

## 4 Random Variations

The differential equations that we've worked with so far in these lectures are deterministic. You set the initial conditions and what then follows is set in stone. I know some people whose lives are like that, but most things in the biological world are not. Instead, important features of our lives are dictated by randomness, the kind of event that, to quote the classics, blindsides you at 4pm on an idle Tuesday.

The purpose of this section is to learn how to incorporate such random fluctuations into our equations. We will do this by studying the evolution of probability distributions over the space of outcomes. Throughout, we will take time to be continuous but the outcomes themselves may be either discrete or continuous. In the latter case, the probability distribution will be governed by the famous Fokker-Planck equation.

### 4.1 Discrete Outcomes

We start by considering the situation where the possible outcomes are discrete. We will build up slowly, first considering just two possible outcomes, then  $\infty$ , then  $\infty^2$ .

#### 4.1.1 Two Outcomes

Suppose that there are just two possible states in our system,  $A$  and  $B$ . We would like to understand the probability  $P(A, t)$  to be in state  $A$  and the related probability  $P(B, t) = 1 - P(A, t)$  to be in state  $B$ .

For this, we need to stipulate the underlying dynamics which tells us how the system evolves between  $A$  and  $B$ . This too will be probabilistic. We will assume that we are dealing with a *Markov process*, meaning that the probability to transition from one state to the other depends only on the current state. For simplicity, we consider the following rules.

- If in state  $A$ , the system has a probability per unit time  $\lambda$  to transition to state  $B$ .
- Once in state  $B$ , the system stays there.

The real purpose of this warm-up example is to understand what we mean by “probability per unit time”. In a short time  $\delta t$ , the probability that we jump from  $A$  to  $B$  is  $\lambda \delta t$ . Equivalently, in the same short time  $\delta t$ , the probability that we remain in state  $A$  is  $(1 - \lambda)\delta t$ .

From this information, we can write down a differential equation that governs the probability. If we know  $P(A, t)$  at time  $t$  then, at time  $t + \delta t$ , the probability is

$$P(A, t + \delta t) = P(A, t)(1 - \lambda)\delta t . \quad (4.1)$$

If we now Taylor expand the left-hand side, we have

$$P(A, t) + \frac{dP(A, t)}{dt}\delta t = P(A, t)(1 - \lambda\delta t) \implies \frac{dP(A, t)}{dt} = -\lambda P(A, t) . \quad (4.2)$$

This is easily solved. If we start off most definitely in state  $A$ , so  $P(A, 0) = 1$ , then we have

$$P(A, t) = e^{-\lambda t} . \quad (4.3)$$

We see that a constant probability per unit time to jump from  $A$  to  $B$  means an exponential depletion of  $A$ .

There are further questions that we can ask of this simple system. We could, for example, ask for the probability distribution  $f(t)$  for the time  $t$  that we make the jump from  $A$  to  $B$ . To get this, we first consider the probability that we made the jump at some time  $t < T$ ,

$$\text{Prob}[t < T] = \int_0^T f(t) dt . \quad (4.4)$$

But this can be identified with the probability that we're in state  $B$  at time  $T$ ,

$$\text{Prob}[t < T] = P(B, T) = 1 - e^{-\lambda T} . \quad (4.5)$$

Equating these two expressions and differentiating (and, perhaps confusingly, replacing the dummy variable  $T$  with  $t$ ), gives

$$f(t) = \frac{dP(B, t)}{dt} = \lambda e^{-\lambda t} . \quad (4.6)$$

This is the probability distribution. It obeys

$$\int_0^\infty f(t) dt = 1 \quad (4.7)$$

as probability distributions should. From this, we can easily round up the usual statistical suspects. The expected time to make the jump is

$$\langle t \rangle = \int_0^\infty t f(t) dt = \frac{1}{\lambda} . \quad (4.8)$$

Meanwhile, the variance is given by

$$\text{Var}(t) = \langle t^2 \rangle - \langle t \rangle^2 = \int_0^\infty t^2 f(t) dt - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \quad (4.9)$$

We see that the standard deviation  $\sigma(t) = \sqrt{\text{Var}(t)} = 1/\lambda$  is the same as the mean. This is telling us that fluctuations are important in this system. If we knew only about the average time to jump, this wouldn't agree particularly well with observations in any given case.

#### 4.1.2 Discrete Population Size

With this simple example under our belts, let's now turn to a situation where the outcomes are labelled by  $n \in \mathbb{N} = \{0, 1, 2, \dots\}$ . (As an aside: mathematicians can't make up their minds whether or not zero is a natural number. Here I have decided for them.) We can think of  $n$  as labelling the population size.

Again, we need to specify the dynamics of the system. Here we take constant probability per unit time  $\lambda$  to jump from  $n$  to  $n + 1$ . This is known as a *Poisson process*.

(You might reasonably argue that it is unrealistic for a population to jump from  $n = 0$  to  $n = 1$ . You might, for that matter, argue that it's equally unrealistic for most populations to jump from  $n = 1$  to  $n = 2$ . If you're worried, think "immigration" rather than "birth".)

Again, we can translate this statement into a differential equation. We have, for  $n \geq 1$ ,

$$P(n, t + \delta t) = (1 - \lambda \delta t)P(n, t) + \lambda \delta t P(n - 1, t). \quad (4.10)$$

Here the first term captures the probability that we remain in state  $n$ , while the second captures the probability that we jump up from state  $n - 1$ . Taylor expanding the left-hand side then gives us the differential equation

$$\frac{dP(n, t)}{dt} = \lambda [P(n - 1, t) - P(n, t)]. \quad (4.11)$$

Equations like (4.11) (or (4.2)) that govern the evolution of a probability distribution are called, rather pompously, *master equations*. In the present case, it is a differential equation in  $t$  and a difference equation in  $n$ .

## The Generating Function

The most systematic way to solve the master equation (4.11) is to introduce the *generating function*

$$\phi(s, t) = \sum_{n=0}^{\infty} s^n P(n, t) = \langle s^n \rangle \quad (4.12)$$

where we think of  $s \in [0, 1]$ , ensuring convergence. The generating function is a lovely object that cleanly captures many of the things we most care about in the distribution. For example, the average population size is

$$\langle n(t) \rangle = \sum_{n=0}^{\infty} n P(n, t) = \left. \frac{\partial \phi(s, t)}{\partial s} \right|_{s=1} . \quad (4.13)$$

Similarly,

$$\begin{aligned} \langle n^2(t) \rangle &= \sum_{n=0}^{\infty} n^2 P(n, t) \\ &= \sum_{n=0}^{\infty} n(n-1) P(n, t) + \sum_{n=0}^{\infty} n P(n, t) \\ &= \left. \frac{\partial^2 \phi(s, t)}{\partial s^2} \right|_{s=1} + \left. \frac{\partial \phi(s, t)}{\partial s} \right|_{s=1} . \end{aligned} \quad (4.14)$$

Combining these, we can extract the standard deviation.

If we know the generating function then we can easily reconstruct the probability distribution by differentiating

$$P(n, t) = \frac{1}{n!} \left. \frac{\partial^n \phi(s, t)}{\partial s^n} \right|_{s=0} . \quad (4.15)$$

Finally, the generating function satisfies a boundary condition at  $s = 1$  that comes from the observation that

$$\phi(1, t) = \sum_{n=0}^{\infty} P(n, t) = 1 . \quad (4.16)$$

There may be an additional boundary condition at  $t = 0$  coming from an initial condition on the probability distribution.

