Topic 2: Discrete-Time/Discrete-State Markov Chains

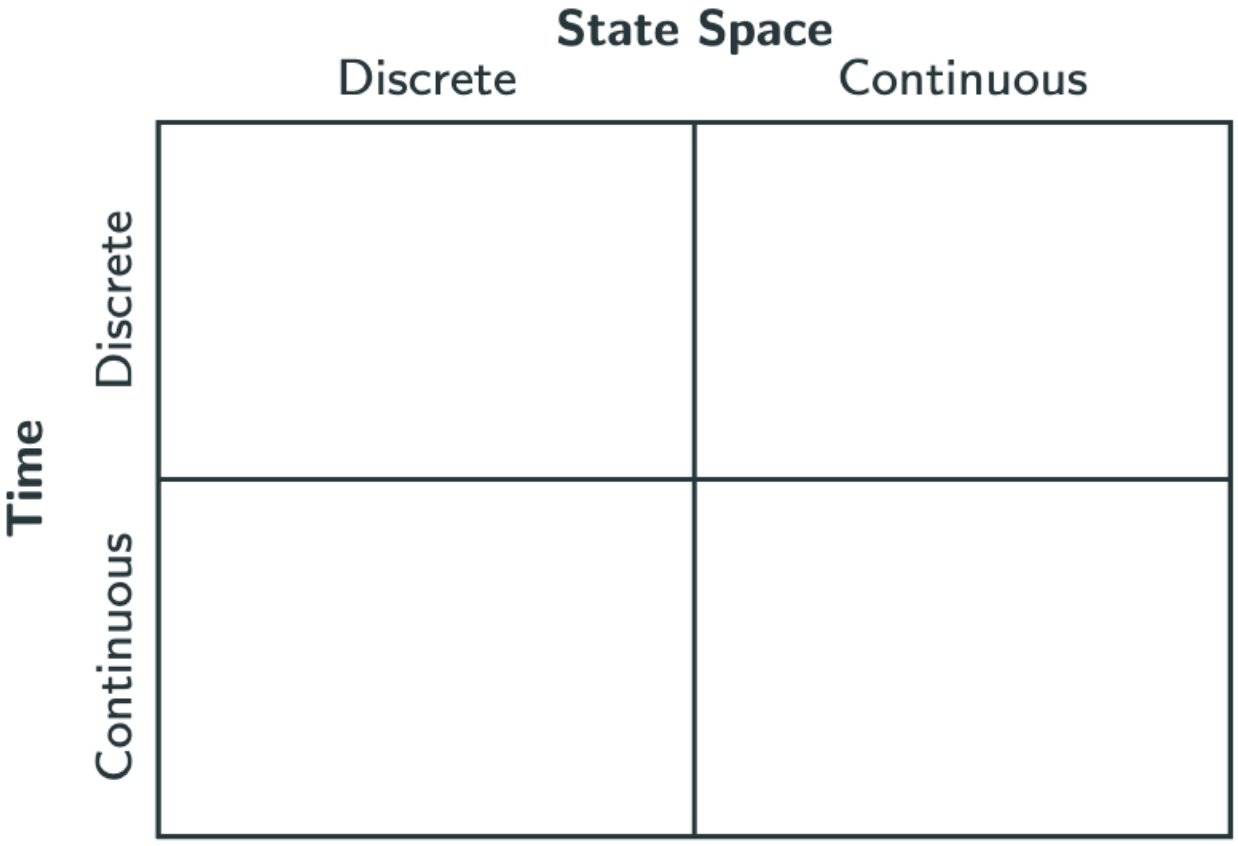
Learning Objectives

- Describe different types of stochastic processes depending on their state and natural measurement of time. Give an example each from biology.
- Define the Markov property in discrete time
- Use a transition probability matrix to describe a discrete-time discrete-space Markov process (DTDS-MC)
    - List the properties of a transition rate matrix.
    - In this course, we will focus almost exclusively on time-homogenous processes, what does this mean and what is an example of a time-dependent process?
- Propose and justify a DTDS-MC model for a biological process.
- Characterize the states of a DTDS-MC as \textbf{transient}, \textbf{absorbing}, or \textbf{recurrent}. Use these mathematical characterizations of states to draw biological conclusions.
- Analyze a DTDS-MC and use these analyses to draw biological conclusions:
    - Use first-step analyses to find the absorption probabilities and time to absorption
    - Derive the stationary distribution
    - Numerical iterate a stochastic process
- Simulate DTDS-MC, calculate their moments through time and use these moments to draw biological conclusions
- Branching processes and their analysis
    - What is an example of a branching process?
    - Analyze branching processes using the probability of extinction
- Describe neutral genetic drift
    - What is the Wright-Fisher model?
    - What is the Moran model?

Lecture 2.1 Intro to DTDS Markov Chains

Characterization of Stochastic Processes

A **stochastic process** is a sequence of random variables X_t, where the index of the sequence has the interpretation of time. The **state space** of the stochastic process is the domain of X and the index t may either be discrete or continuous.



