Stochastic population dynamics: branching processes

Erol Akçay

Biol 417 Spring 2019

February 18, 2019

So far, we have dealt entirely with deterministic models of population dynamics, in which if I tell you the parameters of the model and the initial state of the population, you know exactly how the population will look like at all points in the future.[1] This is true even for the discrete logistic model when it exhibits chaotic behavior: *if* you know the initial conditions precisely, you can predict what is going to happen.

[1] Assuming that the parameters don't change.

Obviously, real world is somewhat more complicated. In particular, nature is stochastic (at least for our purposes) so even if you know what is likely to happen *in expectation*, each particular population is a *realization* of a stochastic process. Stochastic models are fundamentally different than deterministic ones. In a deterministic model, we can hope to predict the trajectory of state variables given enough information, but in a stochastic model, we can only describe some statistics about state variables in the future even with a detailed specification of the initial conditions. In what follows, we follow the same path as the deterministic model, first considering continuous time, then discrete.

The birth-death process in continuous time

As mentioned above, the state variables in a stochastic model are fundamentally different than in a deterministic model. In particular, we cannot really follow the population size of a given population, since that will fluctuate randomly. Instead, the best we can hope to do is to follow the *probability distribution* of population sizes: i.e., if we start the population, at a given size at a given time, what is the probability that we will find the population at some other size at a later time? This is the question we will focus on with a very simple model called a birth-death process.

To construct this model, suppose the probability of finding the population at size n at time t is given by $p_n(t)$. Suppose that the probability that there is a death between t and some (slightly) later time t' is given by $bn(t'-t)$, where b is a constant signifying the per capita birth rate. We assume that b is independent of n, which means we are modeling a case without density dependence.[2] Likewise, suppose that the probability of a death occurring between t and t' is $dn(t'-t)$, where d is the death rate. Similar to what we did in deterministic continuous time model, we take the time interval $t'-t$ to be so small that the probability of two births, two deaths, or a birth and a death during that time interval is negligible. We now want to write the probability of being at population size n

[2] In the next section, we will allow the birth- and death-rates to depend on population size.

at time t'. There are three ways the population can end up at size n at time t': (i) it can start at $n-1$ at t, and experience a birth event, the probability of which is $p_{n-1}(t)b(n-1)(t'-t)$, (ii) it can start at $n+1$ and experience a death event, which has probability $p_{n+1}(t)d(n+1)(t'-t)$, and (iii) it can start at n, and experience neither a birth nor a death event, which happens with probability $p_n(t)(1-b-d)n(t'-t)$. Thus, we can write

$$p_n(t') = p_{n-1}(t)b(n-1)(t'-t) + p_{n+1}(t)d(n+1)(t'-t) + p_n(t)(1-(b+d)n(t'-t)) .$$

(1)

If we subtract $p_n(t)$ from both sides, divide by $(t'-t)$ and take the limit $t' \to t$, we end up with the instantaneous change in $p_n(t)$:

$$\frac{dp_n}{dt} = \lim_{t' \to t} \frac{p_n(t') - p_n(t)}{t' - t} = p_{n-1}(t)b(n-1) + p_{n+1}(t)d(n+1) - p_n(t)(b+d)n$$

(2)

We thus have a system of ordinary differential equations (ODEs) describing the time evolution of the probability of finding the population at any given size. But how do we go about solving this system? Note that our system of ODEs is technically infinite-sized, because there is no bound on n and we have one equation for each n. This seems like a impossibly daunting object to handle, but it turns out we can handle it just fine in a lot of interesting cases.

First, let us consider the somewhat artifical case of no deaths, i.e. $d = 0$, signifying a population that can only grow.[3] Suppose you initiate this population at some population size n_0 for $t = 0$, so that $p_{n_0}(0) = 1$. Then, the equation for n_0 reduces to:

$$\frac{dp_{n_0}}{dt} = -bn_0 p_{n_0}(t) ,$$

(3)

which is basically the exponential equation from before (since b and n_0 are constants), so its solution is given by: $p_{n_0}(t) = e^{-bn_0 t}$. In other words, the probability of finding the population at the initial size will decrease exponentially. The equation for $n_0 + 1$ is then:

