial conditions matching the expected value at  $t_{j-1}$ . We consider the difference between these local trajectories, defined in (2.1), and the expected value at the end of each subinterval.

The intuition of the literature has been that these birth-death processes behave like the differential equation, at least locally. In that respect, the minimum result one might be interested in is that the stochastic solutions stay ‘close’ to local solutions of the ODE on each small time interval. We derive such local bounds describing the potential difference between the expected value at the end of a time interval and the value of the local ODE solution at this same time, given that this local ODE starts at the expected value on the previous time step.

An assumption necessary for the derivation of these bounds is that we have compact support within a particular basin of attraction. Compact support is necessary for Lipschitz continuity of the model and must hold for all families of the solutions, including stochastic simulations over the entire time interval as well as each subinterval, the expected value over the entire time interval, and solutions of the ODE over the entire time interval and subintervals. This assumption restricts blow-up behavior for the ODE as well as the variance in the stochastic simulations.

We first define the local ODE solutions on the interval  $[t_{j-1}, t_j]$ ,

$$\begin{cases} \dot{\tilde{y}}^j = \tilde{y}^j g(\tilde{y}^j) = \tilde{y}^j (g_1(\tilde{y}^j) - g_2(\tilde{y}^j)), & [t_{j-1}, t_j] \\ \tilde{y}^j(t_{j-1}) = \tilde{y}_{j-1}^j = E(Y_{j-1}), & j = 1, 2, \dots, n \end{cases} \quad (2.1)$$

where  $g_1(\tilde{y}^j) > 0, g_2(\tilde{y}^j) > 0$ . We use the notational convention that the superscript gives the interval on which the local solution is defined and the subscript gives the specific time at which the function is evaluated. Note that  $\tilde{y}^0$  is only defined at time  $t_0$  such that  $\tilde{y}_0^0 = E(Y_0)$ . Fig. 2.1 shows a picture of these local solutions, the expected value of the discrete stochastic model, and the solution of the ODE over the entire time.

Now recall the total expected value of  $Y_n$ ,

$$E(Y_n) = E(Y_{n-1}) + \int_{t_{n-1}}^{t_n} E(Y_{n-1}g(Y_{n-1})) dt. \quad (2.2)$$

We can describe the deterministic solution to (2.1) on the  $n^{\text{th}}$  interval at time  $t_n$  as

$$\tilde{y}_n^n = \tilde{y}_{n-1}^n + \int_{t_{n-1}}^{t_n} \tilde{y}^n g(\tilde{y}^n) dt = E(Y_{n-1}) + \int_{t_{n-1}}^{t_n} \tilde{y}^n g(\tilde{y}^n) dt. \quad (2.3)$$

The first theorem bounds the degree to which the solution of a local ODE and the expected value of the discrete stochastic model can diverge on each small time interval.

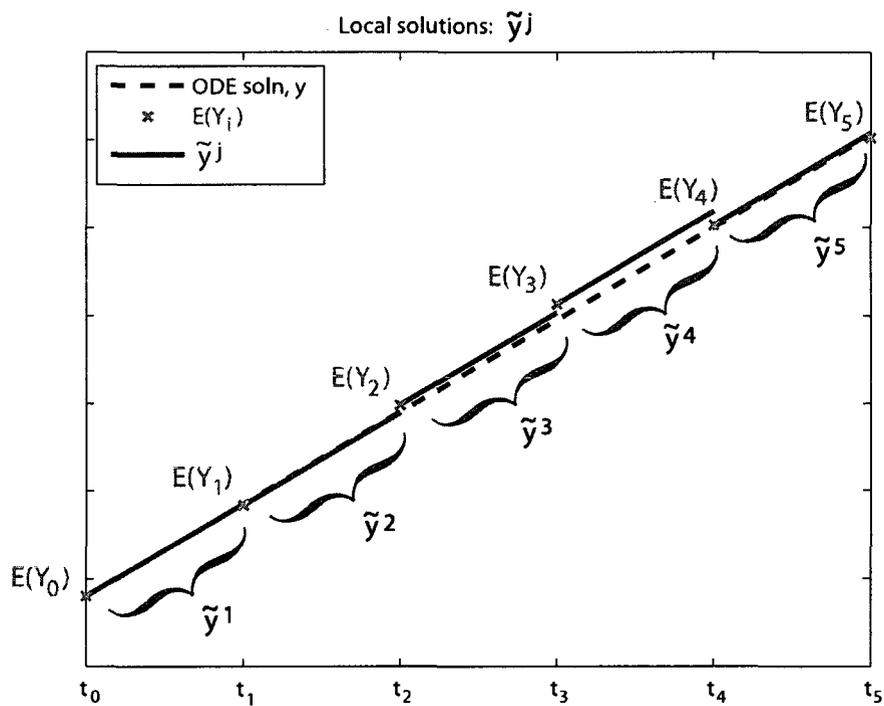


Figure 2.1: Local solutions,  $\tilde{y}^j$ , defined on  $[t_{j-1}, t_j]$  with  $\tilde{y}^j(t_{j-1}) = E(Y_{j-1})$ , which solve the ODE (2.1).

**Theorem 2.1.1** *Local Bound on One Time Step*

The difference on each time step  $[t_{n-1}, t_n]$  is bounded by

$$|E(Y_n) - \tilde{y}_n^n| \leq I + II, \quad n = 1, 2, 3, \dots$$

with

$$I = [Y_{n-1}g(Y_{n-1})] \equiv |E(Y_{n-1}g(Y_{n-1})) - E(Y_{n-1})g(E(Y_{n-1}))| \Delta t, \quad (2.4)$$

and

$$II = L_f \Delta t |E(Y_{n-1}) - \tilde{y}_{n-1}^{n-1}| + L_f \int_{t_{n-1}}^{t_n} |\tilde{y}^n - \tilde{y}_{n-1}^{n-1}| dt,$$

where  $\tilde{y}_0^0 \equiv E(Y_0)$ .

Term I is an expression that measures the difference resulting from the fact that the expectation does not commute with nonlinear generators, i.e. it is the difference between the expectation of the forcing and the forcing of the expectation. If the expected value and the local solution on each time interval begin at the same value,  $E(Y_{n-1})$ , so that the local solution is truly local, then the bound measures only how the stochastic model and the ODE can diverge over one subinterval. We show later that this difference is bounded by the variance of the discrete probability model.

Term II measures how much the solution of the ODE can diverge from its initial condition. This same expression arises in analysis of the explicit Euler finite difference scheme and reflects the stability properties of the ODE in the crudest terms. Based on two nearby trajectories of  $y$ , this term describes the maximum possible divergence between the trajectory on one subinterval compared to its initial value.

In the linear case of exponential growth, term I is zero so the two models only differ depending on the potential growth of the ODE solution from its initial condition.

In the logistic case, whose bound is shown in Fig. 2.2, both the nonlinearity and  $[Yg(Y)]$  increase with time, making the total bound increase on each time step. Fig. 2.3 shows the corresponding growth of  $[Yg(Y)]$  for these simulations. Note that the computations of these bounds use unbiased estimators to calculate the expectation. See Appendix A for a discussion.

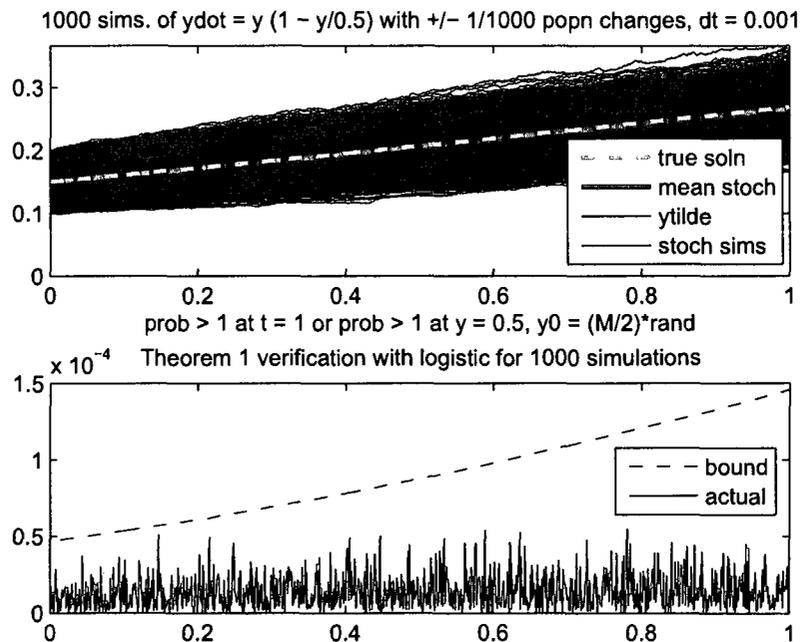


Figure 2.2: The first plot shows the simulations and expected value for the small time step model, as well as the solution to the logistic ODE. The second plot shows the difference  $|E(Y_n) - y_n|$  using an unbiased estimator for the expectation and numerical solution of  $y$  as well as the bound of Theorem 2.1.1.

### Proof of Theorem 2.1.1

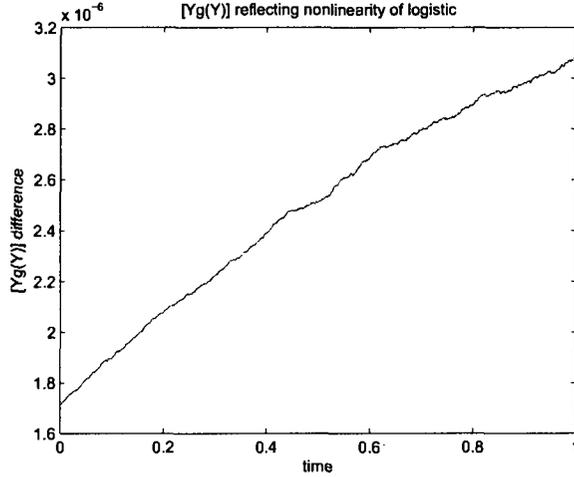


Figure 2.3: This plot shows the nonlinear effects of the logistic model by plotting the difference  $[Yg(Y)]$ , which was calculated along with the simulations in Fig. 2.2.

$$|E(Y_n) - \tilde{y}_n^n| = \left| \int_{t_{n-1}}^{t_n} (E(Y_{n-1}g(Y_{n-1})) - \tilde{y}^n g(\tilde{y}^n)) dt \right|$$

Adding and subtracting the integral  $\int_{t_{n-1}}^{t_n} E(Y_{n-1})g(E(Y_{n-1}))dt$  gives

$$|E(Y_n) - \tilde{y}_n^n| = \left| \int_{t_{n-1}}^{t_n} (E(Y_{n-1}g(Y_{n-1})) - E(Y_{n-1})g(E(Y_{n-1}))) dt + \int_{t_{n-1}}^{t_n} (E(Y_{n-1})g(E(Y_{n-1})) - \tilde{y}^n g(\tilde{y}^n)) dt \right| \quad (2.5)$$

By Lipschitz continuity of  $f$ , the second integral is bounded by

$$\begin{aligned} |E(Y_n) - \tilde{y}_n^n| &\leq \left| \int_{t_{n-1}}^{t_n} (E(Y_{n-1}g(Y_{n-1})) - E(Y_{n-1})g(E(Y_{n-1}))) dt \right| \\ &\quad + L_f \int_{t_{n-1}}^{t_n} |E(Y_{n-1}) - \tilde{y}^n| dt \\ &= \bar{I} + \bar{II}. \end{aligned} \quad (2.6)$$

Now,  $\bar{I}$  is equivalent to  $I$  if we just notice that all the terms are constant on  $[t_{n-1}, t_n]$ .

Term  $\bar{II}$  describes how far the solution  $\tilde{y}^n$  varies from its initial local value of  $E(Y_{n-1})$ . We rewrite this term by adding and subtracting the previous local integral of  $\tilde{y}^{n-1}$  at time  $t_{n-1}$ ,  $\tilde{y}_{n-1}^{n-1}$ ,

$$\begin{aligned}\bar{T} &\leq L_f \int_{t_{n-1}}^{t_n} |E(Y_{n-1}) - \tilde{y}_{n-1}^{n-1}| dt + L_f \int_{t_{n-1}}^{t_n} |\tilde{y}^n - \tilde{y}_{n-1}^{n-1}| dt \\ &= L_f \Delta t |E(Y_{n-1}) - \tilde{y}_{n-1}^{n-1}| + L_f \int_{t_{n-1}}^{t_n} |\tilde{y}^n - \tilde{y}_{n-1}^{n-1}| dt\end{aligned}$$

The first term simply describes the difference between the ending point of the previous local solution,  $\tilde{y}_{n-1}^{n-1}$ , and the starting point of the next solution,  $\tilde{y}_{n-1}^n = E(Y_n)$ .

Putting this back together with (2.6) and recalling the notation of (2.4) gives the result:

$$|E(Y_n) - \tilde{y}_n^n| \leq [Y_{n-1}g(Y_{n-1})] + L_f \Delta t |E(Y_{n-1}) - \tilde{y}_{n-1}^{n-1}| + L_f \int_{t_{n-1}}^{t_n} |\tilde{y}^n - \tilde{y}_{n-1}^{n-1}| dt. \quad (2.7)$$

Defining  $\tilde{y}_0^0 \equiv E(Y_0)$  implies this bound holds for  $n = 1, 2, 3, \dots$

□

Equation (2.7) holds for all  $n = 1, 2, 3, \dots$  and depends on the difference bound at the previous time step,  $|E(Y_{n-1}) - \tilde{y}_{n-1}^{n-1}|$ . This previous difference appears on the right hand side of the current bound and hints at the potential for a Gronwall argument in the accumulation of these bounds. We use this to advantage in the next theorem, but note that this argument is only valid for transient behavior.

Given the local bound in Theorem 2.1.1, we naturally ask if this means the expected value of the discrete stochastic model remains close to the ODE solution for some number of steps. The next result shows a bound between the expected value and a *set* of local solutions. Note, however, that this set of local solutions can not be expected to remain close to a global solution of the ODE. Hence it does not provide the desired bound.

**Theorem 2.1.2** *Accumulated Local Bound*

For any time  $t_n \geq t_0$ ,

$$\begin{aligned} |E(Y_n) - \tilde{y}_n^n| &\leq \sum_{j=1}^n (L_f \Delta t)^{n-j} [Y_{j-1} g(Y_{j-1})] \\ &\quad + L_f \sum_{j=1}^n (L_f \Delta t)^{n-j} \int_{t_{j-1}}^{t_j} |\tilde{y}^j - \tilde{y}_{j-1}^{j-1}| dt. \end{aligned} \quad (2.8)$$

**Proof of Theorem 2.1.2**

We begin with (2.7) and use a discrete Gronwall argument. In the first step, substitute the difference at time  $t_{n-1}$  to obtain

$$\begin{aligned} |E(Y_n) - \tilde{y}^n(t_n)| &\leq [Y_{n-1} g(Y_{n-1})] + L_f \Delta t |E(Y_{n-1}) - \tilde{y}_{n-1}^{n-1}| \\ &\quad + L_f \int_{t_{n-1}}^{t_n} |\tilde{y}^n - \tilde{y}_{n-1}^{n-1}| dt \\ &\leq [Y_{n-1} g(Y_{n-1})] + L_f \int_{t_{n-1}}^{t_n} |\tilde{y}^n - \tilde{y}_{n-1}^{n-1}| dt \\ &\quad + L_f \Delta t [Y_{n-2} g(Y_{n-2})] \\ &\quad + L_f^2 (\Delta t)^2 |E(Y_{n-2}) - \tilde{y}_{n-2}^{n-2}| + L_f^2 \Delta t \int_{t_{n-2}}^{t_{n-1}} |\tilde{y}^{n-1} - \tilde{y}_{n-2}^{n-2}| dt. \end{aligned}$$

Substituting the difference at  $t_{n-2}$  gives

$$\begin{aligned} |E(Y_n) - \tilde{y}_n^n| &\leq [Y_{n-1} g(Y_{n-1})] + L_f \Delta t [Y_{n-2} g(Y_{n-2})] + L_f^2 (\Delta t)^2 [Y_{n-3} g(Y_{n-3})] \\ &\quad + L_f \int_{t_{n-1}}^{t_n} |\tilde{y}^n - \tilde{y}_{n-1}^{n-1}| dt + L_f^2 \Delta t \int_{t_{n-2}}^{t_{n-1}} |\tilde{y}^{n-1} - \tilde{y}_{n-2}^{n-2}| dt \\ &\quad + L_f^3 (\Delta t)^2 \int_{t_{n-3}}^{t_{n-2}} |\tilde{y}^{n-2} - \tilde{y}_{n-3}^{n-3}| dt \\ &\quad + L_f^3 (\Delta t)^3 |E(Y_{n-3}) - \tilde{y}_{n-3}^{n-3}|. \end{aligned}$$

Continuing in this manner gives the result,

$$\begin{aligned} |E(Y_n) - \tilde{y}^n(t_n)| &\leq \sum_{j=1}^n [Y_{j-1} g(Y_{j-1})] (L_f \Delta t)^{n-j} \\ &\quad + L_f \sum_{j=1}^n \int_{t_{j-1}}^{t_j} |\tilde{y}^j - \tilde{y}_{j-1}^{j-1}| dt (L_f \Delta t)^{n-j}, \end{aligned} \quad (2.9)$$

where the last term  $|E(Y_0) - \tilde{y}_0^0|$  is zero since we define  $\tilde{y}_0^0 \equiv E(Y_0)$ .

□

We next investigate each term in this cumulative bound further. This focuses on different sources for the size of the bound and may potentially improve or at least get

more information about the bound, depending on the nature of the ODE. We look at the first sum in (2.8) and expand  $[Y_{j-1}g(Y_{j-1})]$ . This yields a very interesting and intuitive bound in terms of variance and the second moment of  $Y$ .

**Theorem 2.1.3**  $[Yg(Y)]$  *Bounded by Variance and Second Moment*

*The unbiased estimator for  $[Yg(Y)]$  is bounded on each time step by*

$$\begin{aligned} Est([Y_{j-1}g(Y_{j-1})]) \leq & \left(\frac{N-1}{N}\right)^{1/2} L_g \Delta t \left( \frac{1}{N-1} \sum_{i=1}^N |\{Y_{j-1}\}_i - E(\{Y_{j-1}\}_i)|^2 \right)^{1/2} \\ & \cdot \left( \frac{1}{N} \sum_{i=1}^N |\{Y_{j-1}\}_i|^2 \right)^{1/2}, \end{aligned} \tag{2.10}$$

where  $\{Y_{j-1}\}_i$  is the  $i^{\text{th}}$  simulation at time step  $t_{j-1}$ .

Note that this bound is precisely the square root of the unbiased estimator for variance of  $Y_{j-1}$  multiplied by the square root of the unbiased estimator of its second moment, which has small variance for large  $N$  [32]. The multiplication of these two gives a nice intuitive relationship. If the variance is large, but the magnitude of  $|Y_{j-1}|$  is small, the impact of non-linearity is relatively small on the difference between the stochastic model's expected value and the solution of the ODE. Similarly, if the magnitude of  $|Y_{j-1}|$  is large, but the variance is small, the impact may also be small. On the other hand, if both the magnitude and variance are large, the stochastic model may behave quite differently from the solution of the differential equation. Note, in the logistic case, the variance is relatively small for the scaling constant  $M = 1000$  and time step  $\Delta t = 0.001$ .

In Fig. 2.4, we show the bound and actual difference,  $[Y_{j-1}g(Y_{j-1})]$ , for the logistic model in 1.10 with a single initial value. In Fig. 2.5 we show the bound for random

initial conditions. Note that the actual difference as well as the bound in this figure are similar to Fig. 2.4, except for early time values where the bound is higher. This follows as a range of initial conditions give our probability model higher initial variance and an increased potential for a large difference between the expected value and the deterministic solution.

