

Part II Mathematical Biology - Section 2

Lecturer: Dr Julia Gog (jrg20@cam.ac.uk)

Lent Term 2016, this version updated February 14, 2017

Contents

2	Stochastic systems	63
2.0	Preliminaries	63
2.0.0	Revision: discrete probabilities and generating functions	63
2.0.1	Why bother?	64
2.0.2	The first step	65
2.1	Discrete population sizes	67
2.1.1	Single populations	67
2.1.2	Extinction	76
2.1.3	Multiple populations	79
2.2	Continuous population sizes	83
2.2.1	Fokker-Planck for a single variable	83
2.2.2	Multivariate Fokker-Planck	87

2 Stochastic systems

2.0 Preliminaries

2.0.0 Revision: discrete probabilities and generating functions

We will be considering discrete systems where the population size takes non-negative integer values $(0, 1, 2, 3, \dots)$. This could be representing the number of individuals, cells, or perhaps molecules. We will usually¹ write the random variable as N and particular values it takes as n .

$$P(n, t) = \text{Probability}(N = n \text{ at time } t)$$

For clarity, often we will use the simpler notation of p_n to mean $P(n, t)$ when it is unambiguous. As N can only take positive integer values, summing probabilities over all of these values:

$$\sum_{n=0}^{\infty} p_n = 1.$$

We will also be making extensive use of generating functions:

$$\phi(s, t) = \sum_{n=0}^{\infty} s^n p(n, t) = \sum_{n=0}^{\infty} s^n p_n = \langle s^N \rangle$$

Where the angle brackets $\langle \cdot \rangle$ denote the expected value. Given a generating function, we can recover p_n by looking for the coefficient of s^n . Recall some of the useful properties of generating functions, particularly by evaluating at special values of s :

$$\phi(1, t) = \sum_{n=0}^{\infty} p_n = 1$$

$$\phi(0, t) = p_0 = \text{Probability that the population has died out}$$

We can also differentiate with respect to s :

$$\left. \frac{\partial}{\partial s} \right|_t \phi = \phi_s = \sum_{n=0}^{\infty} n s^{n-1} p_n$$

and from now on take partial derivatives with respect to s to mean t treated as constant, and vice-versa. And using this derivative of the generating function we can find the mean by evaluating at $s = 1$:

$$\phi_s(1, t) = \sum n p_n = \langle N \rangle = \text{mean pop.} = \mu$$

¹I will try to stick to this throughout these notes just for the sake of clarity (capitals for the random variable, lower case for the number it takes in an equation, usually as an index being summed over), but usually easy to interpret what is meant in any given expression, so not essential to be super careful about all this.

(where μ is just the most commonly used letter for the mean). We do not need to be super careful about putting limits on the sums so long as we say $p_n = 0$ for $n < 0$. Differentiating a second time then evaluating at $s = 1$:

$$\phi_{ss}(1, t) = \sum n(n-1)p_n = \langle N(N-1) \rangle = \langle N^2 \rangle - \langle N \rangle = \langle N^2 \rangle - \mu$$

and recall the variance is given by

$$\text{Var}(N) = \langle (N - \mu)^2 \rangle = \langle N^2 - 2\mu N + \mu^2 \rangle = \langle N^2 \rangle - 2\mu \langle N \rangle + \mu^2 = \langle N^2 \rangle - \mu^2$$

We know μ from the generating function already, so we now have variance:

$$\text{Var}(N) = \phi_{ss}(1, t) + \mu - \mu^2.$$

2.0.1 Why bother?

So far, we have only considered deterministic systems, and also we have focussed on continuous variables for things like population size. We know that we are doing this for convenience of course. The number of individuals in a population should be a non-negative integer. To some extent, all biological processes are subject to random fluctuations. However by taking these simplifications, we are in a realm where we have a great deal of mathematical machinery we can use (differential equations, and dynamical systems in general). It means we can write down models and explore them without much difficulty. We can also understand them clearly enough to gain insights which often extend far beyond the particular model we have in front of us. We might be confident that some of our results (for example the concept of R_0 in disease dynamics and hence the idea of a vaccination threshold) will hold even if we were to add more layers of complexity to the models.

However, what have we missed by ignoring random fluctuations and forcing population numbers to be continuous? Perhaps the most obvious problem is that populations that are decaying and getting very small still cannot ever reach zero, e.g.:

$$\frac{dN}{dt} = -\lambda N \quad \implies \quad N = N_0 e^{-\lambda t}$$

($\lambda > 0$) and even as t becomes very large, still N is positive. Does this matter? Well, yes it does if something changes later and the population size, which should have been zero, can claw its way back from tiny numbers to being large.

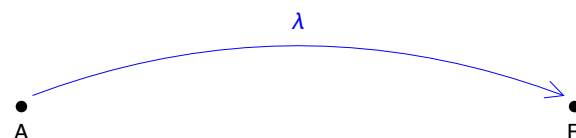
A classic case of this was in a model for rabies in foxes proposed by Murray and others. In a published paper: the scenario of a rabies introduced to the south of England was explored by a deterministic diffusion model. As one might expect, rabies spreads out almost radially over a few years. However it doesn't end there: after some years, rabies *reappears* from the same initial location and spreads out again (though less dramatically). This second introduction was not put in there by the modellers, but

just happens in the simulations. This was explained by Mollison in a paper in 1991²: the density of infected foxes at the origin location never actually died down to zero, though it did get to a very small value: 10^{-18} of a fox per square kilometre. Anyone reasonable would describe this as essentially *no* rabid foxes, but mathematically it is still a quantity that can grow back when conditions allow (renewed susceptibility when the fox population has turned over for a few years). This Mollison called the ‘atto-fox’, and this has become a useful term to describe this modelling issue³. Note that this is not a numerical issue from inaccuracy in computation, but one which comes from the very nature of continuous population deterministic models. Does this mean that original model was worthless? I would argue it still has use, but attention to results from simulations should stop after the first wave.

So although we do not always need to use stochastic models, we should be vigilant for atto-foxes and other artifactual effects in our deterministic models. And to watch out for these effects, we must be aware of what they can be. As well as this issue with small numbers not going to zero, we can already guess (from thinking about excitable systems already studied) that in some special situations, small fluctuations can lead to large effects. Further, we also do not always expect ‘average’ behaviour to in fact be representative of typical behaviours. In addition, we will see in the later parts of this section that small fluctuations in two quantities are not always just independent noise: they can actually covary if the dynamics are interlinked.

2.0.2 The first step

Consider first this very simple system: we have states A and B only, and we are always in exactly one of those states. We jump from A to B at probability rate λ . We can make a diagram like this:



Where dots (or nodes) denote states, arrows are the possible ‘jumps’ and the arrows are labelled with the rates. There are two ways to proceed from here. The less formal approach (taken in lectures) is simply to go from the diagram to consider flows of probability, an idea which might be familiar from continuous time Markov Chains from Ib . For completeness, the slightly more formal way is included in these notes [in blue](#). If the more formal approach helps you understand things better than great, if not then then you can safely skip it. In both approaches, we introduce notation $P(A, t)$ to mean probability we are in state A at time t .

²Mollison, D. *Dependence of epidemic and population velocities on basic parameters*. 1991, Mathematical biosciences 107 (2)

³I have also heard this phenomena described as ‘nano-hawks’

More formal What does probability rate λ actually mean? Essentially it is a probability per unit time of making a jump so that for small δt :

$$P(\text{jump by } t + \delta t \mid \text{in } A \text{ at time } t) = \lambda \delta t + \mathcal{O}(\delta t^2)$$

One could justify this by essentially saying this is the definition of probability rate (a bit like a probability density function $f(x) dx$). But perhaps more intuitive is to think of it as being an instantaneous rate λ until it happens. It might have already happened in δt so then the rate is no longer λ . In fact that helps us to see why the correction is order δt^2 .

We can build up a differential equation for the probability by considering the probability of being in A at time $t + \delta t$ conditional on each of the possibilities at time t , and what happens in the time period between of length δt :

$$\begin{aligned} P(A, t + \delta t) &= P(A, t) P(\text{stayed in } A \text{ for } \delta t) + P(B, t) P(\text{went from } B \text{ to } A \text{ in } \delta t) \\ &= P(A, t) [1 - P(\text{left } A \text{ during } \delta t)] + P(B, t) [0] \\ &= P(A, t) [1 - \lambda \delta t] + \mathcal{O}(\delta t)^2 \end{aligned}$$

Then strategic rearrange and divide by δt :

$$\frac{P(A, t + \delta t) - P(A, t)}{\delta t} = -\lambda P(A, t) + \mathcal{O}(\delta t)$$

and taking the limit $\delta t \rightarrow 0$:

$$\frac{d}{dt} P(A, t) = -\lambda P(A, t)$$

Alternatively, working straight from the diagram, it would not be unreasonable to go directly for the same equation immediately, i.e.

$$\frac{d}{dt} P(A, t) = -\lambda P(A, t).$$

We can build a similar differential equation for $P(B, t)$, but we do not need it as in fact we must be in A or B , so $P(A, t) + P(B, t) = 1$

We can solve this if we have an initial condition. Let us suppose we are at A initially so $P(A, 0) = 1$:

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

and hence also

$$P(B, t) = 1 - e^{-\lambda t}.$$

This rather unsurprising result lets us quickly check the distribution of times to jump. Let T be the time when we jump from state A to B . Then we can find the cumulative distribution for T by noting we are in B if T has passed already:

$$P(T < t) = P(B, t) = 1 - e^{-\lambda t}$$

We can use this to find the probability density function:

$$f(t) = \frac{d}{dt}(1 - e^{-\lambda t}) = \lambda e^{-\lambda t}$$

which by now we see is the exponential distribution. And as always:

$$\int_0^{\infty} f(t) dt = 1.$$

We can find the expected jump time

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

and the variance of the jump time

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

and we could square root this for the standard deviation σ and we see that it is equal to the mean.

2.1 Discrete population sizes

2.1.1 Single populations

We now study the pure import model. There is a probability rate λ of adding one individual to the population, no matter what the current population size is. We start with zero population. Again, represent states as dots and jumps as arrows:



And writing $P(n, t)$ to mean probability of the population size (N) being n at time t , then the initial condition just says that $P(0, 0) = 1$ and $P(n, 0) = 0$ for $n > 0$.

Again we can set things up by considering where we are at $t + \delta t$ starting from t :

$$P(n, t + \delta t) = P(n, t) \times (1 - \lambda \delta t + \mathcal{O}(\delta t^2)) + P(n - 1, t) \times (\lambda \delta t + \mathcal{O}(\delta t^2)) + \mathcal{O}(\delta t^2).$$

In theory we should be including a term for every possible previous state on the right hand side here, but that last $\mathcal{O}(\delta t^2)$ takes care of all of the others. For example, starting from $n - 2$ we would have to jump *twice* during δt to end up in n . Other states are even more remote, or even impossible to get to n : that last $\mathcal{O}(\delta t^2)$ is a big bin for them all.