$$\frac{dp_{n_0+1}}{dt} = bn_0 p_{n_0}(t) - b(n_0 + 1)p_{n_0+1}(t) ,$$

(4)

but now we know $p_{n_0}(t)$, and can directly substitute it, to get:

$$\frac{dp_{n_0+1}}{dt} = bn_0 e^{-bn_0 t} - b(n_0 + 1)p_{n_0+1}(t) ,$$

(5)

which is solved by $p_{n_0+1} = n_0 \left(e^{bt} - 1 \right) e^{-b(n_0+1)t}$. In a similar manner, one can write out the equation for p_{n_0+2}, substitute p_{n_0+1}, solve, and so by induction can obtain the time trajectory of all p_n for $n > n_0$, to get:

$$p_n(t) = \binom{n-1}{n_0-1} e^{-bn_0 t}(1 - e^{-bt})^{n-n_0}$$

(6)

[3] Obviously this cannot be a good long-term model of any population, since it would blow up, but one might imagine cases (e.g., long-lived organisms) where in the short run it may be an adequate approximation. Here, though, we are mainly using it to illustrate a technique of solving the master equation.

However, when $d > 0$, this method of analysis is not available, since what happens to p_n now depends both on what happens to p_{n-1} and p_{n+1}. Solving "from below" therefore does not work, since we would also simultaneously solve from above. We need a different method. To find one, let us take a brief detour to pick up a very useful tool in analyzing stochastic processes.

Probability generating functions

If X is a discrete random variable that can take non-negative integer values with probabilities $p_x = \text{Prob}(X = x)$, then the series

$$F(z) = \sum_{i=0}^{\infty} p_i z^i$$

is called the *probability generating function* (PGF) of the random variable X. The PGF $F(z)$ is a very useful mathematical object. Firstly, one can directly express the probabilities of each event in terms of $F(z)$ by taking its derivative and evaluating at $z = 0$:[4]

[4] Hence, the name "Probability Generating Function."

$$F(0) = p_0 = \text{Prob}(X = 0)$$

$$\left. \frac{dF}{dz} \right|_{z=0} = \sum_{i=0}^{\infty} p_i i 0^{i-1} = p_1 = \text{Prob}(X = 1)$$

$$\vdots$$

$$\left. \frac{1}{j!} \frac{d^j F}{dz^j} \right|_{z=0} = p_j = \text{Prob}(X = i)$$

Likewise, the moments of X (i.e., mean, variance, etc.) are also readily expressed in terms of the PGF and its derivatives – you just have to evaluate the derivatives at $z = 1$ (and do some algebra). For the first two moments, the mean and the variance of a probability distribution, we have:

$$\left. \frac{dF}{dz} \right|_{z=1} = \sum_{i=0}^{\infty} i p_i = \text{E}(X)$$

$$\left. \frac{d^2 F}{dz^2} \right|_{z=1} = \sum_{i=0}^{\infty} i(i-1) p_i = \text{E}(X^2) - \text{E}(X) = \text{Var}(X) + \text{E}(X)^2 - \text{E}(X)$$

$$(7)$$

Analyzing the birth-death process using probability generating functions

The above properties of PGFs indicate that function $F(z)$ carries in itself all the information about the probability distribution of a discrete random variable, such as population size. So, if we could write down an equation that gives us for the PGF for population size changes over time, then we could deduce from it the time-dynamics of p_n.

Let us therefore write the time-derivative of $F(z,t) = \sum_{n=0}^{\infty} p_n(t)z^n$: [5]

$$\frac{\partial F}{\partial t} = \sum_{n=0}^{\infty} \frac{dp_n}{dt} z^n = \sum_{n=0}^{\infty} \left(p_{n-1}b(n-1) + p_{n+1}d(n+1) - p_n(b+d)n \right) z^n \ .$$
(8)

Now, each of the terms in the parentheses are products of a constant, p_i, and i. This suggests that we can write these as derivatives of $F(z,t)$ with respect to z, since $F'(z) = \sum_{i=0}^{\infty} ip_i z^{i-1}$. Doing so, we end up with:

$$\frac{\partial F}{\partial t} = \frac{\partial F}{\partial z} \left(bz^2 - (b+d)z + d \right) \ .$$
(9)

In other words, we now have a partial differential equation of first order for the probability generating function. If the population starts at size n_0 at $t = 0$, we have the initial condition:

$$F(z,0) = z^{n_0} \ .$$
(10)

This partial differential equation can be solved using the method of characteristic curves.

