Math 605: Stochastic Models in Biology

Fall 2013

University of Wisconsin at Madison

Instructor: David F. Anderson

What is this course?

This course is

1. an introduction to stochastic processes, at Math 632 level,
2. with an added focus on computational techniques and
3. applications arising from biology.

The main mathematical models of study are

- discrete time Markov chains,
- branching processes,
- point processes,
- continuous time Markov chains,
- and diffusion processes (those incorporating Brownian motion).

We will focus on computational methods for the different models.

The software package of choice will be Matlab.

Our focus

The main mathematical models of study are

- discrete time Markov chains,
- branching processes,
- point processes,
- continuous time Markov chains,
- and diffusion processes (those incorporating Brownian motion).

- We will pay special attention to the continuous time Markov chain models that arise in the study of population processes, such as cellular level biochemical models.

- Our perspective will differ from that of other texts on the subject in that we will primarily attempt to understand these processes via the random time change representation of Thomas Kurtz, which will be used to derive the relevant approximations (including the standard mass-action kinetics model for deterministic dynamics) and simulation strategies.

Prerequisites

We require:

1. A solid understanding of calculus.
2. Linear algebra: eigenvalues, eigenvectors, matrix multiplications, etc.
3. A basic knowledge of differential equations.
4. A basic understanding of probability: sample spaces, random variables, expectations, conditional probability.

I will begin the class with a brief introduction to probability before discussing stochastic processes and stochastic models.

Stochastic versus deterministic models

Before proceeding too far, it is important to understand the basic terms "stochastic" and "deterministic."

A process is *deterministic* if its future is completely determined by its present and past. Examples of deterministic processes include solutions to differential and difference equations.

Example

The initial value problem

$$\dot{x}(t) = 3x(t) \qquad x(0) = 2,$$

has the solution $x(t) = 2e^{3t}$. □

Example

Consider the difference equation

$$F_1 = F_2 = 1$$
$$F_n = F_{n-1} + F_{n-2}, \quad \text{for } n > 2.$$

Then $\{F_n\}_{n=1}^{\infty}$ is the well known Fibonacci sequence: $\{1, 1, 2, 3, 5, 8, \dots\}$. □

Stochastic versus deterministic models

On the other hand, a *stochastic process* is a random process evolving in time.

Informally, this means that even if you have full knowledge of the state of the system (and it's entire past), you can not be sure of it's value at future times.

More formally, a stochastic process is a collection of random variables, $X(t)$ or X_t, indexed by time.

Example

Consider rolling a fair, six-sided die many times, and for $k \in \{1, 2, \dots\}$, let Z_k be the outcome of the kth roll. Let

$$X_n = \sum_{k=1}^{n} Z_k.$$

- Thus, X_n is the accumulated total of the first n rolls.
- Knowing $X_1 = 3$ only tells you that $X_2 \in \{4, \dots, 9\}$, with equal probability.
- Note that time, indexed here by n, is discrete in that we only update the system after each roll of the die. □

Stochastic versus deterministic models

Example

Consider a frog sitting in a pond with k lily pads, labeled 1 through k.

1. The frog starts the day by sitting on a randomly chosen pad (for example, they could be chosen with equal probability).
2. However, after a random amount of time, the frog will jump to another pad, also randomly chosen.
3. Letting $t = 0$ denote the start of the day, we let $X(t) \in \{1, \ldots, k\}$ denote the lily pad occupied by the frog at time t. In this example, time is naturally continuous.

However, if we are only interested in which lily pad the frog is on after a given number of jumps,

1. then we may let Z_n denote the lily pad occupied by the frog *after the nth jump*,
2. with Z_0 defined to be the starting lily pad.
3. The process Z_n is discrete in time.
4. The processes Z_n and $X(t)$ are clearly related, and Z_n is usually called the *embedded discrete process associated with* $X(t)$.