As before, rearrange and divide by δt to prepare for a derivative:

$$\frac{P(n, t + \delta t) - P(n, t)}{\delta t} = -\lambda P(n, t) + \lambda P(n - 1, t) + \mathcal{O}(\delta t).$$

Take the limit $\delta t \rightarrow 0$.

Or again, work directly from the diagram:

$$\frac{d}{dt}P(n, t) = \lambda P(n - 1, t) - \lambda P(n, t)$$

and at this point it would be sensible to use the shorter notation $p_n = P(n, t)$:

$$\dot{p}_n = \lambda(p_{n-1} - p_n) \quad (1)$$

and we can say this holds for all $n \geq 0$ by the nice convention of saying $p_n = 0$ for $n < 0$.

Equation (1) is an example of a *master equation*, which is just a fancy way of saying a differential equation for p_n . It is a slightly curious object though as it is (continuous) differential equation in t , but a (discrete) difference equation in n .

In this case, we see that the master equation gives a system of differential equations that we can actually fully solve explicitly.

Exercise 28: Set $p_0 = 1$, $p_n = 0$ for $n > 0$ at $t = 0$. Solve the master equation inductively to get

$$p_n = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Generally we cannot fully solve in this way, in which case it is often a good idea to try a generating function. Recall:

$$\phi(s, t) = \sum_{n=0}^{\infty} s^n p_n = \langle s^N \rangle$$

Then we can differentiate ϕ with respect to t (keeping s fixed), and substitute in using the master equation:

$$\begin{aligned} \frac{\partial \phi}{\partial t} &= \sum s^n \dot{p}_n = \sum s^n \lambda (p_{n-1} - p_n) = \lambda \sum s^n p_{n-1} - \lambda \sum s^n p_n \\ &= \lambda \sum s^{n+1} p_n - \lambda \sum s^n p_n \\ &= \lambda s \phi - \lambda \phi \\ &= (s - 1) \lambda \phi \end{aligned}$$

And we can integrate this (remember that s was treated as a constant):

$$\phi = A(s) e^{(s-1)\lambda t}$$

To resolve $A(s)$, consider $t = 0$:

$$\phi(s, 0) = \sum s^n p_n(0) = p_0(0) + s p_1(0) + s^2 p_2(0) + s^3 p_3(0) + \dots$$

and $p_0(0) = 1$ and the rest of the $p_n(0) = 0$ for $n > 0$, so $\phi(s, 0) = 1$. And from this we see that $A(s) = 1$ for all s . So here we have the full solutions for the generating function

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

Great. So what can we do with ϕ ? Well, it encodes all of the p_n , so we could use it to recover the p_n by taking the coefficient of s^n . This we could do by Taylor series (recognition or differentiating n times and evaluating at $s = 0$):

$$\begin{aligned} p_n &= \text{Coef of } s^n \text{ in } \phi = \text{Coef of } s^n \text{ in } e^{-\lambda t} e^{\lambda t s} \\ &= \text{Coef of } s^n \text{ in } e^{-\lambda t} \left(1 + (\lambda t s) + \frac{1}{2!} (\lambda t s)^2 + \dots + \frac{1}{n!} (\lambda t s)^n + \dots \right) \\ &= e^{-\lambda t} \frac{(\lambda t)^n}{n!} \end{aligned}$$

Often we are not very interested in p_n itself, but rather want things like the mean and variance, which can be found directly without having to go via the p_n . For the mean:

$$\mu = \langle N \rangle = \left. \frac{\partial \phi}{\partial s} \right|_{s=1} = \lambda t e^{(s-1)\lambda t} \Big|_{s=1} = \lambda t$$

and for the variance:

$$\langle N(N-1) \rangle = \left. \frac{\partial^2 \phi}{\partial s^2} \right|_{s=1} = (\lambda t)^2 e^{(s-1)\lambda t} \Big|_{s=1} = (\lambda t)^2$$

so

$$\sigma^2 = \text{Var}(N) = \langle N^2 \rangle - \langle N \rangle^2 = \left. \frac{\partial^2 \phi}{\partial s^2} \right|_{s=1} + \mu - \mu^2 = (\lambda t)^2 + (\lambda t) - (\lambda t)^2 = \lambda t.$$

This could be used to say something about the size of small fluctuations. Typical fluctuations are of the order of the standard deviation $\sigma = \sqrt{\lambda t}$. Typical population sizes are of the order of the mean $\mu = \lambda t$, so relative to population size, fluctuations go like $(\lambda t)^{-1/2}$ so fluctuations become less important for this system as time goes on.

end of
lecture 12

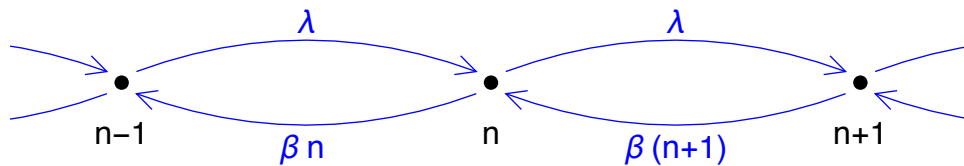
One further note on this simple import system: if we were to construct an analogous deterministic system, it would surely be

$$\frac{dN}{dt} = +\lambda$$

and with $N(0) = 0$ the solution would be $N = \lambda t$. This agrees with the mean of the stochastic system. Note: this agreement does not always hold⁴.

⁴You might like to try and identify what is special about this system. It might help to think about for what functions f is it true that $\langle f(N) \rangle = f(\langle N \rangle)$.

Import and death model Here we use the same model for import (probability rate λ of adding a new individual, including when at $n = 0$). Now we add a *per capita* probability rate of death β so the total rate is βn :



By now, we should be able to build the master equation by considering ‘flows’ of probability between states (the approach with δt would be fine but rather tedious). Focussing on state n in the diagram above, we can see two arrows going in and two going out, giving us the four terms:

$$\begin{aligned}\dot{p}_n &= \lambda p_{n-1} - \lambda p_n + \beta(n+1)p_{n+1} - \beta n p_n \\ &= \lambda(p_{n-1} - p_n) + \beta[(n+1)p_{n+1} - n p_n]\end{aligned}$$

Here, we go straight for the equation satisfied by the generating function:

$$\frac{\partial \phi}{\partial t} = \sum s^n \dot{p}_n = \lambda \sum s^n p_{n-1} - \lambda \sum s^n p_n + \beta \sum s^n (n+1) p_{n+1} - \beta \sum s^n n p_n$$

and as always, it is a good idea to split into separate sums, so we can shift the index on each one separately, aiming for p_n in each:

$$\frac{\partial \phi}{\partial t} = \lambda \sum s^{n+1} p_n - \lambda \sum s^n p_n + \beta \sum n s^{n-1} p_n - \beta \sum n s^n p_n$$

The first two sums are much as before ($s\phi$ and ϕ), but for the latter two, we must differentiate ϕ with respect to s to get the bonus n in the coefficients to appear:

$$\frac{\partial \phi}{\partial s} = \phi_s = \sum n s^{n-1} p_n$$

so we can immediately recognise the third sum as ϕ_s . The fourth one is the same thing again but with a spare factor of s :

$$\frac{\partial \phi}{\partial t} = \lambda s \phi - \lambda \phi + \beta \phi_s - \beta s \phi_s$$

and we can factorise⁵ out $(s - 1)$ to get

$$\phi_t = (s - 1) [\lambda \phi - \beta \phi_s] \quad (2)$$

We should also specify an initial condition. Again we will take $N = 0$ initially so that $\phi(s, 0) = 1$.

⁵This keeps happening and should not be a surprise. We know that $\phi = 1$ for $s = 1$, so $\phi_t = 0$ for $s = 1$.

Exercise 29: Get from the dots, arrows and rates diagram above to equation (2) without lecture notes

We can actually fully solve for ϕ again in this case. Take a strategic form:

$$\phi = e^{(s-1)f(t)}$$

where $f(t)$ is a function yet to be determined and substitute into equation (2):

$$(s-1)f'(t)\phi = (s-1)[\lambda\phi - \beta f(t)\phi]$$

and then it is clear that this form was designed to cancel the $(s-1)\phi$ and we're left with a DE in purely f :

$$f'(t) = \lambda - \beta f(t) \quad \implies \quad f(t) = \frac{\lambda}{\beta} + Ae^{-\beta t}$$

The initial condition that $\phi = 1$ at $t = 0$ corresponds to $f(0) = 0$, and this determines A, which fully determines f and hence we have

$$\phi(s, t) = e^{\frac{\lambda}{\beta}(s-1)(1-e^{-\beta t})}$$

and the double exponential might not be pretty, but it is nice to have fully solved for ϕ .

Exercise 30: Use this ϕ to show that the mean is given by

$$\langle N \rangle = \frac{\lambda}{\beta}(1 - e^{-\beta t})$$

and the variance is actually the same

$$\text{Var}(N) = \frac{\lambda}{\beta}(1 - e^{-\beta t})$$

Suppose we are now interested in what happens to this system at large time, so we are looking for a steady state solution and not worrying about initial condition. We could use our solution and let $t \rightarrow \infty$. However, to illustrate a more generally useful technique we will suppose that we do not actually have the solution for ϕ . We can go back to equation (2) and seek the solution that is independent of time (i.e. set $\phi_t = 0$):

$$0 = (s-1)[\lambda\phi - \beta\phi_s]$$

which is just a DE for ϕ in s , which we can solve:

$$\phi_s = \frac{\lambda}{\beta}\phi \quad \implies \quad \phi = Ae^{\frac{\lambda}{\beta}s}$$

and we cannot use the time initial condition here, but need something for a fixed value of s . The one to go for here is always $\phi = 1$ at $s = 1$. So $A = e^{-\frac{\lambda}{\beta}}$ and we have the steady state solution for ϕ :

$$\phi = e^{\frac{\lambda}{\beta}(s-1)}$$

and we see this is reassuringly consistent with what we would get with $t \rightarrow \infty$ in the general solution above.