Solving eq 9 by the method of characteristics

The intuition behind the method of characteristic curves is as follows: $F(z,t)$ describes a surface in a three-dimensional space. We want to find curves on this surface where F stays constant so that we can write the partial differential equation as a collection of ordinary differential equations; the collection of all such curves then gives us the surface. To find these curves, we make a change of variables from (z,t) to some other pair (u,v), where v will determine which curve we are on and u where in that curve. If $b \neq d$[6], we can write:

$$\frac{dF}{du} = \frac{\partial F}{\partial t}\frac{dt}{du} + \frac{\partial F}{\partial z}\frac{dx}{du} \ .$$
(11)

Now if we select $u = t + $ Constant, such that $\frac{dt}{du} = 1$, then we can see that (9) is equivalent to two ordinary differential equations:

$$\frac{dz}{du} = - \left(bz^2 - (b+d)z + d \right)$$
(12)

$$\frac{dF}{du} = 0 \ .$$
(13)

In other words, F is constant along the curve described by the variable u. Without loss of generality, we can set $u = t$ (since that only amounts to setting the initial point of u). We then set v^{n_0} to be the value of F along the curve traced by u:

$$F(0,v) = v^{n_0} \ ,$$
(14)

which gives us the initial value for z at $u = 0$:

$$z(0, v) = v \, . \tag{15}$$

Since $\frac{dF}{du} = 0$ and $F(0, v) = v^{n_0}$, we have

$$F(u, v) = v^{n_0} \, , \tag{16}$$

which means if we solve (12) and invert it to solve for v, we are done. To do that, write (12) as:

$$(bz - d)(1 - z)du = dz \, , \tag{17}$$

which after a bit of rearranging gives us:

$$\left(\frac{1}{1 - z} + \frac{b}{bz - d} \right) dz = (b - d)du \, . \tag{18}$$

Now we have the sum of $-1/x dx$ plus $1/y dy$ on the left hand side (with $x = 1 - z$ and $y = bz - d$), which we know are derivatives of $\ln x$ and lny, so integration gives us:

$$- \ln(1 - z) + \ln(bz - d) = \ln \left(\frac{bz - d}{1 - z} \right) = (b - d)u + C \, , \tag{19}$$

where C is a constant. Exponentiating both sides (and setting $e^C = c$ yields:

$$\left| \frac{bz - d}{1 - z} \right| = ce^{(b-d)u} \, . \tag{20}$$

Our initial conditions say that when $u = t = 0$, $z = v$, so we have for the constant c:

$$\left| \frac{bv - d}{1 - v} \right| = c \, , \tag{21}$$

which substituting back into equation (20) and also substituting $t = u$ gives:

$$\left| \frac{bz - d}{1 - z} \right| = \left| \frac{bv - d}{1 - v} \right| e^{(b-d)t} \tag{22}$$

Finally, we can solve for v and get:

$$v = \frac{d(1 - z)e^{(b-d)t} - (d - bz)}{b(1 - z)e^{(b-d)t} - (d - bz)} \, , \tag{23}$$

which means:

$$F(z, t) = \left[\frac{d(1 - z)e^{(b-d)t} - (d - bz)}{b(1 - z)e^{(b-d)t} - (d - bz)} \right]^{n_0} \, . \tag{24}$$

The nice thing about the PGFs is that we are done.[7] To get and p_i, we simply need to differentiate (24) the appropriate number of times and evaluate it at $z = 0$, which would be tedious by hand, but is a breeze with a computer algebra system like Mathematica or sympy.

[7] in principle, at least.