Example: Bacterial Growth

Let's consider two oversimplified models for bacterial growth (by *growth* here, I mean the growth of the size of the colony, not of an individual bacterium):

- one deterministic
- one stochastic.

We suppose

- there are 10 bacteria at time zero.
- each bacteria divides at an "average" rate of once per three hours.

Deterministic model: a "reasonable" model would be

$$\frac{d}{dt}x(t) = \frac{1}{3}x(t) \qquad x(0) = 10, \tag{1}$$

with solution

$$x(t) = 10e^{t/3},$$

where the units of t are hours.

Example: Bacterial Growth

Stochastic Model: Without going into the finer details, assume

1. Each bacteria divides after a random (independent, exponential) amount of time with an average wait of 3 hours.

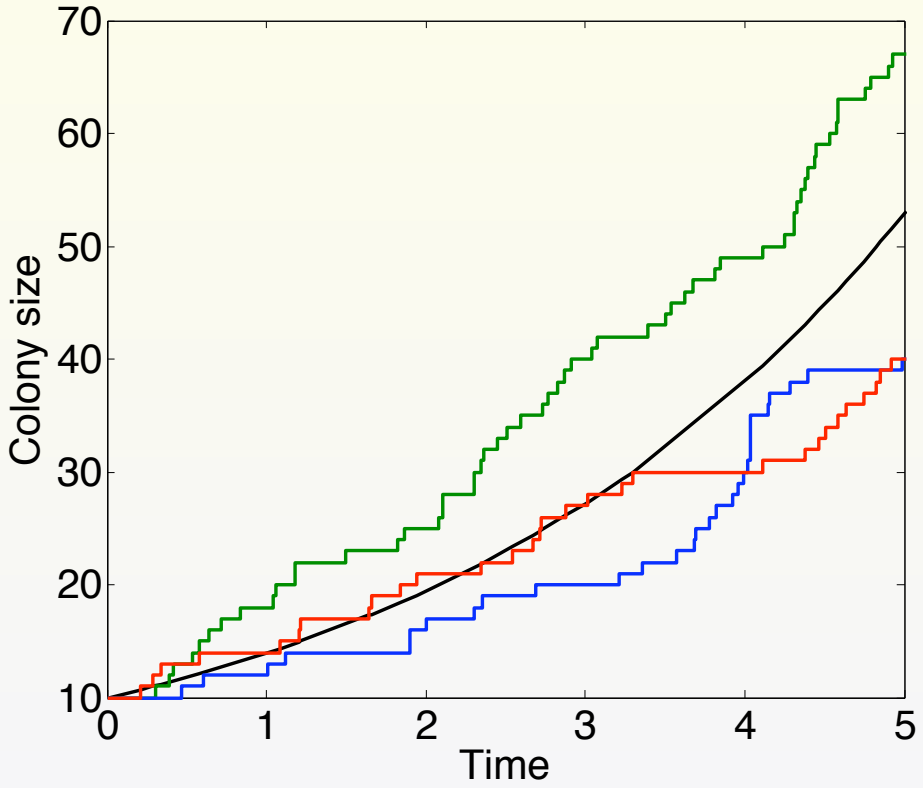
Similar to equation (1) for the deterministic model, it is possible to write down systems of equations describing the time evolution of model