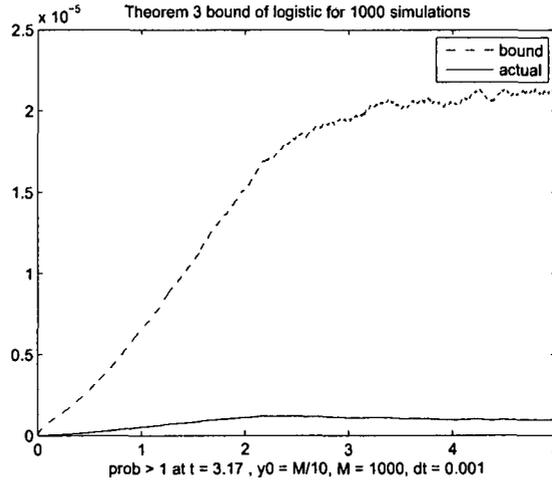


Figure 2.4: This plot shows the actual difference  $[Yg(Y)]$  and the bound from Theorem 2.1.3 for the same initial value in each simulation.

### Proof of Theorem 2.1.3

Recall the definition of (2.4)

$$[Y_{n-1}g(Y_{n-1})] \equiv |E(Y_{n-1}g(Y_{n-1})) - E(Y_{n-1})g(E(Y_{n-1}))| \Delta t.$$

Now since  $E(Y_{j-1})$  is a deterministic value, we can write

$$E(Y_{j-1})g(E(Y_{j-1})) = E(Y_{j-1}g(E(Y_{j-1}))).$$

Substituting this into the difference equation gives

$$\begin{aligned} [Y_{j-1}g(Y_{j-1})] &= |E(Y_{j-1}g(Y_{j-1})) - E(Y_{j-1}g(E(Y_{j-1})))| \Delta t \\ &= |E(Y_{j-1}(g(Y_{j-1}) - g(E(Y_{j-1}))))| \Delta t. \end{aligned} \tag{2.11}$$

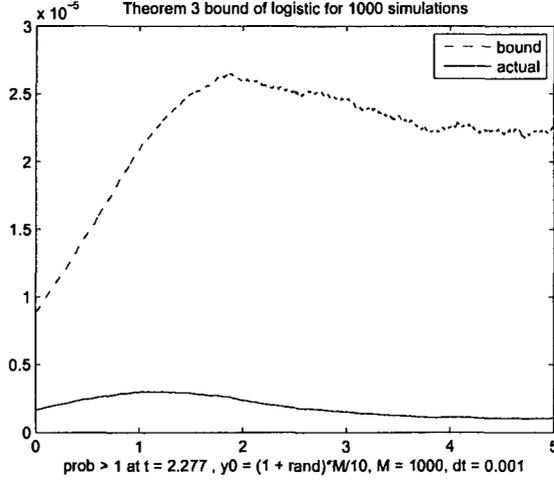


Figure 2.5: We plot the difference  $[Y_{j-1}g(Y_{j-1})]$  and the bound from Theorem 2.1.4 for  $Y_0 \in U([0.11, 0.21])$ .

Using an unbiased estimator to approximate the expectation

$$\begin{aligned} [Y_{j-1}g(Y_{j-1})] &\approx \left| \frac{1}{N} \sum_{i=1}^N \{Y_{j-1}\}_i (g(\{Y_{j-1}\}_i) - g(E(\{Y_{j-1}\}_i))) \right| \Delta t \\ &\equiv Est([Y_{j-1}g(Y_{j-1})]), \end{aligned} \quad (2.12)$$

for  $N$  simulations. See Appendix A for more on unbiased estimators.

Bringing the absolute value and  $1/N$  inside the sum, we use Hölder's inequality to obtain

$$\begin{aligned} Est([Y_{j-1}g(Y_{j-1})]) &\leq \sum_{i=1}^N \left| \frac{1}{\sqrt{N}} (g(\{Y_{j-1}\}_i) - g(E(\{Y_{j-1}\}_i))) \cdot \frac{1}{\sqrt{N}} \{Y_{j-1}\}_i \right| \Delta t \\ &\leq \left( \sum_{i=1}^N \left| \frac{1}{\sqrt{N}} (g(\{Y_{j-1}\}_i) - g(E(\{Y_{j-1}\}_i))) \right|^2 \right)^{1/2} \\ &\quad \cdot \left( \sum_{i=1}^N \left| \frac{1}{\sqrt{N}} \{Y_{j-1}\}_i \right|^2 \right)^{1/2} \Delta t. \end{aligned} \quad (2.13)$$

Simplifying and using the Lipschitz continuity of  $g$ , we have

$$\begin{aligned} Est([Y_{j-1}g(Y_{j-1})]) &\leq \left( \frac{1}{N} \sum_{i=1}^N L_g^2 |\{Y_{j-1}\}_i - E(\{Y_{j-1}\}_i)|^2 \right)^{1/2} \\ &\quad \cdot \left( \frac{1}{N} \sum_{i=1}^N |\{Y_{j-1}\}_i|^2 \right)^{1/2} \Delta t, \end{aligned}$$

or

$$\begin{aligned} Est([Y_{j-1}g(Y_{j-1})]) &\leq L_g \Delta t \left( \frac{1}{N} \sum_{i=1}^N |\{Y_{j-1}\}_i - E(\{Y_{j-1}\}_i)|^2 \right)^{1/2} \\ &\quad \cdot \left( \frac{1}{N} \sum_{i=1}^N |\{Y_{j-1}\}_i|^2 \right)^{1/2}. \end{aligned}$$

Multiplying by  $(\frac{N-1}{N-1})^{1/2}$  and rearranging gives the result with unbiased estimators for variance and the second moment of  $Y_{j-1}$ ,

$$\begin{aligned} Est([Y_{j-1}g(Y_{j-1})]) &\leq \left(\frac{N-1}{N}\right)^{1/2} L_g \Delta t \left( \frac{1}{N-1} \sum_{i=1}^N |\{Y_{j-1}\}_i - E(\{Y_{j-1}\}_i)|^2 \right)^{1/2} \\ &\quad \cdot \left( \frac{1}{N} \sum_{i=1}^N |\{Y_{j-1}\}_i|^2 \right)^{1/2} \end{aligned}$$

□

Another interesting relationship to observe is the bound that results if we choose not to use the Lipschitz constant of  $g$  on equation (2.13). Using the same argument as in the proof, we arrive at the following bound

$$\begin{aligned} Est([Y_{j-1}g(Y_{j-1})]) &\leq \left(\frac{N-1}{N}\right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^N |\{Y_{j-1}\}_i|^2 \right)^{1/2} \\ &\quad \cdot \left( \frac{1}{N-1} \sum_{i=1}^N |g(\{Y_{j-1}\}_i) - g(E(\{Y_{j-1}\}_i))|^2 \right)^{1/2} \Delta t, \end{aligned}$$

which gives another perspective of the bound and the degree to which the function  $g(E(Y_{j-1}))$  varies from  $E(g(Y_{j-1}))$ . Writing the bound in this way allows one to use known behavior of  $g(y)$  (such as nonlinearity or stability properties) to analyze the bound, instead of the properties of  $y$  or  $f(y)$ , which may or may not be more easily determined.

Next, observe that we can also use  $f(y)$  with Lipschitz constant  $L_f$  instead  $yg(y)$  for this analysis. The following theorem shows that using  $f(y)$  gives a bound just in terms of variance, absorbing the second moment in the Lipschitz constant,  $L_f$ .

**Theorem 2.1.4**  $[Yg(Y)]$  Bounded by Variance

The unbiased estimator for  $[Yg(Y)]$  is bounded on each time step by

$$Est([Y_{j-1}g(Y_{j-1})]) \leq \left(\frac{N-1}{N}\right)^{1/2} L_f \Delta t \left(\frac{1}{N-1} \sum_{i=1}^N |\{Y_{j-1}\}_i - E(\{Y_{j-1}\}_i)|^2\right)^{1/2}. \quad (2.14)$$

The bound in Fig. 2.7 for the logistic case has a larger initial value than the bound in Fig. 2.6, due to the variance already present in the initial  $Y_0$  values. Also notice that the bound and  $[Yg(Y)]$  in this figure increases as the variance of the stochastic model increases, then reaches a relatively stable value.

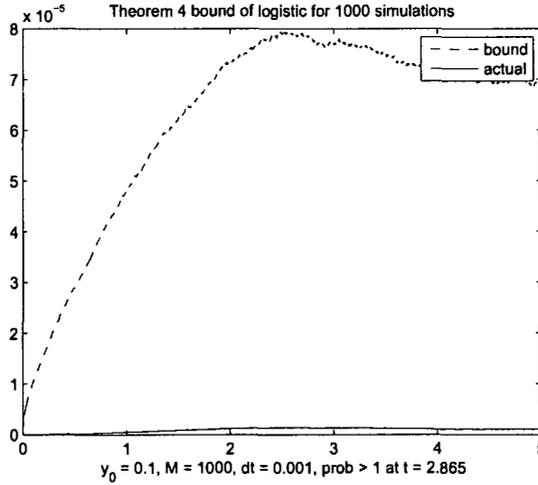


Figure 2.6: This plot shows the actual difference  $[Yg(Y)]$  and the bound from Theorem 2.1.4 for the same initial value in each simulation.

**Proof of Theorem 2.1.4**

Writing (2.4) in terms of  $f$  gives

$$[Y_{j-1}g(Y_{j-1})] = |E(f(Y_{j-1})) - f(E(Y_{j-1}))| \Delta t.$$

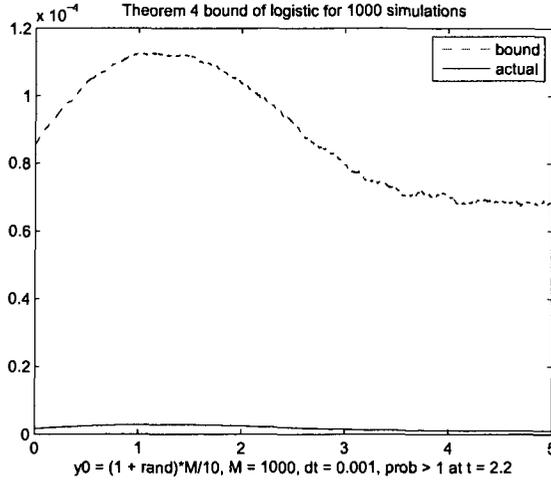


Figure 2.7: Here we show the actual difference  $[Yg(Y)]$  and the bound from Theorem 2.1.4 for the logistic model with initial values chosen from  $U([0.11, 0.21])$ .

Again realizing that  $f(E(Y_{j-1}))$  is a deterministic value, we substitute  $f(E(Y_{j-1})) = E(f(E(Y_{j-1})))$  into the above equation. This gives

$$[Y_{j-1}g(Y_{j-1})] = |E(f(Y_{j-1}) - f(E(Y_{j-1})))| \Delta t.$$

Writing this using an unbiased estimator for expected value gives

$$\begin{aligned} Est([Y_{j-1}g(Y_{j-1})]) &= \left| \frac{1}{N} \sum_{i=1}^N (f(\{Y_{j-1}\}_i) - f(E(\{Y_{j-1}\}_i))) \right| \Delta t \\ &\leq \sum_{i=1}^N \left| \frac{1}{N} (f(\{Y_{j-1}\}_i) - f(E(\{Y_{j-1}\}_i))) \right| \Delta t \\ &= \sum_{i=1}^N \left| \frac{1}{\sqrt{N}} (f(\{Y_{j-1}\}_i) - f(E(\{Y_{j-1}\}_i))) \cdot \frac{1}{\sqrt{N}} \right| \Delta t. \end{aligned}$$

Using Hölder's inequality again gives

$$\begin{aligned} Est([Y_{j-1}g(Y_{j-1})]) &\leq \left( \sum_{i=1}^N \left| \frac{1}{\sqrt{N}} (f(\{Y_{j-1}\}_i) - f(E(\{Y_{j-1}\}_i))) \right|^2 \right)^{1/2} \\ &\quad \cdot \left( \sum_{i=1}^N \left| \frac{1}{\sqrt{N}} \right|^2 \right)^{1/2} \Delta t, \end{aligned} \tag{2.15}$$

or, equivalently,

$$Est([Y_{j-1}g(Y_{j-1})]) \leq \left( \frac{1}{N} \sum_{i=1}^N |f(\{Y_{j-1}\}_i) - f(E(\{Y_{j-1}\}_i))|^2 \right)^{1/2} \Delta t.$$

Using the Lipschitz continuity of  $f$  and multiplying by the constant  $\left(\frac{N-1}{N}\right)^{1/2}$  gives the result

$$Est(\{Y_{j-1}g(Y_{j-1})\}) \leq \left(\frac{N-1}{N}\right)^{1/2} L_f \Delta t \left(\frac{1}{N-1} \sum_{i=1}^N |\{Y_{j-1}\}_i - E(\{Y_{j-1}\}_i)|^2\right)^{1/2}.$$

□

Again, observe that if we choose not to use the Lipschitz condition on  $f$ , we arrive at a bound that gives a sort of variance of  $f(E(Y))$  from the expected value  $E(f(Y))$ , due to nonlinearity. Equation (2.15) is precisely this relation.

As with the  $yg(y)$  form, the bound in the above theorem is a function of the variance of the stochastic model. Using either  $L_g$  or  $L_f$  in the bound of  $Est(\{Y_{j-1}g(Y_{j-1})\})$ , we may use the second moment and variance bounds to advantage. For example, knowing the variance a priori may indicate where the model is most sensitive to changes or an instability occurs. Given this information, we may be able to devote more resources to a particular time and use more coarse time steps in areas where the variance is small. In general, increasing the time step will increase the efficiency of the implementation. For some models, such as 1.13, this creates more problems.

Next, we combine these results into a single bound and observe that it is  $O(\Delta t)$ .

**Theorem 2.1.5** *Total Local Bound*

For  $[t_0, t_n]$ ,

$$\begin{aligned} Est(|E(Y_n) - \tilde{y}^n(t_n)|) &\leq \left(\frac{N-1}{N}\right)^{1/2} L_f \Delta t \sum_{j=1}^n (L_f \Delta t)^{n-j} \left(\frac{1}{N-1} \sum_{i=1}^N |\{Y_{j-1}\}_i - E(\{Y_{j-1}\}_i)|^2\right)^{1/2} \\ &\quad + L_f \sum_{j=1}^n (L_f \Delta t)^{n-j} \int_{t_{j-1}}^{t_j} |\tilde{y}^j - \tilde{y}_{j-1}^{j-1}| dt \end{aligned} \quad (2.16)$$

and

$$|E(Y^n) - \tilde{y}_n^n| = O(\Delta t).$$

This bound is useful in its description of the differences between these two models, but it remains relatively constant over the time interval. This follows as the bound is the same as Theorem 2.1.2, though slightly larger with the variance bound included. We show this in Fig. 2.8.

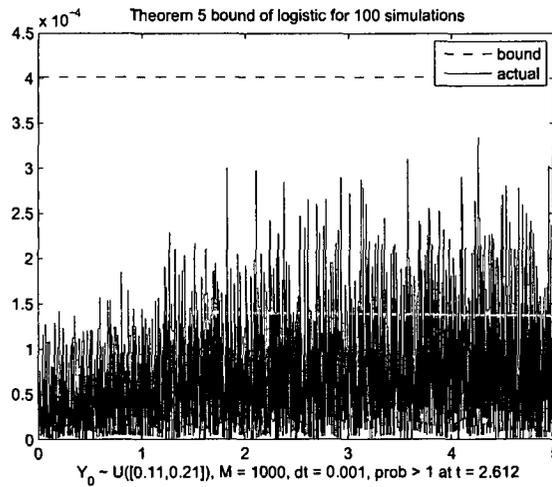


Figure 2.8: This plot shows the actual difference and the bound from Theorem 2.1.4 for the logistic model with initial values chosen from on  $U([0.11, 0.21])$ .

### Proof of Theorem 2.1.5

The bound in (2.16) follows by substituting bounds from Theorem 2.1.2 and Theorem 2.1.4.

Next, we show the difference bound to be  $O(\Delta t)$ . We consider each term in (2.16) and the conditions for it to be less than or equal to some constant  $C$  times  $\Delta t$ .

Starting with the first term, we need to satisfy the inequality:

$$\left(\frac{N-1}{N}\right)^{1/2} L_f \Delta t \sum_{j=1}^n (L_f \Delta t)^{n-j} \left(\frac{1}{N-1} \sum_{i=1}^N |\{Y^{j-1}\}_i - E(\{Y^{j-1}\}_i)|^2\right)^{1/2} \leq C \Delta t. \quad (2.17)$$

The variance of  $Y^{j-1}$  is assumed to be finite, so we define the variance on the  $j^{\text{th}}$  time step to be

$$K_j = \left(\frac{1}{N-1} \sum_{i=1}^N |\{Y^{j-1}\}_i - E(\{Y^{j-1}\}_i)|^2\right)^{1/2}.$$

Substituting gives

$$\left(\frac{N-1}{N}\right)^{1/2} L_f \Delta t \sum_{j=1}^n (L_f \Delta t)^{n-j} K_j \leq C \Delta t.$$

Canceling  $\Delta t$  from both sides and absorbing the constants on the left hand side into

$C$  such that  $\tilde{C} = \frac{C\sqrt{N}}{L_f\sqrt{N-1}}$  gives the equivalent condition to be satisfied,

$$\sum_{j=1}^n (L_f \Delta t)^{n-j} K_j \leq \tilde{C}. \quad (2.18)$$

Factoring out  $K_{\max} = \max_j K_j$  from the LHS of (2.18) gives

$$LHS \leq K_{\max} \left( \frac{K_n}{K_{\max}} + \frac{L_f K_{n-1}}{K_{\max}} \Delta t + \dots + \frac{L_f^{n-1} K_1}{K_{\max}} \Delta t^{n-1} \right).$$

We now divide both sides by  $K_{\max}$  and note that  $\frac{K_j}{K_{\max}} \leq 1$  for all  $j$ , by definition.

Absorbing  $K_{\max}$  into the constant  $\tilde{C}$  gives the geometric series

$$LHS \leq 1 + L_f \Delta t + (L_f \Delta t)^2 + \dots + (L_f \Delta t)^{n-2} \leq \frac{1 - (L_f \Delta t)^{n-1}}{1 - L_f \Delta t}.$$

If  $L_f \Delta t < 1$ , then this is bounded uniformly in  $n$ .

For the second term in (2.16), we need to satisfy the inequality

$$L_f \sum_{j=1}^n (L_f \Delta t)^{n-j} \int_{t_{j-1}}^{t_j} |\tilde{y}^j - \tilde{y}_{j-1}^{j-1}| dt \leq C \Delta t. \quad (2.19)$$

In the next step, we assume the expected value, initial conditions, and solutions of the ODE remain in a compact set. Given this assumption, we can bound the integral and set it equal to  $K_{\max}\Delta t$ ,

$$\int_{t_{j-1}}^{t_j} |\tilde{y}^j - \tilde{y}_{j-1}^{j-1}| dt \leq \max_{[t_{j-1}, t_j]} |\tilde{y}^j - \tilde{y}_{j-1}^{j-1}| \Delta t = K_{\max}\Delta t.$$

Making this substitution gives the revision of (2.19),

$$L_f \sum_{j=1}^n (L_f \Delta t)^{n-j} K_{\max} \Delta t \leq C \Delta t.$$

Again, we define  $\tilde{C}$  to absorb the constants such that  $\tilde{C} = \frac{C}{L_f K_{\max}}$  and cancel  $\Delta t$  from both sides.