With this in mind, let's now return to our master equation (4.11) and use it to construct a differential equation for the generating function. We have

$$\begin{aligned}
\frac{\partial \phi(s, t)}{\partial t} &= \sum_{n=0}^{\infty} s^n \frac{\partial P(n, t)}{\partial t} \\
&= \lambda \sum_{n=0}^{\infty} s^n [P(n-1, t) - P(n, t)] \\
&= \lambda \sum_{n=0}^{\infty} [s^{n+1} P(n, t) - s^n P(n, t)] \\
&= \lambda(s-1)\phi(s, t) .
\end{aligned} \tag{4.17}$$

where, in the second line, we've used  $P(-1, t) = 0$ . This is a differential equation in  $t$ . We can simply integrate it, treating  $s$  as a constant to get

$$\phi(s, t) = \phi(s, 0)e^{\lambda(s-1)t} . \tag{4.18}$$

The function  $\phi(s, 0)$  is fixed by the initial probability distribution at time  $t = 0$ . We will take this to be  $P(n, 0) = \delta_{n,0}$ , meaning that everything kicks off at  $n = 0$  and  $P(0, 0) = 1$ . This gives  $\phi(s, 0) = 1$  and we have

$$\phi(s, t) = e^{\lambda(s-1)t} . \tag{4.19}$$

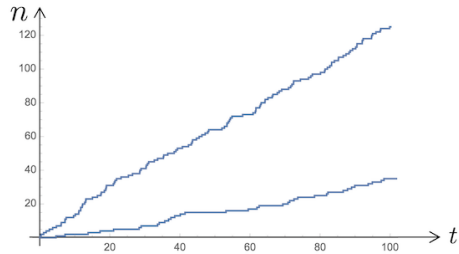
The associated probability density is then given by, using (4.15), by

$$P(n, t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} . \tag{4.20}$$

This is the *Poisson distribution*. The expectation and standard deviation can be computed from (4.13) and (4.14) and are given by

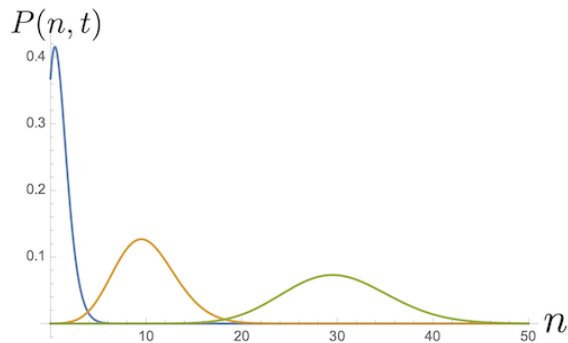
$$\langle n(t) \rangle = \lambda t \quad \text{and} \quad \sigma = \sqrt{\lambda t} . \tag{4.21}$$

We see that  $\sigma/\langle n \rangle = 1/\sqrt{\langle n \rangle}$ , meaning that fluctuations get less important over time as the population grows. Two Poisson processes, one with  $\lambda = 1.3$  and the other with  $\lambda = 0.3$ , are shown in the figure.



A plot of the probability distribution for different times  $t$  is shown in Figure 51. We can get an analytic handle on the evolution of the probability distribution if we invoke Stirling's approximation,

$$n! \approx \sqrt{2\pi n} n^n e^{-n} . \tag{4.22}$$



**Figure 51.** The march of probability, plotted here (for  $\lambda = 1$ ) for times  $t = 1$  (in blue),  $t = 10$  (in orange), and  $t = 30$  (in green).

This is proved, for example, in the lectures on [Statistical Physics](#). With this approximation, the probability distribution (4.21) can be written as

$$P(n, t) \approx \frac{1}{\sqrt{2\pi}} e^{-\lambda t} e^{g(n, t)} \quad \text{with} \quad g(n, t) = n + n \log(\lambda t) - n \log n - \frac{1}{2} \log n. \quad (4.23)$$

The function  $g(n, t)$  has a maximum at  $\partial g / \partial n = 0$  which, you can check, is given at large  $n$  by  $n = n^* \approx \lambda t$ . Expanding about this maximum gives us an approximate expression for the exponent

$$\begin{aligned} g(n, t) &\approx g(n^*) + \frac{1}{2} (n - n^*)^2 \frac{\partial^2 g}{\partial n^2} + \dots \\ &= \lambda t - \frac{(n - \lambda t)^2}{2\lambda t} - \frac{1}{2} \log(\lambda t). \end{aligned} \quad (4.24)$$

This then translates into a late time, large  $n$ , expression for the probability distribution:

$$P(n, t) \approx \frac{1}{\sqrt{2\pi\lambda t}} e^{-(n-\lambda t)^2/2\lambda t}. \quad (4.25)$$

We see that, at late times, the probability distribution settles down to a Gaussian distribution, marching forwards with linear growth  $n \approx \lambda t$ .

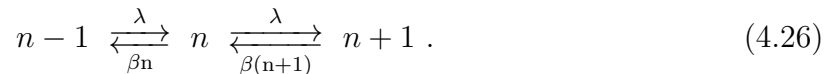
### 4.1.3 Birth and Death Again

We find ourselves turning once again, like the great Russian novelists, to the grand questions of life and death. This time, with a probabilistic slant.

We'll stick with our population model with states given by  $n \in \mathbb{N}$ . Now the dynamics includes the possibility for birth (or immigration) which increases  $n$  by one, and death which decreases  $n$  by one. The probability per unit time for these is:

- For  $n \rightarrow n + 1$ , we take a constant rate,  $\lambda$ .
- For  $n \rightarrow n - 1$ , we take a constant rate *per capita*, so that the rate is  $\beta n$ .

To avoid continually writing out these words, we summarise this in a reaction-like diagram



We will skip the step of writing  $P(n, t + \delta t)$  and just jump immediately to the master equation for the probability distribution which is

$$\frac{dP(n, t)}{dt} = -(\lambda + \beta n)P(n, t) + \lambda P(n - 1, t) + \beta(n + 1)P(n + 1, t) . \quad (4.27)$$

You can trace the origin of each of these terms to the diagram above. We'll take this equation to hold for  $n \geq 0$ , with the proviso that  $P(n = -1, t) = 0$ .

The generating function  $\phi(s, t)$  is again defined by (4.12). It obeys

$$\begin{aligned} \frac{\partial \phi(s, t)}{\partial t} &= \sum_{n=0}^{\infty} s^n \frac{\partial P(n, t)}{\partial t} \\ &= \sum_{n=0}^{\infty} s^n [ -\lambda P(n, t) + \lambda P(n - 1, t) - \beta n P(n, t) + \beta(n + 1)P(n + 1, t) ] \\ &= \sum_{n=0}^{\infty} [ -s^n \lambda + s^{n+1} \lambda - s^n \beta n + s^{n-1} \beta n ] P(n, t) \end{aligned} \quad (4.28)$$

where, in the final line, we've shifted the summation variable to gather all terms of the form  $P(n, t)$ . It's simple to write the first two terms using the generating function: they are proportional to  $\phi(s, t)$  and  $s\phi(s, t)$  respectively. For the second two terms, we have an extra factor of  $n$  in the sum. This arises by differentiating the generating function

$$\frac{\partial \phi(s, t)}{\partial s} = \sum_{n=0}^{\infty} n s^{n-1} P(n, t) . \quad (4.29)$$

In this way, having transition rates that are proportional to  $n$ , like the  $\beta$  rates above, leads to a partial differential equation for the generating function,

$$\frac{\partial \phi(s, t)}{\partial t} = (s - 1) \left( \lambda \phi(s, t) - \beta \frac{\partial \phi(s, t)}{\partial s} \right) . \quad (4.30)$$

Note that we have an overall factor of  $s - 1$ . This should be expected because, as shown in (4.16), we have  $\phi(1, t) = 1$  for all  $t$ , so  $\partial \phi / \partial t$  should vanish at  $s = 1$ .