1. Evolution of individual sample paths – instance of experiment (like the ODE model)

2. Evolution of the distribution (probability of being in certain states)

Example: Bacterial Growth - evolution of sample paths

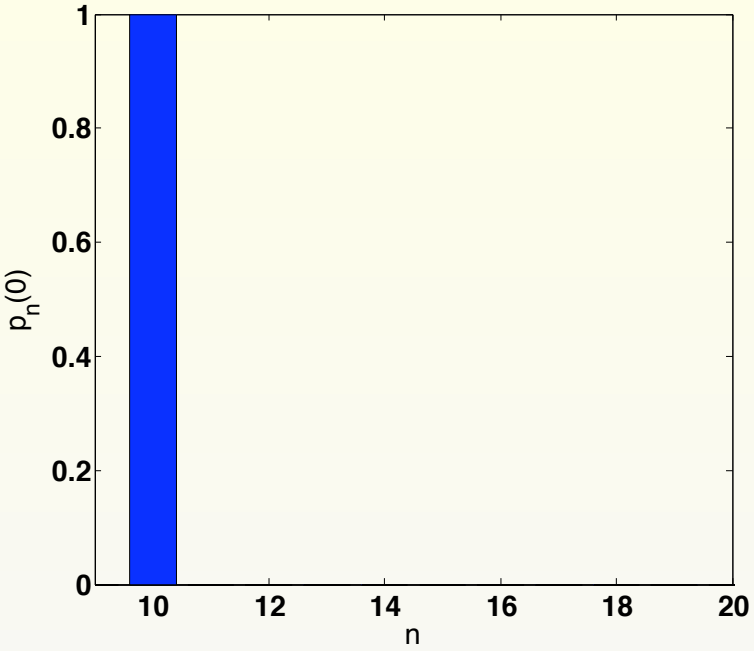
- Below is a plot of the solution of the deterministic system versus three different realizations of the stochastic system.

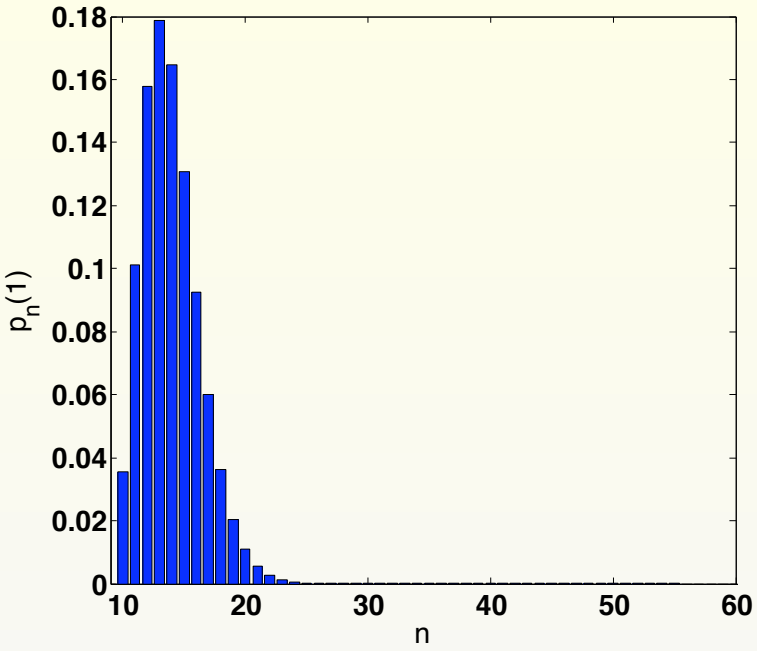


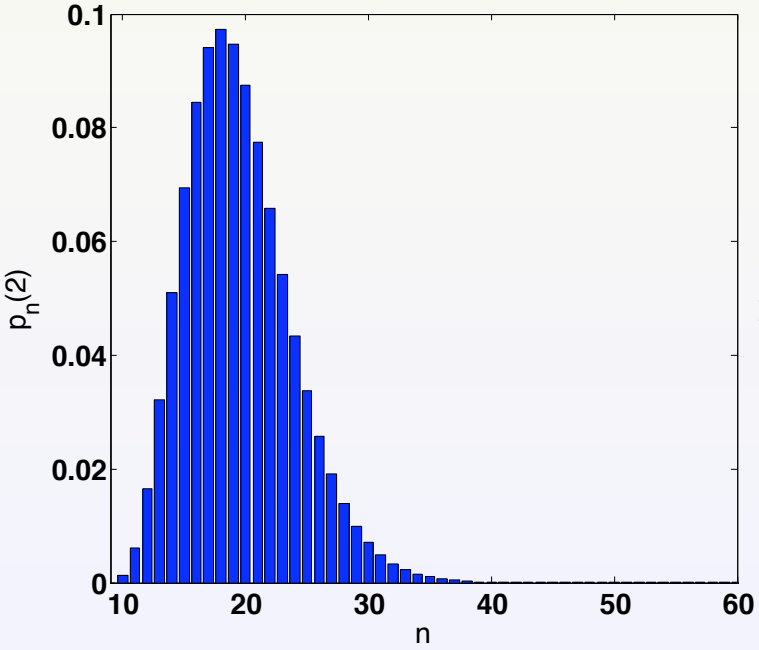
- Stochastic realizations/experiments appear to follow the deterministic system in a “noisy” way.