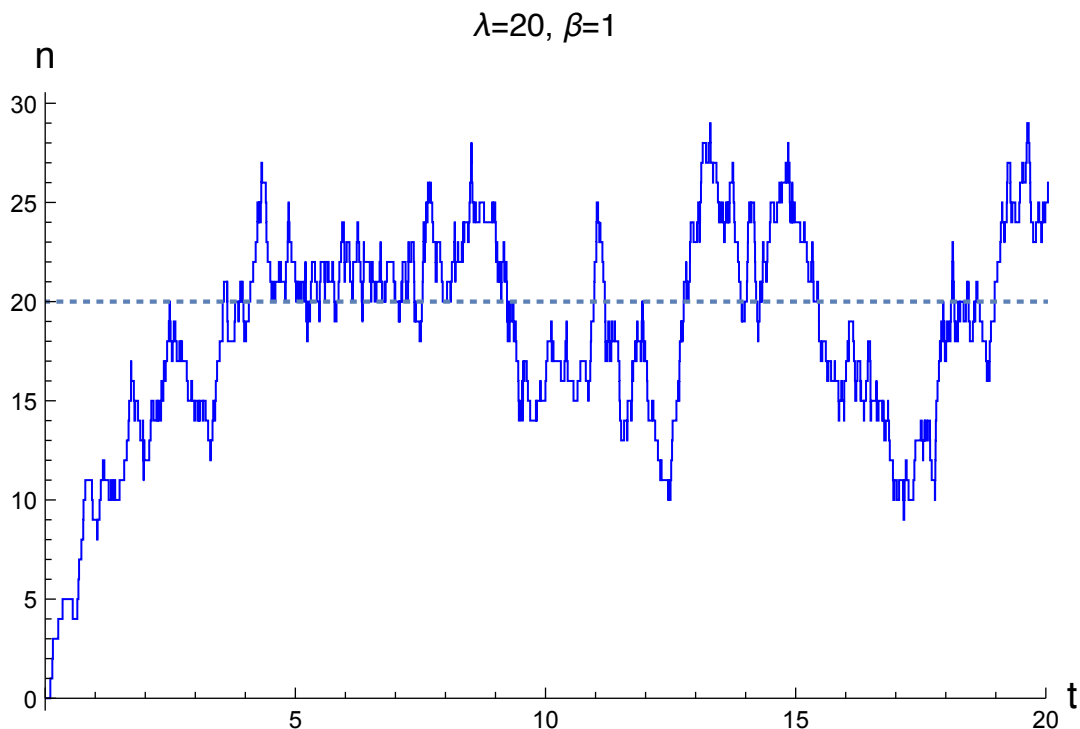
We can also quickly see the steady state p_n by pulling out the appropriate coefficient of s^n :

$$p_n = \text{Coef of } s^n \text{ in } \phi = e^{-\frac{\lambda}{\beta}} \left(\frac{\lambda}{\beta} \right)^n \frac{1}{n!}$$

(which we recognise as the Poisson distribution).

Aside on simulations

As always, it helps build intuition for what this is all about by looking at the output of simulations to get a feel for how these things behave. Here's one for this import and death model from this last section with $\lambda = 20$ and $\beta = 1$:



The dotted line is $\lambda/\mu = 20$, what we expect the mean to settle to. Indeed you can see that it heads quickly away from the initial condition and bumbles around near the mean.

You could even say the standard deviation looks to be about 4 or 5, which is consistent too with long term behaviour ($\sigma = \sqrt{20}$).

This programming approach is certainly not examinable, but it is actually very easy to simulate models like these. The *Gillespie algorithm* is both simple to code and fast to run. It relies on some nice properties of the exponential function. Suppose current time is t_i and population size is n_i . Then from the model we have some probability rates for jumps to other states. Each of these possible jumps happens at exponentially distributed time, but only the first one is actually the one that happens. The nice property is that the time for the *first* of a bunch of exponential distributions is itself an exponential distribution, with rate just sum of the others. Simulating an exponential distribution is super easy: just pick uniform random real in $(0, 1)$, take minus log of it, and divide by rate.

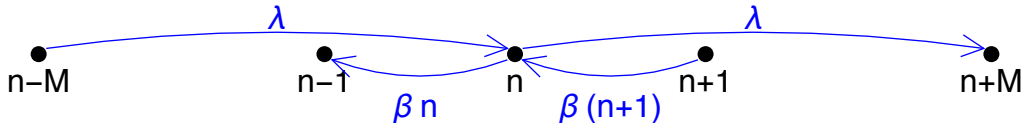
Then which event is it that happens? Also easy to run: just pick another uniform random and use it to choose event by weighting the unit interval in proportion to the rates of each possible event. Then you have $t_{i+1} = t_i + \Delta t$ and n_{i+1} is whatever it is after the chosen event.

Then rinse and repeat a number of times, but how many? Can either do a set number of events (some max for i), or until time is above some threshold (some max for t), but need to check some event is possible, i.e. total rate has not hit zero ($r_{tot} \neq 0$). That last possibility can't happen in this model, but could in others later, if we are in some stuck state (extinction, below). Here it is as pseudo code for this model: start with $i = 0$, choose some n_0 and probably $t_0 = 0$:

- Loop back to here, got n_i, t_i at some step i
- Compute rates for events $r_b = \lambda, r_d = \beta n_i$. Total rate $r_{tot} = r_b + r_d$.
- Generate a couple of uniform $(0, 1)$ random numbers: $random1, random2$
- Time to next event $\Delta t = -\log(random1)/r_{tot}$
- Choose which event: if $random2 < r_b/r_{tot}$ it is a birth, else a death
- update variables: $t_{i+1} = t_i + \Delta t$ and $n_{i+1} = n_i + 1$ if birth, $n_{i+1} = n_i - 1$ if death
- $i = i + 1$, loop back

Variance and step size

A generalisation of the import and death model allows us to see how the variance of the steady state depends on step size. The death rate is still β per capita, but now we suppose that M individuals are added each time, rather than just one, but still at rate λ . (Take M to be some positive integer of course.) How do we expect the mean and variance in steady state to change?



This works much as above, so let us jump straight to the master equation:

$$\dot{p}_n = \lambda (p_{n-M} - p_n) + \beta [(n+1) p_{n+1} - n p_n]$$

and then find the PDE satisfied by the generating function ϕ .

Exercise 31: Show

$$\frac{\partial \phi}{\partial t} = \lambda (s^M - 1) \phi - \beta (s - 1) \phi_s$$

We could work directly with the generating function for example to find steady state:⁶

$$\phi_t = 0 \implies \phi_s = \frac{\lambda}{\beta} \left(\frac{s^M - 1}{s - 1} \right) \phi$$

And this would work well for any given M , as we'd just have a nice polynomial in that bracket. But to illustrate another more general approach, suppose we couldn't progress with the generating function directly, then we could instead think of working with the master equation directly to find differential equations (in time) for the moments (expected N to some power):

$$\begin{aligned} \frac{d}{dt} \langle N \rangle &= \sum n \dot{p}_n \\ &= \lambda \sum n p_{n-M} - \lambda \sum n p_n + \beta \sum n(n+1) p_{n+1} - \beta \sum n^2 p_n \\ &= \lambda \sum (n+M) p_n - \lambda \sum n p_n + \beta \sum (n-1)n p_n - \beta \sum n^2 p_n \\ &= \lambda \sum M p_n + \beta \sum (-n) p_n \\ &= \lambda M - \beta \langle N \rangle \end{aligned}$$

where we have used the master equation to sub in for \dot{p}_n , and then expanded to write one sum per p term, then shifted the index on each sum so the p term is p_n , then it tidies from there. So this is a nice⁷ differential equation in $\langle N \rangle$. For initial condition, we supposed $N = 0$ at $t = 0$, so also $\langle N \rangle = 0$ at $t = 0$, and we can then easily solve:

⁶Do note that in stochastic models, steady state means that the p_n are constant, not N . The general behaviour of the system has settled into something, not that the population size has settled to a fixed value.

⁷actually this all seems rather too nice, and you might be suspicious that there was something special here. Indeed there is! The birth and death rates were only linear in n , which means the coefficients in the master equation are only linear. The equation for time derivative for each moment will end up being just in terms of that moment and lower. If birth or death rates were quadratic in n , we would find $\frac{d}{dt} \langle N \rangle$ depends on $\langle N^2 \rangle$ and so on up, which means we no longer have a closed system.

$$\langle N \rangle = \frac{\lambda M}{\beta} (1 - e^{-\beta t})$$

and we can also see $\langle N \rangle \rightarrow \lambda M / \beta$ as $t \rightarrow \infty$.

A similar approach with higher moments will also work:

$$\begin{aligned} \frac{d}{dt} \langle N^2 \rangle &= \sum n^2 \dot{p}_n \\ &= \lambda \left(\sum n^2 p_{n-M} - \sum n^2 p_n \right) + \beta \left(\sum n^2 (n+1) p_{n+1} - \sum n^3 p_n \right) \\ &= \lambda \sum ((n+M)^2 - n^2) p_n + \beta \sum ((n-1)^2 n - n^3) p_n \\ &= \lambda \sum (2nM + M^2) p_n + \beta \sum (-2n^2 + n) p_n \\ &= \lambda (2\langle N \rangle M + M^2) + \beta (-2\langle N^2 \rangle + \langle N \rangle) \\ &= \lambda M^2 + (2\lambda M + \beta) \langle N \rangle - 2\beta \langle N^2 \rangle \end{aligned}$$

and this also gives us a sensible DE as we already have solved to get $\langle N \rangle$ as a function of time. If we write $\langle N^2 \rangle = X$ then we have a DE in this form:

$$\dot{X} = -2\beta X + a + b e^{\beta t}$$

where the coefficients a and b are some constants that can be deduced from above.

Exercise 32: Show that at steady state

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

Note, you have a choice here. Fast method (which would be fine) is to just use the differential equation at steady state to find the equilibrium value directly. Full method (good for extra practice) is to fully solve the DE to get $\langle N^2 \rangle$ as a function of t , and then let $t \rightarrow \infty$.

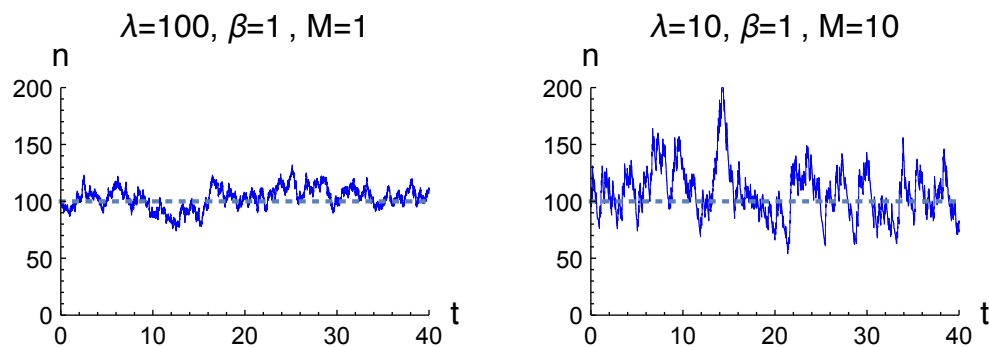
Either way... show also that if $M \gg 1$,

$$\text{var}(N) = \frac{\lambda}{2\beta} M^2.$$

Intuition and effect of step size: now think about the effect of changing M . Remember that the model is to add M individuals at rate λ and they die at some per capita rate β . It is intuitive that the population will head towards some balance between introduction and death: if the population is much larger than this balance then death rate

will dominate, if much smaller then the birth rate will dominate. So, this matches with expected N being $\lambda M/\beta$. We can also imagine that once the system is near that, then it will just stay vaguely near, but some random fluctuations coming from the stochastic nature of the model.