We're left with (4.30) to solve. To do this, we make the (not immediately obvious) ansatz

$$\phi(s, t) = \exp((s - 1)f(t)) \quad (4.31)$$

for some to-be-determined function  $f(t)$ . Our ansatz automatically obeys the constraint  $\phi(1, t) = 1$ . Substituting into (4.30), we see that all  $s$ -dependence happily drops out and we are left with the a differential equation only for  $f(t)$ :

$$\frac{df}{dt} = \lambda - \beta f(t) . \quad (4.32)$$

We take the initial condition  $P(0, 0) = 1$  or, equivalently,  $\phi(s, 0) = 1$ . This requires  $f(0) = 1$  and the equation above has solution

$$f(t) = \frac{\lambda}{\beta} (1 - e^{-\beta t}) . \quad (4.33)$$

So our generating function takes the double-exponential form

$$\phi(s, t) = \exp\left(\frac{\lambda}{\beta} (s - 1)(1 - e^{-\beta t})\right) . \quad (4.34)$$

With this in hand, we can now compute various expectation values. The average population size is

$$\langle n(t) \rangle = \left. \frac{\partial \phi}{\partial s} \right|_{s=1} = \frac{\lambda}{\beta} (1 - e^{-\beta t}) . \quad (4.35)$$

The variance can be computed from (4.14) to be

$$\begin{aligned} \sigma^2(t) &= \langle n^2 \rangle - \langle n \rangle^2 \\ &= \left. \frac{\partial^2 \phi}{\partial s^2} \right|_{s=1} + \left. \frac{\partial \phi}{\partial s} \right|_{s=1} - \left( \left. \frac{\partial \phi}{\partial s} \right|_{s=1} \right)^2 = \frac{\lambda}{\beta} (1 - e^{-\beta t}) . \end{aligned} \quad (4.36)$$

So, again we have  $\sigma / \langle n \rangle = 1 / \sqrt{\langle n \rangle}$ .



In the limit  $t \rightarrow \infty$ , the system settles down to a steady state. In the present case, we can extract this straightforwardly from the generating function (4.34). However, it is rare that we can find an exact expression for the generating function. Nonetheless, it's often possible to get the steady state by returning to the differential equation that governs the generating function, in this case (4.30). In the steady state, we have  $\partial\phi/\partial t = 0$  and so

$$\frac{\partial\phi}{\partial s} = \frac{\lambda}{\beta}\phi \quad \implies \quad \phi(s) = \exp\left(\frac{\lambda}{\beta}(s-1)\right) \quad (4.37)$$

where we've used the boundary condition  $\phi(s=1) = 1$  to fix the overall normalisation. Translated to a steady-state probability distribution, this is

$$P(n) = \frac{1}{n!} \left(\frac{\lambda}{\beta}\right)^n e^{-\lambda/\beta} . \quad (4.38)$$

This is again a Poisson distribution. Note that for these kinds of stochastic models, the steady state means that we have a constant probability distribution, rather than a constant  $n$ .

### More Offspring Means More Variation

We can make a simple change to the model above, and suppose that a birth results in  $M$  new individuals. In this case, the master equation (4.27) is replaced by

$$\frac{dP(n,t)}{dt} = -(\lambda + \beta n)P(n,t) + \lambda P(n-M,t) + \beta(n+1)P(n+1,t) . \quad (4.39)$$

You can rerun the steps above to find the new equation governing the generating function,

$$\frac{\partial\phi(s,t)}{\partial t} = \lambda(s^M - 1)\phi(s,t) - \beta(s-1)\frac{\partial\phi(s,t)}{\partial s} . \quad (4.40)$$

Now this equation is harder to solve. We could restrict ourselves to look for long-time steady state solutions with  $\partial\phi/\partial t = 0$ , so that we have to solve

$$\frac{\partial\phi}{\partial s} = \frac{\lambda}{\beta} \frac{s^M - 1}{s-1} \phi(s) . \quad (4.41)$$

This is somewhat easier to solve. For example, if  $M = 2$ , then we have

$$\phi(s) = \exp\left(\frac{\lambda}{\beta}\left(s + \frac{s^2}{2} - \frac{3}{2}\right)\right) . \quad (4.42)$$

And, from this, we can then reconstruct the steady state probability distribution.

Alternatively, we could just jump immediately to what we're most interested in: the expectation  $\langle n(t) \rangle$  and the variance, which follows from  $\langle n^2(t) \rangle$ . We can derive equations that govern both of these quantities. For the expectation, we have

$$\begin{aligned}
\frac{d\langle n(t) \rangle}{dt} &= \sum_{n=0}^{\infty} n \frac{dP(n, t)}{dt} \\
&= \sum_{n=0}^{\infty} n [ -(\lambda + \beta n)P(n, t) + \lambda P(n - M, t) + \beta(n + 1)P(n + 1, t) ] \\
&= \sum_{n=0}^{\infty} [ -\lambda n - \beta n^2 + \lambda(n + M) + \beta(n - 1)n ] P(n, t) \\
&= \sum_{n=0}^{\infty} [ \lambda M - \beta n ] P(n, t) \\
&= \lambda M - \beta \langle n(t) \rangle .
\end{aligned} \tag{4.43}$$

We see that we get a simple differential equation for  $\langle n(t) \rangle$  which we can now just solve. Using the initial condition  $\langle n(0) \rangle = 0$ , the solution is

$$\langle n(t) \rangle = \frac{\lambda M}{\beta} (1 - e^{-\beta t}) . \tag{4.44}$$

This takes the same functional form as our previous result (4.35), but with the birth rate  $\lambda$  now increased to  $\lambda M$ . That makes sense. However, there's more to be seen if we look at the variance. This too obeys its own differential equation,

$$\begin{aligned}
\frac{d\langle n^2(t) \rangle}{dt} &= \sum_{n=0}^{\infty} n^2 \frac{dP(n, t)}{dt} \\
&= \sum_{n=0}^{\infty} n^2 [ -(\lambda + \beta n)P(n, t) + \lambda P(n - M, t) + \beta(n + 1)P(n + 1, t) ] \\
&= \sum_{n=0}^{\infty} [ -\lambda n^2 - \beta n^3 + \lambda(n + M)^2 + \beta(n - 1)^2 n ] P(n, t) \\
&= \sum_{n=0}^{\infty} [ \lambda M^2 + (2\lambda M + \beta)n - 2\beta n^2 ] P(n, t) \\
&= \lambda M^2 + (2\lambda M + \beta)\langle n(t) \rangle - 2\beta \langle n^2(t) \rangle .
\end{aligned} \tag{4.45}$$

We already have an expression for  $\langle n(t) \rangle$ , so this is a differential equation for  $\langle n^2(t) \rangle$ . Things are easier if we look at the steady state distribution. Here we have  $\langle n \rangle = \lambda M / \beta$  and so

$$\langle n^2 \rangle = \frac{1}{2\beta} \left( \lambda M^2 + \frac{\lambda M}{\beta} (2\lambda M + \beta) \right) . \tag{4.46}$$

Note that there is an  $M^2$  term, as well as a term linear in  $M$ . For  $M \gg 1$ , this quadratic term dominates and we have variance

$$\text{var}(n) \approx \frac{\lambda}{2\beta} M^2 . \quad (4.47)$$

While the average population scales as  $\lambda M$ , the variation scales as  $\lambda M^2$ . This too makes sense: a birth now gives a jump of  $M$  in the population, rather than just one, and so the jumps around the mean value are larger.

### Non-Linear Growth Rates

For the model above, we have birth and death rates that were either constant or proportional to  $n$ . And this was reflected in the nice differential equation (4.43) that we derived for the expectation  $\langle n(t) \rangle$ . We might wonder if the stochastic growth rates that we enter into the master equation always arise in the equation for the expectation value in this way. The answer, sadly, is no. Things are less pleasant when the rates depend non-linearly on  $n$ .