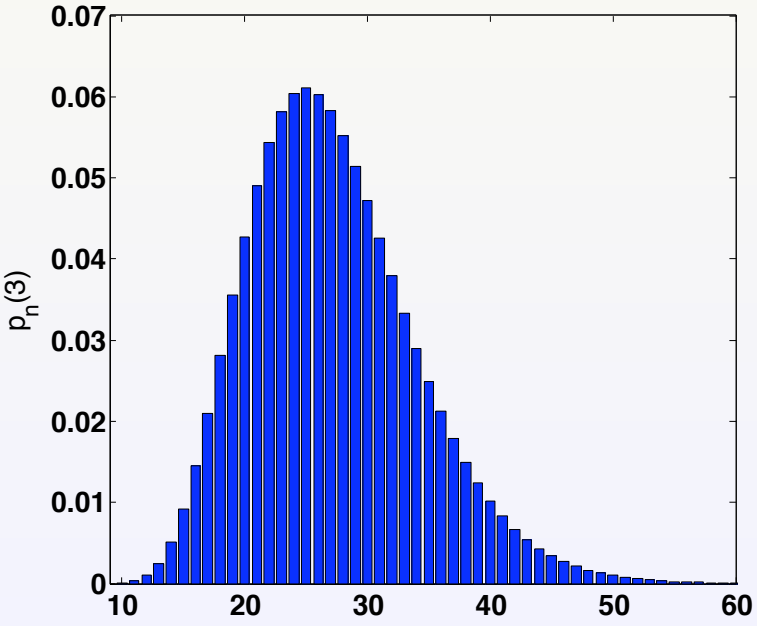
- It is clear that the behavior of a single realization or experiment of the stochastic system can not be predicted with absolute accuracy.

Example: population growth - evolution of distribution









Example: Bacterial Growth and Death

Now suppose that we change the model "slightly" in that:

1. we allow bacteria to die as well as divide.
2. we suppose we begin with only two bacteria.

We suppose that they die after about five hours.

Our new deterministic model could be

$$\dot{x}(t) = \frac{1}{3}x(t) - \frac{1}{5}x(t) = \frac{2}{15}x(t), \qquad x(0) = 2,$$