Suppose now that we make M ten times as big, and shrink λ by a corresponding factor of 10. Then intuitively we have the same average rate of individuals being added, and indeed we see $\langle N \rangle$ is unchanged (proportional to λM). However, we see that the variance (proportional to λM^2 for large M) will increase, and indeed ten times as much if we started with large M . Hopefully, this will also seem intuitive: we've made the stochastic process more 'lumpy' by having a rarer but larger step. This idea will be useful later!



Small M on left, large M on the right, also changing λ to keep the mean constant. There's clearly more variability with the larger M .

end of
lecture 13

2.1.2 Extinction

The examples above had a constant import rate, so even when $n = 0$ there is a constant immigration into the system. It turns out that this is crucial to get non-trivial steady states. In models with a *closed* population, no immigration is possible, so there is no way to get from $n = 0$ to $n = 1$, and once the population has gone extinct, it stays that way forever. In this section, we will build up some typical properties we might expect from a single population stochastic model, and we see that extinction is inevitable.

Here are three reasonable properties for a closed population model:

1. No births/immigration from $n = 0$
2. Death is possible for all $n > 0$
3. For large n , the death rate is much larger than the birth rate



Setting b_n and d_n to be birth and death rates respectively from state n , we can translate these properties to something in terms of these rates:

1. $b_0 = 0$
2. $d_n > 0$ for all $n > 0$
3. Either $b_n = 0$ for some n , or $d_n/b_n \gg 1$ for large n

We now divide into two cases to consider extinction:

(i) If $b_n = 0$ for some $n > 0$: then the state space is actually finite $(0, 1, 2, 3 \dots n-1, n)$ so we can use a result from IB Markov Chains: if we have a finite state space and all states communicate with $n = 0$, and $n = 0$ is absorbing (i.e. once we are there, we stay there), then as $t \rightarrow \infty$, $p_0 \rightarrow 1$. In other words, extinction is inevitable.

(i) If $b_n > 0$ for all $n > 0$: then we seem to have an infinite state space, but actually the same result turns out to be true. Perhaps this is almost intuitive if we remember that one of our conditions is that the death rate far outweighs birth rate large for large n , so it should behave as if the state space is effectively finite. But, here's a slightly more mathematical argument (which of course uses that condition of deaths outweighing births). We set up q_n the probability of extinction eventually, given we start at n . Then we can immediately have that $q_0 = 1$ (this is instant game over: we started at zero). For $n > 0$, we set up a recurrence, again using law of total probability: if we're at $n > 0$ then either a birth or a death next. If there's a birth, then we will be at $n+1$ and the probability of ultimate extinction starting from here is q_{n+1} . Similarly for death next. So:

$$q_n = P(\text{birth next}) q_{n+1} + P(\text{death next}) q_{n-1}$$

and what is the probability of birth or death next? We have the rates b_n and d_n (and know neither is zero in this case), so:

$$P(\text{birth next}) = \frac{b_n}{b_n + d_n}, \quad P(\text{death next}) = \frac{d_n}{b_n + d_n}.$$

Subbing these in, and multiplying through by total rate $b_n + d_n$:

$$(b_n + d_n)q_n = b_n q_{n+1} + d_n q_{n-1}$$

Rearrange to split b_n and d_n :

$$\begin{aligned} b_n(q_n - q_{n+1}) &= d_n(q_{n-1} - q_n) \\ (q_n - q_{n+1}) &= \frac{d_n}{b_n}(q_{n-1} - q_n) \end{aligned}$$

Note that we can't immediately solve this recurrence with only one initial condition, but we can argue through that there is only one feasible solution. Notice the bracket terms are just difference in consecutive q_i , and this recurrence equation holds for all $n > 0$, so chase the bracket on the right down:

$$\begin{aligned} (q_n - q_{n+1}) &= \frac{d_n}{b_n}(q_{n-1} - q_n) \\ &= \frac{d_n}{b_n} \frac{d_{n-1}}{b_{n-1}}(q_{n-2} - q_{n-1}) \\ &= \frac{d_n}{b_n} \frac{d_{n-1}}{b_{n-1}} \cdots \frac{d_1}{b_1}(q_0 - q_1) \\ &= \left(\prod_{i=1}^n \frac{d_i}{b_i} \right) (q_0 - q_1) \end{aligned}$$

And again this holds for all n . Notice the LHS is the difference of two probabilities, so must be between -1 and 1 for all n . Consider the product on the RHS and taking larger and larger n . We can make this product as large as we like⁸. Obviously $(q_0 - q_1)$ doesn't depend on n , but that equation holds for all n , so the only possibility we are left with is $q_0 - q_1 = 0$. So $q_1 = q_0 = 1$.

Now by the above equation, $q_{n+1} - q_n = 0$ for all $n > 0$, so finally we have that $q_n = 1$ for all n . Extinction is inevitable, starting from any n . This completes the second case (infinite state space, but actually effectively finite state space).

It can be shown that a similar result holds for multiple populations. Think about a vector for replacing n , where the elements are population numbers for different species. We should still have similar conditions: the vector of zeros being absorbing, deaths in each species being possible if any members are present, and very large populations being squished down by death rate outweighing the birth rate.

Treat this blue paragraph as starred, and ignore if it doesn't help you... It is intuitive that the same sort of argument as above should work, but here's a slightly different way to think about it (not given in lectures): divide up the infinite state space into three: (i) extinction (all populations zero, a single state), (ii) all populations within sensible size (some very large but finite state space), (iii) at least one population ridiculously large (the rest of the infinite state space). Then think of the possible transitions between these three states. State (iii) has death rates pushing us back towards (ii) and by carefully choosing what we mean by 'ridiculously large' we could make conditions on

⁸and actually this tells us more precisely what we mean by death rate outweighing birth rate for large n . We can see that it would be enough to say $d_n \geq R b_n$ for all $n > \hat{n}$, for some \hat{n} and some $R > 1$.

birth/death rates that give finite expected time until we are back in (ii). From state (ii) we can go to (iii) or (i), and as there are finitely many states, we pop out one way or the other in finite time. Obvious in (i) we stay there. So now we have something similar to a three state Markov chain with single absorbing state, and the other two states communicate with it. Hence, we will end up there with probability 1.

However, for single or multiple populations, the expected time to extinction could be extremely large⁹

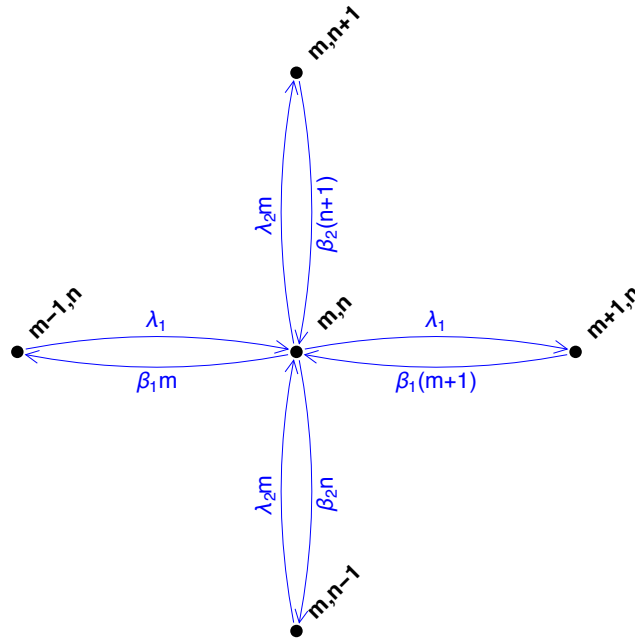
2.1.3 Multiple populations

Here we are going to look at a stochastic model with two populations: wildebeest (represented by m) and flies (n). This model is a bit unusual in that we are going to have the wildebeest affecting the dynamics of the flies, but not vice versa. This is not typical, but means we can handle the mathematics fairly easily, and be able to see the consequence of the interactions clearly on the flies.

The wildebeest have population size m , and ‘births’ are rate λ_1 (i.e. this is the import model above, so no longer worry about extinction) and deaths at rate $\beta_1 m$ (so β_1 *per capita*). The flies have population size n , ‘birth’ rate $\lambda_2 m$, and death rate $\beta_2 n$. The weird bit here is that the birth rate is proportional to the number of wildebeest. This is not because wildebeest give birth to flies, but interpret it here as a kind of import rate, and the flies are attracted by the wildebeest. The more wildebeest there are, the higher the rate of flies that join the system.

The current state of the system is given by (m, n) , which we can think of as corresponding to grid coordinates. Then we could do our arrows diagram like this:

⁹It might seem alarming that any stochastic model like this leads to total extinction always, and one might start fretting that life on earth could in principle be represented by such a system, and complications like time-dependence from days and seasons could be ironed out. However, there is a big difference between it being a mathematical inevitability, and it being something to actually lose sleep over. Expected time could be orders of magnitude larger than the age of our solar system. The ‘end of life on earth’ is not on my personal worry list. Extinction of individual species, however, is. If this also interests you, please *do* go and think and read more on this. Here’s one starting point: <https://www.durrell.org/wildlife/wildlife/durrell-index/explore/>.



For the master equation, really best to use the simpler notation $p_{m,n}$ rather than $P(m, n, t)$, which would be very cumbersome.

$$\begin{aligned} \dot{p}_{m,n} = & \lambda_1 [p_{m-1,n} - p_{m,n}] + \beta_1 [(m+1)p_{m+1,n} - m p_{m,n}] \\ & \lambda_2 [m p_{m,n-1} - m p_{m,n}] + \beta_2 [(n+1)p_{m,n+1} - n p_{m,n}] \end{aligned}$$

And using a similar method as above (lots of sums and shifting the indices carefully, we can find differential equations for the time evolution of moments. Note that the sums are over both population variables (m and n). (I'm going to reduce the number of steps I take each time, so if you are struggling to follow, try to put an extra step in, following an earlier example.) Let us start with the expected number of wildebeest, $\langle M \rangle$:

$$\begin{aligned} \frac{d}{dt} \langle M \rangle = \sum_{m,n} \dot{p}_{m,n} = & \lambda_1 [\langle M+1 \rangle - \langle M \rangle] + \beta_1 [\langle (M-1)M \rangle - \langle M^2 \rangle] \\ & \lambda_2 [\langle M^2 \rangle - \langle M^2 \rangle] + \beta_2 [\langle MN \rangle - \langle MN \rangle] \end{aligned}$$

And this tidies up very nicely:

$$\frac{d}{dt} \langle M \rangle = \lambda_1 - \beta_1 \langle M \rangle$$

and this is exactly as we'd expect, given that we already studied the import and death model above, and the dynamics of the wildebeest are just that. The flies just don't matter to the wildebeest.

Now move on to the expected number of flies, $\langle N \rangle$:

$$\begin{aligned}\frac{d}{dt}\langle N \rangle &= \lambda_1 \langle N - N \rangle + \beta_1 \langle MN - MN \rangle \\ &\quad \lambda_2 \langle M(N + 1) - MN \rangle + \beta_2 \langle (N - 1)N - N^2 \rangle \\ &= \lambda_2 \langle M \rangle - \beta_2 \langle N \rangle\end{aligned}$$

Which interesting just appears to have the parameters from the birth and death processes that directly affect the flies, but of course the birth rate includes expected number of wildebeest (which is itself a function of time, remember).