For example, we might try to cook up something akin to the logistic equation by taking a birth rate proportional to  $\lambda n$  and a death rate proportional to  $\beta n^2$ . It's straightforward to write down the corresponding master equation,

$$\frac{dP(n, t)}{dt} = -(\lambda n + \beta n^2)P(n, t) + \lambda(n-1)P(n-1, t) + \beta(n+1)^2P(n+1, t) . \quad (4.48)$$

We can then retrace our steps that led to (4.43). This time we have

$$\begin{aligned} \frac{d\langle n(t) \rangle}{dt} &= \sum_{n=0}^{\infty} n \frac{dP(n, t)}{dt} \\ &= \sum_{n=0}^{\infty} [-\lambda n^2 - \beta n^3 + \lambda n(n+1) + \beta(n-1)n^2]P(n, t) \\ &= \sum_{n=0}^{\infty} [\lambda n - \beta n^2]P(n, t) \\ &= \lambda \langle n(t) \rangle - \beta \langle n^2(t) \rangle . \end{aligned} \quad (4.49)$$

But that's not so useful: the equation for  $\langle n \rangle$  requires us to know something about  $\langle n^2 \rangle$ . And the equation for  $\langle n^2 \rangle$  will need us to know about  $\langle n^3 \rangle$  and so on. The set of equations doesn't close and to make progress we need to make some approximation about these higher order moments, or turn to numerical simulation. Nonetheless, although we can't solve such models completely, as we now show, there are some things that we can say.

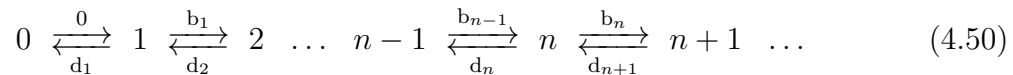
#### 4.1.4 Extinction

*“Do one calculation every day that scares you.”*

Eleanor Roosevelt.

In stochastic models, with populations fluctuating up and down, it’s quite possible that the population fluctuates to zero and stays there. We would like to know the likelihood of such an extinction event.

Extinction isn’t possible in the simple model above since, even if the population does die down to zero, it can still grow again. (We suggested at the time that this may be due to immigration rather than laziness when designing the model.) But we can make things more realistic by considering a model where the birth and death rates depend on the population size  $n$ . We write these as  $b_n$  and  $d_n$  respectively, and then take  $b_0 = 0$ , meaning that if the population hits  $n = 0$  it stays there. That’s extinction. This is summarised in the following reaction diagram:



Now we ask: what’s the probability of extinction? Or, more precisely: suppose that the population sits at some healthy number  $n$ . What’s the probability  $Q_n$  that it will eventually become extinct? We will see that, under one further reasonable assumption, this probability is necessarily one.

Our strategy is to set up a recurrence relation for  $Q_n$ . The probability of extinction for a population  $n$  can be related to

$$\begin{aligned} Q_n &= \text{Prob}(\text{birth next}) Q_{n+1} + \text{Prob}(\text{death next}) Q_{n-1} \\ &= \frac{b_n}{b_n + d_n} Q_{n+1} + \frac{d_n}{b_n + d_n} Q_{n-1} . \end{aligned} \quad (4.51)$$

Rearranging, gives the recurrence relation

$$Q_{n+1} - Q_n = \frac{d_n}{b_n} (Q_n - Q_{n-1}) = \left( \prod_{i=1}^n \frac{d_i}{b_i} \right) (Q_1 - Q_0) . \quad (4.52)$$

All the  $n$  dependence on the left-hand side sits in that product. The next question that we want to ask is: does the product converge for large  $n$ ? The answer, in any realistic model, is no! Our requirement for realism is that as the population swells, the death

rate exceeds the birth rate. Specifically, we require that there exist an integer  $N$  and a number  $R > 1$  such that

$$\frac{d_n}{b_n} \geq R \quad \text{for all } n > N. \quad (4.53)$$

In this case, the product in (4.52) can get arbitrarily large as  $n$  gets large. But the left-hand side is a difference of probabilities, so  $(Q_{n+1} - Q_n) \in [-1, +1]$ . This means that the only way (4.52) can be satisfied for very large  $n$  is if

$$Q_{n+1} - Q_n = Q_1 - Q_0 = 0 \quad (4.54)$$

for all  $n$ . In particular, we must have  $Q_n = Q_0$  but if the population is at  $n = 0$  then it's already extinct and so  $Q_0 = 1$ . We learn that  $Q_n = 1$  for all  $n$ . Closed systems go extinct under reasonable assumptions. All men must die.

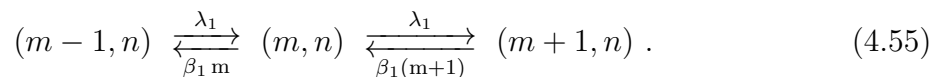
There is one glimmer of light in this calculation: we didn't yet compute how long we've got left! Happily, it turns out that the expected lifetime of a species can be very large.

#### 4.1.5 Multiple Populations: Wildebeest and Flies

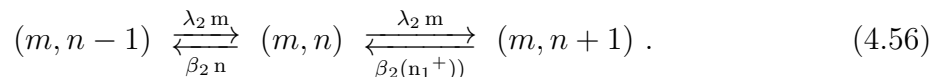
We can extend the ideas above to multiple populations. Here we describe a simple system which, to add some colour, we will think of as a population  $m$  of wildebeest and a population  $n$  of flies. The birth and death rates are taken to be

- The wildebeest have a birth rate  $\lambda_1$  and death rate  $\beta_1 m$ .
- The flies have a birth rate (or, said differently, an import rate) of  $\lambda_2 m$  and a death rate of  $\beta_2 n$ . Note that the "birth rate" is proportional to the number of wildebeest, which might sound slightly weird, but we should interpret this as wildebeest attracting flies from elsewhere into the system.

We can capture this in two reaction diagrams (or, alternatively in one 2d reaction diagram). The growth of the wildebeest population is described by



Meanwhile, the growth of the flies is described by



The system is slightly unusual in that the wildebeest population is unaffected by the flies, but the flies care about the wildebeest. Note, also, that everything is linear which will make the system tractable.

To keep our equations looking vaguely reasonable, we will write the probability that we are in the state  $(m, n)$  as  $p_{m,n}(t)$ . We can read off the master equation from the reaction diagrams above:

$$\begin{aligned} \frac{dp_{m,n}}{dt} = & \lambda_1 [p_{m-1,n} - p_{m,n}] + \beta_1 [(m+1)p_{m+1,n} - mp_{m,n}] \\ & + \lambda_2 [mp_{m,n-1} - mp_{m,n}] + \beta_2 [(n+1)p_{m,n+1} - np_{m,n}] . \end{aligned} \quad (4.57)$$

From this, we can compute the evolution of the average population size. For wildebeest, we have

$$\begin{aligned} \frac{d\langle m \rangle}{dt} = & \sum_{m,n} mp_{m,n} \\ = & \lambda_1 [\langle m+1 \rangle - \langle m \rangle] + \beta_1 [\langle (m-1)m \rangle - \langle m^2 \rangle] \\ & + \lambda_2 [\langle m^2 \rangle - \langle m^2 \rangle] + \beta_2 [\langle mn \rangle - \langle mn \rangle] \\ = & \lambda_1 - \beta_1 \langle m \rangle . \end{aligned} \quad (4.58)$$

For the flies, we have

$$\begin{aligned} \frac{d\langle n \rangle}{dt} = & \sum_{m,n} np_{m,n} \\ = & \lambda_1 [\langle n \rangle - \langle n \rangle] + \beta_1 [\langle mn \rangle - \langle mn \rangle] \\ & + \lambda_2 [\langle m(n+1) \rangle - \langle mn \rangle] + \beta_2 [\langle (n-1)n \rangle - \langle n^2 \rangle] \\ = & \lambda_2 \langle m \rangle - \beta_2 \langle n \rangle . \end{aligned} \quad (4.59)$$

