Lecture 9: Stochastic Processes

Introduction to Stochastic Processes:

If we take a measurement of the state of a system, our measurement will always have some noise. We usually assume that the noise is due to measurement error, while the underlying process that defines the state of the system is deterministic. However, many systems are inherently stochastic, so that the actual system state, and not just our measurement of it, is stochastic. We call this situation a **stochastic process**, and must use **stochastic differential equations (SDEs)** to model it mathematically.

The dependent variables in a stochastic process are **random variables**.

We will focus on two common types of stochastic processes:

Brownian dynamics is the stochastic movement of small particles that are subjected to diffusive forces as well as deterministic forces. For Brownian dynamics, the dependent variables are the positions of the particle or particles. Position can vary continuously to take on any real value within the geometry of the system, so these are **continuous random variables**. We refer to continuous stochastic processes as **diffusive processes**.

Stochastic reaction equations are the stochastic reactions between discrete particles, such as molecules, cells, or organisms. The dependent variables are the number of particles in a category, so can only take on discrete values such as nonnegative integers, and thus are referred to as **discrete random variables**. Examples include (bio)chemical reactions, events involving cells such as infection of a blood cell by a virus, and cell division, differentiation or death, or even events involving organisms, such as becoming infected or resistant to a disease, or predator-prey interactions. Because the discrete random variables either do not change at all or jump in value suddenly, we call these **jump processes**. The number of particles are random variables and the number of events is a random process because we only know the rate at which we expect reactions to occur, but we don't know when any given particle will react.

All stochastic processes involve deterministic and stochastic terms. For example, we can write $\frac{d\vec{x}}{dt} = \vec{f}(\vec{x},t) + \vec{g}(\vec{x},t)$, where $\vec{f}(\vec{x},t)$ is the deterministic term and $\vec{g}(\vec{x},t)$ is the stochastic term. This is the general form for an SDE. Note that we use the vector notation for the dependent variable $\vec{x}$ as well the functions of the SDE to remind us that we can have multiple dependent variables and will then need a SDE for each. This week and next week we will learn how to simulate SDEs.

Deciding whether to use SDEs vs ODEs or PDEs.

Since all particles are subjected to the random collisions of molecules that cause diffusive forces, and all reactions occur stochastically for a given particle, all processes are actually stochastic processes. However, SDEs are very expensive, since stochastic simulations are slower than deterministic, and since we get a different result every time we solve the identical equations, so we need to run many simulations and obtain statistics on the solutions to understand the system behavior. Therefore, we only want to use SDEs if we need them. An obvious question is

how the solution $\vec{x}(t)$ to the SDE $\frac{d\vec{x}}{dt} = \vec{f}(\vec{x},t) + \vec{g}(\vec{x},t)$ compares to the solution to the deterministic equation $\frac{d\vec{x}}{dt} = \vec{f}(\vec{x},t)$. If the accumulated effect of $\vec{g}(\vec{x},t)$ is negligible compared to the value of $\vec{x}(t)$, or more specifically, is negligible compared to our measurement error, then we can ignore the stochastic terms and use ODEs. How do we estimate the effects of the stochastic terms to decide this?

For Brownian motion, the Particle Reynold's Number, which is a ratio of inertia to diffusion, tells us when we can neglect diffusion. The general form of the particle Reynold's number is usually given as $Re_p = \frac{\rho V_p d_p}{\mu}$, where ρ and μ are the density and viscosity of the solution, while V_p and d_p are the velocity and the diameter of the particle. If $Re_p >> 1$, we can ignore diffusion. Alternatively, if we have enough particles, we can model the number of particles as a function of location, instead of the location of each particle. That is, we can model a diffusive process as a PDE, as we did recently. However, we can only do this if there is a large number of particles. For example, we discussed in class a particle subjected to deterministic gravitational force and a stochastic diffusion forces. We assumed there were a lot of particles, and used PDEs to model the positions of particles over time without ignoring diffusion. However, this would not have told us the position of any given particle. For that, we would need SDEs.