As before, we can bound the sum

$$\sum_{j=1}^n (L_f \Delta t)^{n-j} \leq \frac{1 - (L_f \Delta t)^{n-1}}{1 - L_f \Delta t}.$$

For  $L_f \Delta t < 1$ , this sum is bounded uniformly in  $n$  and condition (2.19) is satisfied.

Note that  $L_f \Delta t$  can always be made less than 1 if we choose  $\Delta t < \frac{1}{L_f}$  and when  $L_f$  is large, we require small time steps.

□

## 2.2 Global Difference Bounds

We now derive a global estimate for the difference of the expected values of the discrete stochastic model and the solution of the ODE continuum model. This is a so-called transient or short time result that uses the simplest notion of stability. In this analysis, we consider any end time  $T = t_n > t_0$  and compare the expected value of the discrete stochastic model at this time with the solution of the ODE starting with initial condition  $y_0 = E(Y_0)$ .

Using the integral solution of  $y_n$  and  $E(Y_n)$ , we consider the absolute distance between the expected value and the ODE solution at time  $t_n$  and obtain a bound using only knowledge from the previous time step.

### Theorem 2.2.1 Global Difference Bound Over One Time Step

For any time  $t_n > t_0$ ,

$$|E(Y_n) - y_n| \leq (1 + L_f \Delta t) |E(Y_{n-1}) - y_{n-1}| + [Y_{n-1} g(Y_{n-1})] + L_f |y'|_{\infty}^n (\Delta t)^2, \quad (2.20)$$

where  $|y'|_{\infty}^n$  indicates the infinity norm on the  $n^{\text{th}}$  interval,  $[t_{n-1}, t_n]$ .

In the linear case, each of the  $[Y_{n-1} g(Y_{n-1})]$  terms are zero, so the bound is only dependent on the stability of the ODE at that time and its accumulation from the previous time.

Fig. 2.9 shows the bound of Theorem 2.2.1 in the logistic case for  $M = 1000$ ,  $\Delta t = 0.001$ . This global bound is much larger than the local bounds and grows exponentially from the compounding of the previous difference over a large time interval.

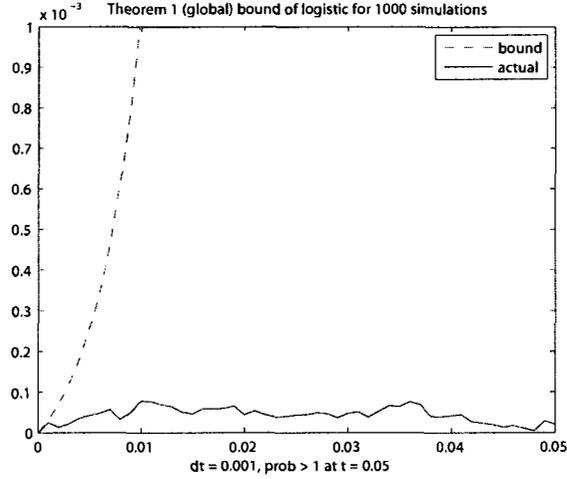


Figure 2.9: Plotting the bound of Theorem 2.2.1 and the actual difference  $|E(Y_n) - y_n|$  for the logistic model.

An interesting thing to note is that the bulk of the true difference between the two models is attributed to the  $[Yg(Y)]$  terms, which are of the same order of magnitude of  $|E(Y_n) - y_n|$ . The stability terms involving the infinity norm over this interval are an order of magnitude smaller than  $|E(Y_n) - y_n|$ . The difference on the previous time step is also an order of magnitude smaller than the true difference if we ignore the effect of the  $[Yg(Y)]$  and stability terms. This gives a better idea of the importance of these nonlinearity terms in the difference between the two models.

### Proof of Theorem 2.2.1

We consider the difference of integral equations for  $y_n$  and  $E(Y_n)$  and use the triangle inequality to obtain

$$|E(Y_n) - y_n| \leq |E(Y_{n-1}) - y_{n-1}| + \left| \int_{t_{n-1}}^{t_n} (E(Y_{n-1})g(Y_{n-1})) - yg(y) dt \right|.$$

Adding and subtracting the integral  $\int_{t_{n-1}}^{t_n} E(Y_{n-1})g(E(Y_{n-1}))dt$  gives

$$\begin{aligned}
|E(Y_n) - y_n| &\leq |E(Y_{n-1}) - y_{n-1}| \\
&\quad + \left| \int_{t_{n-1}}^{t_n} (E(Y_{n-1}g(Y_{n-1})) - E(Y_{n-1})g(E(Y_{n-1}))) dt \right. \\
&\quad \left. + \int_{t_{n-1}}^{t_n} (E(Y_{n-1})g(E(Y_{n-1})) - yg(y)) dt \right|
\end{aligned} \tag{2.21}$$

By Lipschitz continuity of  $f$ , the second integral is bounded by

$$\begin{aligned}
|E(Y_n) - y_n| &\leq |E(Y_{n-1}) - y_{n-1}| \\
&\quad + \left| \int_{t_{n-1}}^{t_n} (E(Y_{n-1}g(Y_{n-1})) - E(Y_{n-1})g(E(Y_{n-1}))) dt \right| \\
&\quad + L_f \int_{t_{n-1}}^{t_n} |E(Y_{n-1}) - y| dt \\
&= |E(Y_{n-1}) - y_{n-1}| + [Y_{n-1}g(Y_{n-1})] \\
&\quad + L_f \int_{t_{n-1}}^{t_n} |E(Y_{n-1}) - y| dt \\
&= A + B + C.
\end{aligned} \tag{2.22}$$

Term C describes how far the solution  $y$  varies from the local value of  $E(Y_{n-1})$ . We rewrite this term by adding and subtracting the integral of  $y$  at time  $t_{n-1}$  to give

$$\begin{aligned}
C &\leq L_f \int_{t_{n-1}}^{t_n} |E(Y_{n-1}) - y_{n-1}| dt + L_f \int_{t_{n-1}}^{t_n} |y(t) - y_{n-1}| dt \\
&= L_f \Delta t |E(Y_{n-1}) - y_{n-1}| + L_f \int_{t_{n-1}}^{t_n} |y(t) - y_{n-1}| dt.
\end{aligned}$$

Putting this back together with (2.22) gives

$$\begin{aligned}
|E(Y_n) - y_n| &\leq (1 + L_f \Delta t) |E(Y_{n-1}) - y_{n-1}| + [Y_{n-1}g(Y_{n-1})] \\
&\quad + L_f \int_{t_{n-1}}^{t_n} |y(t) - y_{n-1}| dt.
\end{aligned}$$

We can bound the last integral using the infinity norm on the interval  $[t_{n-1}, t_n]$ , where  $|y|_\infty = \sup_{\hat{t}} y(\hat{t})$  and  $\hat{t} \in [t_{n-1}, t_n]$ .

This gives a simplified version of the difference bound on the interval  $[t_{n-1}, t_n]$ ,

$$\begin{aligned}
|E(Y_n) - y_n| &\leq (1 + L_f \Delta t) |E(Y_{n-1}) - y_{n-1}| + [Y_{n-1}g(Y_{n-1})] \\
&\quad + L_f |y'|_\infty^n (\Delta t)^2.
\end{aligned}$$

□

Now (2.20) holds for  $n = 1, 2, 3, \dots$  with the desired bound on the previous time step appearing on the right hand side. Using a discrete Gronwall argument, again, we show several results from the accumulation of these bounds. To make sense of

these global bounds, we state each in succession and then show the proof as stages in a single derivation.

**Theorem 2.2.2** *Global Bound - Stage 1*

$$\begin{aligned}
|E(Y_n) - y_n| \leq & (1 + L_f \Delta t)^{n-1} \int_{I_1} |E(Yg(Y)) - E(Y)g(E(Y))| dt \\
& + (1 + L_f \Delta t)^{n-2} \int_{I_2} |E(Yg(Y)) - E(Y)g(E(Y))| dt \\
& + \dots + \int_{I_n} |E(Yg(Y)) - E(Y)g(E(Y))| dt \\
& + (1 + L_f \Delta t)^{n-1} L_f |y'|_{\infty}^1 (\Delta t)^2 + (1 + L_f \Delta t)^{n-2} L_f |y'|_{\infty}^2 (\Delta t)^2 \\
& + \dots + L_f |y'|_{\infty}^n (\Delta t)^2,
\end{aligned} \tag{2.23}$$

where we  $|E(Yg(Y)) - E(Y)g(E(Y))|$  is piecewise constant on each subinterval  $I_1, I_2, \dots, I_n$ .

This bounds the difference between the expected value of the discrete stochastic model and the solution of the differential equation in a way that shows the potential effect of early differences, which accumulate over time. In general, this formulation is the closest equivalent to the true bound of global errors as it allows the difference to vary widely on some intervals though not on every one. Also note that this bound is almost identical to the global bound on a single time step. See Fig. 2.10 and Fig. 2.9. In this bound, we have simply written the accumulation of the bound term by term. In the following bounds, we simplify but also increase the bound by considering the maximum difference attained over each subinterval and assume it holds for all subintervals.

The weighting of (2.23) places a higher penalty on initial nonlinearities since poor approximations at early intervals may continue to have an effect at every interval afterward. As discussed previously, these nonlinearities correspond to the variance of the discrete probability model. A large initial variance causes simulations to immediately diverge from the expected value and behave similarly to other trajectories of

the ODE. Recall the effect of a range of initial values on the behavior of the logistic model, as shown in Fig. 1.3. (This result is similar to the effect of approximating an ODE solution with the forward Euler numerical method. In the case of exponential growth, approximations that fail to capture the nonlinearity of the initial trajectory fall to lower trajectories and track that behavior.)

Note that in the exponential growth example, we have only the accumulation of growth from the previous term as the nonlinearity terms are zero. Since the ODE is exponentially growing, however, the bound is still relatively large.

In the logistic case, the difference  $[Yg(Y)]$  is on the order of  $10^{-7}$  with a small increase over the initial stage of nonlinearity and increased variance. This is orders of magnitude smaller than the constant multipliers  $(1 + L_f\Delta t)^k$ , for the appropriate value of  $k$ , so it is difficult to see the direct effect of the nonlinear behavior. However, it is the accumulation of these differences that give the exponential growth we see in Fig. 2.10.

**Theorem 2.2.3** *Global Bound - Stage 2*

$$|E(Y_n) - y_n| \leq e^{L_f T} \int_0^T |E(Yg(Y)) - E(Y)g(E(Y))| dt + e^{L_f T} |y'|_{\infty} \Delta t \quad (2.24)$$

Assuming maximum nonlinearity and growth on each subinterval, we can see at most exponential behavior in the difference between the expected value of the discrete stochastic model and the solution of the differential equation.

In the linear case, the nonlinearity term is again zero, but we do obtain an exponential upper bound that is  $O(\Delta t)$  for short time  $T$  and finite  $|y'|_{\infty}$ .

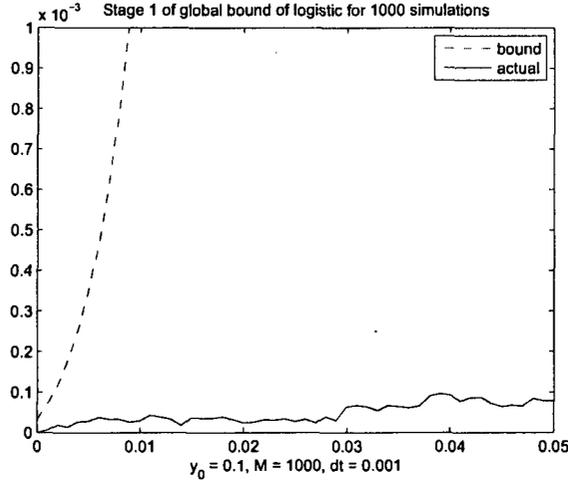


Figure 2.10: This plot shows the bound of Theorem 2.2.2 and the true difference  $|E(Y_n) - y_n|$  for the logistic ODE.

We plot this bound for the logistic case in Fig. 2.11. By using the maximum differences and growth on each interval, we have increased the bound by over three orders of magnitude from Stage 1.

Alternatively, we may consider the nonlinear effects of the model in terms of an average value, given in the following bound.

**Theorem 2.2.4** *Global Bound - Stage 3*

$$|E(Y_n) - y_n| \leq \frac{e^{L_f T}}{L_f} \max_j \frac{1}{\Delta t} \int_{I_j} |E(Yg(Y)) - E(Y)g(E(Y))| dt + e^{L_f T} |y'|_{\infty} \Delta t \quad (2.25)$$

or, equivalently,

$$|E(Y_n) - y_n| \leq \frac{1}{L_f \Delta t} e^{L_f T} \max_j [Y_{j-1} g(Y_{j-1})] + e^{L_f T} |y'|_{\infty} \Delta t \quad (2.26)$$

Writing the bound this way shows the effect of the nonlinearity in terms of an average value on each subinterval. Recall from the local analysis that these terms are

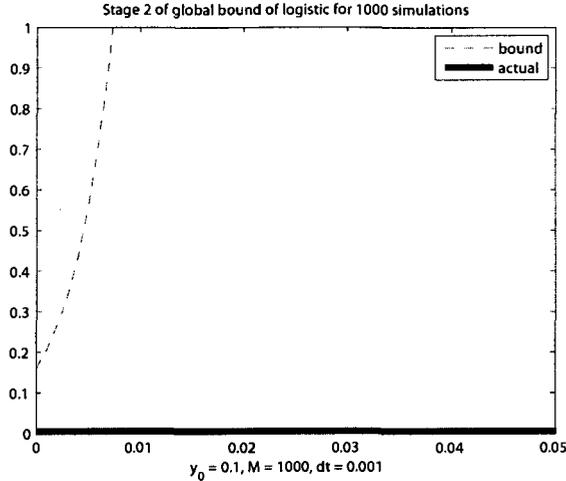


Figure 2.11: This plot shows the bound of Theorem 2.2.3 and the difference  $|E(Y_n) - y_n|$ , which is now four orders of magnitude smaller than the bound, for the logistic model.

bounded by the variance of the stochastic model. Taking only the maximum average value over the subintervals instead of the sum as in Theorem 2.2.3, the bound and the true difference are again of the same order of magnitude for a short time. See Fig. 2.12.

Note that in the linear case,  $E(Y_{n-1}g(Y_{n-1}))$  is exactly equal to  $E(Y_{n-1})g(E(Y_{n-1}))$ , so the term  $[Y_{j-1}g(Y_{j-1})] = |E(Y_{n-1}g(Y_{n-1})) - E(Y_{n-1})g(E(Y_{n-1}))|$  is zero and the bound of Theorem 2.2.4 is just  $O(\Delta t)$  for finite  $T$ . This follows our intuition and matches the results obtained in [28].

We show the bound for the logistic case in Fig. 2.12. Since the nonlinear  $[Yg(Y)]$  terms are relatively small and not being accumulated as a function of the previous time step, the true difference of the logistic is also  $O(\Delta t)$ .

### Proof of Stages 1 and 2

For the first accumulated global result of Theorem 2.2.2, we recall Theorem 2.2.1

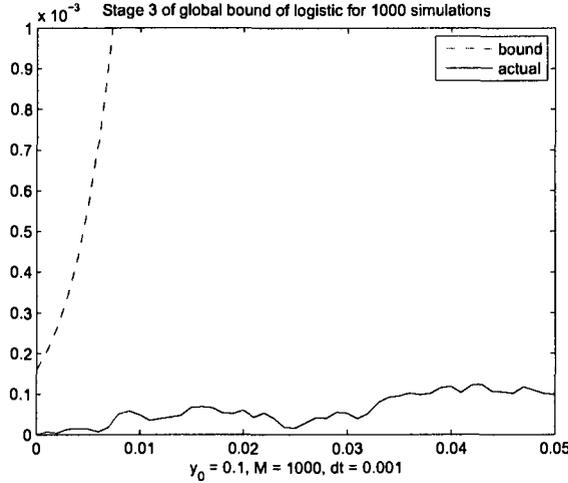


Figure 2.12: This plot shows the bound of Theorem 2.2.4 and the difference  $|E(Y_n) - y_n|$  for the logistic model.

in integral form:

$$|E(Y_n) - y_n| \leq (1 + L_f \Delta t) |E(Y_{n-1}) - y_{n-1}| + \int_{t_{n-1}}^{t_n} |E(Yg(Y)) - E(Y)g(E(Y))| dt + L_f |y'|_{\infty}^n (\Delta t)^2$$

We make the first substitution at the previous time step,  $|E(Y_{n-1}) - y_{n-1}|$ , into

(2.20). This yields

$$\begin{aligned} |E(Y_n) - y_n| \leq & (1 + L_f \Delta t)^2 |E(Y_{n-2}) - y_{n-2}| \\ & + (1 + L_f \Delta t) \int_{I_{n-1}} |E(Yg(Y)) - E(Y)g(E(Y))| dt \\ & + \int_{I_n} |E(Yg(Y)) - E(Y)g(E(Y))| dt \\ & + (1 + L_f \Delta t) L_f |y'|_{\infty}^{n-1} (\Delta t)^2 + L_f |y'|_{\infty}^n (\Delta t)^2. \end{aligned} \quad (2.27)$$

Continuing in this manner gives (2.23) of Theorem 2.2.2:

$$\begin{aligned} |E(Y_n) - y_n| \leq & (1 + L_f \Delta t)^{n-1} \int_{I_1} |E(Yg(Y)) - E(Y)g(E(Y))| dt \\ & + (1 + L_f \Delta t)^{n-2} \int_{I_2} |E(Yg(Y)) - E(Y)g(E(Y))| dt \\ & + \dots + (1 + L_f \Delta t) \int_{I_{n-1}} |E(Yg(Y)) - E(Y)g(E(Y))| dt \\ & + \int_{I_n} |E(Yg(Y)) - E(Y)g(E(Y))| dt \\ & + (1 + L_f \Delta t)^{n-1} L_f |y'|_{\infty}^1 (\Delta t)^2 + (1 + L_f \Delta t)^{n-2} L_f |y'|_{\infty}^2 (\Delta t)^2 \\ & + \dots + (1 + L_f \Delta t) L_f |y'|_{\infty}^{n-1} (\Delta t)^2 + L_f |y'|_{\infty}^n (\Delta t)^2 \end{aligned}$$

since the last term  $|E(Y_0) - y_0|$  is zero.

For the second global bound, we note that the coefficient in each term of the above equation is less than or equal to  $(1 + L_f \Delta t)^{n-1}$ . Similarly, the infinity norm on each interval is less than or equal to the infinity norm over the entire interval. This gives the simplification

$$|E(Y_n) - y_n| \leq (1 + L_f \Delta t)^{n-1} \int_0^T |E(Yg(Y)) - E(Y)g(E(Y))| dt + L_f |y'|_\infty (\Delta t)^2 \sum_{i=0}^{n-1} (1 + L_f \Delta t)^i. \quad (2.28)$$

The sum of coefficients from the last  $n$  terms form a geometric series of  $(1 + L_f \Delta t)$ .

Using the formula

$$1 + x + \dots + x^n = \frac{1 - x^{n+1}}{1 - x},$$

this gives

$$((1 + L_f \Delta t)^{n-1} + \dots + 1) L_f |y'|_\infty (\Delta t)^2 = \frac{1 - (1 + L_f \Delta t)^n}{1 - (1 + L_f \Delta t)} L_f |y'|_\infty (\Delta t)^2.$$

We use the exponential approximation for  $L_f \Delta t < 1$  given in Appendix A to obtain

$$RHS \leq \frac{1 - e^{L_f T}}{-L_f \Delta t} L_f |y'|_\infty (\Delta t)^2 = (e^{L_f T} - 1) |y'|_\infty \Delta t$$

or

$$RHS \leq e^{L_f T} |y'|_\infty \Delta t,$$

where  $n\Delta t = T$ . We keep this term in our bound, but note that since  $y'$  is bounded, this term goes to zero as  $\Delta t$  goes to zero, for finite  $T$ .