Examples processes

1. Discrete-Time Discrete-Space Process

Branching processes are an example of a discrete-time discrete-state Markov process

---

Example: yeast

Budding or Brewer's Yeast *Saccharomyces cerevisiae* is an incredibly useful model organism for experimental evolution. As implied by its name it is used to make beer and wine and divides by budding. Suppose that an evolution experiment begins with a single yeast cell, each time step a yeast cell may either bud with probability b or die with probability $d = 1 - b$. The number of yeast at time step $t \in \mathbb{Z}^+$ is described by a discrete-time branching process.

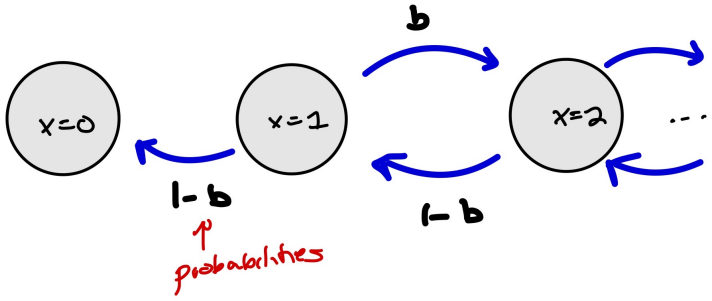
What is the state space here?

number of yeast cells, N, an integer

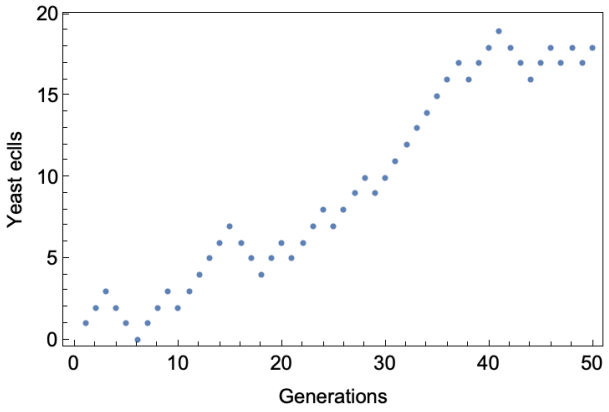
What is the time index here?

generations, t, an integer

Draw a transition diagram of this process



Draw an example Trajectory.



We call a single realization of a stochastic process a "trajectory"

---

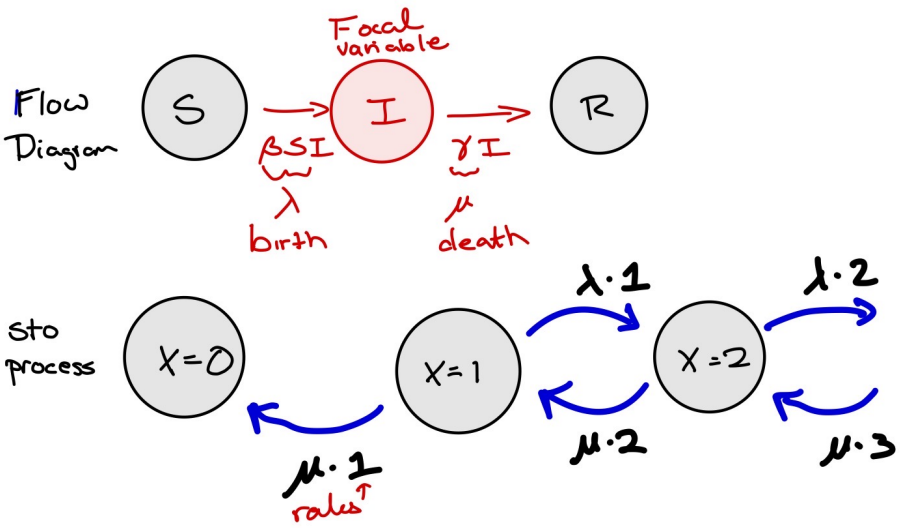
2. Continuous Time Discrete Space Process

Birth-death processes are an example of continuous Time Discrete Space Process. These are mathematical models used to describe 'population' dynamics, where each individual gives 'birth' (creating a new 'individual') at a rate λ and dies at a rate d over time. These processes find application in various biological contexts.

Example: emergent infectious diseases

The initial spread of infections can be described by a birth-death process where an "individual" is a case, "birth" is transmission and "death" is recovery. The state space here is the number of infections. The state space is $X_t \in \{0, 1, \ldots \infty\}$ where $t \in [0, \infty)$.

Draw a determistic flow diagram and transition diagram of this process



Discussion: What are some other examples of birth-death processes? What is the individual and what is the "state space'' of each? What do birth and death mean here?

2. Discrete-Time Continuous Space Process

We often use 'discrete-time' because of how data is collected. This is the case for **First Order Auto-regressive Processes**.

Example: wastewater

Detection of virus in wastewater has been widely used to monitor the prevalence of COVID-19, especially in the absence of any other good case data. Rainfall and urban wastewater in Vancouver are not separated such that the amount of wastewater entering a system changes day to day.