Exercise 33: Continue and get the equivalent equations for the second moments. Should get:

$$\begin{aligned}\frac{d}{dt}\langle M^2 \rangle &= \lambda_1 (2\langle M \rangle + 1) + \beta_1 (-2\langle M^2 \rangle + \langle M \rangle) \\ \frac{d}{dt}\langle MN \rangle &= \lambda_2 \langle M^2 \rangle + \lambda_1 \langle N \rangle - (\beta_1 + \beta_2) \langle MN \rangle \\ \frac{d}{dt}\langle N^2 \rangle &= \lambda_2 (2\langle MN \rangle + \langle M \rangle) + \beta_2 (-2\langle N^2 \rangle + \langle N \rangle)\end{aligned}$$

And supposing we are interested in steady state, we could set the time derivatives to zero and solve for any of these moments. For the population means:

$$\langle M \rangle = \frac{\lambda_1}{\beta_1}, \quad \langle N \rangle = \frac{\lambda_2}{\beta_2} \langle M \rangle = \frac{\lambda_1 \lambda_2}{\beta_1 \beta_2}$$

The variance of the wildebeest will be similar to the import and death model above (hence will be λ_1/β_1) and hence not particularly interesting. The variance of the flies might be more interesting, as the wildebeest dynamics affect it. And using the last equation in the exercise above, we have that:

$$\begin{aligned}\langle N^2 \rangle &= \frac{\lambda_2}{\beta_2} \left(\langle MN \rangle + \frac{1}{2} \langle M \rangle \right) + \frac{1}{2} \langle N \rangle \\ &= \frac{\lambda_2}{\beta_2} \langle MN \rangle + \langle N \rangle\end{aligned}$$

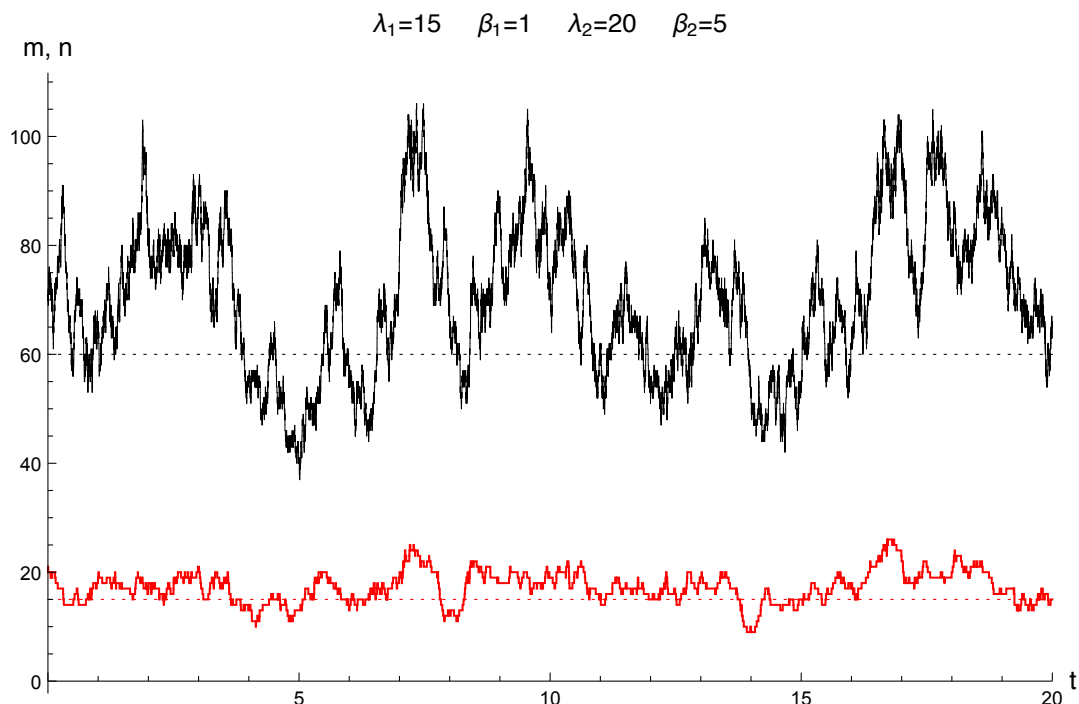
And use this to look at $var(N)$:

$$\begin{aligned}var(N) &= \langle N^2 \rangle - \langle N \rangle^2 \\ &= \frac{\lambda_2}{\beta_2} \langle MN \rangle + \langle N \rangle - \langle N \rangle^2\end{aligned}$$

$$\begin{aligned}
\text{var}(N) &= \langle N \rangle + \frac{\lambda_2}{\beta_2} (\langle MN \rangle - \langle M \rangle \langle N \rangle) \\
&= \langle N \rangle + \frac{\lambda_2}{\beta_2} \text{cov}(M, N)
\end{aligned}$$

recognising the second term as the covariance between M and N . We could then describe the flies variance as having two components: intrinsic variability from the first term, and extrinsic variability from the second term. The first term is just the fluctuations which come from the randomness of the flies own birth and death process directly. If we somehow totally fixed the number of wildebeest and looked at the flies dynamics in isolation, the variance would just be the first term. The second term is to do with the number of wildebeest changing and thus moving the flies birth rate around.

Actually, if we look at a typical output of a simulation, it is possible to see that both populations have fluctuations of course, but if the wildebeest numbers take a bit of a detour from the mean, then flies seem to follow suit. If we didn't already know, we could guess from this that the populations are interacting somehow.



One output of a Gillespie simulation of the model described above. The wildebeest M are plotted in red and the flies N in black. The means (as computed above) for each population is marked with dotted horizontal line of appropriate colour. Look for a place where the red curve moves a bit away from the mean, and check what happens to the black curve. It should be that the flies population starts to drift in the same direction (from increased or decreased birth rate) and thus flies curve is very variable: it has its own intrinsic variability, but also it is being shoved around by the wildebeest fluctuations.

We will come back to these ideas again below, thinking about covariances in general. Here, we can see that we would expect M and N to be covarying positively.

end of
lecture 14

2.2 Continuous population sizes

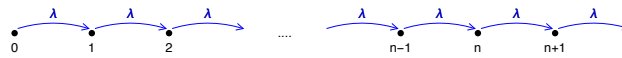
2.2.1 Fokker-Planck for a single variable

So far for stochastic models we have thought of the population size n as an integer. Hence we are dealing with master equations which are discrete in n (but continuous in time). Here, we recognise that these distributions ‘look like’ a continuous distribution, in some sense. This is particularly when the width of our distribution is much larger than our typical step size (usually ± 1). By convention and for clarity, when we think about continuous population sizes, we will write x rather than n , but these are really the same thing. By making an approximation, we replace our master equation (discrete in n) with a Fokker-Planck equation (continuous in x). Watch carefully below for the moment of approximation!

To do all this, we should set up some general notation for a master equation. Let $W(n, r)$ be the ‘jump’ rate from n to $n + r$. In other words, $W(n, r)$ give the rates of moving from n , taking a step size of r (which can be negative). Very often, W is non-zero only for a very small set of r , such as $r \in \pm 1$. So far, the biggest set of r is $r \in \{-1, 1, M\}$.

It is useful to have a couple of examples in hand that we already have fully cracked while working through this, just to see how things might look in practice:

Example 1: Pure import (see above). Take $N = 0$ at $t = 0$.

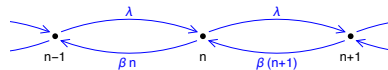


Then

$$W(n, 1) = \begin{cases} \lambda & \text{if } r = 1 \\ 0 & \text{otherwise} \end{cases}$$

and we already know that $\langle N \rangle = \text{var}(N) = \lambda t$.

Example 2: Import and death (see above). Also take $N = 0$ at $t = 0$.



Then

$$W(n, 1) = \begin{cases} \lambda & \text{if } r = 1 \\ \beta n & \text{if } r = -1 \\ 0 & \text{otherwise} \end{cases}$$

and we already know that $\langle N \rangle = \text{var}(N) = \frac{\lambda}{\beta} (1 - e^{-\beta t})$.

Now we have this general notation, we can write a general master equation:

$$\frac{\partial}{\partial t} P(n, t) = \sum_r [W(n - r, r) P(n - r, t) - W(n, r) P(n, t)]$$

where the sum over r is a finite number of terms (usually very small). We are using the full notation here just for extra safety. The partial derivative wrt time is just to make clear that n is fixed. We are about to have some derivatives wrt x appear, so this might matter.

Looking at the square bracket in the master equation, it is essentially of the form $[f(n-r) - f(n)]$ where $f(n) = W(n, r) P(n, t)$. Change the n to x and consider Taylor expansion about x :

$$f(x-r) - f(x) = \left(f(x) - r f'(x) + \frac{1}{2} r^2 f''(x) + \dots \right) - f(x) = -r f'(x) + \frac{1}{2} r^2 f''(x) + \dots$$

so applying that to the actual square bracket:

$$[W(x-r, r) P(x-r, t) - W(x, r) P(x, t)] = -r \frac{\partial}{\partial x} W(x, r) P(x, t) + \frac{1}{2} r^2 \frac{\partial^2}{\partial x^2} W(x, r) P(x, t) + \dots$$

and if we drop the \dots , then we are making an approximation that those terms are small, i.e. $|r| \ll |x|$. This is where the approximation happens. We have ditched $r^3 \frac{\partial^3}{\partial x^3}$ and higher. So, after this approximation, the master equation has become

$$\begin{aligned} \frac{\partial P}{\partial t} &= \sum_r -r \frac{\partial}{\partial x} W(x, r) P(x, t) + \sum_r \frac{1}{2} r^2 \frac{\partial^2}{\partial x^2} W(x, r) P(x, t) \\ \frac{\partial P}{\partial t} &= - \sum_r r \frac{\partial}{\partial x} W P + \frac{1}{2} \sum_r r^2 \frac{\partial^2}{\partial x^2} W P \\ \frac{\partial P}{\partial t} &= - \frac{\partial}{\partial x} \left(\sum_r r W \right) P + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left(\sum_r r^2 W \right) P \end{aligned}$$

where in the last step, the differential and sum were reordered (not a problem as the sum is over a finite number of terms). Now all the r are contained in sums within the brackets. The bracket terms depend on x but not t . Their form is actually fully determined by the model: these functions of x are just made out of the W . Call them $A(x)$ and $B(x)$ respectively, i.e. define:

$$A(x) = \sum_r r W \quad \text{and} \quad B(x) = \sum_r r^2 W$$

We will see below that A can be interpreted as a ‘mean drift’ or advection, some sort of average movement. B is something to do with spreading, or diffusion (indeed it looks that way in the equation below). So finally, we have the Fokker-Planck equation (FPE):

