

## Conversion of a discrete model to an SDE

The techniques we will introduce today require a stochastic differential equation of the following form:

$$dx = \mu(x, t) dt + \sigma(x, t) dW_t,$$

where  $\mu(x, t)$  reflects the rate of change of the deterministic system and  $\sigma(x, t)$  is the standard deviation of the magnitude of the oscillations. For Poisson processes we assume  $\sigma(x, t) = \sqrt{\mu(x, t)}$ , but how do we find  $\sigma(x, t)$  more generally? For example, what happens if our system has multiple interdependent variables?

The following example to demonstrate this technique is based on McKane and Newman (2005). They studied a variant of the Lotka-Volterra predator-prey model. You may have heard of a popular application of this model to the sale of lynx and hare pelts to the Hudson Bay Company in the 1800s and early 1900s (see figure below). The story of its origin is even more interesting than that, however. Lotka originally developed the model in the context of autocatalytic chemical reactions. At the time (1910) there was a general resistance to the notion that a chemical reaction system could show oscillations, even in the presence of unlimited reagents; when Lotka published his model he noted that no such reaction scheme was known to chemistry. When Belousov and Zhabotinsky discovered such a chemical reaction system in the 1950s/60s, the general attitude toward the possibility of oscillatory chemical systems was so poor that they fought to get their work published. Volterra also did not originally study pelts, but rather took an interest due to his son-in-law's involvement in the fishing industry.

For a predator  $x$  and prey  $y$ , the deterministic system is:

$$\frac{dx}{dt} = p_1xy - d_1x \quad \frac{dy}{dt} = (b - d_2)y - (p_1 + p_2)xy$$

where  $b$  is the birth rate of prey,  $d_1$  and  $d_2$  are the rates of death by natural causes,  $p_1 + p_2$  is the rate at which  $x$  kills  $y$ , and  $p_1 < p_1 + p_2$  is the birth rate of  $x$  due to the predation. Using fixed point analysis discussed earlier in the course, we could show that  $x$  and  $y$  oscillate around the fixed point but that any deviation will change the amplitude of their trajectory, just like the simple harmonic oscillator we studied earlier. Common sense tells us that this is not the way predator-prey cycles work in the real world: the real system likely approaches a limit cycle.

A modification of this approach attempts to limit the prey population by introducing a carrying capacity:

$$\frac{dx}{dt} = p_1xy - d_1x \quad \frac{dy}{dt} = by\left(1 - \frac{y}{K}\right) - d_2y - (p_1 + p_2)xy$$

Now  $y$  will not grow larger than  $K$ . Unfortunately, this system doesn't approach a limit cycle: it shows damped oscillations. Like an example on your problem set seven, McKane and Newman were able to show that noise drives oscillations in the latter system. They used both a discrete model and its Langevin equivalent, providing a great opportunity to show the interconversion.

Let's begin with the discrete model. The authors suppose that the environment permits up to  $N$  organisms: physically we can envision that water or some other shared resource is limiting for growth. Some of the  $N$  "sites" are empty ( $E$ ), others occupied by predators ( $X$ ) or prey ( $Y$ ). Then the "reactions" and their propensities are given by the table below. Note that the system size  $\Omega = N$  and the

propensity is  $Nr(\mathbf{x}, N)$ . While it seems a bit odd, we divide the number of each animal by  $N$  to get its “concentration,” determine the rate of the reaction from this value, then multiply through by  $N$  again to convert back to appropriate units. “Reactions” with two reactants therefore have a factor of  $1/N$  in their propensity.

Reaction	Stoichiometry column $s$	Propensity	$ss^T$
$Y + E \xrightarrow{b} Y + Y$	$s_1 = (0, 1)^T$	$P_1 = \frac{by(N-x-y)}{N}$	$s_1 s_1^T = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$
$X \xrightarrow{d_1} E$	$s_2 = (-1, 0)^T$	$P_2 = d_1 x$	$s_2 s_2^T = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$
$Y \xrightarrow{d_2} E$	$s_3 = (0, -1)^T$	$P_3 = d_2 y$	$s_3 s_3^T = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$
$X + Y \xrightarrow{p_1} X + X$	$s_4 = (1, -1)^T$	$P_4 = \frac{p_1 xy}{N}$	$s_4 s_4^T = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$
$X + Y \xrightarrow{p_2} X + E$	$s_5 = (0, -1)^T$	$P_5 = \frac{p_2 xy}{N}$	$s_5 s_5^T = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$