The amount of wastewater on day $d = \{1, 2...365\}$ can be modelled using a autoregressive process:

$$X_d = f\left(X_{(d-1)}\right) + \epsilon_d$$

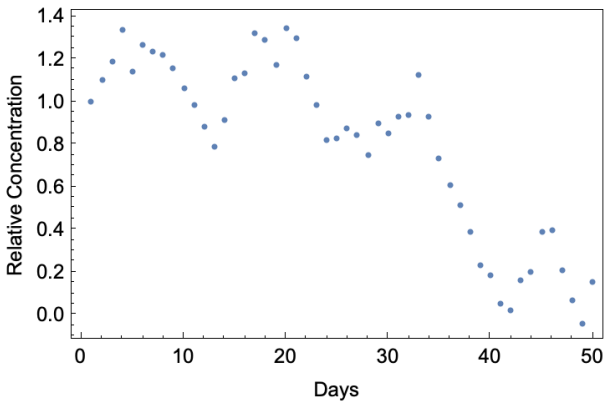
where $f(X)$ is some function, often assumed to be linear or polynomial, and $\epsilon_d \sim N(0, \sigma)$ adds noise.

1. **What is the state/time-space here?**

State: the relative concentration of focal virus versus the control (pepper mosaic virus).

Time: days or weeks

2. **Draw an example trajectory.**



3. Continuous Time Continuous Space Process

The Brownian motion model of trait evolution, also known as the Brownian motion model or the continuous random walk model, is widely used in evolutionary biology to describe how a continuous trait evolves. This model is named after the concept of Brownian motion in physics, where particles undergo random, continuous motion. In the context of trait evolution, the Brownian motion model assumes the following:

- The trait of interest is continuous and can take on a range of values, such as body size, beak length, or metabolic rate. It is often assumed that the trait follows a normal distribution.
- Trait evolution is driven by random neutral evolution. Hence this is a useful null model.
- The change in the trait value at one point in time does not affect or predict the change at another point in time. In models with multiple traits changes in one trait value are independent of the others.
- The rate of trait evolution is constant over time. This means that the variance in trait values increases linearly with time.
- The trait $Z(t)$ can be describe by:

$$Z(t + \Delta t) = Z(t) + N(0, \sigma \Delta t)$$

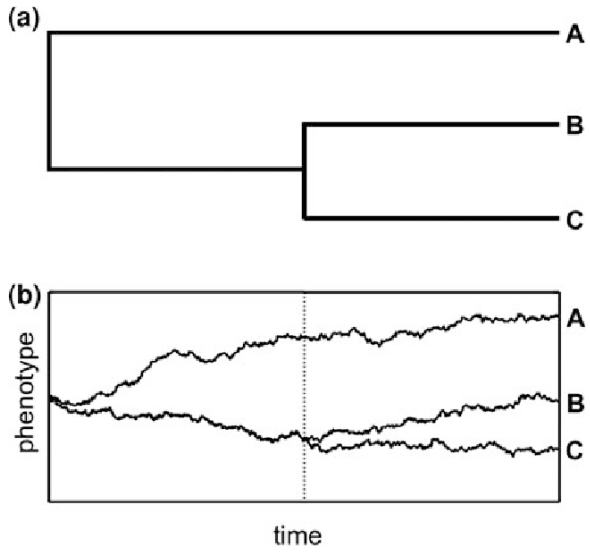


Figure from Symonds and Blomberg 2014

Markov Property

In this course, we are going to focus exclusively on **Markov** processes. The Markov property states that the future of the process is determined solely by its present state and is not influenced by how it arrived at that state from previous states. For the discrete-time Markov process X_n (here X is the state and n is the time step index) we have:

$$P\left(X_{n+1} = x | X_n = x_n, X_{(n-1)} = x_{(n-1)}, ..., X_0 = x_0\right) = P\left(X_{(n+1)} = x | X_n = x_n\right)$$

In other words, Markov processes don't have any "momentum".

We define the **one-step transition probability** as:

$$p_{ij}(n) = \Pr(X_{n+1} = j | X_n = i)$$

In addition to the Markov property, we will also mostly consider **time-homogenous** processes where this one-step transition probability, p_{ij} is independent of time, n.

Discussion: What is an example of a stochastic process that is naturally Markovian? How about a process that is non-Markovian?

Most things in the real world are likely non-markovian, but they can often be well approximated by Markovian processes (with the transition probabilities adjusted)

Drawing DTDS-MC graphically

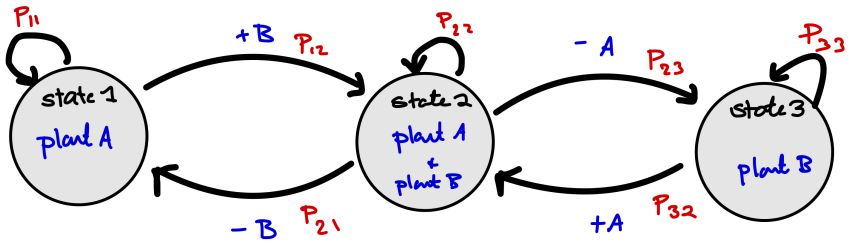
We can often draw a DTDS-MC as a graph. We represent each state with a circle and the probabilities p_{ij} with arrows.

Example: 2.1 Species coexistence

In ecology, one common application of Markov chains is in modelling the (co)occurrence of species in a community. Let's consider a simple example of a Markov chain modelling the coexistence of two hypothetical species Plant A and Plant B in an ecological reserve.