$\frac{\partial P}{\partial t} = - \frac{\partial}{\partial x} (A P) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (B P)$	Fokker-Planck Equation (FPE)
---	------------------------------

Example 2 continued:

$$A = \sum_r rW(x, r) = (-1)(\beta x) + (+1)(\lambda) = \lambda - \beta x$$

$$B = \sum_r r^2W(x, r) = (-1)^2(\beta x) + (+1)^2(\lambda) = \lambda + \beta x$$

And just sub these in to get the FPE:

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}((\lambda - \beta x)P) + \frac{1}{2}\frac{\partial^2}{\partial x^2}((\lambda + \beta x)P)$$

The FPE can be used to see how expected values of x , x^2 or indeed any function of x evolve in time. Here we will do expectation of a general function $f(x)$:

$$\begin{aligned}\frac{d}{dt}\langle f(x) \rangle &= \frac{d}{dt} \int f(x)P(x, t)dx \\ &= \int f(x) \frac{\partial P}{\partial t} dx \quad \text{now sub using FPE} \\ &= \int f(x) \left[\frac{\partial}{\partial x}(AP) + \frac{1}{2} \frac{\partial^2}{\partial x^2}(BP) \right] dx\end{aligned}$$

and at this point, sensible to think about where we would like to get to. Integral wrt x of something times P is the expected value of the something. So we'd like to dig the P out from under the partial x derivatives. The way to do that is integrating by parts (once on first time, twice on second). There are some boundary terms from the parts integration, but all involve P or $\frac{\partial P}{\partial x}$ as $x \rightarrow |\infty|$, and we assume that these tend to zero.

$$\begin{aligned}\frac{d}{dt}\langle f(x) \rangle &= \int f(x) \frac{\partial}{\partial x}(AP) dx + \frac{1}{2} \int f(x) \frac{\partial^2}{\partial x^2}(BP) dx \\ &= - \int \left(\frac{\partial}{\partial x} f(x) \right) (AP) dx + \frac{1}{2} \int \left(\frac{\partial^2}{\partial x^2} f(x) \right) (BP) dx \\ \langle f(x) \rangle' &= - \int A f' P dx + \frac{1}{2} \int B f'' P dx \\ \langle f(x) \rangle' &= \langle A f' \rangle + \frac{1}{2} \langle B f'' \rangle\end{aligned}$$

And now we can put in particular $f(x)$ and learn about the time evolution of the mean and variance. With $f(x) = x$ (so $f' = 1$, $f'' = 0$):

$$\langle x \rangle' = \langle A \rangle$$

and this shows how A therefore gives the drift of the mean. This is not hugely surprising as A is just sum of step size times step probability, so in some sense it captures the 'average' behaviour of the system.

Now to work towards the variance, set $f(x) = x^2$ (so $f' = 2x$, $f'' = 2$):

$$\begin{aligned}\langle x^2 \rangle' &= \langle A 2x \rangle + \frac{1}{2} \langle B 2 \rangle \\ &= 2 \langle Ax \rangle + \langle B \rangle\end{aligned}$$

so the variance time evolution is:

$$\begin{aligned}\frac{d}{dt} \text{var}(x) &= \frac{d}{dt} (\langle x^2 \rangle - \langle x \rangle^2) \\ &= \langle x^2 \rangle' - 2 \langle x \rangle \langle x \rangle' \\ &= 2 \langle Ax \rangle + \langle B \rangle - 2 \langle x \rangle \langle A \rangle \\ &= \langle B \rangle + 2 \text{cov}(A, x)\end{aligned}$$

So we can see that the rate of change involves A and how it covaries with x , but also the B term is now involved, and hence B contributes to how the population size distribution spreads, but not to its mean.

Example 2 continued further:

$$\langle x \rangle' = \langle A \rangle = \lambda - \beta \langle x \rangle$$

$$\begin{aligned}\frac{d}{dt} \text{var}(x) &= \langle B \rangle + 2 \langle Ax \rangle - 2 \langle x \rangle \langle A \rangle \\ &= \lambda + \beta \langle x \rangle + 2 \langle \lambda x - \beta x^2 \rangle - 2(\lambda - \beta \langle x \rangle) \langle x \rangle \\ &= \lambda + \beta \langle x \rangle - 2\beta \text{var}(x)\end{aligned}$$

which can easily be solved for steady states:

$$\langle x \rangle = \frac{\lambda}{\beta}, \quad \text{var}(x) = \frac{\lambda}{\beta}.$$

Interestingly the mean and variance at steady state match what we had for the full master equation, even though Fokker-Planck is an approximation. If we asked what higher terms would be if we continued the approximation earlier, we can see why this might be. The higher terms would involve third and higher derivatives with respect to x . After doing integration by parts, these end up on f . If we're looking for mean and variance, we'd use $f = x$ and $f = x^2$, so *third and higher derivatives would vanish*. So, that gets the first two moments. What happens for higher moments?

Exercise 34: Continue Example 1. Use the FPE to find $\langle x \rangle$, $\langle x^2 \rangle$, $\langle x^3 \rangle$ and $\langle x^4 \rangle$ (explicitly as functions of time).

Go back to the full master equation and find $\langle n \rangle$, $\langle n^2 \rangle$, $\langle n^3 \rangle$ and $\langle n^4 \rangle$.

Checkpoint:

$$\langle x^3 \rangle = (\lambda t)^3 + 3(\lambda t)^2, \quad \langle n^3 \rangle = (\lambda t)^3 + 3(\lambda t)^2 + (\lambda t)$$

You should find the first two moments match, but the next two do not, but actually they are a reasonable match for large t , why is this?

2.2.2 Multivariate Fokker-Planck

All of this can be extended to multivariate systems (for multiple populations). Much of the algebra works through in similar ways, just sometimes with vectors and matrices now. We do however get a nice result at the end of this, i.e. a relatively simple equation to find how populations covary with each other at steady state. We started to think about this with the wildebeest and flies example above (section 2.1.3). Actually we will use that example again to illustrate things in this section.

So now start setting up the language and equations: we write vector \mathbf{x} to give the population sizes. The ‘jump’ is now also a vector (\mathbf{r}) and if an event leads to only one population changing size, then the vector \mathbf{r} will be zeros except for one entry. We allow it to be more general though: for example our ‘populations’ might be stages of the same species e.g. caterpillars and butterflies, in which case we could reasonably have $\mathbf{r} = (-1, +1)$ for the event of one turning into the other. Anyway, there will only be a small (and certainly finite) set of \mathbf{r} to consider.

As before, we let $W(\mathbf{x}, \mathbf{r})$ be the rate at which we make a jump from \mathbf{x} of size \mathbf{r} , i.e. ending up at $\mathbf{x} + \mathbf{r}$.

Then our general master equation¹⁰ is given by

$$\frac{\partial}{\partial t} P(\mathbf{x}, t) = \sum_{\mathbf{r}} [W(\mathbf{x} - \mathbf{r}, \mathbf{r}) P(\mathbf{x} - \mathbf{r}, t) - W(\mathbf{x}, \mathbf{r}) P(\mathbf{x}, t)]$$

and again the square bracket is expanded using Taylor series, but now for a function of a vector. Here is the relevant expansion for a general F , expanding around \mathbf{x} and hoping \mathbf{r} is small (at least when compared with \mathbf{x}):

$$F(\mathbf{x} - \mathbf{r}) = F(\mathbf{x}) - r_i \frac{\partial F}{\partial x_i} + \frac{1}{2} r_i r_j \frac{\partial^2 F}{\partial x_i \partial x_j} + \dots$$

¹⁰for 1D we were a bit more careful about using N for when we were thinking about population size as discrete and x when continuous. By now, we know where this is going so just launch straight into \mathbf{x} here

where we are using summation convention. We could also write this in vectors using divergences, but this more explicit form seems like a sensible idea while we are working through the detail below.

Hence our master equation becomes, approximating to the first two terms only (so correct up to r^2 in effect):

$$\frac{\partial P}{\partial t} = \sum_{\mathbf{r}} -r_i \frac{\partial}{\partial x_i} (W P) + \frac{1}{2} \sum_{\mathbf{r}} r_i r_j \frac{\partial^2}{\partial x_i \partial x_j} (W P)$$

where we are still using summation convention. Now reorder sums and integrals to bunch up the sums into $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ terms:

$$\frac{\partial P}{\partial t} = - \frac{\partial}{\partial x_i} \underbrace{\left(\sum_{\mathbf{r}} r_i W \right)}_{A_i(\mathbf{x})} P + \frac{1}{2} \frac{\partial^2}{\partial x_i \partial x_j} \underbrace{\left(\sum_{\mathbf{r}} r_i r_j W \right)}_{B_{ij}(\mathbf{x})} P$$

where we will see that A_i give components of the vector of average movement. The matrix $\underline{\mathbf{B}}$ is trickier, but we can already see that it is a symmetric matrix (same if we swap i and j). So in summary we have our multidimensional FPE:

$$\boxed{\frac{\partial P}{\partial t} = - \frac{\partial}{\partial x_i} (A_i P) + \frac{1}{2} \frac{\partial^2}{\partial x_i \partial x_j} (B_{ij} P) \quad \text{Fokker-Planck Equation (FPE)}}$$

Exercise 35: Find $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ for wildebeest and flies.

Much as we did for one dimension, we can use the FPE to find time evolution of any function of \mathbf{x} , but here we will explicitly look at components of first and second moments.

$$\begin{aligned} \frac{d}{dt} \langle x_m \rangle &= \frac{d}{dt} \int x_m P(\mathbf{x}, t) dx_1 dx_2 \dots dx_n \\ &= \int x_m \frac{\partial P}{\partial t} dV \quad \text{now sub using FPE} \\ &= \int -x_m \frac{\partial}{\partial x_i} (A_i P) + \frac{1}{2} x_m \frac{\partial^2}{\partial x_i \partial x_j} (B_{ij} P) dV \end{aligned}$$

and again we use parts (or divergence theorem if thinking in terms of vectors) to liberate the P from the derivatives. Again, assume that P and its derivatives go to zero at the boundaries:

$$\begin{aligned}
\frac{d}{dt}\langle x_m \rangle &= \int + (A_i P) \underbrace{\frac{\partial}{\partial x_i} x_m}_{=\delta_{im}} + \frac{1}{2} (B_{ij} P) \underbrace{\frac{\partial^2}{\partial x_i \partial x_j} x_m}_{=0} dV \\
&= \int (A_i P) \delta_{im} dV = \int A_m P dV = \langle A_m \rangle
\end{aligned}$$