At this point we have everything we need for a Gillespie simulation, but let's see what this system would look like in Langevin form. The mean change in  $(x, y)^T$  would be:

$$\mu(x, y, t) = \sum_{i=1}^5 s_i P_i = \begin{pmatrix} \frac{p_1 xy}{N} - d_1 x \\ \frac{by(N-x-y)}{N} - d_2 y - \frac{(p_1+p_2)xy}{N} \end{pmatrix}$$

Similarly, the variance in change in  $(x, y)^T$  is:

$$\sigma^2(x, y, t) = \sum_{i=1}^5 s_i s_i^T P_i = \begin{pmatrix} d_1 x + \frac{p_1 xy}{N} & -\frac{p_1 xy}{N} \\ -\frac{p_1 xy}{N} & \frac{by(N-x-y)}{N} + d_2 y + \frac{(p_1+p_2)xy}{N} \end{pmatrix}$$

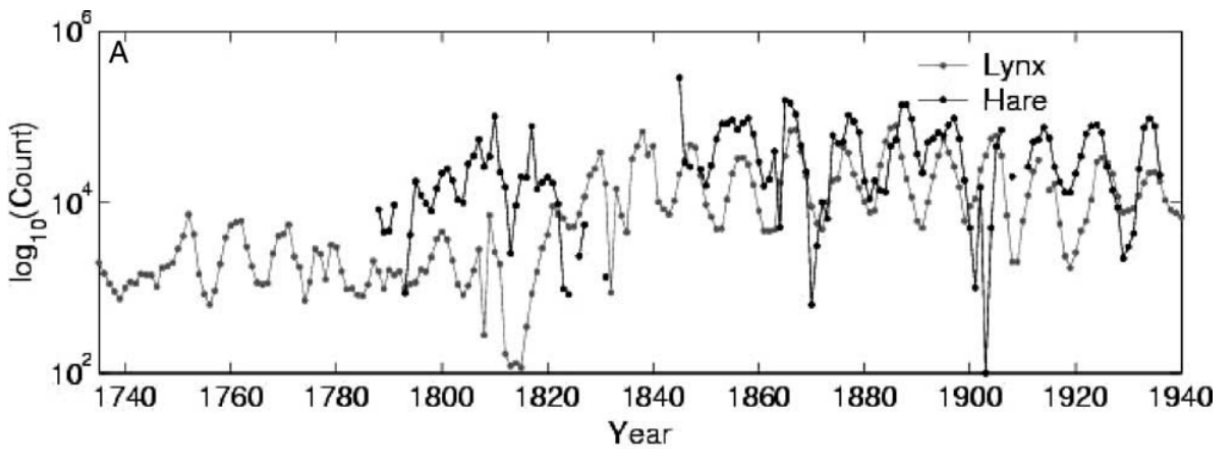
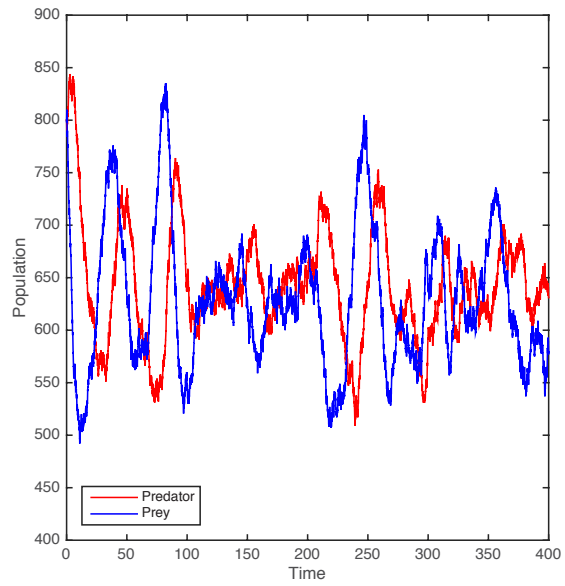
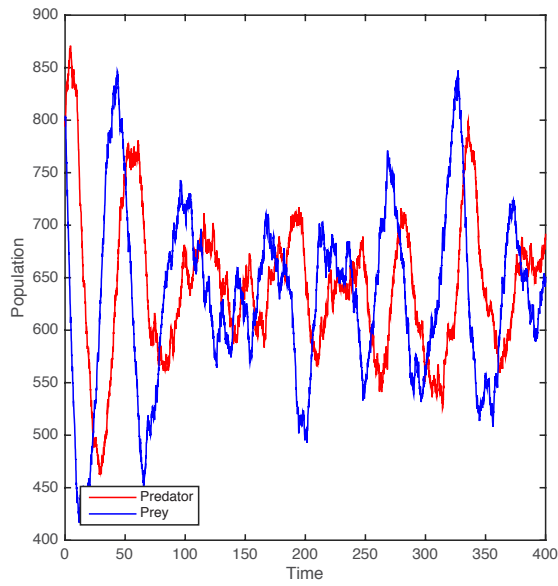
The Langevin form is therefore:

$$d \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{p_1 xy}{N} - d_1 x \\ \frac{by(N-x-y)}{N} - d_2 y - \frac{(p_1+p_2)xy}{N} \end{pmatrix} dt + \begin{pmatrix} d_1 x + \frac{p_1 xy}{N} & -\frac{p_1 xy}{N} \\ -\frac{p_1 xy}{N} & \frac{by(N-x-y)}{N} + d_2 y + \frac{(p_1+p_2)xy}{N} \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix}$$

where  $W_t^1$  and  $W_t^2$  are Wiener processes. To simulate the system, we use the following update strategy:

$$\Delta \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{p_1 xy}{N} - d_1 x \\ \frac{by(N-x-y)}{N} - d_2 y - \frac{(p_1+p_2)xy}{N} \end{pmatrix} \Delta t + \begin{pmatrix} d_1 x + \frac{p_1 xy}{N} & -\frac{p_1 xy}{N} \\ -\frac{p_1 xy}{N} & \frac{by(N-x-y)}{N} + d_2 y + \frac{(p_1+p_2)xy}{N} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \sqrt{\Delta t}$$

where  $\eta_1, \eta_2 \sim \mathcal{N}(0, 1)$  are random numbers drawn independently. Both types of simulations give approximately equivalent results (Langevin on the left, Gillespie on the right, data on the sale of lynx and hare pelts reproduced from Pineda-Krch et al. (2007):



By determining the power spectrum of e.g. the predator from these equations, McKane and Newman showed that there is a resonant frequency for this system: random noise that happens to occur at that frequency gets amplified.

## Fokker-Planck equation

Two more handy tricks, both concerning the probability density function of states  $p(x, t)$ , are worth knowing about should you ever need them. For a differential equation of the form

$$dx = \mu(x, t) dt + \sigma(x, t) dW_t,$$

an approximate expression for the time evolution of  $p(x, t)$  is given by the Fokker-Planck equation:

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial}{\partial x} [\mu(x, t)p(x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [\sigma^2(x, t)p(x, t)]$$

Why is this so? Consider the continuous version of the chemical master equation (aka the forward Kolmogorov equation), with  $r(a, b)$  the rate at which the system changes by an amount  $a$  from an initial state  $b$ :

$$\frac{\partial p(x, t)}{\partial t} = \int [p(s, t)r(x-s, s) - p(x, t)r(s-x, x)] ds$$

Introducing a new variable  $y = x - s$ , this becomes:

$$\frac{\partial p(x, t)}{\partial t} = \int [p(x - y, t)r(y, x - y) - p(x, t)r(y, x)] dy$$

Consider the Taylor expansion of the term on the left in the integrand:

$$p(x - y, t)r(y, x - y) = p(x, t)r(y, x) - y \frac{\partial}{\partial x} [p(x, t)r(y, x)] + \frac{y^2}{2!} \frac{\partial^2}{\partial x^2} [p(x, t)r(y, x)] + \dots$$

Keeping the first three terms (e.g. assuming that  $p$  is slowly varying), this becomes:

$$\begin{aligned} \frac{\partial p(x, t)}{\partial t} &= \int \left[ -y \frac{d}{dx} [p(x, t)r(y, x)] + \frac{y^2}{2!} \frac{d^2}{dx^2} [p(x, t)r(y, x)] \right] dy \\ &= -\frac{\partial}{\partial x} p(x, t) \left[ \int y r(y, x) dy \right] + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[ p(x, t) \int y^2 r(y, x) dy \right] \end{aligned}$$

which is nothing other than the Fokker-Planck equation with

$$\mu(x, t) = \int y r(y, x) dy \quad \sigma^2(x, t) = \int y^2 r(y, x) dy$$

These integrals are sometimes called the first and second *jump moments*: their equality to  $\mu(x, t)$  and  $\sigma^2(x, t)$  can be seen by calculating  $\langle \Delta x \rangle$  and  $\langle \Delta x^2 \rangle$  from the Langevin equation (see lecture notes from Scott if interested).