For stochastic reaction equations, consider the reaction $A + B \rightarrow C$ at rate k_{on}. We expect the reaction to occur at a rate $[A][B]k_{on}$, expressed in units of concentration per time. We can multiply this by the volume and a discrete time interval to ask how many reactions occur during that time interval, so we expect to get $\lambda = [A][B]k_{on}VT$ reactions occurring. How many reactions actually occur? To answer this question, we need to recall the following. The **Central Limit Theorem** states that, the sum of N random variables, each of which is distributed with mean μ and standard deviation σ, is normally distributed, with mean $\mu_N = N\mu$ and standard deviation n $\sigma_N = \sqrt{N}\sigma$. Thus the fractional standard deviation of the sum of N random variables is $\frac{1}{\sqrt{N}}\frac{\sigma}{\mu}$. This means that the process becomes more accurate as the number of molecules or events increases. In general, $\frac{\sigma}{\mu} < 1$ for most processes, so our fractional noise due to the inherent stochastic process is $< \frac{1}{\sqrt{N}}$. In short, if we have N > 100 particles, we can usually ignore the stochastic terms if we can tolerate 10% error, while N > 10,000 particles provides 1% error if we use ODEs.

The above calculations just tell us whether there is so little noise that the deterministic simulation provides an accurate description of the real system. However, even when the deterministic simulation fails to predict *measurable variability* in an inherently stochastic simulation, ODEs will still be sufficient for our purposes if the *mean of the stochastic simulation approaches the deterministic simulation, AND we only care about the mean, not the distribution or the error*. The engineer needs to decide whether the mean is sufficient information, or whether statistics are needed. For example, if we are simulating the number of fluorescent proteins in each cell, but are measuring the fluorescence of a population of cells, we are only measuring the mean value. In contrast, if we are analyzing images with many cells and calculating the fluorescence of each cell, and want to understand if measured cell-to-cell variation is stochastic or likely to reflect different localized conditions, then we need statistics. If the mean is sufficient

information, then we still need to determine whether the ODE will provide this information. While we will not prove this here, it can be shown that the mean of an SDE solution approaches the solution to the accompanying ODE when the equation is linear. While this may be true for some nonlinear systems, it cannot be assumed to be true in general for nonlinear systems.

In summary, we need to use SDEs when the stochastic terms are significant, AND we need to know the statistics of the solution OR the system is nonlinear so that the ODE may give fundamentally different behavior. We will see this latter situation when we study the stochastic form of the Lotka Volterre predator-prey model.

Continuous random variables and probability distributions

If X can be any real (or complex) number in a finite or infinite interval, than X is a continuous random variable. We refer to little x as a specific realization (value) of the random variable X. We define the **probability density function** $p(x)$ as the probability that $X = x$. For a continuous random variable, $p(x)$ has units of inverse the units of X. (eg, units 1/m if x is a position). Because the sum or integral of all probabilities is always 1, we know the following, which we can often use to scale p(x):

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

We can also define the cumulative distribution function $P(a) = prob(X \leq a) = \int_{-\infty}^{a} p(x)dx$. While $p(x)$ can have any shape, $P(x)$ always increases monotonically from 0 to 1.

We define the **expectation** $E(X) = \langle X \rangle = \mu_X$ of a random variable to be the mean value, which is calculated as:

$$\langle X \rangle = \int_{-\infty}^{\infty} xp(x)dx$$

In general, the **expectation of any function** of a random variable is:

$$E\big(f(X)\big) = \langle f(X) \rangle = \int_{-\infty}^{\infty} f(x)p(x)dx$$

Since integration is a linear operator, so is expectation: $\langle aX_1 + bX_2 \rangle = a\langle X_1 \rangle + b\langle X_2 \rangle$. However, nonlinear combinations of random variables cannot be calculated the same way: $\langle X^2 \rangle = \int_{-\infty}^{\infty} x^2 p(x)dx \neq \langle X \rangle^2$.

The variability of a random variable is described by the **variance**,

$$Var(X) = \sigma^2 = \langle (X - \mu_X)^2 \rangle$$

We can use the linear property of expectation to show that $\sigma^2 = \langle (X - \mu_X)^2 \rangle = \sigma^2 = \langle X^2 - 2X\mu_X + \mu_X^2 \rangle = \langle X^2 \rangle - 2\langle X \rangle \mu_X + \mu_X^2 = \langle X^2 \rangle - 2\mu_X^2 + \mu_X^2$. Thus $\sigma^2 = \langle X^2 \rangle - \mu_X^2$.

The **standard deviation** is the square root of the variance: $\sigma = \sqrt{(\langle X^2 \rangle - \mu_X^2)}$

There are three common distributions you should know for continuous variables.
The **uniform distribution** takes an equal probability between a and b, and zero probability anywhere else. The probability is uniform everywhere, and must integrate to 1, so

$$p(x) = \frac{1}{b-a}, for\ a \le x \le b$$

$$p(x) = 0, for\ x\ \le a, or\ x \ge b$$

You can use the definition of expectation and variance to show that

$$\mu_X = \frac{a+b}{2}, \sigma_X = \frac{b-a}{\sqrt{12}}$$

An example is the position of a particle in a confined space in absence of an external force.