We see again that the wildebeest population (4.58) doesn't depend on the flies, while the converse is not true. The steady state is given by

$$\langle m \rangle = \frac{\lambda_1}{\beta_1} \quad \text{and} \quad \langle n \rangle = \frac{\lambda_2}{\beta_2} \langle m \rangle = \frac{\lambda_1 \lambda_2}{\beta_1 \beta_2} . \quad (4.60)$$

We can also look at the fluctuations, starting by computing quadratic expectations. We've already computed  $\langle m^2 \rangle$  in our previous birth/death model. (It's given by (4.45) after setting  $M = 1$ .) We have

$$\frac{d\langle m^2 \rangle}{dt} = \lambda_1 + (2\lambda_1 + 1)\langle m \rangle - 2\beta_1 \langle m^2 \rangle . \quad (4.61)$$

We can similarly compute the evolution of  $\langle n^2 \rangle$  and  $\langle mn \rangle$ . They are given by

$$\begin{aligned} \frac{d\langle n^2 \rangle}{dt} = & \lambda_2 \langle m \rangle + \beta_2 \langle n \rangle + 2\lambda_2 \langle mn \rangle - 2\beta_2 \langle n^2 \rangle \\ \frac{d\langle mn \rangle}{dt} = & \lambda_1 \langle n \rangle + \lambda_2 \langle m^2 \rangle - (\beta_1 + \beta_2) \langle mn \rangle . \end{aligned} \quad (4.62)$$

In the steady state (4.60), variance of the flies is then given by

$$\begin{aligned}
 \text{var}(n) &= \langle n^2 \rangle - \langle n \rangle^2 \\
 &= \frac{1}{2} \langle n \rangle + \frac{\lambda_2}{2\beta_2} \langle m \rangle + \frac{\lambda_2}{\beta_2} \langle mn \rangle - \langle n \rangle^2 \\
 &= \langle n \rangle + \frac{\lambda_2}{\beta_2} [\langle mn \rangle - \langle n \rangle \langle m \rangle] .
 \end{aligned} \tag{4.63}$$

The term in square brackets is the *covariance* between the variables  $m$  and  $n$ ,

$$\text{cov}(m, n) = \langle mn \rangle - \langle m \rangle \langle n \rangle . \tag{4.64}$$

We learn that the fluctuations of the flies has two terms: an intrinsic fluctuation in the birth and death rates of the flies, proportional to  $\langle n \rangle$ , and an additional fluctuation proportional to  $\text{cov}(m, n)$  that tracks the fluctuations in wildebeest.

## 4.2 Meet the Fokker-Planck Equation

In the previous section, we studied various examples of the “master equation”, which governs how a probability distribution over a discrete set of outcomes evolves. In this section, we would like to generalise this idea to describe a probability distribution over a continuous set of outcomes. The simplest example is a probability distribution  $P(\mathbf{x}, t)$  of some substance distributed spread over some spatial coordinate  $\mathbf{x}$ . The resulting equation is called the *Fokker-Planck equation*.

We’re going to derive the Fokker-Planck equation starting from our discrete master equation. If the width of the probability distribution is much broader than then size between the spacing, then it makes sense to approximate the discrete variable with a continuous variable.

For all our examples above, we thought of the discrete variable  $n \in \mathbb{N}$  as the population size. There will be times when we want to keep that interpretation, but we might also want to think of  $n$  as labelling the position of some object that is restricted to lie on a lattice. (For example, such a set-up arises in [Solid State Physics](#) when we think of an electron moving in a solid.) In this case, we could relabel  $n = x$  to denote position. We will adopt this notation below.

Suppose that the hopping rate to jump from site  $n$  to site  $n + r$  is given by  $W(n, r)$ . Here  $r \in \mathbb{Z}$  can be positive or negative. Then the master equation for the probability distribution over sites  $x = n$  is given by

$$\frac{\partial P(x, t)}{\partial t} = \sum_{r \in \mathbb{Z}} \left[ W(x - r, r) P(x - r, t) - W(x, r) P(x, t) \right] . \tag{4.65}$$

Here the first term captures the fact that the particle could hop from any site to  $x$ , while the second term captures the fact that it could hop away to any site. Typically, this hopping rate will be “short range”, meaning that  $W(x, r)$  drops off quickly as  $r$  gets large.

Now, the term in the square brackets looks like  $f(x - r) - f(x)$  where  $f(x) = W(x, r)P(x, t)$ . For  $r$  small, we Taylor expand

$$f(x - r) = f(x) - r \frac{df}{dx} + \frac{r^2}{2} \frac{d^2 f}{dx^2} + \dots \quad (4.66)$$

We apply this Taylor expansion to the master equation (4.65) and drop the  $\dots$  terms. The decision to truncate the Taylor expansion after the second derivative is important and will have consequence below. We’re left with

$$\begin{aligned} \frac{\partial P(x, t)}{\partial t} &= \sum_{r \in \mathbb{Z}} \left[ -r \frac{\partial}{\partial x} (W(x, r) P(x, t)) + \frac{r^2}{2} \frac{\partial^2}{\partial x^2} (W(x, r) P(x, t)) \right] \\ &= -\frac{\partial}{\partial x} (u(x) P(x, t)) + \frac{\partial^2}{\partial x^2} (D(x) P(x, t)) . \end{aligned} \quad (4.67)$$

This is the *Fokker-Planck equation*<sup>12</sup>. It involves two functions,  $u(x)$  and  $D(x)$ , given by

$$u(x) = \sum_{r \in \mathbb{Z}} r W(x, r) \quad \text{and} \quad D(x) = \frac{1}{2} \sum_{r \in \mathbb{Z}} r^2 W(x, r) . \quad (4.69)$$

We assume that  $W(x, r)$  drops off quickly enough at large  $r$  so that both of these sums converge. You can read more about the Fokker-Planck equation, viewed from a slightly different perspective, in the lectures on [Kinetic Theory](#).

The total probability is necessarily conserved, with  $\int dx P(x, t) = 1$  for all time. Things that are conserved obey a continuity equation, and probability is no exception. We can recast the Fokker-Planck equation in this form, writing

$$\frac{\partial P}{\partial t} + \frac{\partial J}{\partial x} = 0 \quad \text{with} \quad J = uP + \frac{\partial}{\partial x} (DP) . \quad (4.70)$$

---

<sup>12</sup>It’s not uncommon to see the Fokker-Planck equation written as

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x} (AP) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (BP) \quad (4.68)$$

with the obvious relation  $A(x) = u(x)$  and  $B(x) = 2D(x)$ . But, as we’ll see soon, the variables  $u$  and  $D$  are more evocative than  $A$  and  $B$ .



Written in this way, we see that the current  $J$  takes the form that we anticipated in Section 3. The first term corresponds to advection, with  $u(x)$  some background flow that carries the probability with it. The second term gives rise to diffusion. We'll see these interpretations borne out in what follows.

From the definitions (4.69), we see that for positive rates  $W(x, r) > 0$ , it might be possible for the advection  $u(x)$  to vanish, but it will never be possible for the diffusion  $D(x)$  to vanish. In this sense, diffusion is really the essential element that arises from stochastic processes.

### Evolving Moment by Moment

We can look at how various expectation values change with the Fokker-Planck equation. The average position  $\langle x(t) \rangle$  evolves as

$$\frac{d\langle x \rangle}{dt} = \int dx x \frac{\partial P(x, t)}{\partial t} = - \int dx x \frac{\partial(uP)}{\partial x} + \int dx x \frac{\partial^2(DP)}{\partial x^2} . \quad (4.71)$$

We integrate by parts, using the fact that any normalised probability distribution must vanish asymptotically. For the second term, we can integrate by parts twice to get zero. So only the first term contributes, giving

$$\frac{d\langle x \rangle}{dt} = \int dx uP = \langle u(x) \rangle . \quad (4.72)$$

We see that the time evolution of the mean depends on the average of  $u(x)$ , and not on the function  $D(x)$ . This confirms what we said above: the function  $u(x)$  acts like advection, governing the overall drift of the probability distribution.