State 1: Plant A only
State 2: Plant A and B
State 3: Plant B only

1. Draw this stochastic process assuming only one or zero species can be lost/gained in a given year.



The Transition Probability Matrix

A DTDS-MC can be represented by its **transition probability matrix** giving $p_{ij}\ \forall i\ \&j$. Suppose $X_t \in \{0, 2, ...N\}$, in this class we will define the transition probability matrix as a square matrix with $N+1$ (because of 0) rows and $N+1$ columns that give the probability of going from row/state i to column/state j. We could switch how we label rows and columns but this choice will be convenient for a few reasons.

We then represent the state of the system as a row vector, $V(t)$, giving the probability the system is in state i at time t.

$$V(t) = \begin{bmatrix} \Pr(X_t = 0) & \Pr(X_t = 2) & \dots & \Pr(X_t = N) \end{bmatrix}$$

Example 2.2: Species coexistence cont

Let's assume the following transition rate matrix

$$P = \begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.4 & 0.5 & 0.1 \\ 0 & 0.3 & 0.7 \end{bmatrix}$$

1. What is the probability that the system goes from having both species (state 2) to only species A (state 1)?

$$p_{2,1} = 0.4$$

2. What is the initial state vector assuming 30% of the ecosystems (e.g., marshes) start have only species A and 40% only have species B?

$$V(0) = \begin{bmatrix} 0.3 & 0.3 & 0.4 \end{bmatrix}$$

The transition probability matrix has several useful properties:

1. All elements of $\mathbf{P}$ are non-negative
2. Row Sums: $\sum_j P_{ij} = 1$ (you have to go somewhere)

Discussion: Is this true for the example given above?

3. Column Sums: $\sum_i P_{ij}$: probability of transitioning to state i from all possible starting states j}
4. $\mathbf{P}^k$ gives the k^{th} step transition probabilities from X_n to X_{n+k}.
5. **Chapman-Kolmogorov Eq.:**

$$p_{ij}^{n+m} = \sum_k p_{ik}^n p_{kj}^m$$

Example 2.3: Species coexistence cont

Python: Lecture2_1.ipynb

1. What is the probability that an ecosystem where both species currently coexist has only one species 2 years later, how about 3 years later?

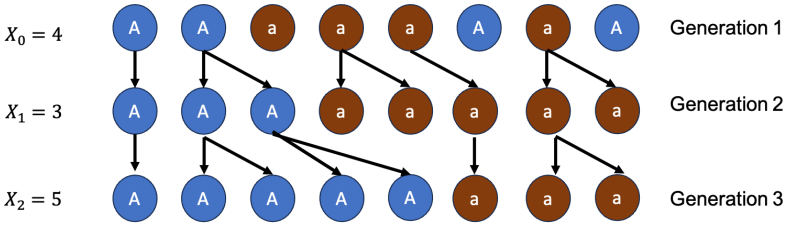
2. Use the Chapman-Kolmogorov equation to calculate the probability of being in state 1 (species A only) in 5 years.

Lecture 2.2 The Wright-Fisher Model

The Wright-Fisher Model

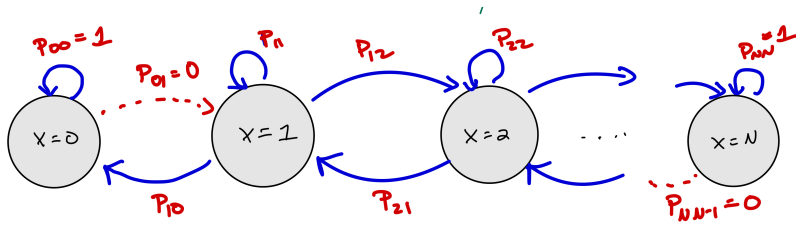
The Wright-Fisher model is a model of **neutral genetic drift**. It models a population with N gene copies as inherited across **discrete generations**. Genes have one of two variant alleles, A or a. The stochastic process then follows the number of A alleles in the population, N_A. Note that $N_a = N - N_A$

We often draw the WF model graphically in the following way.



Each offspring picks a parent at random. This ensures that parents can have several offspring (genes) but each offspring gene can only have one parent.

We can represent the resulting process with the following transition diagram:



The transition probabilities of the WF model are given by the Binomial Distribution with success probability $p = \frac{i}{N}$:

$$\begin{aligned} P(X_{t+1} = j | X_t = i) = p_{ij} &= \binom{N}{j} \frac{i}{N}^j \left(\frac{N-i}{N}\right)^{N-j} \\ &= \underbrace{\binom{N}{j} p^j (1-p)^{N-j}}_{\mathcal{B}_j(N,p)} \end{aligned}$$