Using this approximation on the first term of (2.28), we achieve the bound (2.24) of Theorem 2.2.3:

$$|E(Y_n) - y_n| \leq e^{L_f T} \int_0^T |E(Yg(Y)) - E(Y)g(E(Y))| dt + e^{L_f T} |y'|_\infty \Delta t.$$

□

### Proof of Stage 3

Again, we recall Equation (2.23) of Theorem 2.2.2 combined with the exponential bound of the sum of  $L_f|y'|_\infty\Delta t$  terms:

$$\begin{aligned} |E(Y_n) - y_n| \leq & (1 + L_f\Delta t)^{n-1} \int_{I_1} |E(Yg(Y)) - E(Y)g(E(Y))|dt \\ & + (1 + L_f\Delta t)^{n-2} \int_{I_2} |E(Yg(Y)) - E(Y)g(E(Y))|dt \\ & + \dots + (1 + L_f\Delta t) \int_{I_{n-1}} |E(Yg(Y)) - E(Y)g(E(Y))|dt \\ & + \int_{I_n} |E(Yg(Y)) - E(Y)g(E(Y))|dt + e^{L_f T} |y'|_\infty \Delta t. \end{aligned}$$

Bounding each of the integrals by the maximum, we have

$$\begin{aligned} |E(Y_n) - y_n| \leq & (1 + L_f\Delta t)^{n-1} \max_j \int_{I_j} |E(Yg(Y)) - E(Y)g(E(Y))|dt \\ & + (1 + L_f\Delta t)^{n-2} \max_j \int_{I_j} |E(Yg(Y)) - E(Y)g(E(Y))|dt \\ & + \dots + (1 + L_f\Delta t) \max_j \int_{I_j} |E(Yg(Y)) - E(Y)g(E(Y))|dt \\ & + \max_j \int_{I_j} |E(Yg(Y)) - E(Y)g(E(Y))|dt + e^{L_f T} |y'|_\infty \Delta t. \end{aligned}$$

Again using the sum of the geometric series of  $(1 + L_f\Delta t)$  and assuming  $L_f\Delta t < 1$ , we have the bound

$$|E(Y_n) - y_n| \leq \frac{e^{L_f T}}{L_f\Delta t} \max_j \int_{I_j} |E(Yg(Y)) - E(Y)g(E(Y))|dt + e^{L_f T} |y'|_\infty \Delta t.$$

Rewriting this gives the bound in terms of the maximum average value of  $|E(Yg(Y)) - E(Y)g(E(Y))|$  over each interval

$$|E(Y_n) - y_n| \leq \frac{e^{L_f T}}{L_f} \max_j \frac{1}{\Delta t} \int_{I_j} |E(Yg(Y)) - E(Y)g(E(Y))|dt + e^{L_f T} |y'|_\infty \Delta t.$$

Simply noting that  $|E(Yg(Y)) - E(Y)g(E(Y))|$  is constant over each interval gives Equation (2.26).

□

## Chapter 3

# Extension to Higher Dimensions

We now consider the model of a vector of populations  $\mathbf{y}$ , where each component represents a particular class of the total population, e.g. Kermack-McKendrick models. After understanding the formulation of the scalar discrete probability model, we find the analysis for systems proceeds in a straight forward way if we consider component-wise probabilities.

As before, the probabilities for the discrete stochastic model are determined from an ODE model,

$$\begin{cases} \dot{\mathbf{y}} = f(\mathbf{y}) \\ \mathbf{y}(0) = \mathbf{y}_0 \end{cases}, \quad (3.1)$$

where  $\mathbf{y} = (y^1, y^2, \dots, y^m)^T \in \mathbb{R}^m$ . We define the solution of the  $k^{\text{th}}$  dimension at time  $t_n$  as  $y_n^k$  to keep notation as consistent as possible with the scalar case.

Recall that for the formulation in one dimension, the assumption that  $f$  has the form  $f(y) = yg(y)$ , where  $g$  is interpreted as a probability, is essential. In higher dimensions, the analogous assumption is

$$f(\mathbf{y}) = G(\mathbf{y})\mathbf{y},$$

where  $G$  is a diagonal  $m \times m$  matrix with entries  $g_{ii}(\mathbf{y}) = g_{ii}^+(\mathbf{y}) - g_{ii}^-(\mathbf{y})$  for  $g_{ii}^+(\mathbf{y}), g_{ii}^-(\mathbf{y}) >$

0. The diagonal form of  $G$  is necessary if we assume component-wise event probabilities as we discuss shortly. Note that this assumption implies the homogeneity condition wherein

$$f^k(y^1, \dots, y^{k-1}, 0, y^{k+1}, \dots, y^m) = 0 \quad \forall y^1, \dots, y^{k-1}, y^{k+1}, \dots, y^m \quad \forall k = 1, 2, \dots, m.$$

Before defining the associated probability model, we state a few more assumptions.

### 3.0.1 Assumptions for the System Model

We assume the time step  $\Delta t$  is small enough so that only one event in each population may occur on the interval  $(t, t + \Delta t]$ . Note that there are other possible ways to consider events for a system model. For example, a net zero change in the population over a time interval of length  $\Delta t$ , may result if no event occurs *or* if more than one event are allowed to occur in the same time interval and the same number of births and deaths occur in  $[t_{j-1}, t_j]$ .

We assume Lipschitz continuity, where

$$\|f(\mathbf{x}) - f(\mathbf{y})\| \leq L_f \|\mathbf{x} - \mathbf{y}\|, \tag{3.2}$$

for some constant  $L_f > 0$ ,  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ . For the remainder of the thesis, it is understood that  $\|\cdot\|$  refers the  $\ell^2$  norm and  $\|\cdot\|_\infty$  is the  $\ell^\infty$  norm. We also assume each  $g_{ii}$  is Lipschitz continuous with constant  $L_{g_{ii}}$ ,  $i = 1, 2, \dots, m$ .

In our analysis, we also need  $g_{ii}(\mathbf{y})$  to be finite over the domain of  $f$ . If this is not the case, we impose an artificial bound similar to that described in the scalar case. If any of these probabilities become larger than 1, the probability model is no longer valid beyond that point.

### 3.1 The System Probability Model

We first consider the conditional probability of an event (birth or death) occurring in the  $k^{\text{th}}$  class for a single individual on the small time interval  $(t, t + \Delta t]$ . By the assumption that  $G$  is diagonal, the probability of an increase in  $Y^k$  is given by

$$P(Y^k(t + \Delta t) = Y_{n-1}^k + 1 | Y^k(t) = Y_{n-1}^k) = g_{kk}^+(\mathbf{Y}_{n-1})\Delta t.$$

Similarly, we have the probability of a death or no event occurring for an individual in the  $k^{\text{th}}$  class in a small time interval  $(t, t + \Delta t]$ :

$$P(Y^k(t + \Delta t) = Y_{n-1}^k - 1 | Y^k(t) = Y_{n-1}^k) = g_{kk}^-(\mathbf{Y}_{n-1})\Delta t,$$

$$P(Y^k(t + \Delta t) = Y_{n-1}^k | Y^k(t) = Y_{n-1}^k) = 1 - (g_{kk}^+(\mathbf{Y}_{n-1}) + g_{kk}^-(\mathbf{Y}_{n-1}))\Delta t.$$

Each of the individuals in a particular class,  $Y^k$ ,  $k = 1, 2, \dots, m$  are categorical random variables with three event probabilities, as described above. As explained in the scalar case, the categorical distribution is a generalization of the Bernoulli random variable which has only two possible outcomes. An extension of the binomial theorem, we have the following probability for 1 birth, 0 deaths, and for the remaining  $Y_{n-1}^k - 1$  individuals, neither event occurs.

$$\begin{aligned} P(Y^k(t + \Delta t) = Y_{n-1}^k + 1 | Y^k(t) = Y_{n-1}^k) &= \frac{Y_{n-1}^k}{1!0!(Y_{n-1}^k - 1)!} (g_{kk}^+ \Delta t)^1 (g_{kk}^- \Delta t)^0 (1 - (g_{kk}^+ + g_{kk}^-)\Delta t)^{Y_{n-1}^k - 1 - 0} \\ &= Y_{n-1}^k (g_{kk}^+ \Delta t) (1 - (g_{kk}^+ + g_{kk}^-)\Delta t)^{Y_{n-1}^k - 1} \\ &= Y_{n-1}^k (g_{kk}^+(\mathbf{Y}_{n-1})\Delta t) + o(\Delta t), \end{aligned}$$

assuming finite  $g_{kk}^+(\mathbf{Y}_{n-1})$  and  $g_{kk}^-(\mathbf{Y}_{n-1})$ .

Note that the diagonal form of  $G$ , in particular, motivates this definition of a population of categorical random variables. The  $k^{\text{th}}$  row gives the probability for

an event to occur in the  $k^{th}$  class, but it is unclear how we choose an individual for which that event occurred. Componentwise probabilities make it possible to choose an individual specifically from the  $Y^k$  population where the change is given by a vector of zeros with a 1 in the  $k^{th}$  component. A non-diagonal probability matrix  $G$  means we have to multiply the row of probabilities by the vector  $\mathbf{Y}$ . This process, though, somehow indicates an increase in the population as an increase of +1 for each component in the entire vector. What exactly this means is unclear for a birth-death process.

Additionally, we must be wary of this impact on the assumption that only one event may occur on the small time interval of length  $\Delta t$ . If a change of +1 in the population is described by a *vector* of ones, we can no longer claim the probability of more than one event occurring to be  $o(\Delta t)$ .

For a system of ODEs with non-diagonal matrix  $G$  of probabilities, it is unclear that a birth-death process of categorical random variables is the correct discrete probability model to use. Some other stochastic model may be better suited to this type of system ODE.

Assuming diagonal  $G$  and categorical random variables, we have the following probabilities for one death or no events for the entire  $Y^k$  population,  $k = 1, 2, \dots, m$ ,

$$P(Y^k(t + \Delta t) = Y_{n-1}^k - 1 | Y^k(t) = Y_{n-1}^k) = Y_{n-1}^k g_{kk}^-(\mathbf{Y}_{n-1}) \Delta t + o(\Delta t)$$

and

$$P(Y^k(t + \Delta t) = Y_{n-1}^k | Y^k(t) = Y_{n-1}^k) = 1 - Y_{n-1}^k (g_{kk}^+(\mathbf{Y}_{n-1}) + g_{kk}^-(\mathbf{Y}_{n-1})) \Delta t + o(\Delta t)$$

These event probabilities hold for each  $k = 1, 2, \dots, m$ . We define the vector of event probabilities where  $\mathbf{o}(\Delta t)$  is an  $m \times 1$  vector with each entry  $o(\Delta t)$ :

$$P(\mathbf{Y}(t + \Delta t) = \mathbf{Y}_{n-1} + 1 | \mathbf{Y}(t) = \mathbf{Y}_{n-1}) = G^+(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}\Delta t + \mathbf{o}(\Delta t)$$

$$P(\mathbf{Y}(t + \Delta t) = \mathbf{Y}_{n-1} - 1 | \mathbf{Y}(t) = \mathbf{Y}_{n-1}) = G^-(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}\Delta t + \mathbf{o}(\Delta t)$$

$$P(\mathbf{Y}(t + \Delta t) = \mathbf{Y}_{n-1} | \mathbf{Y}(t) = \mathbf{Y}_{n-1}) = I - (G^+(\mathbf{Y}_{n-1}) + G^-(\mathbf{Y}_{n-1}))\mathbf{Y}_{n-1}\Delta t + \mathbf{o}(\Delta t)$$

### 3.2 Expected Value

With this formulation and using component-wise probabilities, we follow the scalar model formulation and determine the expected value of the change (of +1) in each of the  $m$  populations over a small time interval  $(t, t + \Delta t]$ . This is then used to define the expected value of the population at time  $t_n$ . We first consider the  $k^{th}$  class then extend to the full vector  $\mathbf{Y}$ . Also recall that expectation is component-wise for total and conditional expectation of vectors. e.g.  $E(\mathbf{Y}) = (E(Y^1), E(Y^2), \dots, E(Y^m))^T$ .

$$E(Y_n^k - Y_{n-1}^k | \mathbf{Y}_{n-1}) = E \left( \sum_{i=1}^{Y_{n-1}^k} (+1) \text{Cat}(g_{kk}^+ \Delta t) + \sum_{i=1}^{Y_{n-1}^k} (-1) \text{Cat}(g_{kk}^- \Delta t) \right)$$

The terms in each sum are independent of  $i$ , so we remove the sum and multiply by  $Y_{n-1}^k$ .

$$E(Y_n^k - Y_{n-1}^k | \mathbf{Y}_{n-1}) = E(Y_{n-1}^k (\text{Cat}(g_{kk}^+ \Delta t) - \text{Cat}(g_{kk}^- \Delta t)))$$

The conditional value  $Y_{n-1}^k$  is treated as a constant that we can pull outside the expected value.

$$E(Y_n^k - Y_{n-1}^k | \mathbf{Y}_{n-1}) = Y_{n-1}^k E(\text{Cat}(g_{kk}^+(\mathbf{Y}_{n-1})\Delta t) - \text{Cat}(g_{kk}^-(\mathbf{Y}_{n-1})\Delta t))$$

Linearity of expectation and using the fact that the expected value of a categorical variable is its probability of success, we have

$$E(Y_n^k - Y_{n-1}^k | \mathbf{Y}_{n-1}) = Y_{n-1}^k (g_{kk}^+(\mathbf{Y}_{n-1})\Delta t - g_{kk}^-(\mathbf{Y}_{n-1})\Delta t)$$

or

$$E(Y_n^k - Y_{n-1}^k | \mathbf{Y}_{n-1}) = Y_{n-1}^k g_{kk}(\mathbf{Y}_{n-1})\Delta t.$$

Solving for  $E(Y_n^k | \mathbf{Y}_{n-1})$  gives

$$E(Y_n^k | \mathbf{Y}_{n-1}) = E(Y_{n-1}^k | \mathbf{Y}_{n-1}) + Y_{n-1}^k g_{kk}(\mathbf{Y}_{n-1})\Delta t$$

or

$$E(Y_n^k | \mathbf{Y}_{n-1}) = E(Y_{n-1}^k) + Y_{n-1}^k g_{kk}(\mathbf{Y}_{n-1})\Delta t.$$

To remove the condition on  $\mathbf{Y}_{n-1}$ , we now treat  $\mathbf{Y}_{n-1}$  as a random variable (abusing notation) and use the law of total expectation, taking the expected value again.

$$E(Y_n^k) = E(Y_{n-1}^k) + E(Y_{n-1}^k g_{kk}(\mathbf{Y}_{n-1})) \Delta t \quad (3.3)$$

This expectation holds for each component of  $\mathbf{Y}$ ,  $k = 1, 2, \dots, m$ , giving the matrix equation

$$E(\mathbf{Y}_n) = E(\mathbf{Y}_{n-1}) + E(G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}) \Delta t. \quad (3.4)$$

## 3.3 System Examples

### 3.3.1 SI Model of Fox Rabies

In the following example, we consider an SI model for fox rabies, originally described by R. M. Anderson, et. al. [1]. The differential equation describing this system is given in (3.5) and may also be used to describe a predator-prey interaction such as the Lotka-Volterra model. The motivation of parameters is slightly different for a predator-prey model, but we omit this example and consider analysis of the fox rabies example and pneumonic plague. Fox rabies dynamics are described by

$$\begin{cases} \frac{dS}{dt} = -\beta SI + rS - \mu S \\ \frac{dI}{dt} = \beta SI - (\mu + \alpha)I \\ S(0) = S_0 \quad I(0) = I_0, \end{cases} \quad (3.5)$$

where  $\beta$  is the contact parameter between susceptible and infected foxes,  $r$  is the birth rate,  $\mu$  is the natural death rate, and  $\alpha$  is the death rate due to the disease with  $1/\alpha$  the mean infective period.

Rabies is a viral infection that affects many animals, including bats, foxes, coyotes, wolves, and rats. Human infection is rare, but is usually fatal in both humans and some animals, including foxes but excluding bats, which are probably the ancestral host for this disease. The above model is specific to foxes, so we assume no recovered class. Because of the relatively quick death rate (high  $\alpha$ ), there is assumed to be no birth rate associated with infected individuals. In 1981, Anderson modeled the epidemic of fox rabies which spread throughout Europe in 1939 [1]. In [8], F. Brauer and C. Castillo-Chavez simplified this model to the system of ODEs given above. We show the details of this probability model below, given that the total population at time  $t$  is the sum  $N(t) = S(t) + I(t)$ .

In matrix notation,  $f(\mathbf{y}) = G(\mathbf{y})\mathbf{y}$ , we let  $y_1(t) = S(t)$  and  $y_2(t) = I(t)$  and define the matrix  $G$  by

$$G = \begin{bmatrix} -\beta I + r - \mu & 0 \\ 0 & \beta S - (\mu + \alpha) \end{bmatrix}.$$

This gives the following probabilities for the stochastic model:

$$P((S(t + \Delta t), I(t + \Delta t) = (S_{n-1} + 1, I_{n-1}) | (S(t), I(t))) = rS\Delta t$$

$$P((S(t + \Delta t), I(t + \Delta t) = (S_{n-1} - 1, I_{n-1}) | (S(t), I(t))) = (\beta I + \mu)S\Delta t$$

$$P((S(t + \Delta t), I(t + \Delta t) = (S_{n-1}, I_{n-1} + 1) | (S(t), I(t))) = \beta SI\Delta t$$

$$P((S(t + \Delta t), I(t + \Delta t) = (S_{n-1}, I_{n-1} - 1) | (S(t), I(t))) = (\mu + \alpha)I\Delta t$$

and

$$\begin{aligned} P((S(t + \Delta t), I(t + \Delta t) = (S_{n-1}, I_{n-1}) | (S(t), I(t))) \\ = 1 - ((r + \beta I + \mu)S + (\beta S + \mu + \alpha)I) \Delta t. \end{aligned}$$

Note that the parameter values in (3.5) are given in terms of total numbers of population, so we do not need to rescale the model. The result of this, however, is that we must use a time step at least as small as  $1 \times 10^{-6}$  to maintain a total probability less than one. In Fig. 3.1, we plot the stochastic simulations, the expected value, and the solution to (3.5) using parameter values from the epidemic of 1939,  $\beta = 80$ ,  $r = 1.0$ ,  $\mu = 0.5$ ,  $\alpha = 73$  [1]. We use initial conditions to simulate the potential for an epidemic given one infected individual in a wholly susceptible population ( $R_0$ ):  $S(0) = 50$  and  $I(0) = 1$ . According to the  $R_0$  formula given in [8],  $R_0 = \beta S / (\mu + \alpha)$ , these parameters and initial conditions give  $R_0 > 1$  and should result in the epidemic that we see in Fig. 3.1.