Uniform distribution (Left) probability density function. (Right) cumulative distribution function, for the uniform distribution. Figure from http://mathworld.wolfram.com/

The **exponential distribution** decreases exponentially with a characteristic length $\frac{1}{\lambda}$, for $x > 0$ and zero probability x < 0. The probability is $p(x) = Cexp(-\lambda x)$ and we can show that $C = \lambda$ in order to integrate to 1. Thus:

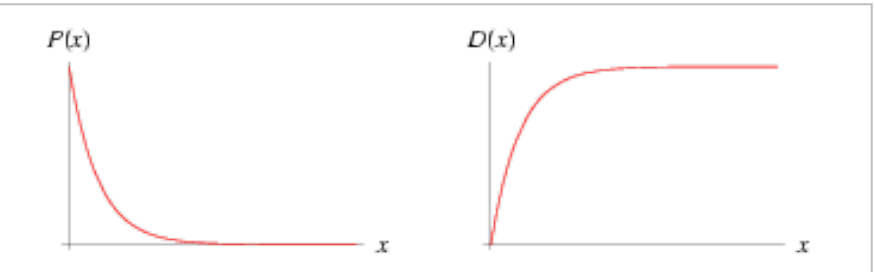
$$p(x) = \lambda \exp(-\lambda x), for\ x \ge 0$$

$$p(x) = 0, for\ x\ \le 0$$

You can use the definitions of expectation and variance to show that

$$\mu_X = \frac{1}{\lambda}, \sigma_X = \frac{1}{\lambda}$$

An example is the position of an oxygen molecule in the earth's atmosphere.



Exponential Distribution. (Left) probability density function. (Right) cumulative distribution function. Figure from http://mathworld.wolfram.com/

The **normal (Gaussian) distribution** is a bell–shaped curve. The probability density function is defined for all values of x:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Although the calculus is not trivial, one can use the definitions of expectation and variance to show that

$$\mu_X = \mu, \sigma_X = \sigma$$

Normal (Gaussian) Distribution. (Left) probability density function. (Right) cumulative distribution function. Figure from http://mathworld.wolfram.com/

The Langevan Equation: Modeling Brownian Dynamics

Recall that the Newtonian equation for motion in a deterministic system in its general form is the second-order differential equation, $m\frac{d^2\vec{x}}{dt^2} + \gamma\frac{d\vec{x}}{dt} = F(\vec{x}, t)$, where γ is the damping coefficient, m is the mass, and $F(x, t)$ is any deterministic force due.

If Brownian motion on the particles is important, then there is also a stochastic force in this equation, which is referred to as the **Langevan force**, and usually indicated by $L(t)$. That is, the velocity and acceleration display random fluctuations due to collisions of the particle with the surrounding fluid. The addition of the Langevan force to the equation of motion is referred to as the **Langevan equation**, which is an example of a stochastic differential equation.

$$m\frac{d^2\vec{x}}{dt^2} + \gamma\frac{d\vec{x}}{dt} = \vec{F}(\vec{x}, t) + \vec{L}(t)$$

The Langevan force in one dimension, which we will call here L(t), is independent of the Langevan forces in other dimensions and has the following properties:

1. The average fluctuation is zero: $\langle L(t)\rangle = 0$
2. Fluctuations and different times are uncorrelated: $\langle L(t_1)L(t_2)\rangle = 0$
3. The magnitude of fluctuations in constant over time: $\langle L(t)^2\rangle = \Gamma$

Numeric methods to solve SDEs: the Wiener Process.

Two additional assumptions are made that allow a numeric approximation of the Langevin equation:

1. $L(t)$ follows a gaussian (normal) distribution
2. $L(t)$ is white noise, meaning has equal intensity at all frequency bands.

In these conditions, the Wiener Process $W(\Delta t)$ can be defined as the integral of the Langevan force, which thus gives the accumulated noise in a finite time interval of length Δ:

$$W(\Delta t) = \int_0^{\Delta t} L(t)dt \,, or\ equivalently: dW(t) = L(t)$$

This means that if we jump forward by a timestep Δt, we have a function $W(\Delta t)$ that will add a stochastic term to our finite difference equation. We do not show this here, but it can be shown that $W(\Delta t)$ is normally distributed with mean 0 and variance σ^2 that is proportional to Δt. That is, the standard deviation in $W(\Delta t)$ goes with the square root of Δt.