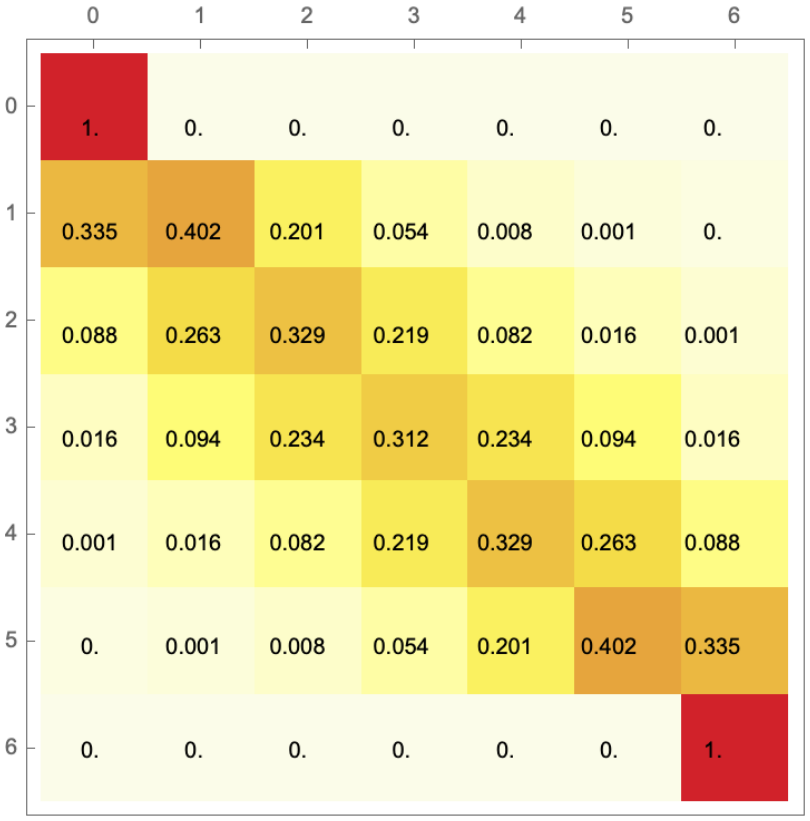
The value p here is known as the **allele frequency** or in other words, the frequency of the 'A' allele in the population.

Note: a note on notation, in many cases the value of N in the above model is replaced with $2N$ to reflect that there are two copies of each gene in a *diploid* individual. Hence using this alternative notation N is the number of individuals and $2N$ is the number of genes. The equations given above are for the *haploid* model.

Example 2.4: Wright-Fisher Model

1. Consider a population with $N = 6$ haploid individuals, what does the transition probability matrix look like?

Python: Lecture2_2.ipynb



2. Suppose that the population starts with an allele frequency of $p = 0.5$, what is the probability that the frequency is $p = 0.5$ in the next generation?

$$p_{3,3} = 0.312$$

Note the other two most likely states to be in are $j = 2, p = 0.33$ and $j = 4, p = 0.66$.

$$p_{3,2} = p_{3,4} = 0.219$$

3. Suppose that you start with 0 copies of the 'A' allele, what is the probability that you end up in each of the other states in the next generation?

$$\Pr(X_1 = x | X_0 = 0) = \begin{cases} 1 & x = 0 \\ 0 & \text{otherwise} \end{cases}$$

Discussion: why does this make sense?

Characterization of States

In a discrete-state discrete-time stochastic process, various characterizations can be assigned to individual states based on their behaviour and properties within the process.

Absorbing State: An absorbing state is a state, a, from which the process cannot leave once it enters. Once the process reaches an absorbing state, it remains in that state indefinitely with probability 1. Absorbing states are often used to represent terminal or

absorbing outcomes in a process.

A state is absorbing if:

$$P_{a,a} = 1$$

Transient State: A transient state is one that, once entered, has a non-zero probability of eventually transitioning to an absorbing state or reaching a terminal state. Transient states do not have a long-term impact on the process, and the process is expected to leave them eventually.

A state is transient if:

$$P_{i,a}^{k} > 0$$

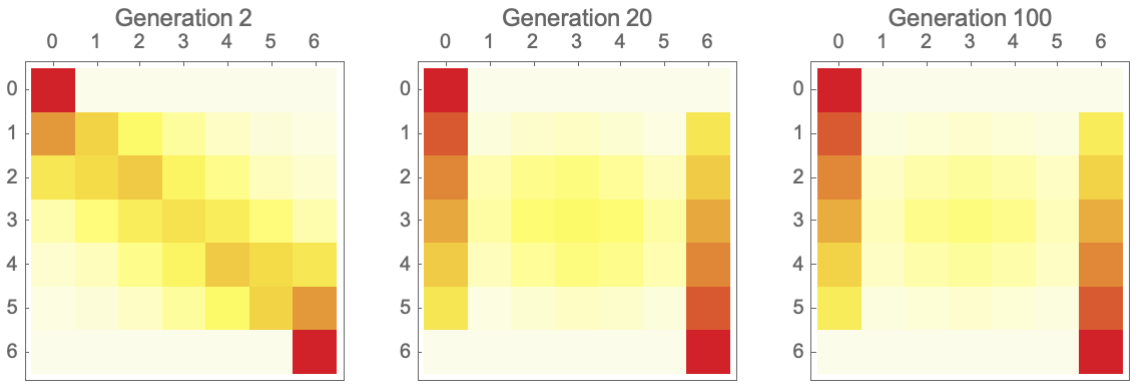
Recurrent State: A recurrent state is one where, if the process enters it, it will return to that state with probability 1. A recurrent state is revisited infinitely often.

Example 2.5: Wright-Fisher Model

1. In a haploid WF with N gene copies, characterize the states of the WF model as absorbing, transient, or recursive.

Absorbing $X = 0$ & $X = N$

Transient $0 < X < N$, to see this let's look at a P^k, which tells us where the system will likely be after k generations.