Within the first few time steps, most of the susceptible population has become extinct while the infection persists in the rest of the population. For a few simulations, though, the infection immediately dies out and the susceptible class remains

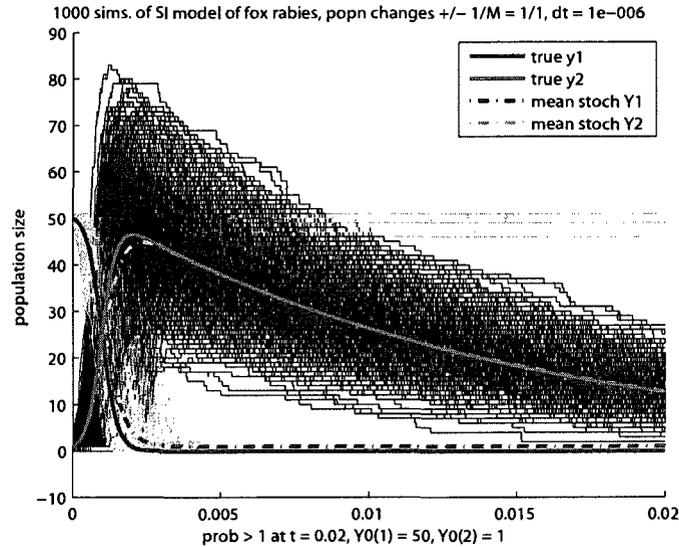


Figure 3.1: Simulations for fox rabies with  $\Delta t = 1 \times 10^{-6}$  and  $S(0) = 50$ ,  $I(0) = 1$ .

relatively stable with zero probability of infection. This gives an expected value for the susceptible population just above zero. As discussed in Section 2.1, the discrete probability model is no longer valid once the distance between a simulation and a steady state becomes less than one increment of the discrete probability model. In this case, a distance less than 1 between simulations and the steady state  $y^* = 0$  somehow implies that less than one individual has survived.

### 3.3.2 SI Model of pneumonic plague

Also from [8], we now consider a simplified Kermack-McKendrick model of the pneumonic form of the bubonic plague. The bubonic plague is a vector borne disease passed between rats via fleas. The pneumonic form, however, is passed via direct contact and is modeled by the system below. In this model, we assume an infection rate  $\beta$ , equal natural birth and death rates in the absence of the disease, and death

rate  $\gamma$  specific to the disease.

$$\begin{cases} \frac{dS}{dt} = -\beta SI \\ \frac{dI}{dt} = (\beta S - \gamma)I \end{cases} \quad S(0) = S_0 \quad I(0) = I_0 \quad (3.6)$$

For parameter values consistent with the data from the Eyam plague epidemic in 1665 [8], we plot 1000 simulations in Fig. 3.2. The vertical line gives the separation between short and long time behavior after which the expected value of the infected population is less than one individual and the two models should not be compared.

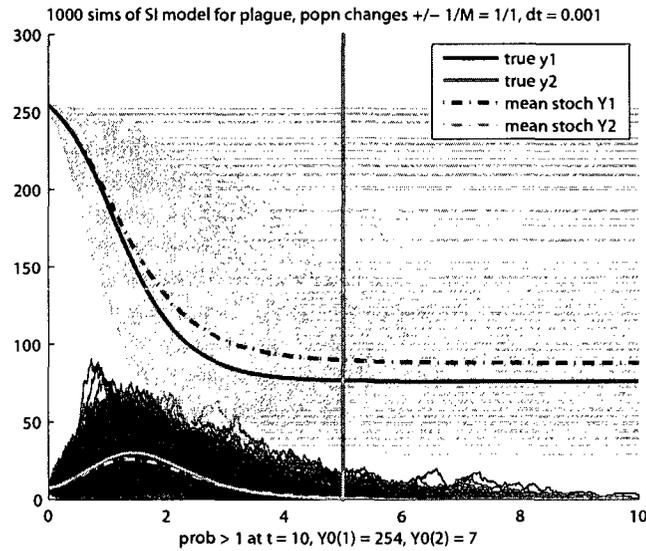


Figure 3.2: Simulations for bubonic plague using parameter values  $\beta = 0.0178$ ,  $\gamma = 2.73$ ,  $S(0) = 254$ ,  $I(0) = 7$ . The vertical line indicates the separation between short and long time behavior.

As with the fox rabies example, the infection dies out immediately for some simulations. As a result, the corresponding susceptible population cannot become infected and is only subject to natural birth and death events. The difference between the expected value and deterministic solution of the susceptible population, in particular, is large due to this increasing variance among simulations. Again, this potential

difference is reflected by the moment and variance bounds proven in Chapter 4.

# Chapter 4

## System Difference Bounds

### 4.1 Local Difference Bounds

As with the scalar formulation, we are interested in the local and global difference bounds between the expected value of the stochastic model and the solution to the associated differential equation. These bounds are quite similar to the scalar case, but require norms on the differences. While the proofs are similar, we give most of the details for clarification and completeness.

Beginning with local bounds for a system of ODEs, we first define the set of local solutions,  $\tilde{\mathbf{y}}^j$ , on each time interval  $(t_{j-1}, t_j]$ ,  $j = 1, 2, \dots, n$ . Following the same notational conventions as in the scalar case,  $\tilde{\mathbf{y}}^j$  solves the system,

$$\begin{cases} \dot{\tilde{\mathbf{y}}}^j = G(\tilde{\mathbf{y}}^j)\tilde{\mathbf{y}}^j = G^+(\tilde{\mathbf{y}}^j) - G^-(\tilde{\mathbf{y}}^j) & [t_{j-1}, t_j] \\ \tilde{\mathbf{y}}_{j-1}^j = E(\mathbf{Y}_{j-1}), & j = 1, 2, \dots, n, \end{cases} \quad (4.1)$$

where  $G^+(\tilde{\mathbf{y}}^j), G^-(\tilde{\mathbf{y}}^j) > 0$  and  $G$  is a diagonal matrix in  $\mathbb{R}^{m \times m}$ .

We utilize the component-wise attributes of our system and the expected value to consider the  $k^{\text{th}}$  dimension in our analysis. Since  $G$  is diagonal, we may also describe the deterministic solution as a vector of integrals for each  $k = 1, 2, \dots, m$ . In the description of this vector of integrals, we abuse notation and describe, for any

$\mathbf{x} \in \mathbb{R}^m$ ,

$$\int_{t_{n-1}}^{t_n} \mathbf{x} dt = \left( \int_{t_{n-1}}^{t_n} x^1 dt, \int_{t_{n-1}}^{t_n} x^2 dt, \dots, \int_{t_{n-1}}^{t_n} x^m dt \right).$$

From Section 3.2, the total expected value of  $\mathbf{Y}_n$  in integral form is given by

$$E(\mathbf{Y}_n) = E(\mathbf{Y}_{n-1}) + \int_{t_{n-1}}^{t_n} E(G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}) dt. \quad (4.2)$$

The deterministic solution for  $\tilde{\mathbf{y}}^j$  of (4.1) at  $t_j$  for  $j = n$  is given by

$$\tilde{\mathbf{y}}_n^n = \tilde{\mathbf{y}}_{n-1}^n + \int_{t_{n-1}}^{t_n} G(\tilde{\mathbf{y}}_n)\tilde{\mathbf{y}}_n dt = E(\mathbf{Y}_{n-1}) + \int_{t_{n-1}}^{t_n} G(\tilde{\mathbf{y}}_n)\tilde{\mathbf{y}}_n dt. \quad (4.3)$$

**Theorem 4.1.1** *Local Bound on One Time Step*

*The difference on each time step  $[t_{n-1}, t_n]$  is given by*

$$\|E(\mathbf{Y}_n) - \tilde{\mathbf{y}}_n^n\| \leq I + II$$

with

$$I = [G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}] \equiv \|E(G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}) - G(E(\mathbf{Y}_{n-1}))E(\mathbf{Y}_{n-1})\| \Delta t, \quad (4.4)$$

and

$$II = L_f \Delta t \|E(\mathbf{Y}_{n-1}) - \tilde{\mathbf{y}}_{n-1}^{n-1}\| + L_f \int_{t_{n-1}}^{t_n} \|\tilde{\mathbf{y}}_n - \tilde{\mathbf{y}}_{n-1}^{n-1}\| dt,$$

where  $\tilde{\mathbf{y}}_0^0 \equiv E(\mathbf{Y}_0)$ .

In Fig. 4.1, we show the bound of Theorem 4.1.1 and the difference  $\|E(\mathbf{Y}_n) - \mathbf{y}_n\|$  on each time step. The true difference is orders of magnitude smaller than the bound, so it appears as a line at zero. We show a plot of  $\|E(\mathbf{Y}_n) - \mathbf{y}_n\|$  in Fig. 4.2.

**Proof of Theorem 4.1.1**

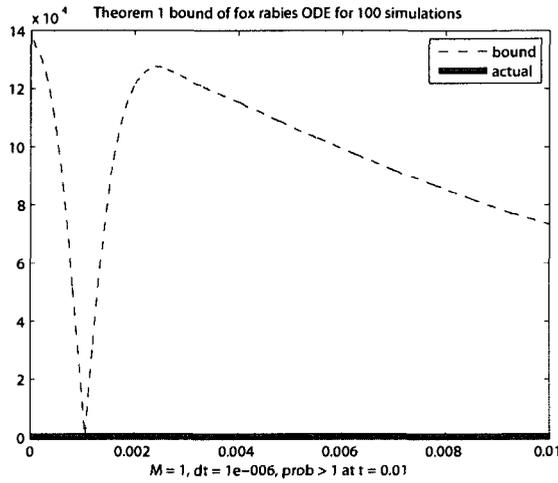


Figure 4.1: The plot gives the bound of Theorem 4.1.1 for the fox rabies example as well as the true difference of  $\|E(\mathbf{Y}_n) - \mathbf{y}_n\|$  at each time step.

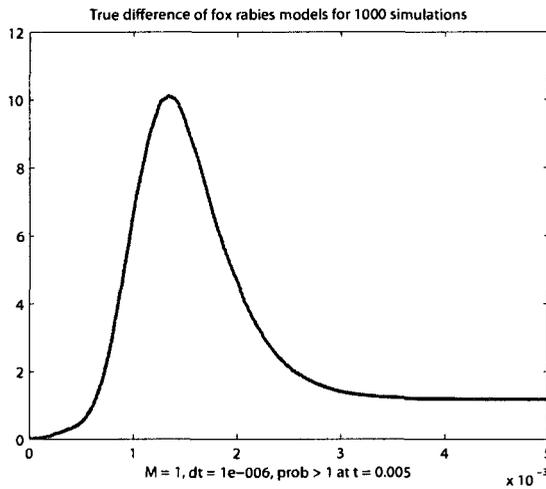


Figure 4.2: This plot zooms in on the true difference  $\|E(\mathbf{Y}_n) - \mathbf{y}_n\|$  between the discrete probability model and the ODE solution for the fox rabies at each time step.

$$\|E(\mathbf{Y}_n) - \tilde{\mathbf{y}}_n^n\| = \left\| \int_{t_{n-1}}^{t_n} (E(G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}) - G(\tilde{\mathbf{y}}_n)\tilde{\mathbf{y}}_n) dt \right\|$$

Adding and subtracting the integral  $\int_{t_{n-1}}^{t_n} G(E(\mathbf{Y}_{n-1}))E(\mathbf{Y}_{n-1})dt$  gives

$$\|E(\mathbf{Y}_n) - \tilde{\mathbf{y}}_n^n\| = \left\| \int_{t_{n-1}}^{t_n} (E(G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}) - G(E(\mathbf{Y}_{n-1}))E(\mathbf{Y}_{n-1})) dt + \int_{t_{n-1}}^{t_n} (G(E(\mathbf{Y}_{n-1}))E(\mathbf{Y}_{n-1}) - G(\tilde{\mathbf{y}}^n)\tilde{\mathbf{y}}^n) dt \right\|. \quad (4.5)$$

The first integral listed above consists of only constants, which simplifies to term I in the theorem. The second integral is bounded by the Lipschitz constant of  $f$  to give

$$\begin{aligned} \|E(\mathbf{Y}_n) - \tilde{\mathbf{y}}_n^n\| &\leq \|E(G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}) - G(E(\mathbf{Y}_{n-1}))E(\mathbf{Y}_{n-1})\| \Delta t \\ &\quad + L_f \int_{t_{n-1}}^{t_n} \|E(\mathbf{Y}_{n-1}) - \tilde{\mathbf{y}}^n\| dt \\ &= I + \overline{II}. \end{aligned} \quad (4.6)$$

We rewrite term  $\overline{II}$  by adding and subtracting the previous local integral of  $\tilde{\mathbf{y}}^{n-1}$  at time  $t_{n-1}$ .

$$\begin{aligned} \overline{II} &\leq L_f \int_{t_{n-1}}^{t_n} \|E(\mathbf{Y}_{n-1}) - \tilde{\mathbf{y}}_{n-1}^{n-1}\| dt + L_f \int_{t_{n-1}}^{t_n} \|\tilde{\mathbf{y}}^n - \tilde{\mathbf{y}}_{n-1}^{n-1}\| dt \\ &= L_f \Delta t \|E(\mathbf{Y}_{n-1}) - \tilde{\mathbf{y}}_{n-1}^{n-1}\| + L_f \int_{t_{n-1}}^{t_n} \|\tilde{\mathbf{y}}^n - \tilde{\mathbf{y}}_{n-1}^{n-1}\| dt \end{aligned}$$

Putting this back together with (4.6) gives the result:

$$\|E(\mathbf{Y}_n) - \tilde{\mathbf{y}}_n^n\| \leq [G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}] + L_f \Delta t \|E(\mathbf{Y}_{n-1}) - \tilde{\mathbf{y}}_{n-1}^{n-1}\| + L_f \int_{t_{n-1}}^{t_n} \|\tilde{\mathbf{y}}^n - \tilde{\mathbf{y}}_{n-1}^{n-1}\| dt. \quad (4.7)$$

Again, defining  $\tilde{\mathbf{y}}_0^0 \equiv E(\mathbf{Y}_0)$  implies this bound holds for  $n = 1, 2, 3, \dots$

□

Again, the assumption of  $f(\mathbf{y}) = G(\mathbf{y})\mathbf{y}$  is vital in the formulation of this proof. As with the scalar case, the difference between the expected value and the deterministic solution over a small time interval is due to a combination of non-linear effects given by term I, which we show is bounded by the variance of the stochastic model, and the

stability of the solution of the ODE on each interval based on two nearby trajectories of  $y$ , shown in term II of Theorem 4.1.1.

As before, since the bound at time  $t_n$  depends on the bound at the previous time step, we use Gronwall's inequality to define the accumulation of these local differences.

**Theorem 4.1.2** *Accumulated Local Bound*

For any time  $t_n \geq t_0$ ,

$$\begin{aligned} \|E(\mathbf{Y}_n) - \tilde{\mathbf{y}}_n^n\| \leq & \sum_{j=1}^n (L_f \Delta t)^{n-j} [G(\mathbf{Y}_{j-1}) \mathbf{Y}_{j-1}] \\ & + L_f \sum_{j=1}^n (L_f \Delta t)^{n-j} \int_{t_{j-1}}^{t_j} \|\tilde{\mathbf{y}}^j - \tilde{\mathbf{y}}_{j-1}^{j-1}\| dt. \end{aligned} \tag{4.8}$$

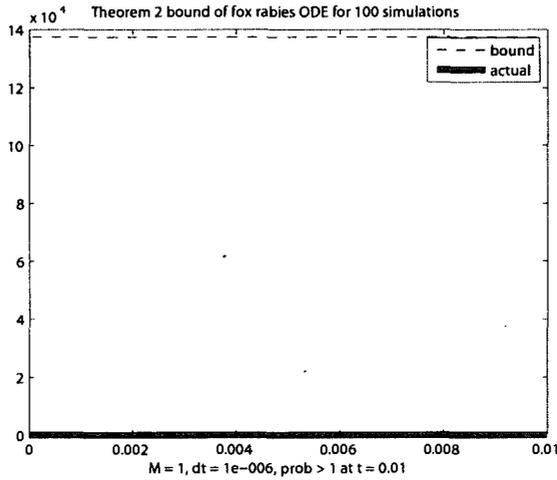


Figure 4.3: This plot gives the bound of Theorem 4.1.2 for the fox rabies example as well as the true difference of  $\|E(\mathbf{Y}_n) - \mathbf{y}_n\|$  at each time step.

**Proof of Theorem 4.1.2**

We begin with (4.7) and implement the first step in the Gronwall argument to obtain

$$\begin{aligned}
\|E(\mathbf{Y}_n) - \tilde{\mathbf{y}}_n^n\| &\leq [G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}] + L_f\Delta t\|E(\mathbf{Y}_{n-1}) - \tilde{\mathbf{y}}_{n-1}^{n-1}\| \\
&\quad + L_f\int_{t_{n-1}}^{t_n} \|\tilde{\mathbf{y}}^n - \tilde{\mathbf{y}}_{n-1}^{n-1}\|dt \\
&\leq [G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}] + L_f\int_{t_{n-1}}^{t_n} \|\tilde{\mathbf{y}}^n - \tilde{\mathbf{y}}_{n-1}^{n-1}\|dt \\
&\quad + L_f\Delta t[G(\mathbf{Y}_{n-2})\mathbf{Y}_{n-2}] \\
&\quad + L_f^2(\Delta t)^2\|E(\mathbf{Y}_{n-2}) - \tilde{\mathbf{y}}_{n-2}^{n-2}\| \\
&\quad + L_f^2\Delta t\int_{t_{n-2}}^{t_{n-1}} \|\tilde{\mathbf{y}}^{n-1} - \tilde{\mathbf{y}}_{n-2}^{n-2}\|dt
\end{aligned}$$

With another substitution of the difference at time  $t_{n-2}$ , we have

$$\begin{aligned}
\|E(\mathbf{Y}_n) - \tilde{\mathbf{y}}_n^n\| &\leq [G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}] + L_f\Delta t[G(\mathbf{Y}_{n-2})\mathbf{Y}_{n-2}] \\
&\quad + L_f^2(\Delta t)^2[G(\mathbf{Y}_{n-3})\mathbf{Y}_{n-3}] \\
&\quad + L_f\int_{t_{n-1}}^{t_n} \|\tilde{\mathbf{y}}^n - \tilde{\mathbf{y}}_{n-1}^{n-1}\|dt \\
&\quad + L_f^2\Delta t\int_{t_{n-2}}^{t_{n-1}} \|\tilde{\mathbf{y}}^{n-1} - \tilde{\mathbf{y}}_{n-2}^{n-2}\|dt \\
&\quad + L_f^3(\Delta t)^2\int_{t_{n-3}}^{t_{n-2}} \|\tilde{\mathbf{y}}^{n-2} - \tilde{\mathbf{y}}_{n-3}^{n-3}\|dt \\
&\quad + L_f^3(\Delta t)^3\|E(\mathbf{Y}_{n-3}) - \tilde{\mathbf{y}}_{n-3}^{n-3}\|.
\end{aligned}$$

Continuing in this manner gives

$$\begin{aligned}
\|E(\mathbf{Y}_n) - \tilde{\mathbf{y}}_n^n\| &\leq \sum_{j=1}^n [G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}](L_f\Delta t)^{n-j} \\
&\quad + L_f\sum_{j=1}^n \int_{t_{j-1}}^{t_j} \|\tilde{\mathbf{y}}^j - \tilde{\mathbf{y}}_{j-1}^{j-1}\|dt(L_f\Delta t)^{n-j},
\end{aligned} \tag{4.9}$$

where the last term is zero since  $\tilde{\mathbf{y}}_0^0 \equiv E(\mathbf{Y}_0)$ .

□

As in the scalar case, we next consider an expansion of the term  $[G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}]$  and observe how the behavior of the probability model effects the bound in terms of its variance and second moment.

**Theorem 4.1.3** *Local  $[G(\mathbf{Y})\mathbf{Y}]$  Bounded by Variance and Second Moment*

*The unbiased estimator for  $[G(\mathbf{Y})\mathbf{Y}]$  is bounded on each time step by*

$$\begin{aligned}
Est([G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}]) &\leq \left(\frac{N-1}{N}\right)^{1/2} L_g\Delta t \left(\frac{1}{N-1} \sum_{i=1}^N |\{\mathbf{Y}_{j-1}\}_i - E(\{\mathbf{Y}_{j-1}\}_i)|^2\right)^{1/2} \\
&\quad \cdot \left(\frac{1}{N} \sum_{i=1}^N |\{\mathbf{Y}_{j-1}\}_i|^2\right)^{1/2}.
\end{aligned} \tag{4.10}$$

We show this bound for the fox rabies example in Fig. 4.4. Here, we notice that the bound has the same behavior as the true difference of  $[G(\mathbf{Y})\mathbf{Y}]$ . This correlation makes the connection between nonlinear effects and the variance and second moment quite clear. Both quantities increase with the nonlinearity of the model as well as the increased variance as the simulations spread between extinct and sustained simulations. Fig. 4.5 shows similar behavior, though it uses the Lipschitz constant of  $f$  and only variance.