We said in the introduction that a stochastic differential equation has the general form

$$\frac{d\vec{x}}{dt} = \vec{f}(\vec{x}, t) + \vec{g}(\vec{x}, t)$$

To solve this numerically, we need to switch to discrete time. The simplest algorithm for the deterministic part of this equation is the first-order Tayler expansion.

$$\vec{x}(t + \Delta\vec{t}) = \vec{x}(t) + \vec{f}(\vec{x}, t)\Delta t$$

This essentially integrates the change in $\vec{x}$ over time, assuming the derivative $\vec{f}(\vec{x}, t)$, is constant. To expand this methods to SDEs, we need to integrate $\vec{g}(\vec{x}, t)$ over time. As long as $\vec{g}(\vec{x}, t)$ is the Langevin force with the 5 assumptions made above, then the integral is $W(\Delta t)$, so our stochastic finite difference equation is:

$$\vec{x}\left(t + \Delta\vec{t}\right) = \vec{x}(t) + \vec{f}(\vec{x}, t)\Delta t + W(\Delta t)$$

Note that the Wiener term may vary with position and time and still fulfill the 5 requirements above as long as it does not change significantly within the time step Δt. It is common to scale $W(\Delta t)$ to be normal distributed with mean 0 and variance Δt, which is indicated in brief as $N(0, \Delta t)$, and then to scale this with a function $\vec{g}(\vec{x}, t)$. This results in

$$\vec{x}\left(t + \Delta\vec{t}\right) = \vec{x}(t) + \vec{f}(\vec{x}, t)\Delta t + \vec{g}(\vec{x}, t)W(\Delta t)$$

In summary, the Wiener Process is the discrete form of the Langevin equation. While the Langevin equation allows analytic solutions to some problems, the Wiener process allows numeric approximations to most problems.

In MATLAB, we use the `randn` function to return a random variable that is Gaussian distributed with mean 0 and variance 1, thus standard deviation 1. `randn(1,N)` returns a vector N such independent random variables. `mu+sigma*randn` returns a random variable with mean mu and standard deviation sigma. Thus, you can define `W = sqrt(delt)*randn(1,N)` to get N values of the normalized Wiener process for a time step delt.

Analyzing SDEs. When we run a stochastic simulation, the result will be different each time. While a single run or comparison of several runs provides information about the nature of the outcome of interest, we only obtain a good understanding by running the simulation many times and taking statistics of the results. Useful statistics include:

- The mean value of an outcome of interest over time. In MATLAB, you can use the `mean` function.
- The standard deviation of an outcome of interest over time. In MATLAB, you can use the `std` function.

- A histogram that shows the complete distribution of an outcome of interest at a given time. In MATLAB, you can use the `hist` or the `histogram` functions.

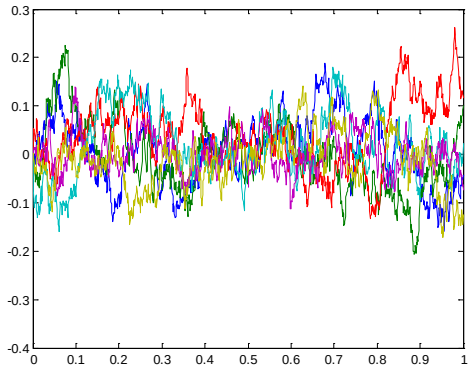
Verifying SDEs. The statistics of multiple simulations should be essentially the same if we use a different time step, as long as the step is not too big for this system. Of course, the statistics (mean and standard deviation) will not be identical for multiple runs, so what do we mean by essentially identical? Each simulation provides a mean, an standard deviation, and a standard error of the mean (standard deviation divided by the square root of the number of simulations). Briefly, the two means should be approximately within one to two times the standard error of the mean. More rigorously, you should not be able to eliminate the null hypothesis that the two means are the same. Another verification that can be used in certain situations is that for a linear system (but not necessarily for a nonlinear system), the mean of the stochastic simulation should be the same as the deterministic simulation.