For the variance, we first look at

$$\frac{d\langle x^2 \rangle}{dt} = - \int dx x^2 \frac{\partial(uP)}{\partial x} + \int dx x^2 \frac{\partial^2(DP)}{\partial x^2} . \quad (4.73)$$

Now both terms survive integration by parts. We have

$$\frac{d\langle x^2 \rangle}{dt} = 2\langle xu(x) \rangle + 2\langle D(x) \rangle . \quad (4.74)$$

The variation is, as usual,  $\text{var}(x) = \langle x^2 \rangle - \langle x \rangle^2$  and obeys the equation

$$\frac{d(\text{var}(x))}{dt} = \frac{d\langle x^2 \rangle}{dt} - 2\langle x \rangle \frac{d\langle x \rangle}{dt} = 2\langle D(x) \rangle + 2\text{cov}(x, u(x)) . \quad (4.75)$$

We see the same kind of behaviour as for our wildebeest problem, with two terms contributing to the variance. The first is the expectation value of  $D(x)$ , the second a covariance between  $x$  and  $u(x)$  given by  $\text{cov}(x, u(x)) = \langle xu(x) \rangle - \langle x \rangle \langle u(x) \rangle$ .

### 4.2.1 Constant Drift and Diffusion

To illustrate these ideas, let's return to the particularly simple model that marches tentatively forward at a constant rate  $\lambda$ . In the discrete case, this gave rise to the Poisson process, governed by the master equation (4.11)

$$\frac{dP(n, t)}{dt} = \lambda [P(n-1, t) - P(n, t)] . \quad (4.76)$$

Comparing to (4.65), we have  $W(n, 1) = \lambda$  for all  $n$ , with  $W(n, r) = 0$  for  $r \neq 1$ . The corresponding Fokker-Planck equation is

$$\frac{\partial P}{\partial t} = -\lambda \frac{\partial P}{\partial x} + \frac{\lambda}{2} \frac{\partial^2 P}{\partial x^2} . \quad (4.77)$$

We can use our results above, with  $u = \lambda$  and  $D = \frac{1}{2}\lambda$ , to compute how the moments evolve. We have

$$\frac{d\langle x \rangle}{dt} = \lambda \quad \Longrightarrow \quad \langle x \rangle = \lambda t \quad (4.78)$$

and

$$\frac{d\langle x^2 \rangle}{dt} = 2\lambda \langle x \rangle + \lambda \quad \Longrightarrow \quad \langle x^2 \rangle = \lambda t + (\lambda t)^2 . \quad (4.79)$$

It's worth pointing out that both of these agree with the corresponding discrete model, where we also had

$$\langle n \rangle = \lambda t \quad \text{and} \quad \langle n^2 \rangle = \lambda t + (\lambda t)^2 . \quad (4.80)$$

It's natural to ask: does the Fokker-Planck equation coincide with the discrete Poisson process? The answer is no: the first two moments coincide, but not higher moments. You can check, for example, that

$$\begin{aligned} \langle x^3 \rangle &= (\lambda t)^3 + 3(\lambda t)^2 \\ \langle n^3 \rangle &= (\lambda t)^3 + 3(\lambda t)^2 + \lambda t . \end{aligned} \quad (4.81)$$

This is a typical feature of the Fokker-Planck equation when compared to a discrete master equation. The fact that the two agree for the first two moments, and then disagree, can be traced to our truncation of the Taylor expansion at the second derivative in (4.67). Note, however, that the two agree for large time, which reflects the fact that  $\langle x \rangle = \lambda t$  and so  $\langle x \rangle \gg r$  for large time.

In this simple case, it's not difficult to solve the Fokker-Planck equation. Motivated by the fact that we have a constant drift  $u = \lambda$ , we introduce the variable  $\xi = x - \lambda t$  and consider the ansatz

$$P(x, t) = G(\xi, t) . \quad (4.82)$$

Then the Fokker-Planck equation (4.77) becomes the diffusion equation

$$\frac{\partial G}{\partial t} = \frac{\lambda}{2} \frac{\partial^2 G}{\partial \xi^2} . \quad (4.83)$$

We've already seen solutions to this equation in Section 3.1. If we start with a delta-function initial condition, then the probability distribution is given by an ever-spreading Gaussian, now with an overall drift set by  $u = \lambda$ ,

$$P(x, t) = \frac{1}{\sqrt{2\pi\lambda t}} e^{-(x-\lambda t)^2/2\lambda t} . \quad (4.84)$$

### Diffusion Revisited

It's straightforward to cook up a situation in which the drift vanishes, but diffusion remains. Suppose that we have a particle that lives on a line, with position  $n \in \mathbb{Z}$ . This time, it bounces back and forth at the same rate  $\lambda$ , so we have  $W(n, 1) = W(n, -1) = \lambda$ . Now we have  $A = 0$  and  $D = \lambda$  and the Fokker-Planck equation coincides with the heat equation that we studied in Section 3,

$$\frac{\partial P}{\partial t} = D \frac{\partial^2 P}{\partial x^2} . \quad (4.85)$$

This illustrates how diffusion arises from underlying randomness. This is what happens, for example, in Brownian motion in which small particles, suspended in a liquid, move in an erratic motion as they are constantly bombarded by surrounding molecules.

#### 4.2.2 Birth and Death Once More

We can also look at the Fokker-Planck equation for our birth and death model with master equation (4.27)

$$\frac{dP(n, t)}{dt} = -(\lambda + \beta n)P(n, t) + \lambda P(n - 1, t) + \beta(n + 1)P(n + 1, t) . \quad (4.86)$$

We compare this to (4.65) to find  $W(n, 1) = \lambda$  and  $W(n, -1) = \beta n$ . We replace the discrete  $n$  with the continuous  $x$  which, in this context, still measure the population. The Fokker-Planck equation for  $P(x, t)$  then becomes

$$\frac{\partial P}{\partial t} = -\frac{\partial(uP)}{\partial x} + \frac{\partial^2(DP)}{\partial x^2} \quad (4.87)$$

with

$$\begin{aligned} u(x) &= W(n, 1) - W(n, -1) = \lambda - \beta x \\ D(x) &= \frac{1}{2}(W(n, 1) + W(n, -1)) = \frac{1}{2}(\lambda + \beta x) . \end{aligned} \quad (4.88)$$

This is one of the simplest examples of the Fokker-Planck equation. Following our expectations above, we have (4.72),

$$\frac{d\langle x \rangle}{dt} = \lambda - \beta \langle x \rangle \implies \langle x \rangle = \frac{\lambda}{\beta} (1 - e^{-\beta t}) . \quad (4.89)$$

This is identical to the discrete birth and death model that we met previously. (See, for example, (4.44) with  $M = 1$ .) Similarly, we have from (4.74)

$$\frac{d\langle x^2 \rangle}{dt} = \lambda + (2\lambda + \beta)\langle x \rangle - 2\beta\langle x^2 \rangle . \quad (4.90)$$

This too agrees with the differential equation (4.45) that governs the discrete model. The variance then obeys

$$\frac{d(\text{var}(x))}{dt} = \lambda + \beta\langle x \rangle - 2\beta\text{var}(x) \implies \text{var}(x) = \frac{\lambda}{\beta} (1 - e^{-\beta t}) . \quad (4.91)$$

where we've used the expression for  $\langle x \rangle$  in (4.89) and implemented the initial condition  $\text{var}(x) = 0$  when  $t = 0$ .

### 4.2.3 Fokker-Planck With More Variables

It's straightforward to generalise the Fokker-Planck equation to include more variables so that we work with the vector  $\mathbf{x} \in \mathbb{R}^d$ . In the context of physics,  $\mathbf{x}$  is a spatial coordinate; in the context of ecology,  $\mathbf{x} = \mathbf{n}$  is a variable that describes the population of  $d$  different species.