Analyzing the birth-death process in discrete time using matrix algebra

In the preceding, we analyzed a simple birth-death process using probability generating functions. Another mode of analysis for the same problem in discrete time is available that uses the machinery of matrix models we have encountered for structured populations before.[8] In particular, suppose the probability that the population is at size n at time t is given by $p_n(t)$.[9] Let us assume again that the time interval is small enough that we can only have one birth or death in one step, so the population can either increase by one, decrease by one, or stay unchanged in any time step, with probabilities b_n, d_n, and $1 - b_n - d_n$, respectively. Notice now that the birth and death probabilities (which are not rates now, but probabilities, so have to be equal to or less than 1) are indexed by n, meaning that we allow them to be density dependent (but I assume they are time-invariant). Then, we can write for $p_n(t + 1)$:

$$\cdots$$
$$p_{n-1}(t + 1) = p_{n-2}(t)b_{n-2} + p_n(t)d_n + p_{n-1}(t)(1 - b_{n-1} - d_{n-1})$$
$$p_n(t + 1) = p_{n-1}(t)b_{n-1} + p_{n+1}(t)d_{n+1} + p_n(t)(1 - b_n - d_n)$$
$$p_{n+1}(t + 1) = p_n(t)b_n + p_{n+2}(t)d_{n+2} + p_{n+1}(t)(1 - b_{n+1} - d_{n+1})$$
$$\cdots$$

Suppose for the moment that there is a maximum population size, m. This is a realistic assumption for many populations that after all live on a finite Earth, and it will simplify some of the exposition.[10] Now we can use the short-hand of matrix algebra to write the same equation, just as we did for structured population dynamics:

$$\mathbf{p}(t + 1) = \mathbf{M}\mathbf{p}(t) \,, \tag{25}$$

where $\mathbf{p}(t)$ is the vector with $m + 1$ elements collecting all the $p_n(t)$ in one place (starting with $p_0(n)$, and $\mathbf{M}$ is a $(m + 1) \times (m + 1)$ square matrix called the "transition matrix", with elements:

$$\mathbf{M} = \begin{pmatrix} 1 - b_0 & d_0 & 0 & 0 & \cdots \\ b_0 & 1 - b_1 - d_1 & d_1 & 0 & \cdots \\ 0 & b_1 & 1 - b_2 - d_2 & d_2 & \cdots \\ 0 & 0 & b_2 & 1 - b_3 - d_3 & \cdots \\ & & \vdots & & \end{pmatrix} \tag{26}$$

Note that we have a probability b_0 of a "birth" happening in a population of size 0. Biologically, this can be thought of the probability an immigrant founding a population where it did not exist (or had gone extinct). If $b_0 = 0$, then $n = 0$ is an **absorbing state**: once the process reaches it, it can never get out. Note that all the columns of the matrix $\mathbf{M}$ add up to 1. This is a reflection of the fact that bs

and ds are now probabilities (as opposed to rates like they were in the continuous time model). A population in each state has to do something, so the probabilities of all the things that it can do have to add up to 1. Such matrices where all columns add up to 1 are called **stochastic matrices**. Stochastic matrices are guaranteed to have at least one leading eigenvalue that is exactly equal to 1.[11] This is a nice fact, since as you might remember from matrix population models, the leading eigenvalue of $\mathbf{M}$ gives the long-term rate of growth of the vector $\mathbf{p}$, but since $\mathbf{p}$ is a probability distribution, its elements should always add up to 1, and thus should not grow or shrink.

Stationary distribution

One important piece of information we frequently want about stochastic processes is their long-term outcome, just as we want to learn about the equilibria of deterministic systems. Obviously, the long term outcome can only be expressed probabilistically in stochastic systems, so we want to learn if a system has a probability distribution of finding the system at some state in the long term. This is called the stationary distribution. If we denote the stationary distribution of $\mathbf{M}$ with π, it should satisfy:

$$\pi = \mathbf{M}\pi \,, \tag{27}$$

in other words, the stationary distribution is a right eigenvector of $\mathbf{M}$ associated with the leading eigenvalue (which is 1). Suppose that the matrix $\mathbf{M}$ doesn't exhibit periodic behavior, which is true in our example. [12] Then, there are two possible outcomes. If there are no absorbing states, then there exists unique leading eigenvalue and a corresponding eigenvector. If there are $k \geq 1$ absorbing states, there are k eigenvalues equal to 1, each with an associated eigenvector that corresponds to an absorbing state (i.e., if state j is an absorbing state, then the eigenvector has a one for its jth element, and zeros elsewhere. Therefore, for aperiodic Markov chains, the system will always end up in one of the absorbing states, if there are any. This gives us our first conclusion: if the population size is at most finite ($\leq m$) and $b_0 = 0$, then eventually the population will go extinct for sure. This makes intuitive sense. For any finite population, there is a positive (even if small) probability that reproduction will fail enough times in succession that everyone will die. Since that only has to happen once for the population to go extinct, and in the long term, there are essentially infinite chances for that sequence of events to happen, it will happen.[13]