There are no recurrent states

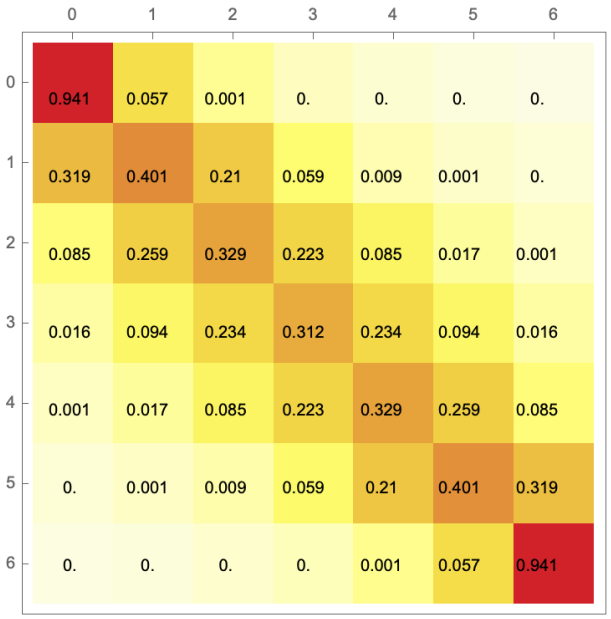
The Wright-Fisher Model with Mutation

We can add the possibility of mutation to this model by assuming that 'A' parents give rise to 'a' offspring with probability u and 'a' parents give rise to 'A' parents with probability v. The new transition probabilities are given by:

$$\begin{aligned} P(X_{t+1} = j | X_t = i) &= \binom{N}{j} \left(\frac{i}{N}(1-u) + \frac{N-i}{N} v \right)^{j} \left(\frac{N-i}{N}(1-v) + \frac{i}{N} u \right)^{N-j} \\ &= \underbrace{\binom{N}{j} \left(p(1-u) + (1-p)v \right)^{j} \left((1-p)(1-v) + pu \right)^{N-j}}_{\mathcal{B}_j(N, p(1-u)+(1+p)v)} \end{aligned}$$

Example 2.6: Wright-Fisher Model with Mutation
Consider a population with $N = 6$ haploid individuals with mutation probabilities $u = v = 0.05$

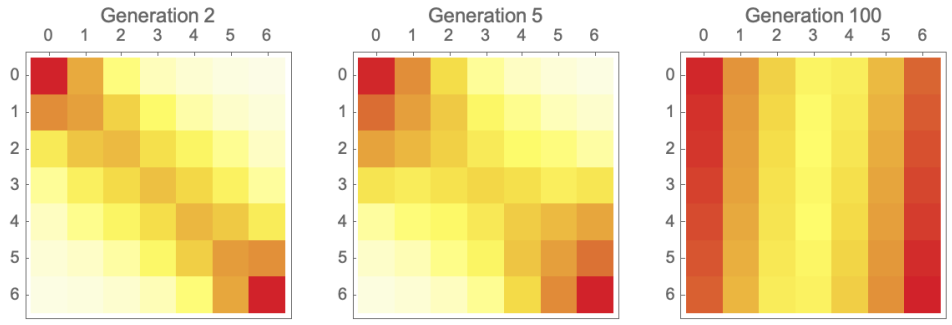
1. What does the transition probability matrix look like?
Python: Lecture2_2.ipynb



Discussion: why does this compare to the case without mutation?

2. Characterize the states of this stochastic model.

All the states are recurrent.



Numerical Iteration

Given a **initial** state (row) vector of probabilities $\vec{X}_0$ we can use the transition probability matrix to iterate the probability through time:

$$X_{n+1} = X_n \mathbf{P} \quad \text{given: } X_0$$

Example 2.7: Wright-Fisher Model

Consider a population with $N = 6$ haploid individuals and no mutation that starts with an allele frequency of $p(0) = 0.5$

1. **What is the initial state vector?**

We start with 3 copies of the 'A' allele.

$$\vec{X}_0 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

2. **What is the state vector in the next generation and the generation after?**

$$\vec{X}_1 = \vec{X}_0 . \mathbf{P} = \begin{bmatrix} 0.02 & 0.09 & 0.23 & 0.31 & 0.23 & 0.09 & 0.02 \end{bmatrix}$$

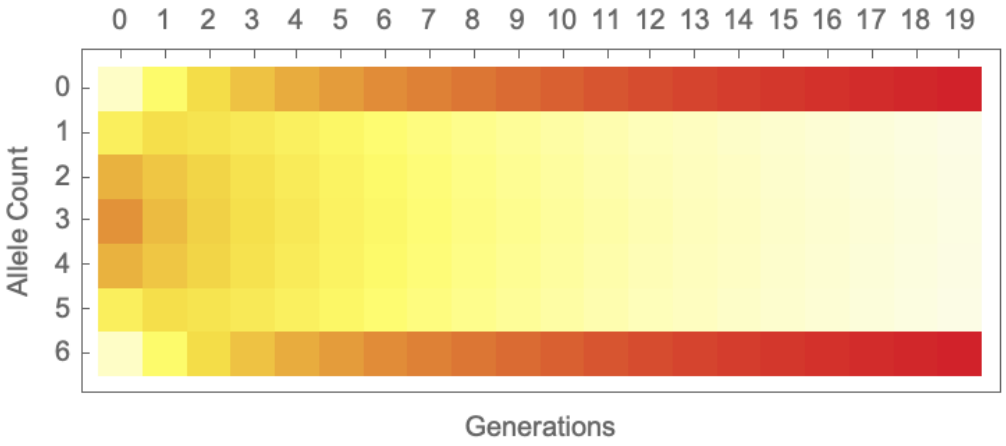
$$\vec{X}_2 = \vec{X}_1 . \mathbf{P} = \begin{bmatrix} 0.07 & 0.13 & 0.19 & 0.21 & 0.19 & 0.13 & 0.07 \end{bmatrix}$$