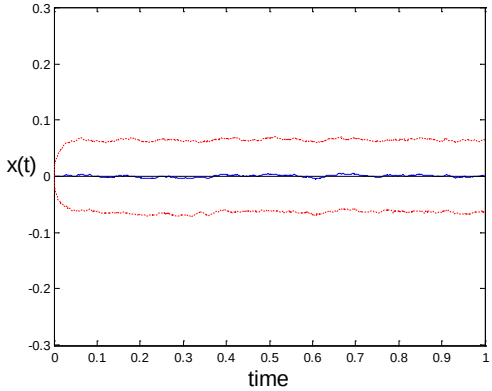
Example of a Brownian dynamics simulation:

Consider a particle attached by a spring at low Reynold's numbers. At low Reynold's numbers, we can neglect inertia and thus acceleration, and Newton's laws of motion simplify to the first order differential equation in position. Note that we have moved the drag coefficient to the right-hand size of the equation to isolate the derivative on the left-hand side:

$$\frac{d\vec{x}}{dt} = \frac{F}{\gamma}(\vec{x}, t) + \frac{L(t)}{\gamma}$$

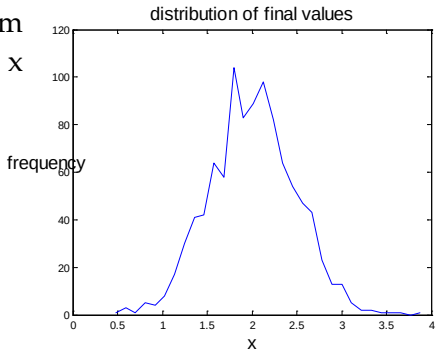
The result is that the particle moves stochastically within a well. Sample trajectories with initial condition x = 0 are shown below on the left, while then mean (blue) and mean +/- standard deviation (red) are shown on the right for multiple runs.





To plot a histogram of locations, we can use the histogram function as shown below, for the final value of the variable x for many runs:

```
[freq,desc]=histogram(x(:,end)')
X = linspace(desc(1),desc(2),desc(3))
plot(X,freq)
```



Summary of Stochastic Processes

1. A stochastic process can be modeled with a **stochastic differential equation** $\frac{d\vec{x}}{dt} = \vec{f}(\vec{x},t) + \vec{g}(\vec{x},t)$, where $\vec{f}(\vec{x},t)$ is the deterministic term and $\vec{g}(\vec{x},t)$ is the stochastic term.
2. The total probability equals one, so $\int_{-\infty}^{\infty} p(x)dx = 1$ for a continuous random variable and $\sum_{-\infty}^{\infty} p(x) = 1$ for a discrete random variable.
3. The **expectation** of a random variable is the sum or integer of the value of the variable times its probability: $\langle X \rangle = \sum_{-\infty}^{\infty} xp(x)$ or $\langle X \rangle = \int_{-\infty}^{\infty} xp(x)dx$
4. The expectation of a function f of a random variables is the sum/integer of that function weighted by the probability: $E\big(f(X)\big) = \sum_{-\infty}^{\infty} f(x)p(x)$ or $E\big(f(X)\big) = \int_{-\infty}^{\infty} f(x)p(x)dx$
5. The **variance** of a random variable is $Var(X) = \sigma^2 = \langle (X - \mu_X)^2 \rangle = \langle X^2 \rangle - \mu_X^2$
6. The **Central Limit Theorem** states that, the sum of N random variables, each of which is distributed with mean μ and standard deviation σ, is normally distributed, with mean $\mu_N = N\mu$ and standard deviation n $\sigma_N = \sqrt{N}\sigma$.
7. The **exponential distribution** is a decay curve; $p(x) = \lambda \exp(-\lambda x), for\ x \geq 0$ and has $\mu_X = \frac{1}{\lambda}, \sigma_X = \frac{1}{\lambda}$.
8. **The normal (Gaussian) distribution** is a bell–shaped curve; $p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ and has $\mu_X = \mu, \sigma_X = \sigma$.
9. The **Langevan force** is the random forces caused by collision with solvent molecules, and has zero mean but a constant magnitude of fluctuation over time, with no correlation between fluctuations at different times.
10. The **Wiener term** is the integral of the Langevan force over a discrete time: $W(\Delta t) = \int_0^{\Delta t} L(t)dt$.
11. The Wiener tem $W(\Delta t)$ is normally distributed with mean 0 and a variance that is proportional to Δt. That is, the standard deviation in $W(\Delta t)$ is proportional to $\sqrt{\Delta t}$.
12. A simple algorithm to simulate a stochastic differential equation calculates the new system state from the old state as follows: $\vec{x}\big(t + \Delta\vec{t}\big) = \vec{x}(t) + \vec{f}(\vec{x},t)\Delta t + \vec{g}(\vec{x},t)W(\Delta t)$. (This is for a SDE in the form defined in point 1.)
13. You can't determine whether the time step is too long by testing how much the system changed *after* implementing the stochastic step, since this will bias the system against larger changes. Instead, you need to use the expected change to evaluate the time step.