Suppose now that $b_0 > 0$ so there is no extinction. How can we find the stationary distribution? Well, we can use a computer system to identify the eigenvectors using a computer, but if m is large, that can be computationally hard. Fortunately, the special structure of the matrix $\mathbf{M}$ in the birth-death model gives us a more direct way. Consider equation (27); the equation for the first

element of vector π, π_0 is:

$$\pi_0 = (1 - b_0)\pi_0 + d_1\pi_1$$

$$\Rightarrow \quad \frac{b_0}{d_1}\pi_0 = \pi_1 \tag{28}$$

Likewise, the equation for π_1 is:

$$\pi_1 = (1 - b_1 - d_1)\pi_1 + b_0\pi_0 + d_2\pi_2$$

$$\pi_1(b_1 + d_1) - b_0\pi_0 = d_2\pi_2$$

$$\Rightarrow \frac{b_1}{d_2}\pi_1 = \frac{b_0}{d_1}\frac{b_1}{d_2}\pi_0 = \pi_2 \ , \tag{29}$$

where for the last step, we substituted from equation (28). We can keep going like this, and write for $n > 0$:

$$\pi_n = \prod_{i=0}^{n-1} \frac{b_i}{d_{i+1}}\pi_0 \ , \tag{30}$$

Defining $\kappa_n = \prod_{i=0}^{n-1} \frac{b_i}{d_{i+1}}$, and $\kappa_0 = 1$, we can divide equation (30) by the sum of all π_n to get the normalized[14] stationary distribution:

$$\pi_n = \frac{\kappa_n}{\sum_{j=0}^{m} \kappa_j} \tag{31}$$

[14] i.e., made to add up to 1.

This expression also confirms why $b_0 > 0$ is crucial: if $b_0 = 0$, all κ_n for n>0 are zero, and only $\pi_0 = \kappa_0 = 1$, consistent with our result above. Interestingly, this calculation also works for when $m \to \infty$, provided that sum in the denominator stays finite as m goes to infinity; otherwise, there is no stationary distribution and the population will eventually reach infinite size. This condition basically means that at some population size either the birth probability should decrease fast enough with population size or the death probability should increase with population size, i.e., there is enough density dependence.

Branching processes

A related type of question to the one we have been concerned so far is to determine the fate of a lineage that starts with a single individual. This problem was first posed by Reverend H.W. Watson and Sir Francis Galton who were interested in the extinction rate of family names.[15] In modern evolutionary theory, branching processes have an important role in determining the invasion success of a new type (e.g., a de novo mutation) in a population, so is an important part of evolutionary analyses.

Let us formulate the question in the following way: at time 0, we start with a single individual. The probability that the individual has i descendants (including itself) is given by q_i for $i \geq 0$.[16] Each subsequent time step, each of the descendants reproduce according to the same probability distribution. The

[15] Which is a very British problem. Galton posed the problem in the *Educational Times* in 1873, but apparently received only one answer that completely missed the point, so he drafted his mathematician friend Reverend Watson, who proposed the iterative solution we'll talk about. But ironically, Watson and Galton, because of an algebraic error, concluded all lineages must go extinct, which is not true. The branching process reappeared in evolutionary biology in the late 1920s and 30s through the work of JBS Haldane and RA Fisher.See Kot (2001) for more detail and further references.

[16] If an individual has zero descendants, it means that individual has died without producing offspring.

question we are then interested in is: what is the probability that, in the long term, the lineage that is founded by a single individual will go extinct? One minus the extinction probability gives us the probability that the lineage will persist, or establish itself in the population.