Hence in vectors we could write:

$$\langle \mathbf{x} \rangle' = \langle \mathbf{A} \rangle$$

and hence we can see that \mathbf{A} represents the ‘mean drift’, i.e. the time derivative of the mean population sizes. Now continue with same approach for $\langle x_m x_n \rangle$, but we can reuse the working above as it is almost the same initially, and jump in after the integration by parts:

$$\begin{aligned}
\langle x_m x_n \rangle' &= \int + (A_i P) \underbrace{\frac{\partial}{\partial x_i} x_m x_n}_{=\delta_{im} x_n + \delta_{in} x_m} + \frac{1}{2} (B_{ij} P) \underbrace{\frac{\partial^2}{\partial x_i \partial x_j} x_m x_n}_{=\delta_{im} \delta_{jn} + \delta_{in} \delta_{jm}} dV \\
&= \int (A_m x_n + A_n x_m) P + \frac{1}{2} (B_{mn} + B_{nm}) P dV \\
&= \int (A_m x_n + A_n x_m) P + B_{mn} P dV
\end{aligned}$$

using that $B_{mn} = B_{nm}$. Hence

$$\langle x_m x_n \rangle' = \langle A_m x_n \rangle + \langle A_n x_m \rangle + \langle B_{mn} \rangle.$$

These second moments can be used to consider the covariances

$$C_{mn} = \text{cov}(x_m, x_n) = \langle x_m x_n \rangle - \langle x_m \rangle \langle x_n \rangle,$$

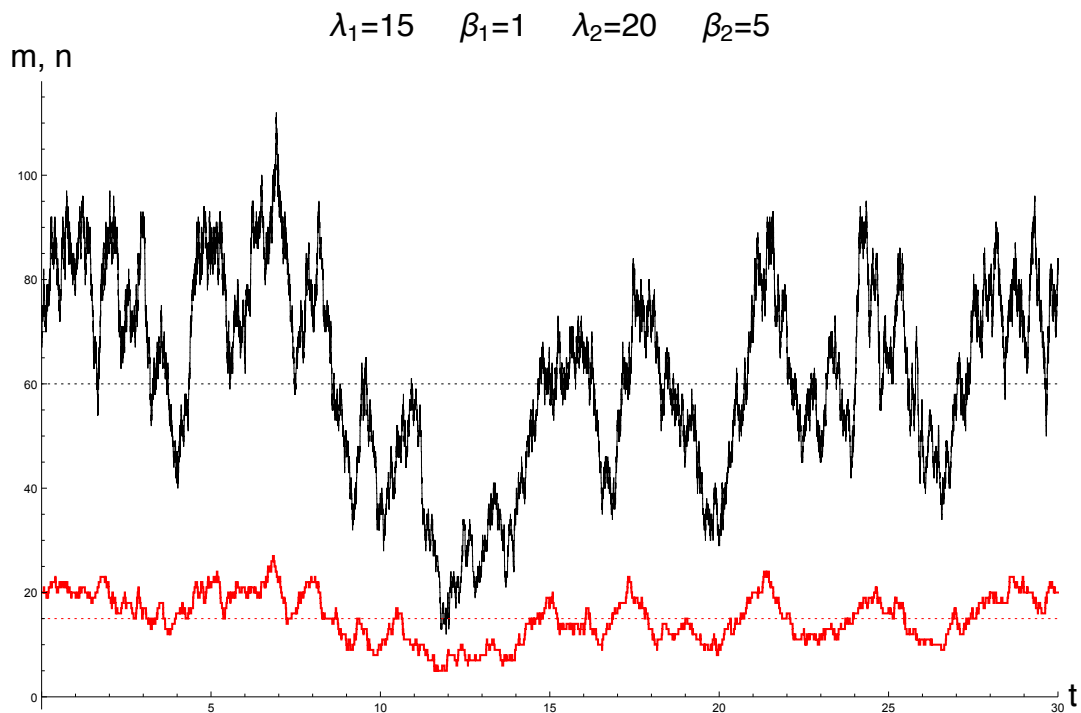
so $C_{ii} = \text{var}(x_i)$, and $\underline{\underline{\mathbf{C}}}$ is a symmetric matrix. Using the above:

$$\begin{aligned}
C'_{mn} &= \langle x_m x_n \rangle' - \langle x_m \rangle' \langle x_n \rangle - \langle x_m \rangle \langle x_n \rangle' \\
&= \langle A_m x_n \rangle + \langle A_n x_m \rangle + \langle B_{mn} \rangle - \langle A_m \rangle \langle x_n \rangle - \langle x_m \rangle \langle A_n \rangle \\
&= \langle A_m x_n \rangle - \langle A_m \rangle \langle x_n \rangle + \langle A_n x_m \rangle - \langle x_m \rangle \langle A_n \rangle + \langle B_{mn} \rangle \\
C'_{mn} &= \text{cov}(A_m, x_n) + \text{cov}(A_n, x_m) + \langle B_{mn} \rangle
\end{aligned}$$

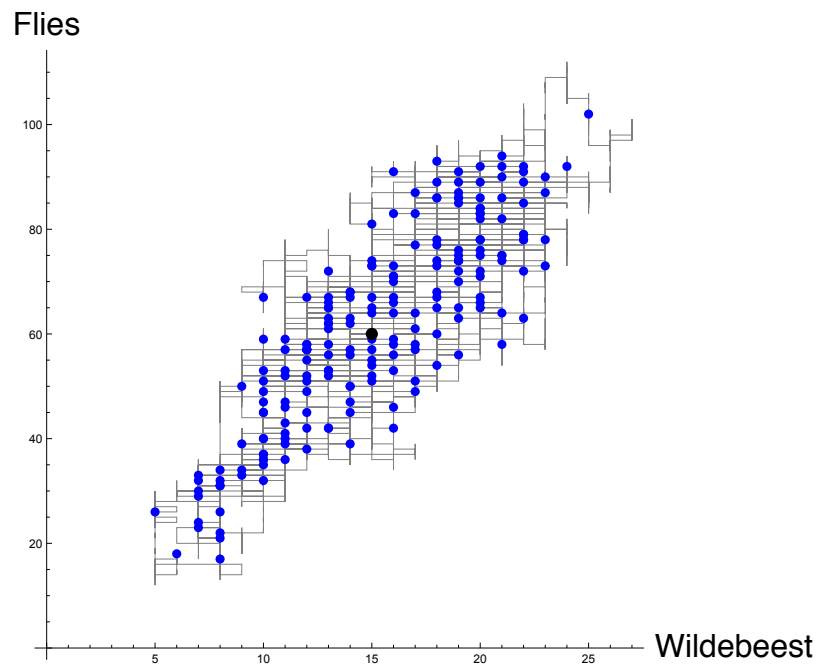
and all this holds for general \mathbf{A} and $\underline{\underline{\mathbf{B}}}$.

Finding behaviour at steady state:

Here is another run of the wildebeest and flies, each plotted against time:



a different way to visualise this would be plotting flies against wildebeest, i.e. exactly the two-dimensional space that the system is moving around in:



and as the tracks go over themselves repeatedly, we can add the blue dots at fixed time intervals¹¹. And by now we are comfortable with thinking of this as steady state: the system is not stuck at fixed \mathbf{x} but rather the probability of being a in a given state has settled to a constant.

We can see that the steady state mean (marked in black) is in the middle of this random splurge, but we can also see that the spread around it has some interesting shape: it is spread out diagonally. This corresponds to saying the wildebeest and flies covary positively.

We would like to find this steady state covariance (and variances) in general (not just wildebeest and flies model). We have an equation above for C' , but it depends on $A_m(\mathbf{x})$ and $B_{mn}(\mathbf{x})$ which in turn depend on \mathbf{x} .

Here's the final idea that we need to do this. Approximate near steady state (summation convention below):

- Approximate A_m as being linear in \mathbf{x} :

$$A_m(\mathbf{x}) = \lambda_m + a_{mk} x_k$$

- Approximate B_{mn} as being constant in \mathbf{x} :

$$B_{mn}(\mathbf{x}) = b_{mn}$$

and in both cases this is usually done by taking approximation (Taylor series) near steady state. All of λ_m , a_{mk} and b_{mn} now are constants. Actually $\underline{\mathbf{a}}$ can be thought of as the Jacobian near the steady state mean.

Then (still with summation convention):

$$\text{cov}(A_m, x_n) = a_{mk} \text{cov}(x_k, x_n) = a_{mk} C_{kn}$$

$$\langle B_{mn} \rangle = b_{mn}$$

and hence:

$$C'_{mn} = a_{mk} C_{kn} + a_{nk} C_{km} + b_{mn}.$$

This looks nice in matrix notation:

$$\underline{\underline{C'}} = \underline{\underline{\mathbf{a}}} \underline{\underline{C}} + (\underline{\underline{\mathbf{a}}} \underline{\underline{C}})^T + \underline{\underline{\mathbf{b}}}$$

¹¹should be careful to use set time period, not fixed number of steps, or will not be sampling uniformly over time

and as $\underline{\underline{C}}$ is symmetric, $(\underline{\underline{a}}\underline{\underline{C}})^T = \underline{\underline{C}}\underline{\underline{a}}^T$. So at steady state, the covariace matrix $\underline{\underline{C}}$ must satisfy:

$$\underline{\underline{a}}\underline{\underline{C}} + \underline{\underline{C}}\underline{\underline{a}}^T + \underline{\underline{b}} = 0 \quad \text{Lyapunov equation}$$

or in components (with summation convention):

$$a_{ik}C_{kj} + a_{jk}C_{ki} + b_{ij} = 0.$$

Exercise 36: Continue with the wildebeest and flies: approximate around steady state, i.e.:

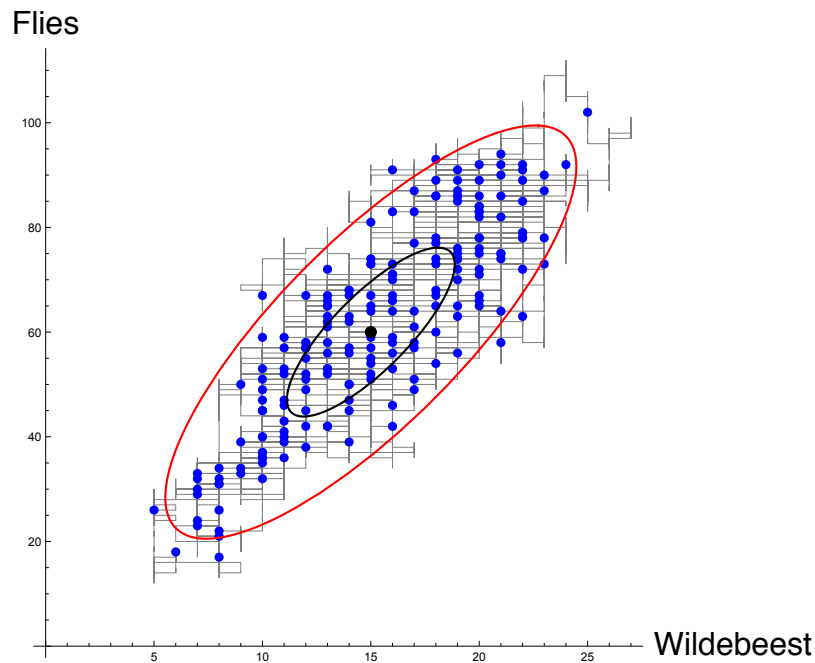
$$x_1 = \frac{\lambda_1}{\beta_1}, \quad x_2 = \frac{\lambda_1\lambda_2}{\beta_1\beta_2}$$

(i.e. where $\mathbf{A} = 0$). Find $\underline{\underline{a}}$, $\underline{\underline{b}}$. By taking components of the Lyapunov equation show that near steady state:

$$C_{11} = \frac{\lambda_1}{\beta_1}, \quad C_{12} = \frac{\lambda_1\lambda_2}{\beta_1(\beta_1 + \beta_2)}, \quad C_{22} = \frac{\lambda_1\lambda_2}{\beta_1\beta_2} \left(1 + \frac{\lambda_2}{\beta_1 + \beta_2}\right).$$

These give the variance of x_1 , covariance between x_1 and x_2 , and the variance of x_2 respectively.