### Drift in population genetics

Suppose we start a culture of bacteria by mixing two strains, one red and one green but otherwise identical, in a 1:1 ratio to contain  $N$  cells total. We allow the culture to grow exponentially then dilute it back to just  $N$  cells<sup>1</sup>. We will assume that no mutations occur during the experiment so that every cell has an equal chance of leaving offspring.

We understand intuitively that the ratio of green to red cells in our culture will fluctuate with each dilution step and will eventually reach either all red or all green, where it will remain: the endpoints will on average be reached faster if the population is small. To understand better how fast this happens, let's rewrite this as a stochastic differential equation in  $x = n/N$ , the fraction of green cells.

Let  $n$  be the number of green cells after the last round of dilution. The probability that we have  $m$  green cells after the next dilution is:

$$P(m | n, N) = \binom{N}{m} \left( \frac{n}{N} \right)^m \left( 1 - \frac{n}{N} \right)^{N-m}$$

i.e., the number of green cells is binomially distributed with parameters  $N$  and  $x = n/N$ . Using known properties of the binomial distribution:

$$\langle m \rangle = Nx = n \quad \langle m^2 \rangle - \langle m \rangle^2 = Nx(1-x) - n^2 = \frac{n}{N} (N - n) - n^2$$

Therefore the first and second moments of change in number of green cells in one round are:

$$\mu = \langle m - n \rangle = 0 \quad \sigma^2 = \langle (m - n)^2 \rangle = \langle m^2 \rangle - 2n \langle m \rangle + \langle m \rangle^2 = \frac{n}{N} (N - n)$$

<sup>1</sup>If we assume that all cells leave the same number of descendants during the growth phase, then dilution is the equivalent of picking balls from an  $N$ -ball urn with replacement.

Now we have everything needed to write our SDE:

$$dx = dW_t \sqrt{\frac{x(1-x)}{N}}$$

The formula applies until  $x$  reaches 0 or 1, at which point it will remain there<sup>2</sup>. The change in  $x$  was indeed a random walk as we suspected, but it may not have been obvious that the step sizes were changing in this way. Plugging into the Fokker-Planck equation, we see that:

$$\frac{\partial p(x, t)}{\partial t} = \frac{1}{2N} \frac{\partial^2}{\partial x^2} [x(1-x)p(x, t)]$$

## Wright's formula

The stationary solutions of the Fokker-Planck equation satisfy:

$$\begin{aligned} 2\mu(x, t)p_{ss}(x, t) &= \frac{\partial}{\partial x} [\sigma^2(x, t)p_{ss}(x, t)] \\ \frac{2\mu(x, t)}{\sigma^2(x, t)} &= \frac{1}{\sigma^2(x, t)p_{ss}(x, t)} \frac{\partial}{\partial x} [\sigma^2(x, t)p_{ss}(x, t)] = \frac{\partial}{\partial x} \ln [\sigma^2(x, t)p_{ss}(x, t)] \\ 2 \int^x \frac{\mu(s, t)}{\sigma^2(s, t)} ds &= \ln [\sigma^2(x, t)p_{ss}(x, t)] + C \\ p_{ss}(x, t) &\propto \frac{1}{\sigma^2(x, t)} \exp \left[ 2 \int^x \frac{\mu(s, t)}{\sigma^2(s, t)} ds \right] \end{aligned}$$

The probability distribution should of course be normalized. This formula for finding stationary solutions is sometimes attributed to Wright (1938). The good news is that you can find stationary solutions using regular calculus, even for an SDE. You will apply this approach in problem set eight to study the spread of disease.

Although we will go no further in this introductory course, I hope this discussion has piqued your interest in the study of stochastic differential equations or, if nothing else, conveyed a respect for their proper husbandry.

---

<sup>2</sup>Calculations of expected persistence times and related statistics are beyond the scope of this course but of obvious relevance to conservation biology and population genetics.