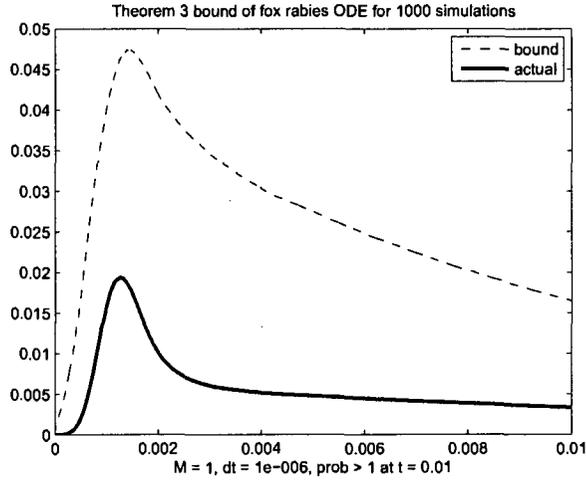


Figure 4.4: The plot gives the bound of Theorem 4.1.3 for the fox rabies example as well as the true difference of the  $[G(\mathbf{Y})\mathbf{Y}]$  terms at each time step.

### Proof of Theorem 4.1.3

Recall the definition of (4.4)

$$[G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}] \equiv \|E(G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}) - G(E(\mathbf{Y}_{j-1}))E(\mathbf{Y}_{j-1})\| \Delta t.$$

Since  $E(\mathbf{Y}_{j-1})$  is a deterministic value, we can write

$$G(E(\mathbf{Y}_{j-1}))E(\mathbf{Y}_{j-1}) = E(G(E(\mathbf{Y}_{j-1}))\mathbf{Y}_{j-1}).$$

This substitution gives

$$\begin{aligned} [G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}] &= \|E(G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}) - E(G(E(\mathbf{Y}_{j-1}))\mathbf{Y}_{j-1})\| \Delta t \\ &= \|E((G(\mathbf{Y}_{j-1}) - G(E(\mathbf{Y}_{j-1})))\mathbf{Y}_{j-1})\| \Delta t. \end{aligned} \quad (4.11)$$

Again, we use an unbiased estimator for the expectation over  $N$  simulations and write

$$Est([G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}]) = \left\| \frac{1}{N} \sum_{i=1}^N (G(\{\mathbf{Y}_{j-1}\}_i) - G(E(\{\mathbf{Y}_{j-1}\}_i))) \{\mathbf{Y}_{j-1}\}_i \right\| \Delta t. \quad (4.12)$$

Bringing the norm inside the sum and rewriting the vector  $\mathbf{Y}$  as the matrix  $\mathbf{Y}^T \cdot I_m$

gives the inequality

$$\begin{aligned} Est([G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}]) &\leq \sum_{i=1}^N \left\| \frac{1}{\sqrt{N}} (G(\{\mathbf{Y}_{j-1}\}_i) - G(E(\{\mathbf{Y}_{j-1}\}_i))) \right. \\ &\quad \left. \cdot \frac{1}{\sqrt{N}} \{\mathbf{Y}_{j-1}\}_i^T I_m \right\| \Delta t. \end{aligned}$$

Since  $\frac{1}{\sqrt{N}} (G(\{\mathbf{Y}_{j-1}\}_i) - G(E(\{\mathbf{Y}_{j-1}\}_i)))$  and  $\frac{1}{\sqrt{N}} \{\mathbf{Y}_{j-1}\}_i^T I_m$  are both contained in  $\ell^2(\mathbb{R}^{m \times m})$ , we use Hölder's inequality in the next step. First recall that the  $\ell^2$  norm of the vector  $\mathbf{Y}_{j-1}$  is given by

$$\|\mathbf{Y}_{j-1}\|_2 = \left( \sum_{i=1}^N \|\mathbf{Y}_{j-1}\|^2 \right)^{1/2}$$

and induces the matrix norm

$$\|G\|_2 = \max_{\|\mathbf{Y}\|_2 \neq 0} \frac{\|G(\mathbf{Y})\mathbf{Y}\|_2}{\|\mathbf{Y}\|_2}.$$

Assuming these norms, Hölder's inequality gives

$$\begin{aligned} Est([G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}]) &\leq \left( \sum_{i=1}^N \left\| \frac{1}{\sqrt{N}} (G(\{\mathbf{Y}_{j-1}\}_i) - G(E(\{\mathbf{Y}_{j-1}\}_i))) \right\|^2 \right)^{1/2} \\ &\quad \cdot \left( \sum_{i=1}^N \left\| \frac{1}{\sqrt{N}} \{\mathbf{Y}_{j-1}\}_i^T I_m \right\|^2 \right)^{1/2} \Delta t. \end{aligned} \quad (4.13)$$

The Lipschitz continuity of  $G$  then gives

$$\begin{aligned} Est([G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}]) &\leq \left( \frac{1}{N} \sum_{i=1}^N L_g^2 \|\{\mathbf{Y}_{j-1}\}_i - E(\{\mathbf{Y}_{j-1}\}_i)\|^2 \right)^{1/2} \\ &\quad \cdot \left( \frac{1}{N} \sum_{i=1}^N \|\{\mathbf{Y}_{j-1}\}_i\|^2 \right)^{1/2} \Delta t \end{aligned}$$

or

$$\begin{aligned} Est([G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}]) &\leq L_g \Delta t \left( \frac{1}{N} \sum_{i=1}^N \|\{\mathbf{Y}_{j-1}\}_i - E(\{\mathbf{Y}_{j-1}\}_i)\|^2 \right)^{1/2} \\ &\quad \cdot \left( \frac{1}{N} \sum_{i=1}^N \|\{\mathbf{Y}_{j-1}\}_i\|^2 \right)^{1/2}. \end{aligned}$$

Multiplying by  $(\frac{N-1}{N-1})^{1/2}$  gives the result.

□

Again, note that if we do not to use the Lipschitz constant of  $G$  in equation (4.13), we may also observe the non-linear term to be bounded by

$$\begin{aligned} Est([G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}]) &\leq \left( \frac{N-1}{N} \right)^{1/2} \left( \frac{1}{N-1} \sum_{i=1}^N \|G(\{\mathbf{Y}_{j-1}\}_i) - G(E(\{\mathbf{Y}_{j-1}\}_i))\|^2 \right)^{1/2} \\ &\quad \cdot \left( \frac{1}{N} \sum_{i=1}^N \|\{\mathbf{Y}_{j-1}\}_i\|^2 \right)^{1/2} \Delta t. \end{aligned}$$

Again, we use  $f(\mathbf{y})$  and the Lipschitz constant  $L_f$  to absorb the second moment, giving a bound in terms of variance only.

**Theorem 4.1.4** *Local  $[G(\mathbf{Y})\mathbf{Y}]$  Bounded by Variance*

$$\begin{aligned} Est([G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}]) &\leq \left( \frac{N-1}{N} \right)^{1/2} L_f \Delta t \\ &\quad \cdot \left( \frac{1}{N-1} \sum_{i=1}^N \|\{\mathbf{Y}_{j-1}\}_i - E(\{\mathbf{Y}_{j-1}\}_i)\|^2 \right)^{1/2} \end{aligned} \quad (4.14)$$

**Proof of Theorem 4.1.4**

Writing (4.4) in terms of  $f$  gives

$$[G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}] = \|E(f(\mathbf{Y}_{j-1})) - f(E(\mathbf{Y}_{j-1}))\| \Delta t.$$

Using the fact that  $E(\mathbf{Y}_{j-1})$  is a deterministic value, we substitute  $f(E(\mathbf{Y}_{j-1})) = E(f(E(\mathbf{Y}_{j-1})))$  into the above equation to give

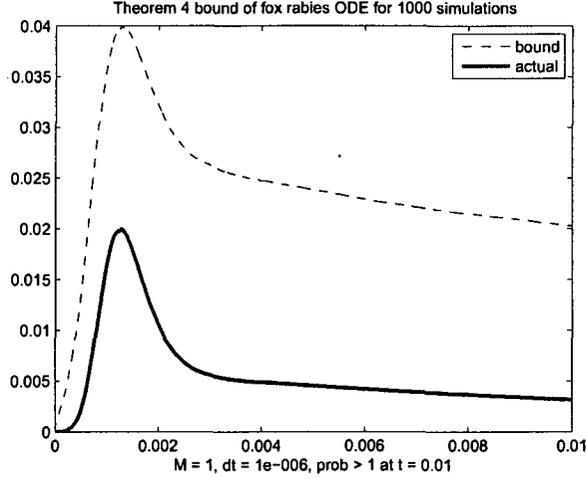


Figure 4.5: The plot gives the bound of Theorem 4.1.4 for the fox rabies example as well as the true difference of the  $[G(\mathbf{Y})\mathbf{Y}]$  terms at each time step.

$$[G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}] = \|E(f(\mathbf{Y}_{j-1}) - f(E(\mathbf{Y}_{j-1})))\| \Delta t.$$

Using the unbiased estimator for expectation gives

$$\begin{aligned} Est([G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}]) &= \left\| \frac{1}{N} \sum_{i=1}^N (f(\{\mathbf{Y}_{j-1}\}_i) - f(E(\{\mathbf{Y}_{j-1}\}_i))) \right\| \Delta t \\ &\leq \sum_{i=1}^N \left\| \frac{1}{N} (f(\{\mathbf{Y}_{j-1}\}_i) - f(E(\{\mathbf{Y}_{j-1}\}_i))) \right\| \Delta t \\ &= \sum_{i=1}^N \left\| \frac{1}{\sqrt{N}} (f(\{\mathbf{Y}_{j-1}\}_i) - f(E(\{\mathbf{Y}_{j-1}\}_i))) I_m \cdot \frac{1}{\sqrt{N}} I_m \right\| \Delta t \end{aligned}$$

Hölder's inequality then implies the following:

$$\begin{aligned} Est([G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}]) &\leq \left( \frac{1}{N} \sum_{i=1}^N \|(f(\{\mathbf{Y}_{j-1}\}_i) - f(E(\{\mathbf{Y}_{j-1}\}_i))) I_m\|^2 \right)^{1/2} \\ &\quad \cdot \left( \frac{1}{N} \sum_{i=1}^N \|I_m\|^2 \right)^{1/2} \Delta t \end{aligned}$$

or, equivalently,

$$Est([G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}]) \leq \left( \frac{1}{N} \sum_{i=1}^N \|f(\{\mathbf{Y}_{j-1}\}_i) - f(E(\{\mathbf{Y}_{j-1}\}_i))\|^2 \right)^{1/2}. \quad (4.15)$$

Using the Lipschitz continuity of  $f$  and multiplying by  $(\frac{N-1}{N-1})^{1/2}$  gives the result.

□

Again, if we choose not to use the Lipschitz condition on  $f$ , Equation (4.15) gives a bound in terms of  $f(E(\mathbf{Y}))$  from the expected value  $E(f(\mathbf{Y}))$ , due to nonlinearity.

Putting the results of Theorem 4.1.2 and Theorem 4.1.4 together for accumulated local bounds, we have the final difference bound below.

**Theorem 4.1.5** *Total Local Bound*

For  $[t_0, t_n]$ ,

$$\begin{aligned}
 & Est(\|E(\mathbf{Y}_n) - \tilde{\mathbf{y}}_n^n\|) \\
 & \leq \left(\frac{N-1}{N-1}\right)^{1/2} L_f \Delta t \sum_{j=1}^n (L_f \Delta t)^{n-j} \left( \frac{1}{N-1} \sum_{i=1}^N \|\{\mathbf{Y}_{j-1}\}_i - E(\{\mathbf{Y}_{j-1}\}_i)\|^2 \right)^{1/2} \\
 & \quad + L_f \sum_{j=1}^n (L_f \Delta t)^{n-j} \int_{t_{j-1}}^{t_j} \|\tilde{\mathbf{y}}^j - \tilde{\mathbf{y}}_{j-1}^{j-1}\| dt
 \end{aligned} \tag{4.16}$$

and

$$\|E(\mathbf{Y}_n) - \tilde{\mathbf{y}}_n^n\| = O(\Delta t).$$

The proof for this theorem follows just as in the scalar case.

## 4.2 Global Difference Bounds

We now consider the short time global bound for the difference between the two models, considering only knowledge gained on the immediate previous time step.

### Theorem 4.2.1 Global Difference Bound Over One Time Step

For any time  $t_n > t_0$ ,

$$\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq (1 + L_f \Delta t) \|E(\mathbf{Y}_{n-1}) - \mathbf{y}_{n-1}\| + [G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}] + L_f \|\mathbf{y}'\|_\infty (\Delta t)^2. \quad (4.17)$$

For the bound of Theorem 4.2.1 and the true difference  $\|E(\mathbf{Y}_n) - \mathbf{y}_n\|$ , see Fig. 4.6. Notice the exponential growth of the bound compared to the true difference which increases over the time of highest nonlinearity and variance of the stochastic model. In the following theorem, we compare the bound to the size of  $[G(\mathbf{Y})\mathbf{Y}]$ .

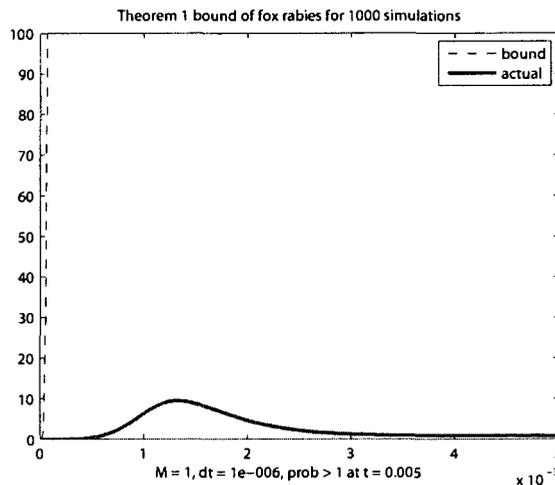


Figure 4.6: This plot shows the bound from Theorem 4.2.1 as well as the true difference  $\|E(\mathbf{Y}_n) - \mathbf{y}_n\|$  for the fox rabies model.

### Proof of Theorem 4.2.1

As before, we use the triangle inequality along with integral forms of  $\mathbf{y}_n$  and  $E(\mathbf{Y}_n)$  to obtain the difference

$$\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq \|E(\mathbf{Y}_{n-1}) - \mathbf{y}_{n-1}\| + \left\| \int_{t_{n-1}}^{t_n} (E(G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}) - G(\mathbf{y})\mathbf{y}) dt \right\|.$$

Adding and subtracting the integral  $\int_{t_{n-1}}^{t_n} G(E(\mathbf{Y}_{n-1}))E(\mathbf{Y}_{n-1})dt$  gives

$$\begin{aligned} \|E(\mathbf{Y}_n) - \mathbf{y}_n\| &\leq \|E(\mathbf{Y}_{n-1}) - \mathbf{y}_{n-1}\| \\ &\quad + \left\| \int_{t_{n-1}}^{t_n} (E(G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}) - G(E(\mathbf{Y}_{n-1}))E(\mathbf{Y}_{n-1})) dt \right\| \\ &\quad + \left\| \int_{t_{n-1}}^{t_n} (G(E(\mathbf{Y}_{n-1}))E(\mathbf{Y}_{n-1}) - G(\mathbf{y})\mathbf{y}) dt \right\|. \end{aligned} \tag{4.18}$$

By Lipschitz continuity of  $f$ , the second integral is bounded by

$$\begin{aligned} \|E(\mathbf{Y}_n) - \mathbf{y}_n\| &\leq \|E(\mathbf{Y}_{n-1}) - \mathbf{y}_{n-1}\| \\ &\quad + \left\| \int_{t_{n-1}}^{t_n} (E(G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}) - G(E(\mathbf{Y}_{n-1}))E(\mathbf{Y}_{n-1})) dt \right\| \\ &\quad + L_f \int_{t_{n-1}}^{t_n} \|E(\mathbf{Y}_{n-1}) - \mathbf{y}\| dt \\ &= \|E(\mathbf{Y}_{n-1}) - \mathbf{y}_{n-1}\| + [G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}] \\ &\quad + L_f \int_{t_{n-1}}^{t_n} \|E(\mathbf{Y}_{n-1}) - \mathbf{y}\| dt \\ &= A + B + C. \end{aligned} \tag{4.19}$$

As in the scalar case, term  $C$  describes how far the solution  $\mathbf{y}$  varies from the local value of  $E(\mathbf{Y}_{n-1})$ . We rewrite this term by adding and subtracting the integral of  $\mathbf{y}$  at time  $t_{n-1}$  to give

$$\begin{aligned} C &\leq L_f \int_{t_{n-1}}^{t_n} \|E(\mathbf{Y}_{n-1}) - \mathbf{y}_{n-1}\| dt + L_f \int_{t_{n-1}}^{t_n} \|\mathbf{y}(t) - \mathbf{y}_{n-1}\| dt \\ &= L_f \Delta t \|E(\mathbf{Y}_{n-1}) - \mathbf{y}_{n-1}\| + L_f \int_{t_{n-1}}^{t_n} \|\mathbf{y}(t) - \mathbf{y}_{n-1}\| dt. \end{aligned}$$

Putting this back together with (4.19) gives

$$\begin{aligned} \|E(\mathbf{Y}_n) - \mathbf{y}_n\| &\leq (1 + L_f \Delta t) \|E(\mathbf{Y}_{n-1}) - \mathbf{y}_{n-1}\| + [G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}] \\ &\quad + L_f \int_{t_{n-1}}^{t_n} \|\mathbf{y}(t) - \mathbf{y}_{n-1}\| dt. \end{aligned}$$

We can bound the last integral as well using the infinity norm where  $\|\mathbf{y}\|_\infty = \sup_{\hat{t}} \|\mathbf{y}(\hat{t})\|$ , in this case for  $\hat{t} \in [t_{n-1}, t_n]$ . The details follow similarly as in the scalar case.

This gives a simplified version of the difference bound on the interval  $[t_{n-1}, t_n]$  and the bound in the theorem

$$\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq (1 + L_f \Delta t) \|E(\mathbf{Y}^{n-1}) - \mathbf{y}_{n-1}\| + [G(\mathbf{Y}_{n-1})\mathbf{Y}_{n-1}] + L_f \|\mathbf{y}'\|_\infty (\Delta t)^2.$$

□

Now (4.17) holds for  $n = 1, 2, 3, \dots$  with the desired bound on the previous time step appearing on the right hand side. Using a discrete Gronwall argument, again, we show several results from the accumulation of these bounds. To make sense of these global bounds, we state each in succession and then show the proof as stages in a single derivation.

**Theorem 4.2.2** *Global Bound - Stage 1*

$$\begin{aligned} \|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq & (1 + L_f \Delta t)^{n-1} \int_{I_1} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\ & + (1 + L_f \Delta t)^{n-2} \int_{I_2} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\ & + \dots + \int_{I_n} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\ & + (1 + L_f \Delta t)^{n-1} L_f \|\mathbf{y}'\|_\infty^1 (\Delta t)^2 + (1 + L_f \Delta t)^{n-2} L_f \|\mathbf{y}'\|_\infty^2 (\Delta t)^2 \\ & + \dots + L_f \|\mathbf{y}'\|_\infty^n (\Delta t)^2, \end{aligned} \tag{4.20}$$

where  $\|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\|$  is piecewise constant on  $m$ -dimensions for each subinterval  $I_1, I_2, \dots, I_n$ .