And just for completeness with the example we have been following, $\underline{\underline{C}}$ can be used to find the confidence ellipse (or error ellipse), and shown here for equivalent of one standard deviation, and also the 95% CI. And this looks pretty good at capturing the spread and directionality of fluctuations from the mean at steady state.



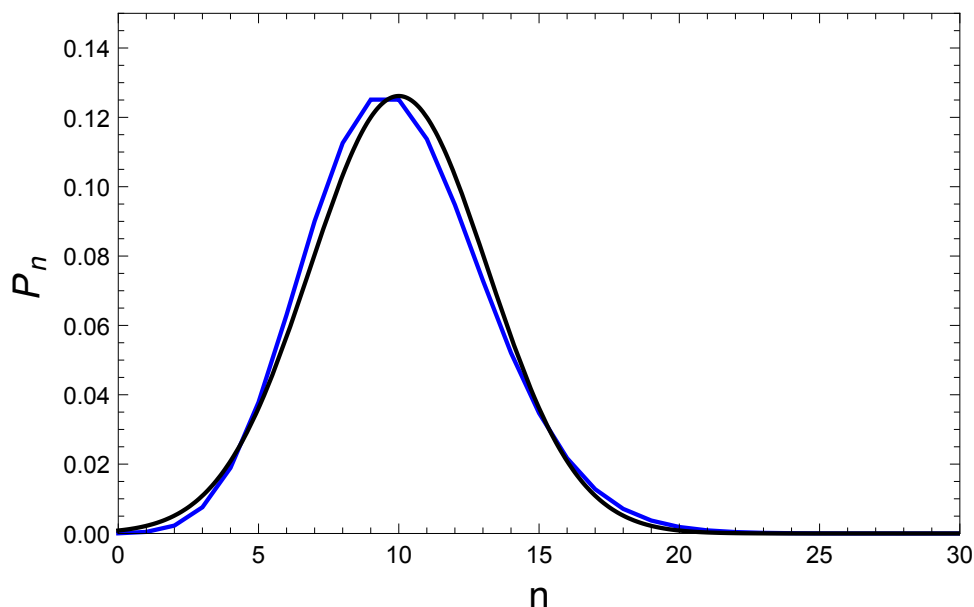
Steady-state vs quasi steady-state

For one small logical jump we can have a big extension to our covariance result. We already know what if zero (for all species) is fixed (no jumps out), then the only true steady state for the full system is total extinction in the long term. However, if we have a solution \mathbf{x} for where the mean is constant in time $\mathbf{A}(\mathbf{x}) = 0$ and it is well away from zero, then we call a spread around that as a 'quasi steady-state'. It is not truly a steady state, but it might look like one for a very very long time, so our above results hold well.

We could in principle be a little more formal about it either by thinking about eigenvectors of the Markov transition matrix (the one with second largest eigenvalue is the quasi-steady-state: the largest corresponds to total extinction). Or we could write about large t conditioning on non-extinction and call that the quasi steady state, but no need to go further for this course: just know that it is often OK to use these results even for systems where extinction is possible.

Example (not done in lectures, ignore if not helpful)

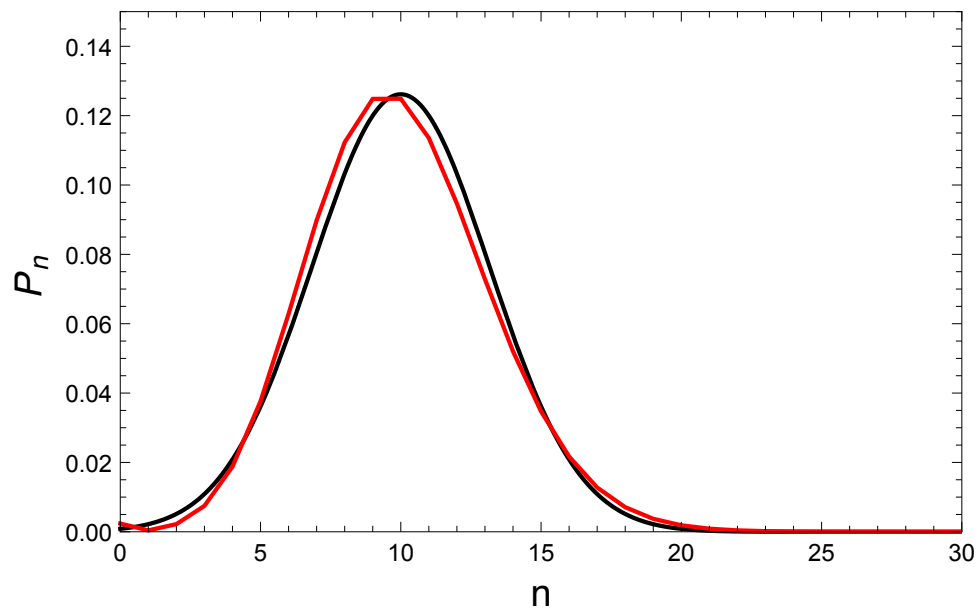
To illustrate what we mean by quasi steady-state, we can use one-dimension. Consider again the simple import and death model (example 2). Set $\lambda = 10, \beta = 1$ for these plots. So we know that the steady state solution has mean and variance 10. Here's an output for large t for the full model in blue, and the normal distribution with the mean and variance from the FPE at steady state in black:



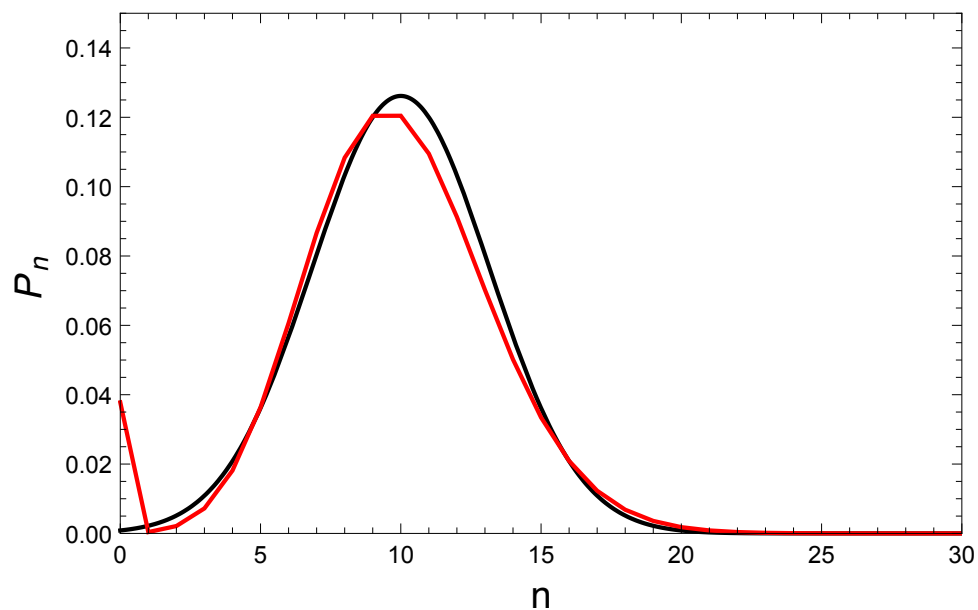
This matches fairly well.

If we now modify the import and death model so there is NO import if $n = 0$, then we in principle know that if we wait long enough, then $P_0 = 1$ and $P_n = 0$ for $n \neq 0$. However,

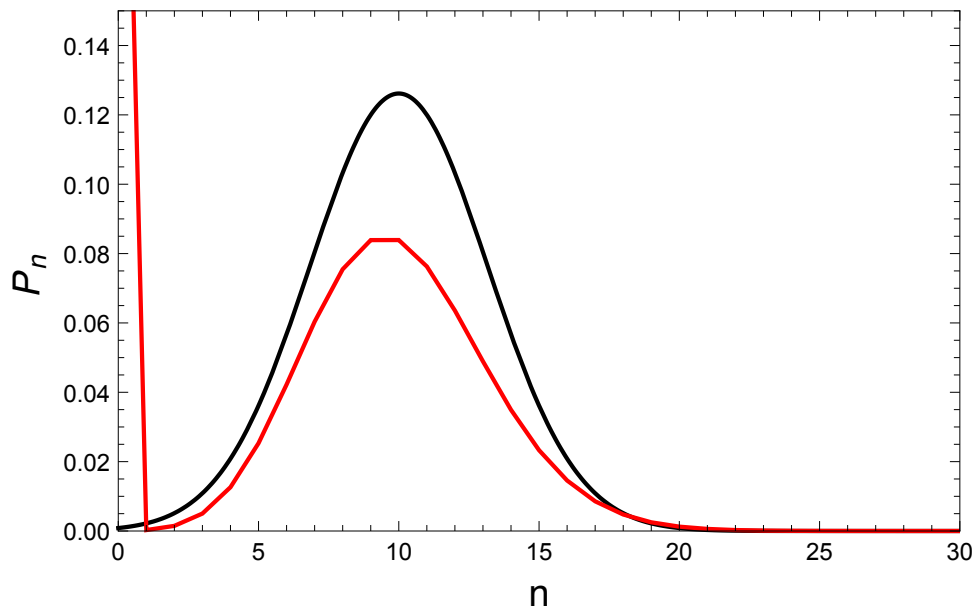
look at what happens for modest time, here $t = 10$ (modified import and death model in red):



It just looks as if there is no modification, hence the steady-state approximation from before is pretty good for early times for this modified system. Indeed if we increase time, here to $t = 100$:



We can see the shape is right, except some of the probability density has 'leaked' out to extinction at $N = 0$. And go further to $t = 1000$:



It still has basically right shape, even though more area has been lost to P_0 . The rest of the curve still keeps its mean and variance.

The FPE for both the original model and this modified model will be almost the same (just A and B will differ at $x = 0$), and the approximation for the variance near steady state from the Lyapunov equation will be the same (from approximating A as linear in x and B as constant near $x = 10$). Unsurprisingly, we've lost the information about extinction being possible. However, we do have the right mean and variance for some large range of times, or even later times if we condition on non-extinction: call this quasi steady-state.