with solution

$$x(t) = 2e^{2t/15}.$$

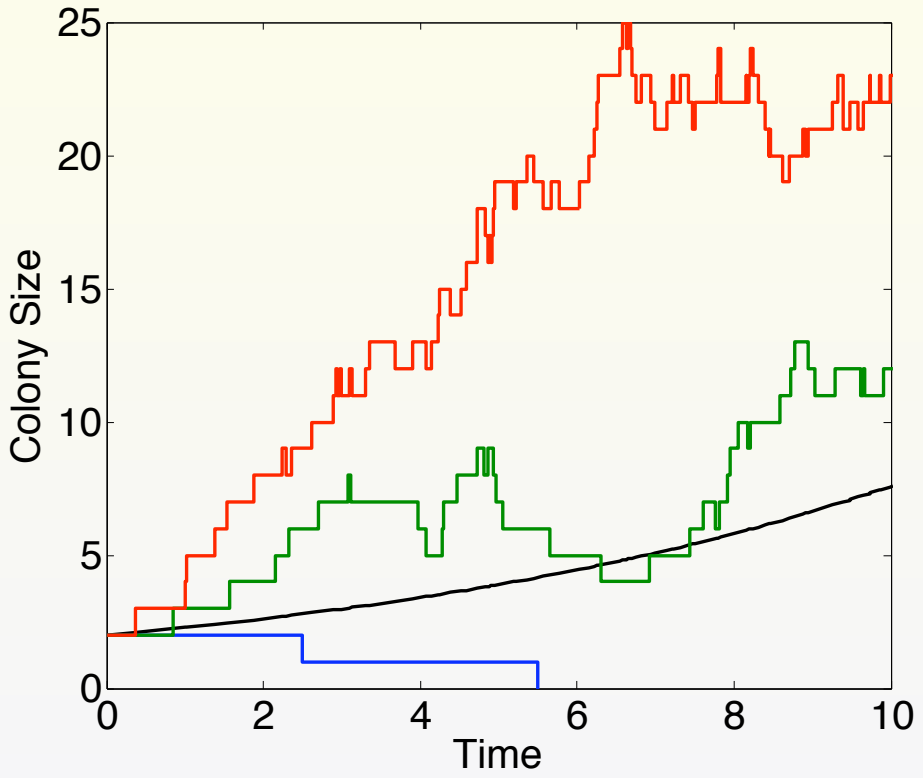
Example: Bacterial Growth and Death

For the stochastic model, we now model the two possible changes to the size of the colony separately. That is, the next event is *either*

1. a growth event (via a division) or
2. a decrease event (via a death).

Example: Bacterial Growth and Death

- Deterministic vs. three realizations/experiments of stochastic system.



- The models now behave *qualitatively* differently:

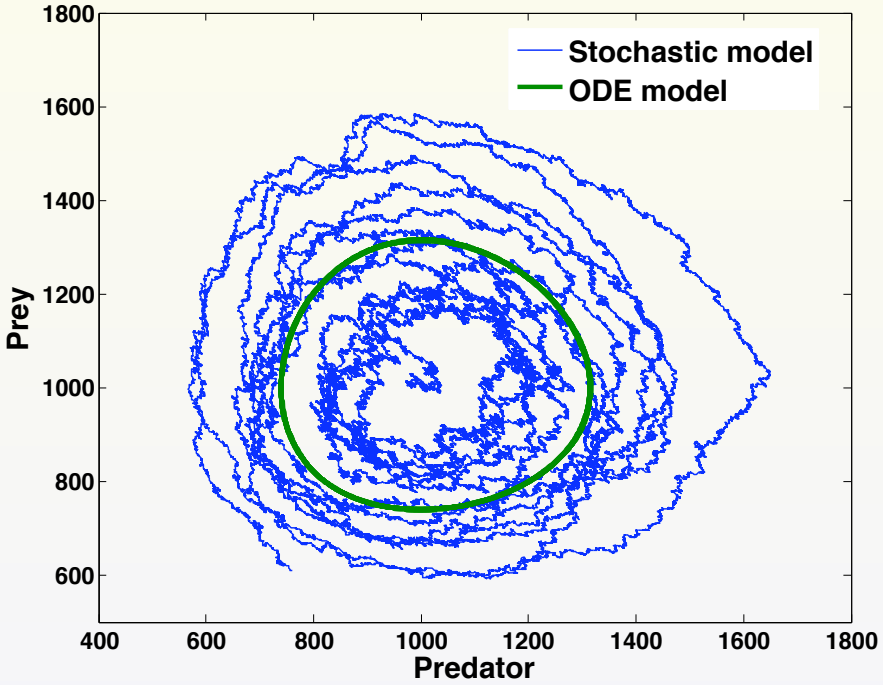
  one of the realizations of the stochastic model (i.e. one of the colonies under observation) has been completely wiped out, *something not possible in the deterministic modeling context*.