3. **What is the state vector after 20 generations?**

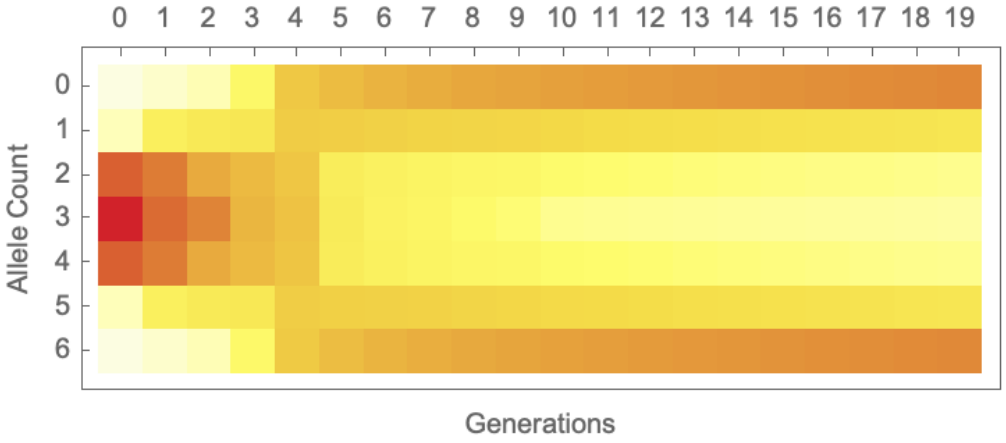
$$\vec{X}_{20} = \vec{X}_0 . \mathbf{P}^{20} = \begin{bmatrix} 0.48 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.48 \end{bmatrix}$$

4. **Plot a heat map showing the dynamics over time.**

Python: Lecture2_2.ipynb



4. **Plot a heat map showing the dynamics of the WF model with mutation over time.**



Lecture 2.3 First-Step analysis and the Stationary Distributions

First-step analysis

We can use the probability of going from i to j in a single step to learn about the probability of absorption and the time to absorption. This approach is called **first-step analysis**.

Let $V_y = \min\{n \geq 0 : X_n = y\}$ be the **first hitting time** of state y.

Consider two different states $a_1 \neq a_2$

Let $h(x) = \Pr_x(V_{a_1} \leq V_{a_2} | X_0 = x)$ be the probability you hit a_1 before you hit a_2 given you start in state x.

We have that:

$$h(i) = \sum_j p_{ij} h(j) \quad \forall i$$
$$h(a_1) = 1 \quad h(a_2) = 0$$

providing a system for equations we can sometimes solve for $h(i)$.

Example 2.8: Wright-Fisher Model

In the (haploid) WF model without mutation, we have two absorbing states $x = 0$ and $x = N = 6$. What is the probability that the A allele fixes given that we start with an allele frequency of 0.5?

Python: Lecture2_2.ipynb

Let $a_1 = \{X = N\}$ and $a_2 = \{X = 0\}$ and hence $h(i)$ is the probability the system fixes given that we start with i copies of the 'A' allele.

We have a system of equations:

$$h(i) = \begin{cases} h(0) = 0 \\ h(i) = \sum_{j=0}^{N} \binom{N}{j} \left(\frac{i}{N}\right)^j \left(\frac{N-i}{N}\right)^{N-j} h(j) \quad \forall i = \{1, \ldots, N-1\} \\ h(N) = 1 \end{cases}$$

Solving we have:

$$h(0) = 0 \quad h(1) = \frac{1}{6} \quad h(2) = \frac{1}{3}$$
$$h(3) = \frac{1}{2} \quad h(4) = \frac{2}{3} \quad h(5) = \frac{5}{6} \quad h(6) = 1$$

Discussion: Does this make sense? What do you think the probability of loss (i.e. $A = \{X = 0\}$ and $B = \{X = N\}$) looks like?

We can use first-step analysis to learn about the time to absorption as well. Let $\mathcal{A}$ by the set of absorbing states of a stochastic process. Let $\tau = \min\{n \ni X_n \notin \mathcal{A}\}$.

Let $g(i) = E(\tau | X_0 = i)$

$$g(i) = \begin{cases} 0 & \forall\, i \in \mathcal{A} \\ 1 + \sum_j p_{ij} g(j) & \forall\, i \notin \mathcal{A} \end{cases}$$

Example 2.9: Wright-Fisher Model

What is the time to fixation/loss in the WF model with population size N given you start with an allele frequency or $p(0) = 0.5$?