For Stage 1 of Theorem 4.2.2 and the true difference  $\|E(\mathbf{Y}_n) - \mathbf{y}_n\|$ , see Fig. 4.7. Notice the exponential growth of the bound compared to the true difference. The true difference increases over the time of highest nonlinearity and variance of the stochastic model. We also plot this bound with the terms  $[G(\mathbf{Y})\mathbf{Y}]$  in Fig. 4.8 and observe that these terms have a significant effect on this bound.

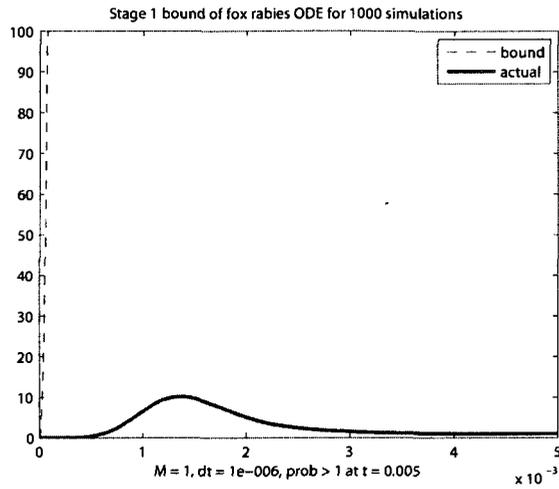


Figure 4.7: This plot shows the bound from Theorem 4.2.2 as well as the true difference  $\|E(\mathbf{Y}_n) - \mathbf{y}_n\|$  for the fox rabies model.

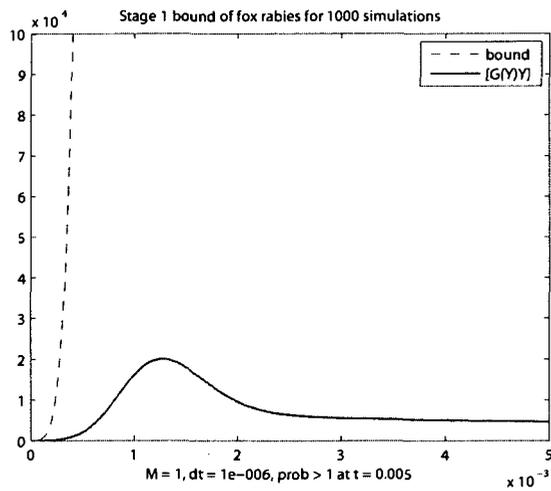


Figure 4.8: This plot shows the bound from Theorem 4.2.2 as well as the difference  $[G(\mathbf{Y})\mathbf{Y}]$ .

**Theorem 4.2.3** *Global Bound - Stage 2*

$$\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq e^{L_f T} \int_0^T \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt + e^{L_f T} \|\mathbf{y}'\|_\infty \Delta t \quad (4.21)$$

We show this bound for the fox rabies model in Fig. 4.9. Notice that the price of bounding the coefficients of Theorem 4.2.2 results in a much higher bound, orders of magnitude larger than the true difference. In the next stage, we reconsider an alternative bound derived from (4.20).

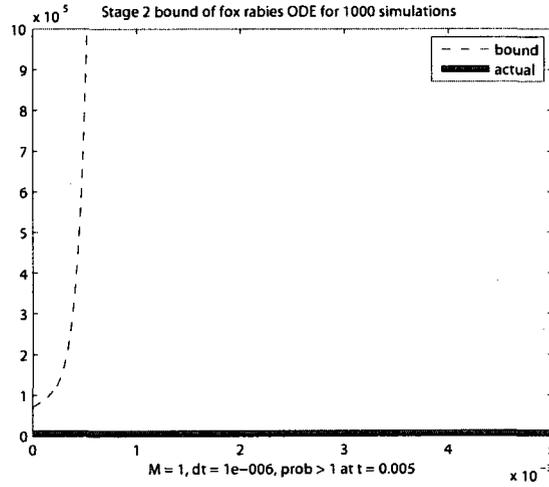


Figure 4.9: This plot shows the bound from Theorem 4.2.3 as well as the true difference  $\|E(\mathbf{Y}_n) - \mathbf{y}_n\|$  for the fox rabies model.

**Theorem 4.2.4** *Global Bound - Stage 3*

$$\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq \frac{e^{L_f T}}{L_f} \max_j \frac{1}{\Delta t} \int_{I_j} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt + e^{L_f T} \|\mathbf{y}'\|_\infty \Delta t \quad (4.22)$$

or, equivalently,

$$\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq \frac{1}{L_f \Delta t} e^{L_f T} \max_j [G(\mathbf{Y}_{j-1})\mathbf{Y}_{j-1}] + e^{L_f T} \|\mathbf{y}'\|_\infty \Delta t \quad (4.23)$$

We show this bound in Fig. 4.10. Notice the exponential behavior drives the size of the bound, while the effects of nonlinearity and variance of the discrete stochastic model are evident in the true difference  $\|E(\mathbf{Y}_n) - \mathbf{y}_n\|$ .

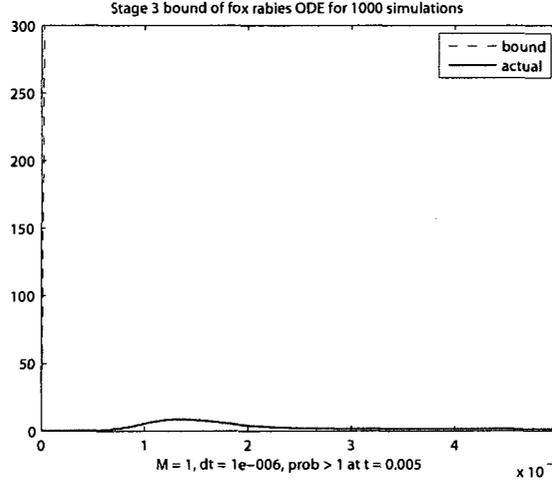


Figure 4.10: This plot shows the bound from Theorem 4.2.4 as well as the true difference  $\|E(\mathbf{Y}_n) - \mathbf{y}_n\|$ .

### Proof of Stages 1 and 2

For the first accumulated global result of Theorem 4.2.2, we recall Theorem 4.2.1 in integral form:

$$\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq (1 + L_f \Delta t) \|E(\mathbf{Y}_{n-1}) - \mathbf{y}_{n-1}\| + \int_{t_{n-1}}^{t_n} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt + L_f \|\mathbf{y}'\|_{\infty}^n (\Delta t)^2$$

We make the first substitution at the previous time step,  $\|E(\mathbf{Y}_{n-1}) - \mathbf{y}_{n-1}\|$ , into

(4.17). This yields

$$\begin{aligned} \|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq & (1 + L_f \Delta t)^2 \|E(\mathbf{Y}_{n-2}) - \mathbf{y}_{n-2}\| \\ & + (1 + L_f \Delta t) \int_{I_{n-1}} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\ & + \int_{I_n} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\ & + (1 + L_f \Delta t) L_f \|\mathbf{y}'\|_{\infty}^{n-1} (\Delta t)^2 + L_f \|\mathbf{y}'\|_{\infty}^n (\Delta t)^2 \end{aligned} \quad (4.24)$$

Continuing in this manner gives (4.20) of Theorem 4.2.2:

$$\begin{aligned}
\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq & (1 + L_f \Delta t)^{n-1} \int_{I_1} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\
& + (1 + L_f \Delta t)^{n-2} \int_{I_2} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\
& + \int_{I_n} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\
& + (1 + L_f \Delta t)^{n-1} L_f \|\mathbf{y}'\|_\infty^1 (\Delta t)^2 + (1 + L_f \Delta t)^{n-2} L_f \|\mathbf{y}'\|_\infty^2 (\Delta t)^2 \\
& + \dots + L_f \|\mathbf{y}'\|_\infty^n (\Delta t)^2
\end{aligned}$$

where the last term  $\|E(\mathbf{Y}_0) - \mathbf{y}_0\|$  is zero.

For the second global bound, we note that the coefficient in each term of the above equation is less than or equal to  $(1 + L_f \Delta t)^{n-1}$ . Similarly, the infinity norm on each interval is less than or equal to the infinity norm over the entire interval. This gives the simplification

$$\begin{aligned}
\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq & (1 + L_f \Delta t)^{n-1} \int_0^T \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\
& + L_f \|\mathbf{y}'\|_\infty (\Delta t)^2 \sum_{i=0}^{n-1} (1 + L_f \Delta t)^i.
\end{aligned} \tag{4.25}$$

Using the formula for the sum of a geometric sequence and the exponential approximation in Appendix A, we can bound (4.25) to give (4.21) of Theorem 4.2.3:

$$\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq e^{L_f T} \int_0^T \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt + e^{L_f T} \|\mathbf{y}'\|_\infty \Delta t.$$

□

### Proof of Stage 3

Again, we recall Equation (4.20) of Theorem 4.2.2 combined with the exponential bound of the sum of  $L_f \|\mathbf{y}'\|_\infty \Delta t$  terms:

$$\begin{aligned}
\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq & (1 + L_f \Delta t)^{n-1} \int_{I_1} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\
& + (1 + L_f \Delta t)^{n-2} \int_{I_2} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\
& + \dots + (1 + L_f \Delta t) \int_{I_{n-1}} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\
& + \int_{I_n} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt + e^{L_f T} \|\mathbf{y}'\|_\infty \Delta t.
\end{aligned}$$

Bounding each of the integrals by the maximum, we have

$$\begin{aligned}
\|E(\mathbf{Y}_n) - \mathbf{y}_n\| &\leq (1 + L_f \Delta t)^{n-1} \max_j \int_{I_j} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\
&\quad + (1 + L_f \Delta t)^{n-2} \max_j \int_{I_j} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\
&\quad + \dots + (1 + L_f \Delta t) \max_j \int_{I_j} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt \\
&\quad + \max_j \int_{I_j} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt + e^{L_f T} \|\mathbf{y}'\|_\infty \Delta t.
\end{aligned}$$

Again using the sum of the geometric series of  $(1 + L_f \Delta t)$  and assuming  $L_f \Delta t < 1$ ,

we have the bound

$$\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq \frac{e^{L_f T}}{L_f \Delta t} \max_j \int_{I_j} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt + e^{L_f T} \|\mathbf{y}'\|_\infty \Delta t.$$

Rewriting this gives the bound in terms of the maximum average value of  $\|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\|$  over each interval

$$\|E(\mathbf{Y}_n) - \mathbf{y}_n\| \leq \frac{e^{L_f T}}{L_f} \max_j \frac{1}{\Delta t} \int_{I_j} \|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\| dt + e^{L_f T} \|\mathbf{y}'\|_\infty \Delta t.$$

Recall that the integral of a vector is taken to be component-wise. Noting that  $\|E(G(\mathbf{Y})\mathbf{Y}) - G(E(\mathbf{Y}))E(\mathbf{Y})\|$  is constant over each interval gives Equation (4.23).

□

# Chapter 5

## Conclusions

There is a standard way to construct a discrete birth-death probability model for an evolution system, in which a continuum ODE model of the system is used to define the probabilities governing the evolution of the stochastic model. Given the significant differences in the dynamical behavior of ODE solutions which are inherently smooth, and stochastic models which are subject to random variation, the question naturally arises about the connection between the two models. In particular, we investigate the validity of using a continuum model to define the evolution of a stochastic model.

We show a consistent way of defining the probabilities for the stochastic model if the ODE has the form  $\dot{y} = f(y) = yg(y)$  in Section 1.2. The deterministic model can then be compared to the expected value of the discrete probability model. For an ODE of this form describing population dynamics, we describe each individual of the population as a categorical random variable. In this formulation, the probability for an event to occur in a population of size  $y$  over a time interval of length  $\Delta t$  is then given by  $yg(y)\Delta t + o(\Delta t)$ , where the type of event (birth or death) depends on the sign of  $g(y)$ . This model is extended to higher dimensions in Chapter 3.

This stochastic model is valid as long as its expected value stays close to the

solution of the differential equation. In Chapters 2 and 4, we analyze and quantify difference bounds for transient behavior, which suggest the models are close locally, but the global difference may be at most exponential in nature. Such a large bound must account for the possibility that the associated probability model may jump across a steady state of the ODE and exhibit divergent behavior while the ODE remains stable.

We explore our results for a number of models, including the illustrative saddle point example given by Equation (1.13). These examples show there can be fundamental differences in the dynamical behavior of the stochastic and ODE model solutions, even when they are close over any given step.

Applying this process of modeling using categorical random variables with probabilities taken from an ODE of the form  $\dot{y} = yg(y)$ , we see that it may be necessary to pursue other types of stochastic models whose expected value may stay closer to the solution of the ODE. Theorems 2.1.3, 2.1.4, 4.1.3, and 4.1.4 indicate that decreasing the variance of a particular stochastic model may be important in keeping the expected value and deterministic solution close.

Our results represent a different approach from another view, which derives some ODE that governs the expected value of the stochastic system. That ODE and the original ODE are *not* the same in general.

# Appendix A

## Mathematical Background

### Theorem A.0.5 *Left-hand Rectangular Rule [2]*

Consider the ODE  $\dot{y} = f(t, y(t))$ , where  $f$  is continuous and differentiable on the interval  $[t_{n-1}, t_n]$ , then  $\int_{t_{n-1}}^{t_n} f(t, y(t)) dt = f(t_{n-1}, y(t_{n-1}))\Delta t + \frac{(\Delta t)^2}{2} f'(\tilde{t}, y(\tilde{t}))$ , for some  $\tilde{t} \in [t_{n-1}, t_n]$ .

### Proof

Using calculus, we have  $\int_{t_{n-1}}^{t_n} f(t, y(t)) dt = y(t_n) - y(t_{n-1})$ .

Now consider the Taylor expansion of  $y(t_n)$  about the point  $t_{n-1}$ ,

$$y(t_n) = y(t_{n-1}) + \Delta t \dot{y}(t_{n-1}) + \frac{(\Delta t)^2}{2!} \ddot{y}(\tilde{t}),$$

for some  $\tilde{t} \in [t_{n-1}, t_n]$ . Writing this in terms of the ODE gives

$$y(t_n) = y(t_{n-1}) + \Delta t f(t_{n-1}, y(t_{n-1})) + \frac{(\Delta t)^2}{2!} f'(\tilde{t}, y(\tilde{t})),$$

for some  $\tilde{t} \in [t_{n-1}, t_n]$ . Substituting this expansion for  $y(t_n)$  gives the result.

□

### Theorem A.0.6 *Exponential approximation*

For an integer  $m$ ,  $e^{mx} = (1+x)^m + o(x)$ , where

$$o(x) = \sum_{k=0}^m \binom{m}{k} \left[ -x^k + \left( \frac{x^2}{2} e^\xi \right)^{m-k} \sum_{j=0}^k \binom{k}{j} x^j \right], \quad (\text{A.1})$$

for  $x \ll 1$  and  $\xi \in [0, 1]$ .

### Proof

We consider the Taylor series expansion of  $e^{mx} = 1 + mx + \frac{(mx)^2}{2!} + \frac{(mx)^3}{3!} + O((mx)^4)$

and subtract the binomial expansion of  $(1+x)^m = \sum_{i=0}^m \binom{m}{i} x^i$ .

$$\begin{aligned} e^{mx} - (1+x)^m &= \left[ 1 + mx + \frac{(mx)^2}{2!} + \frac{(mx)^3}{3!} + O((mx)^4) \right] \\ &\quad - \left[ 1 + mx + \frac{m(m-1)}{2!} x^2 + \frac{m(m-1)(m-2)}{3!} x^3 + O((mx)^4) \right] \\ &= -\frac{m^2}{2} x^2 + O((mx)^3), \end{aligned}$$

which is  $O(x^2)$  if  $x \ll 1$ .

□

### Theorem A.0.7 Gronwall's Inequality (Differentiable Form) [18]

(i) Let  $\eta(\cdot)$  be a nonnegative, absolutely continuous function on  $[0, T]$ , which satisfies for a.e.  $t$  the differential inequality

$$\eta'(t) \leq \phi(t)\eta(t) + \psi(t), \quad (\text{A.2})$$

where  $\phi(t)$  and  $\psi(t)$  are nonnegative, summable functions on  $[0, T]$ . Then

$$\eta(t) \leq e^{\int_0^t \phi(s) ds} \left[ \eta(0) + \int_0^t \psi(s) ds \right], \quad (\text{A.3})$$

for all  $0 \leq t \leq T$ .

(ii) In particular, if

$$\eta'(t) \leq \phi(t)\eta(t) \text{ on } [0, T] \text{ and } \eta(0) = 0,$$

then

$$\eta(t) \equiv 0 \text{ on } [0, T].$$

**Proof**

We first consider

$$\frac{d}{ds} \left( \eta(s) e^{-\int_0^s \phi(r) dr} \right) = \eta(s) \frac{d}{ds} \left( -\int_0^s \phi(r) dr \right) e^{-\int_0^s \phi(r) dr} + \eta'(s) e^{-\int_0^s \phi(r) dr}.$$

By the fundamental theorem of calculus,

$$\frac{d}{ds} \left( -\int_0^s \phi(r) dr \right) e^{-\int_0^s \phi(r) dr} = (-\phi(s)) e^{-\int_0^s \phi(r) dr},$$

which gives

$$\frac{d}{ds} \left( \eta(s) e^{-\int_0^s \phi(r) dr} \right) = e^{-\int_0^s \phi(r) dr} (\eta'(s) - \phi(s) \eta(s)).$$

Rearranging Equation (A.2), we see that

$$\frac{d}{ds} \left( \eta(s) e^{-\int_0^s \phi(r) dr} \right) \leq e^{-\int_0^s \phi(r) dr} \psi(s)$$

for a.e.  $0 \leq s \leq T$ . Integrating both sides of the above equation from 0 to  $t$  gives

$$\eta(t) e^{-\int_0^t \phi(r) dr} - \eta(0) \leq \int_0^t e^{-\int_0^s \phi(r) dr} \psi(s) ds$$

or

$$\eta(t) e^{-\int_0^t \phi(r) dr} \leq \eta(0) + \int_0^t e^{-\int_0^s \phi(r) dr} \psi(s) ds.$$

Since  $\phi(t) > 0$  on  $[0, T]$ , we have  $e^{-\int_0^s \phi(r) dr} \leq 1$ , which gives

$$\eta(t) e^{-\int_0^t \phi(r) dr} \leq \eta(0) + \int_0^t \psi(s) ds.$$

Multiplying by  $e^{\int_0^t \phi(r) dr}$  gives the result

$$\eta(t) \leq e^{\int_0^t \phi(r) dr} \left[ \eta(0) + \int_0^t \psi(s) ds \right].$$

□

**Theorem A.0.8** *Gronwall's Inequality (Integral Form) [18]*

(i) Let  $\xi(t)$  be a nonnegative, summable function on  $[0, T]$  which satisfies for a.e.  $t$  the integral inequality

$$\xi(t) \leq C_1 \int_0^t \xi(s) ds + C_2 \quad (\text{A.4})$$

for constants  $C_1, C_2 \geq 0$ . Then

$$\xi(t) \leq C_2(1 + C_1 t e^{C_1 t}) \quad (\text{A.5})$$

for a.e.  $0 \leq t \leq T$ .