We can follow our earlier definition, starting with (4.65) which, with multiple variables, reads

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = \sum_{\mathbf{r} \in \mathbb{Z}^d} \left[ W(\mathbf{x} - \mathbf{r}, \mathbf{r}) P(\mathbf{x} - \mathbf{r}, t) - W(\mathbf{x}, \mathbf{r}) P(\mathbf{x}, t) \right] . \quad (4.92)$$

Again we Taylor expand  $f(\mathbf{x}) = W(\mathbf{x}, \mathbf{r}) P(\mathbf{x}, t)$  and write

$$f(\mathbf{x} - \mathbf{r}) = f(\mathbf{x}) - r_i \frac{\partial f}{\partial x_i} + \frac{r_i r_j}{2} \frac{\partial^2 f}{\partial x_i \partial x_j} + \dots \quad (4.93)$$

Dropping the ... terms leaves us with the multi-dimensional Fokker-Planck equation

$$\frac{\partial P(x, t)}{\partial t} = -\frac{\partial}{\partial x_i}(u_i(\mathbf{x}) P(\mathbf{x}, t)) + \frac{\partial^2}{\partial x_i \partial x_j}(D_{ij}(\mathbf{x}) P(\mathbf{x}, t)) . \quad (4.94)$$

with our two functions now given by

$$u_i(\mathbf{x}) = \sum_{\mathbf{r} \in \mathbb{Z}^d} r_i W(\mathbf{x}, \mathbf{r}) \quad \text{and} \quad D_{ij}(\mathbf{x}) = \frac{1}{2} \sum_{\mathbf{r} \in \mathbb{Z}^d} r_i r_j W(\mathbf{x}, \mathbf{r}) . \quad (4.95)$$

We see that the advection term now involves a vector function  $\mathbf{u}(\mathbf{x})$ , reflecting its interpretation it plays as a background velocity field. Meanwhile the diffusion term now involves a symmetric matrix  $D(\mathbf{x})$ .

We can again see the meaning of the advection term by computing

$$\begin{aligned} \frac{d}{dt} \langle x_i \rangle &= \int d^d x \, x_i \frac{\partial P(\mathbf{x}, t)}{\partial t} \\ &= \int d^d x \, x_i \left( -\frac{\partial(u_j P)}{\partial x_j} + \frac{\partial(D_{jk} P)}{\partial x_j \partial x_k} \right) = \int d^d x \, u_i P \end{aligned} \quad (4.96)$$

where, in the final equality, we've integrated by parts. We see that we have the obvious generalisation

$$\frac{d}{dt} \langle \mathbf{x} \rangle = \langle \mathbf{u} \rangle . \quad (4.97)$$

Meanwhile, the time derivative of the fluctuations is captured by

$$\begin{aligned} \frac{d}{dt} \langle x_i x_j \rangle &= \int d^d x \, x_i x_j \left( -\frac{\partial(u_k P)}{\partial x_k} + \frac{\partial(D_{kl} P)}{\partial x_k \partial x_l} \right) \\ &= \langle x_i u_j \rangle + \langle u_i x_j \rangle + 2 \langle D_{ij} \rangle . \end{aligned} \quad (4.98)$$

It's useful to define the symmetric *covariance matrix*

$$C_{ij} = \text{cov}(x_i, x_j) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle . \quad (4.99)$$

This contains the variance as its diagonal terms,  $\text{var}(x_i) = C_{ii}$ , with the off-diagonal terms telling us about correlations between different variables. You can check that

$$\frac{dC_{ij}}{dt} = \text{cov}(x_i, u_j) + \text{cov}(x_j, u_i) + 2 \langle D_{ij} \rangle . \quad (4.100)$$

Often, we would like to understand the steady state of a distribution, which means that want to find solutions where the right-hand side of this equation vanishes. But that's not so straightforward because the right-hand side of this equation depends on  $\langle x_i u_j \rangle$  and we don't necessarily have a good handle on this. To illustrate how to proceed, we turn to an example.

#### 4.2.4 Wildebeest and Flies Again

We will revisit the story of the Wildebeest and flies from Section 4.1.5. This is a two-dimensional system, described by  $\mathbf{x} = (m, n)$  where  $m$  is the population of wildebeest and  $n$  the population of flies. We previously derived the master equation (4.57) when treating the population as discrete,

$$\begin{aligned} \frac{dp_{m,n}}{dt} = & \lambda_1 [p_{m-1,n} - p_{m,n}] + \beta_1 [(m+1)p_{m+1,n} - mp_{m,n}] \\ & + \lambda_2 [mp_{m,n-1} - mp_{m,n}] + \beta_2 [(n+1)p_{m,n+1} - np_{m,n}]. \end{aligned} \quad (4.101)$$

Comparing to (4.92), we can read off the non-vanishing values of  $W(\mathbf{x}, \mathbf{r})$ . They are:

- $W(\mathbf{x}, \mathbf{r}) = \lambda_1$  when  $\mathbf{r} = (1, 0)$ .
- $W(\mathbf{x}, \mathbf{r}) = \beta_1 m$  when  $\mathbf{r} = (-1, 0)$ .
- $W(\mathbf{x}, \mathbf{r}) = \lambda_2 m$  when  $\mathbf{r} = (0, 1)$ .
- $W(\mathbf{x}, \mathbf{r}) = \beta_2 n$  when  $\mathbf{r} = (0, -1)$ .

From this, we can read off the functions in the Fokker-Planck equation. The advection velocity is

$$\begin{aligned} \mathbf{u} &= \sum_{\mathbf{r}} \mathbf{r} W(\mathbf{x}, \mathbf{r}) \\ &= \lambda_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta_1 m \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \lambda_2 m \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \beta_2 n \begin{pmatrix} 0 \\ -1 \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 - \beta_1 m \\ \lambda_2 m - \beta_2 n \end{pmatrix}. \end{aligned} \quad (4.102)$$

Meanwhile, the diffusion matrix is

$$\begin{aligned} D &= \frac{1}{2} \left[ \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \beta_1 m \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \lambda_2 m \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \beta_2 n \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right] \\ &= \frac{1}{2} \begin{pmatrix} \lambda_1 + \beta_1 m & 0 \\ 0 & \lambda_2 m + \beta_2 n \end{pmatrix}. \end{aligned} \quad (4.103)$$

Now that we have explicit expressions for  $\mathbf{u}$  and  $D$ , we can return to the question: what does it mean to have a steady state probability distribution for this model?

It's simple to find the steady state for expectation values  $\langle \mathbf{x} \rangle$  because, as we've seen in (4.97), this is given by  $\langle \mathbf{u} \rangle = 0$ , so

$$\langle m \rangle = \frac{\lambda_1}{\beta_1} \quad \text{and} \quad \langle n \rangle = \frac{\lambda_1 \lambda_2}{\beta_1 \beta_2} . \quad (4.104)$$

This agrees with the results from the discrete model (4.60). But now we want to extend this to think about fluctuations, as captured in the covariance matrix  $C_{ij}$ . In the steady state, we want to find solutions to  $dC_{ij}/dt = 0$  and that means that we need to compute  $\text{cov}(x_i, u_j)$  and  $\langle D_{ij} \rangle$ . We see that  $\mathbf{u}$  is linear in  $\mathbf{x} = (m, n)$  and so we can write

$$\mathbf{u} = \boldsymbol{\lambda} + a\mathbf{x} \quad \text{with} \quad \boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ 0 \end{pmatrix} \quad \text{and} \quad a = \begin{pmatrix} -\beta_1 & 0 \\ \lambda_2 & -\beta_2 \end{pmatrix} . \quad (4.105)$$

This gives

$$\text{cov}(x_i, u_j) = a_{jk} \text{cov}(x_i, x_k) = a_{jk} C_{ik} . \quad (4.106)$$