We will again make use of probability generating functions in solving this problem. First, we need to introduce two more properties of PGFs. Consider two independent random variables with integer values, denoted by X and Y, with PGFs $F_X(z) = \sum_{i=0}^{\infty} P_{X,i} z^i$ and $F_Y(z) = \sum_{i=0}^{\infty} P_{Y,i} z^i$, respectively. Then, the PGF of the random variable $X + Y$, F_{X+Y}, is given by:

$$
\begin{aligned}
F_{X+Y}(z) &= \sum_{i=0}^{\infty} P_{X+Y,i} z^i \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{i} P_{X,i} P_{Y,i-j} z^i \\
&= \sum_{j=0}^{\infty} \sum_{i=j}^{\infty} P_{X,i} P_{Y,i-j} z^j z^{i-j} \\
&= \sum_{j=0}^{\infty} P_{X,i} z^j \sum_{i=j}^{\infty} P_{Y,i-j} z^{i-j} = F_X(z) F_Y(z) .
\end{aligned}
\tag{32}
$$

So, the PFG of the sum of two independent random variables is the product of their PGFs. This extends to the sum of any number. This is the first property. To see the second one, consider the sum S_N of N independent, identically distributed random variables with PGF $F(z)$, but where N itself is a random variable (independent from the ones to be summed) with PGF $G(z)$. The PGF of this sum is:

$$
\sum_{n=0}^{\infty} p_n \left[F(z) \right]^n = G(F(z)) ,
\tag{33}
$$

in other words, the PGF of is simply equal to the PGF of N but with its argument replaced by the PGF of the summand random variables. This is the second property.

You can see why the second property is relevant: if we start with a single individual, and want to count the number of descendants it has two time steps down the line, we need to add up the descendants in one time step, but the number of descendants itself is a random variable. Let us denote the PGF of the distribution of descendants of a single individual by $F(z)$, and the PGF of the number of descendants in generation i by $F_i(z)$. Since the PGF for the number of descendants in the first generation and the number of descendants of each in the next generation are the same, we can write for the PGF of the number of descendants two generations down the line:

$$
F_2(z) = F(F_1(z)) = F(F(z))
$$

More generally:

$$F_i(z) = F^i(z) \,, \tag{34}$$

where the subscript between the function name and argument denotes repeated application (or i-fold composition) of the function.

Now we are ready to solve our problem, finding the probability of extinction of a lineage. The probability of extinction at time step i is simply given by $F_i(0) = F^i(0)$. If we take the limit $i \to \infty$, this gives us the long-term probability of extinction. But we don't have to take that limit explicitly. Instead, we can do exactly the same thing we did with the logistic map for discrete time population growth. Since what we do amounts to evaluating the map repeatedly, we can simply use the cobweb diagram, with the starting condition of $z_0 = 0$. If this cobweb diagram converges to a particular z value, that gives us $F^i(0)$, i.e. the probability of extinction of a given lineage.

Equivalently, we can set $F(z) = z$ and solve for z. A popular assumption for the distribution of offspring numbers is the Poisson distribution, where the probability of having n offspring is $\lambda^n e^{-\lambda}/n!$, $\lambda > 0$ being the expected number of offspring. The PGF for a Poisson distribution is $F(z) = e^{\lambda(z-1)}$, and is depicted in the figure on the margin. It can be shown (left as an exercise) that the equation $F(z) = z$ is satisfied for a $0 < z < 1$ if and only if $\lambda > 1$; in other words, that the lineage has positive probability of **not** going extinct only if the expected number of offspring per individual is greater than 1.
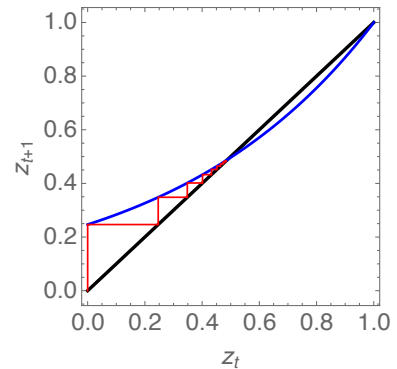
The cobweb diagram for evaluating the asymptotic extinction probability of a lineage. The blue curve depicts the PGF $F(z)$, in this particular, instance, for a Poisson distribution (i.e., $F(z) = e^{\lambda(z-1)}$, where λ is the expected number of offspring). This is the "supercritical" case, where $\lambda > 1$, and there is a unique intersection of $F(z)$ with the 45-degree line (i.e. $F(z) = z$). This z value gives the probability of extinction of the lineage. If the lineage does not go extinct, it will grow to be infinite.

References

Kot, M. 2001. Elements of mathematical ecology. Cambridge University Press.