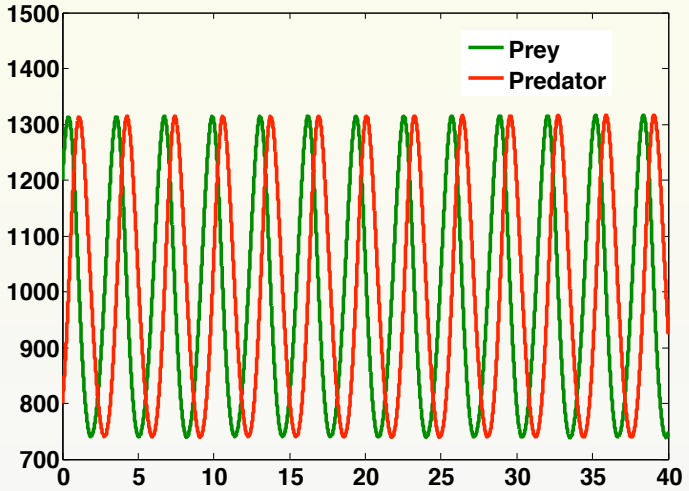
Example: Lotka-Volterra

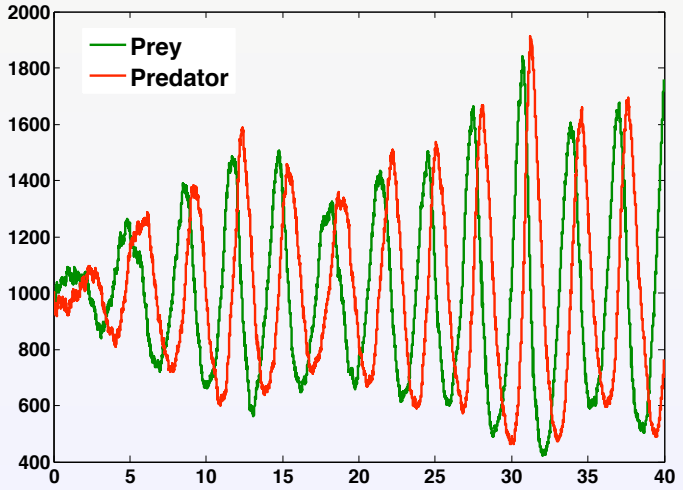
Think of A as a prey and B as a predator.

$$A \xrightarrow{\kappa_1} 2A, \qquad A + B \xrightarrow{\kappa_2} 2B, \qquad B \xrightarrow{\kappa_3} \emptyset,$$

with $A(0) = B(0) = 1000$ and $\kappa_1 = 2$, $\kappa_2 = .002$, $\kappa_3 = 2$.







Behavior is qualitatively different:

1. Deterministic always periodic while stochastic oscillates in random way.
2. Predator and/or prey will end up extinct in stochastic model.

Some questions

We start to see some of the different types of questions that become interesting in the stochastic context as opposed to the deterministic:

1. For a given birth and death rate in bacteria example, what is the probability that the colony will eventually die out?

2. For models in which extinction is eventually guaranteed: what is the expected amount of time before extinction?

3. If we know a stochastic processes X_t neither dies out, nor goes to infinity, and if $a < b$ are real numbers, then what is the probability that the value of the process is between a and b for very large t? That is, what is

$$\lim_{t \to \infty} \text{Prob}\{a \leq X_t \leq b\}?$$

4. How did I make those plots?

How to study stochastic models of intracellular processes?

The stochastic models should not be constructed by simply “adding noise” to the deterministic system.

In fact, the opposite is true; the stochastic models should be developed via first principles or through an understanding of the system.

The deterministic models should actually arise as a consequence of some limiting argument of the stochastic system.

Later in the course, I will show how such limiting arguments can be carried out thereby showing how some of the most well used ODE models in biology came to be.