We also have

$$\langle D_{ij} \rangle = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_1 \lambda_2 / \beta_1 \end{pmatrix} . \quad (4.107)$$

Now we can look for steady state solutions for the covariance matrix  $C_{ij}$ . In steady state, our evolution equation (4.100) becomes

$$\frac{dC}{dt} = aC + Ca^T + 2\langle D \rangle = 0 . \quad (4.108)$$

This is a matrix equation, with  $a$ ,  $C$  and  $\langle D \rangle$  all  $2 \times 2$  matrices. It is an example of a *Lyapunov equation*. The equation is easily solved by writing it out in components and doing some linear algebra. We find (recalling that  $C_{12} = C_{21}$ ).

$$\begin{aligned} -2\beta_1 C_{11} + 2\lambda_1 &= 0 \\ \lambda_2 C_{11} - \beta_2 C_{12} - \beta_1 C_{12} &= 0 \\ 2\lambda_2 C_{12} - 2\beta_2 C_{22} + 2\frac{\lambda_1 \lambda_2}{\beta_1} &= 0 . \end{aligned} \quad (4.109)$$

Rearranging, then gives the variances

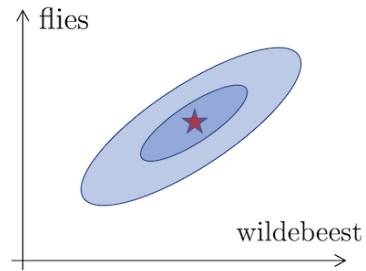
$$\text{var}(m) = C_{11} = \frac{\lambda_1}{\beta_1} \quad \text{and} \quad \text{var}(n) = C_{22} = \frac{\lambda_1 \lambda_2}{\beta_1 \beta_2} \left( 1 + \frac{\lambda_2}{\beta_1 + \beta_2} \right) \quad (4.110)$$

while the covariance is

$$\text{cov}(m, n) = C_{12} = \frac{\lambda_1 \lambda_2}{\beta_1(\beta_1 + \beta_2)} . \quad (4.111)$$

We see that  $C_{12} > 0$  so the wildebeest and flies covary positively: if there is more of one, then there is likely to be more of the other. That's to be expected given our initial assumptions which were that wildebeest attract flies.

With this information, we can plot the range in which we expect to find populations of wildebeest and flies. This is shown schematically in the figure where the mean is shown as a star. Around that, we draw ellipses whose semi-axes are determined by the eigenvectors and eigenvalues of the covariance matrix  $C$ . In the figure, we've sketched ellipses corresponding to one standard deviation and, outside, 95% confidence levels.



### Deriving a General Lyapunov Equation

For our example above, we were lucky because both  $\mathbf{u}(\mathbf{x})$  and  $D(\mathbf{x})$  were linear in the variables  $\mathbf{x}$ . That meant that the steady state condition  $dC_{ij}/dt = 0$  could be expressed entirely in terms of the covariant matrix  $C_{ij}$  and some constant matrices. But that won't always be the case.

To find the steady state for more general, non-linear systems, we typically have to make a (not always justified) approximation. We approximate the advection velocity to be linear and the diffusion matrix to be constant,

$$\mathbf{u} = \boldsymbol{\lambda} + a\mathbf{x} \quad \text{and} \quad D = b \quad (4.112)$$

where both  $a$  and  $b$  are constant matrices. In steady state, the covariance matrix then satisfies the Lyapunov equation

$$aC + Ca^T + b = 0 . \quad (4.113)$$

The solution can again be found using some linear algebra.



### 4.3 An Invitation to Fluctuation and Dissipation

To finish, we can make contact with some basic ideas from elsewhere in physics, notably the lectures on [Statistical Physics](#). We will start by considering something very basic: a particle of mass  $m$ , with Newton force law

$$m\ddot{\mathbf{x}} = -\gamma\dot{\mathbf{x}} - \nabla V + \mathbf{f} . \quad (4.114)$$

The first two terms on the right-hand side are very familiar: the first is a friction term, with the strength of friction dictated by the coefficient  $\gamma$ ; the second is a conservative force arising from a potential  $V(\mathbf{x})$ . The novelty is the third term, consisting of the additional force  $\mathbf{f}$ . This we take to be a random force. You can think of this as arising because the particle is suspended in some liquid, and is being constantly bombarded by the underlying molecules, causing it to bounce back and forth in some random way. This is a famous process known as *Brownian motion*.

There are various ways of dealing with equations like (4.114). The most systematic way is to think of the random force  $\mathbf{f}$  as coming from some probability distribution, and then figuring out how to translate that into a corresponding probability distribution  $P(\mathbf{x})$  for the position of the particle. In this context, (4.114) is known as the *Langevin equation*. You can read more about this in the lectures on [Kinetic Theory](#). Here, instead, we will make direct contact with the Fokker-Planck equation. Our goal is to write down a Fokker-Planck equation for the probability distribution  $P(\mathbf{x}, t)$ .

For this, we should be in the limit where the motion is friction dominated and the acceleration term in (4.114) can be ignored. In this case, we take the average of (4.114) and use the fact that  $\langle \mathbf{f} \rangle = 0$  because the random force is just as likely to hit from any direction. This then tells us that the average velocity of the particle is dictated by the potential

$$\langle \dot{\mathbf{x}} \rangle = -\frac{1}{\gamma} \langle \nabla V \rangle . \quad (4.115)$$

But this is the same kind of equation that we get from the Fokker-Planck equation (4.97) if we set  $\mathbf{u} = -\nabla V/\gamma$ . This suggests that the probability distribution of the particle is governed by a Fokker-Planck equation that takes the form

$$\frac{\partial P}{\partial t} = \frac{1}{\gamma} \nabla \cdot (P \nabla V) + D \nabla^2 P \quad (4.116)$$

for some diffusion constant  $D$ . We have made the additional assumptions here that the matrix  $D_{ij} = D\delta_{ij}$  is diagonal, on grounds of rotational invariance, and, moreover, that  $D$  is independent of  $x$  on grounds of translational invariance. It remains to determine the diffusion constant  $D$  in terms of the variables in the original set-up.

This is where ideas from statistical mechanics come in. First, we look at the equilibrium probability distribution, obeying

$$\nabla \cdot \left( \frac{1}{\gamma} P \nabla V + D \nabla P \right) = 0 . \quad (4.117)$$

We can view this as a differential equation for  $P(\mathbf{x})$ , one that is solved by

$$P(\mathbf{x}) \sim \exp \left( -\frac{\gamma}{D} V(\mathbf{x}) \right) \quad (4.118)$$

up to an overall normalisation that we've ignored. Now suppose that the random force  $\mathbf{f}$  arises because the particle sits in a fluid at temperature  $T$ . Then we know that the probability distribution must take the usual Boltzmann form

$$P(\mathbf{x}) \sim \exp \left( -\frac{1}{k_B T} V(\mathbf{x}) \right) \quad (4.119)$$

with  $k_B$  the Boltzmann constant. (There is no kinetic term in this expression because we're in a friction-dominated environment where we can ignore the  $m\ddot{x}$  term in the original equation of motion.) Equating these two expressions, we learn that the diffusion constant must be given by

$$D = \frac{k_B T}{\gamma} . \quad (4.120)$$

This is the *Einstein relation*. It is the key result in the fourth of his famous collection of 1905 papers. (The one that didn't introduce special relativity or pioneer the idea of the quantum!) The relation is rather surprising: the diffusion constant tells us how much the particle is kicked around by the environment, while the friction term tells us how difficult it is for the particle to plough through the same environment. Remarkably, the two are related.

The Einstein relation is an example of a more general idea known as the *fluctuation-dissipation theorem*, which relates the fluctuations experienced by a system to the dissipation (i.e. friction) experienced by the system. It's an important idea in many areas of physics.