(ii) In particular, if

$$\xi(t) \leq C_1 \int_0^t \xi(s) ds$$

for a.e.  $0 \leq t \leq T$ , then

$$\xi(t) = 0 \text{ a.e.}$$

### Proof

We first let  $\eta(t) = \int_0^t \xi(s) ds$ . Taking the derivative of both sides and using the fundamental theorem of calculus gives  $\eta'(t) = \xi(t) - \xi(0)$ . Substituting the assumption of Equation (A.4) and relabeling constant  $C_2 + \eta(0)$  as  $C_2$  gives

$$\eta'(s) \leq C_1 \int_0^t \xi(s) ds + C_2 = C_1 \eta(t) + C_2.$$

Notice, that this is the assumption needed for the differential form of Gronwall's inequality A.2. With  $\phi(t) = C_1$  and  $\psi(t) = C_2$ , we have its conclusion

$$\eta(t) \leq e^{\int_0^t C_1 ds} \left[ \eta(0) + \int_0^t C_2 ds \right].$$

Simplifying and noting that  $\eta(0) = \int_0^0 \xi(s) ds = 0$  gives

$$\eta(t) \leq C_2 t e^{C_1 t}.$$

Using this equation and assumption (A.4) gives the result

$$\xi(t) \leq C_1 \eta(t) + C_2 \leq C_2 (1 + C_1 t e^{C_1 t}).$$

□

## A.1 Unbiased Estimators and Statistical Properties

The unbiased estimator for the expected value of the population  $Y_{j-1}$  at time  $t_{j-1}$ ,

$j = 1, 2, \dots, n$  is given by

$$\overline{E(Y_{j-1})} = \frac{1}{N} \sum_{i=1}^N \{Y_{j-1}\}_i, \quad (\text{A.6})$$

where  $\{Y_{j-1}\}_i$  is the population  $Y_{j-1}$  for the  $i^{\text{th}}$  simulation,  $i = 1, 2, \dots, N$ .

Similarly, the unbiased estimator for the variance of  $Y_{j-1}$  is given by

$$\overline{Var(Y_{j-1})} = \frac{1}{N-1} \sum_{i=1}^N \left( \{Y_{j-1}\}_i - \overline{E(Y_{j-1})} \right)^2. \quad (\text{A.7})$$

By definition, the unbiased estimator for the second moment is

$$\overline{m_2} = \overline{E((Y_{j-1})^2)} = \frac{1}{N} \sum_{i=1}^N (\{Y_{j-1}\}_i)^2. \quad (\text{A.8})$$

The proof that these estimators are in fact unbiased can be found in most texts on simulations. See [32].

A few statistical properties we consider in the analysis are

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$\text{Var}(X) = \int X^2\rho(X)dX - \left(\int X\rho(X)dX\right)^2$$

If  $X_i$  are uncorrelated, then  $\text{Var}\left(\sum_i X_i\right) = \sum_i \text{Var}(X_i)$

$$\text{Cov}(X, Y) = E(X, Y) - E(X)E(Y)$$

# Appendix B

## Review of Discrete, Continuous, and Stochastic Models

### B.1 Chemical Kinetics and the Gillespie Algorithm

Perhaps the most well-known method for simulating stochastic birth-death processes is the Gillespie algorithm. In 1976 [19] and 1977 [20], Gillespie used physical properties of atoms and mass action assumptions to arrive at a probabilistic model describing chemical systems. Gillespie began using Monte-Carlo techniques and estimated the probability an event (such as a product formed from reactants in a chemical equation) occurred in a small time interval. This was in contrast to the commonly accepted method of the time ([31], [33]) of solving a probabilistic differential-difference equation, called the ‘master equation’. This master equation is equivalent to the Chapman-Kolmogorov equations which describe the probability each product was of a certain size at any time  $t$ , given all the possible ways to arrive at that particular state and the probabilities of getting there. Gillespie describes the evolution of these probabilities in a way similar to that which is done today in probability models [28], [34].

Given some set of  $n$  chemical reactants, there is a finite list of possible products that can be formed. The probabilities each of these events occur in a small time interval,  $\Delta t$ , were found using the physical dimensions of the molecules to determine the probability two molecules come within reacting distance of the molecule of interest. Assuming hard spheres and uniform mixing, he describes a small volume,  $\Delta V$ , wherein two molecules may touch and a reaction may occur. This uniform mixing is a common assumption made in modeling, particularly for large numbers of individuals. The product formed from an interaction depends on the proportion of each atom and a characteristic value associated with each reaction. In the case of chemical kinetics, this is given by the level of inertness. (For something like predator-prey models, this value may describe the nutritional benefit achieved by the predator from consuming the prey. The correlation being that a more healthy individual gives birth with a higher probability.) Gillespie considers proportional volumes in the interactions of these molecules to avoid the zero probability of an interaction as the volume of each molecule goes to zero.

Since analytically solving a 'master equation' or the associated Chapman-Kolmogorov equations was and is a non-trivial matter, Gillespie sought to find a method for which the process of simulating the statistical events and calculating their moments could be done on a computer. In [20], Gillespie describes a Monte-Carlo technique for drawing random event times from an exponential distribution with parameters based on the event probabilities over small time steps, as  $\Delta t$  approaches zero. This Monte-Carlo simulation with exponentially distributed waiting times (for the next event) is now referred to as the Gillespie algorithm. At the time, Gillespie's methods met with some

resistance [33]. However, it finally caught on and was applied to stochastic modeling in many disciplines [6].

Gillespie references methods used in [33] and [17], which show that in the thermodynamic limit (the numbers of molecules of each species and the containing volume all approach infinity in such a way that the molecular concentrations approach finite values), this stochastic formulation approaches the deterministic form of the Liouville (transport) equation from molecular dynamics. R. Dolgoarshinnykh [13] also uses the method in [17] to show convergence of an SIR probability model to a deterministic counterpart. This method involves a Gronwall argument and the law of large numbers, which we discuss further in reference to Dolgoarshinnykh's current analysis. Note that we are not contesting this direction of convergence. Instead, we are concerned with the use of rates from particular deterministic models to form a stochastic model whose expected value does not converge to the solution of the original differential equation. Gillespie notes this inconsistency near chemical instabilities and suggests the stochastic model be used in this case. We explore the reasons for this discrepancy and what to do about it.

## **B.2 Evolutions and biological applications of Gillespie's algorithm**

In current books by Kot [28] and Renshaw [34], Gillespie's algorithm is applied to several ecological population examples, including the simple linear birth process introduced by Yule and Furry, which is extended to a simple birth-death process.

As random number generators and statistics advanced, Gillespie's algorithm be-

came more easily simulated using built-in exponential distribution functions for the random event times and a uniform probability for which event occurred. Kot shows this derivation by assuming event probabilities in a small time step  $\Delta t$  and taking the limit as  $\Delta t$  goes to zero. Using the same assumption from Delbrück [31], this process requires the probability of more than one event occurring in  $\Delta t$  to be first order  $\Delta t$ . Using these probabilities, Kot determines a differential equation whose solution is the probability  $m$  events occur at any time  $t$ . Through induction, it is shown that this probability is given by a negative binomial distribution with probability of success being a negative exponential. Using the law of total probability and independence of events, this gives rise to the waiting time for any future event, given the current population.

Kot relates the probability model back to the solution of a differential equation using the expected value. For the simple linear case, the expected value matches the solution to the differential equation quite well. We use this stochastic value as a measure of comparison as well, but explore non-linear models for which the difference is significantly larger than zero. Kot and Renshaw also consider non-linear predator-prey models with and without spatial variability [28], [34], though not of the specific form we explore. Renshaw also analyzes these birth-death processes for long-time equilibria and how they relate to differential equations. We are most interested in these non-linear models, but for short to intermediate times and of a particular form, which is discussed in section 1.2.

### B.3 Modeling prior to 1970's in chemical kinetics and biology

Next, we take a step back and consider the modeling techniques and applications being used prior Gillespie's algorithm. Among other references, we cite works from McQuarrie [31] and Barucha-Reid [6] which provide reviews of early probability-based modeling in chemical kinetics and biological applications, respectively.

Sydney Chapman and Andrey Kolmogorov independently developed the equations which relate the joint probability distribution function of different sets of coordinates (or states) on a stochastic process [10]. The total probability of moving from state  $i$  to state  $j$  can be found by summing up all probabilities of moving from state  $i$  to some intermediate state  $k$ , multiplied by the probability of moving from this state  $k$  to state  $j$ . This idea is very powerful in probability models. Gillespie uses this formulation in his discussion of the master equation, which is highly descriptive of probabilities for any time  $t$ , but is inherently very difficult to solve. This is partly what motivated him to derive a new method of stochastically simulating these probability models that does not rely on numerical solutions. With a slight change in notation, the concept is also used in the formulation of the basic model for birth-death processes as seen in [28] and [34].

Yule (1924) and Furry (1937) are credited with the first development of a simple linear birth process, which was expanded by Feller and Arley to include deaths in a systematic stochastic Markovian model using probabilities of simple events in a small time step  $\Delta t$  [28], [6]. Kendall [25] continued the simple birth-death model

to include cumulative events - still only one event occurring in a single time step  $\Delta t$ , but considered a new random variable for the number of births (absolute change) occurring up to time  $t$  in addition to the population random variable. Once this framework was laid for these birth-death processes, many applications and extensions were contributed to the biological community in the 1950's, including the logistic model [25], immigration [25], a two-sex model, and the Lotka-Volterra predator-prey model [6]. Non-Markovian time-dependent models were then described by the Bellman-Harris process in 1952 [7]. Kermack and McKendrick published the SIR (susceptible-infected-resistant) disease model in 1927 [27], which was analyzed in the stochastic sense in varying degrees beginning with Bartlett and Kendall in 1956, [5] and [26], respectively. Other applications include mutation processes, the theory of gene frequencies, and radiation effects in biological systems [6]. For a complete review and description of disease models and epidemics, see the text by Brauer and Castillo-Chavez [8].

At about the same time biology was developing these probabilistic models, birth-death processes were being explored in the field of chemical kinetics. In 1940, Kramers described the diffusion of chemical particles using continuous Brownian motion [29]. Delbrück (also in 1940) looked at the rate of a chemical reaction ( $A \rightarrow B$ ), whose rates depended on both reactant and product and could catalyze itself [12]. He assumed that the concentration of A was so large throughout the reaction that it remained constant, thus making the reaction first order (otherwise it is second order if the rate was proportional to two chemicals in the reaction). Contributions by Singer (1953) and Rényi (1954) illustrated the necessity of stochastic models [31]. For small

systems, Singer showed that the fluctuations in the number of reactant species could be the cause of the irreproducibility of some reactions (in addition to the impurities in the chemicals). Soon after, Rényi showed that mass action (the assumption that all particles have the same behavior and are randomly mixed) is only approximately valid and doesn't hold for small systems.

In 1958, Bartholomay published a paper on a unimolecular chemical reaction ( $A \rightarrow B$ ) as a pure death process [3], including details of the derivation involving the Chapman-Kolmogorov equations. In 1959, K. Ishida published a paper on unimolecular reactions as well, but using time dependent rate parameters [23]. In 1958, Bartholomay also published a paper discussing the full details of a general birth-death process with specific application to biological models as Markov chains [4]. This consolidated the ideas of birth-death processes for probability models and likely connected the chemical and biological fields, contributing to the communication between the chemical and biological fields and the use of Gillespie's algorithm in both.

## **B.4 Convergence from Probability to Deterministic Model - R. G. Dolgoarshinnykh**

Disease modeling has been and is currently a very hot topic around the world as new or re-emerging epidemics threaten populations of humans as well as animals [8]. The question of whether or not these birth-death processes are an accurate representation of reality is a very pressing issue. R. G. Dolgoarshinnykh explores the direction of convergence from an SI stochastic model to a deterministic counterpart.

In [13], R. G. Dolgoarshinnykh looks at SIRS epidemic models and shows that the

event when all the infected individuals recover without introducing new infections into the population (related to the deterministic value,  $R_0 < 1$ ) is a large deviations event for a given probability model. In other words, the probability of this event occurring exponentially decays as  $N$  gets large. In [14], she follows the method described in [17] using the Law of Large Numbers (LLN) to show that for large populations of size  $N$  and, under certain conditions, the population proportions of SIRS models are well approximated by trajectories of a deterministic system.

Dolgoarshinnykh assumes a birth-death process typical of Kermack McKendrick compartment models [27], where the portion of population susceptible to a disease is given by  $S$ , the infected individuals by  $I$ , and those resistant to the disease given by  $R$ , where  $S + I + R = N$ . The Markov process for the stochastic model uses population densities,  $s_t = S_t/N$ ,  $i_t = I_t/N$ , and  $r_t = R_t/N = 1 - s_t - i_t$ , and is defined by the probabilities:

$$P^N(S_{t+h}, I_{t+h}) = (S - 1, I + 1)|(S_t, I_t) = (S, I) = N\theta s_i h + o(h)$$

$$P^N(S_{t+h}, I_{t+h}) = (S, I - 1)|(S_t, I_t) = (S, I) = N\rho i h + o(h)$$

$$P^N(S_{t+h}, I_{t+h}) = (S + 1, I)|(S_t, I_t) = (S, I) = N r h + o(h),$$

for positive constants  $\theta$  and  $\rho$ . This model assumes independent events and uses the binomial (called categorical variables for more than two possible outcomes) theorem to define first order event probabilities for each state transition between  $S$ ,  $I$ , and  $R$ . We use these same assumptions in our model, but begin with an ODE of a specific form and consider using its rates to define a birth-death process. At this point, Dolgoarshinnykh does not assume any connection with an ODE.

Ethier and Kurtz describe the random variable  $X(t)$  of a jump Markov process as initial conditions plus the sum of Poisson random variables. Under certain conditions, this defines a martingale (a stochastic process such that the conditional expected value of an observation at some time  $t$ , given all the observations up to some earlier time  $s$ , is equal to the observation at earlier times), which is a function of a general Markov process. The LLN and Gronwall inequality is then used to show that such a stochastic Markov process converges to a solution of some differential equation, for large populations. Dolgoarshinnykh uses this method to show convergence of the SIR model to a solution of an associated ODE.

# Bibliography

- [1] R. M. Anderson, H. C. Jackson, R. M. May, A. M. Smith. *Population dynamics of fox rabies in Europe*. Nature. Vol. 289, pp 765 - 771. (1981)
- [2] K. E. Atkinson. *An Introduction to Numerical Analysis*. 2nd ed. John Wiley & Sons. New York. (1989)
- [3] A. F. Bartholomay. *Stochastic models for chemical reactions: I. Theory of the unimolecular reaction process*. Bull. Math. Biophys. Vol. 20, pp 175-190. (1958)
- [4] A. F. Bartholomay. *On the linear birth and death processes of biology as Markov chains*. Bull. Math. Biophys. Vol. 20, pp 97-118. (1958)
- [5] M. S. Bartlett. *Deterministic and Stochastic Models for Recurrent Epidemics*. Proc. Third Berkeley Symposium on Math. Statistics and Probability. Vol. 4, pp 81-109. (1956)
- [6] A. T. Barucha-Reid. *Elements of the Theory of Markov Processes and Their Applications*. McGraw-Hill Book Company, Inc. New York. (1960)
- [7] R. Bellman and T. E. Harris. *On Age-dependent Binary Branching Processes*. Ann. Math. Vol. 55, pp 280-295. (1952)
- [8] F. Brauer, C. Castillo-Chavez. *Mathematical Models in Population Biology and Epidemiology*. Springer. New York. (2001)
- [9] S. Brault, H. Caswell. *Pod-specific demography of killer whales Orcinus orca*. Ecology. Vol. 74, No. 5, pp. 1444-1454. (1993)
- [10] S. Chapman. *On the Brownian Motion Displacements and Thermal Diffusion of Grains Suspended in a Non-Uniform Fluid*. Proceedings of the Royal Society of London. Series A. Vol. 119, No. 781. pp. 34-54. (May 1, 1928)

- [11] K. Dekker and J. G. Verwer. *Stability of Runge-Kutta methods for stiff nonlinear differential equations*. North-Holland. Amsterdam. (1984)
- [12] M. Delbrück. *Statistical fluctuations in autocatalytic reactions*. J. Chem. Phys. Vol. 8, pp120-124.
- [13] R. G. Dolgoarshinnykh. *Sample Path Large Deviations for SIRS Epidemic Processes*. Download at <<http://www.stat.columbia.edu/~regina/research/>>. Submitted as of June 5, 2009.
- [14] R. G. Dolgoarshinnykh. *Law of Large Numbers for the SIRS Epidemic Processes*. Technical Report. Download at <<http://www.stat.columbia.edu/~regina/research/>>.
- [15] D. Estep. *A Posteriori Error Bounds and Global Error Control for Approximation of Ordinary Differential Equations*. SIAM Journal on Numerical Analysis. Vol. 32, No. 1, pp 1-48. (February 1995)
- [16] D. Estep, D. Neckels. *Fast and reliable methods for determining the evolution of uncertain parameters in differential equations*. Journal of Computational Physics. Vol. 213, pp 530-556. (2006)
- [17] S. N. Ethier, T. G. Kurtz. *Markov Processes Characterization and Convergence*. John Wiley & Sons. New York. (1986)
- [18] L. C. Evans. *Partial Differential Equations; Graduate Studies in Mathematics*. Vol 19. American Mathematical Society. Providence, RI. (1998)
- [19] D. T. Gillespie. *A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions*. Journal of Computational Physics. Vol. 22, pp 403-434. (1976)
- [20] D. T. Gillespie. *Exact Stochastic Simulation of Coupled Chemical Reactions*. The Journal of Physical Chemistry. Vol. 81, No. 25, pp 2340-2361. (1977)
- [21] G. Grimmett, D. Stirzaker. *Probability and Random Processes*. 3rd ed. Oxford University Press. New York. (2001)

- [22] E. Hairer, S.P. Norsett, G. Wanner. *Solving Ordinary Differential Equations*. 2nd ed. Springer. Berlin. (1996)
- [23] K. Ishida. *The Stochastic Model for Unimolecular Gas Reaction*. Bull. Chem. Soc. Japan. Vol. 33, pp 1030-1036. (1959)
- [24] M. J. Keeling and C. A. Gilligan. *Bubonic plague: a metapopulation model of a zoonosis*. Proceedings of the Royal Society of London. Series B. Vol. 267, pp 2219-30. (2000)
- [25] D. G. Kendall. *On the Generalized "Birth and Death" Process*. Ann. Math. Statist. Vol. 19. pp 1-15. (1948)
- [26] D. G. Kendall. *Deterministic and Stochastic Epidemics in Closed Populations*. Proc. Third Berkeley Symposium on Math. Statistics and Probability. Vol. 4, pp 149-165. (1956)
- [27] W. O. Kermack and A. G. McKendrick *A Contribution to the Mathematical Theory of Epidemics*. Proceedings of the Royal Society of London. Series A. Vol. 115, 700-721. (1927)
- [28] M. Kot. *Elements of Mathematical Ecology*. Cambridge University Press. New York. (2001)
- [29] H. A. Kramers. *Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions*. Physica. Vol. 7, No. 4, pp 284-304. (1940)
- [30] A. K. Laird. *Dynamics of tumor growth*. Br. J. Cancer. Vol.13, pp 490-502. (1964)
- [31] D. A. McQuarrie. *Stochastic Approach to Chemical Kinetics*. Journal of Applied Probability. Vol. 4, No. 3, pp. 413-478. (Dec., 1967)
- [32] P. Olofsson. *Probability, Statistics, and Stochastic Processes*. John Wiley and Sons. Hoboken, New Jersey. (2005)
- [33] I. Oppenheim, K. E. Shuler, G. H. Weiss. *Stochastic and Deterministic Formulation of Chemical Rate Equations*. The Journal of Chemical Physics. Vol. 50, No. 1. (Jan. 1969)

- [34] E. Renshaw. *Modelling Biological Populations in Space and Time*. Cambridge University Press. New York. (1991)
- [35] S. M. Ross. *Simulation*. 4th ed. Elsevier Inc. New York. (2006)
- [36] H. M. Taylor and S. Karlin. *An Introduction to Stochastic Modeling*. 3rd ed. Academic Press. San Diego, CA. (1998)