Python: Lecture2_2.ipynb

$$g(i) = \begin{cases} 0 & i = 0 \& i = 6 \\ 1 + \sum_j \binom{N}{j} \left(\frac{i}{N}\right)^j \left(\frac{N-i}{N}\right)^{N-j} g(j) & \end{cases}$$

Solving we have:

$$g(0) = 0 \quad g(1) = 4.62 \quad g(2) = 6.58 \quad g(3) = 7.21$$
$$g(4) = 6.58 \quad g(5) = 4.62 \quad g(6) = 0$$

Stationary Distribution

Definition

The **stationary distribution** of a discrete-time, discrete-space stochastic process is a discrete probability distribution that remains unchanged as the process proceeds over time. In other words, it represents the long-term behaviour of the process, and once reached, it remains constant with each successive time step. The stationary distribution characterizes the probability of the process being in each possible state, regardless of the initial state. It is analogous to the **equilibrium** of a deterministic system.

Formally, let's consider a discrete-time, discrete-space stochastic process with a finite set of states $X_n \in \{1, 2, \ldots, n\}$. The stationary distribution for this process can be represented by the vector: $\vec{\pi} = [\pi_1, \pi_2, \ldots \pi_n]$ which must satisfy

$$\vec{\pi} = \vec{\pi}.\mathbf{P}$$

π_i represents the probability of the process being in state i in the stationary distribution hence we also have that

$$\sum_i \pi_i = 1$$

In simpler terms, π_i tells us the long-term proportion of time that the process is expected to spend in state i as it evolves.

Finding the Stationary Distribution

To find the stationary distribution of a discrete-time discrete-space stochastic process we solve the system:

$$\pi_i = \sum_j P_{ij}\pi_j \quad \forall\, i > 0$$
$$\pi_0 = 1 - \sum_{j>0} \pi_j$$

Once we obtain the solution, always want to check that $\vec{\pi} = \vec{\pi}.\mathbf{P}$

Example 2.10: Ecological state

Recall the model of species coexistence:

State 1: Plant A only
State 2: Plant A and B
State 3: Plant B only

With the transition prob. matrix

$$P = \begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.4 & 0.5 & 0.1 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}$$

1. What is the stationary distribution of this system?

$$\begin{aligned} \pi_2 &= 0.7\pi_1 + 0.4\pi_2 + 0.1\pi_3 \\ \pi_3 &= 0.3\pi_1 + 0.5\pi_2 + 0.1\pi_3 \\ \pi_1 &= 1 - \pi_2 - \pi_3 \end{aligned}$$

Solving we have: $\vec{\pi} = \begin{bmatrix} 0.52 & 0.36 & 0.12 \end{bmatrix}$

As an alternative to the balance equations, we can use eigenanalysis to find the stationary distribution. We want the long-term outcome of the stochastic process, so:

$$X(t) = X(0).\mathbf{P}^t$$

for very large t. Note that we can diagonalize the matrix $\mathbf{P} = \mathbf{A}.\mathbf{D}.\mathbf{A}^{-1}$. Where $\mathbf{A}$ is the matrix of eigenvectors (as columns), $\mathbf{D}$ is a diagonal matrix of eigenvalues, and $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$.

$$X(t) = X(0).\underbrace{\mathbf{A}.\mathbf{D}.\cancel{\mathbf{A}^{-1}}.\cancel{\mathbf{A}}.\mathbf{D}.\mathbf{A}^{-1}\ldots.}_{t\times} = \mathbf{A}.\mathbf{D}^t.\mathbf{A}^{-1}$$

Fortunately, it's easy to raise a diagonal matrix to a power:

$$\mathbf{D}^t = \begin{bmatrix} \lambda_1^t & 0 & \ldots & 0 \\ 0 & \lambda_2^t & 0 & \ldots \\ & & \ddots & \\ 0 & 0 & \ldots & \lambda_N^t \end{bmatrix}$$

We can order the eigenvalues in any way we want to so let $\lambda_1 \geq \lambda_2$ etc. Then for large t we have:

$$\lim_{t\to\infty} \mathbf{D}^t = \begin{bmatrix} \lambda_1^t & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots \\ & & \ddots & \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

Letting $\vec{U}_1$ be the leading (right) eigenvector and $\vec{V}_1$ be the leading (left) eigenvector we have.

$$X(t) = \underbrace{\begin{bmatrix} X(0) \end{bmatrix} \cdot \begin{bmatrix} \vec{U}_1 \end{bmatrix} \lambda_1^t}_{\text{constant}} \cdot \underbrace{\begin{bmatrix} \vec{V}_1 \end{bmatrix}}_{\text{row vec}}$$

So V_1 is the stationary distribution when normallized so that $\sum \vec{V}_1 = 1$

Example 2.11: Wright-Fisher Model with mutation

Python: Lecture2_3.ipynb

1. Find the stationary distribution of the WF model with mutation given $u = v = 0.05$.

We will use the eigenvector method to obtain:

$$\pi = \begin{bmatrix} 0.181 & 0.14 & 0.121 & 0.117 & 0.121 & 0.14 & 0.181 \end{bmatrix}$$

2. Repeat this for $u = v = 0.005$.

$$\pi = \begin{bmatrix} 0.44 & 0.03 & 0.02 & 0.02 & 0.02 & 0.03 & 0.44 \end{bmatrix}$$

Discussion: Does this make sense? What do you think would happen if $u \neq v$?

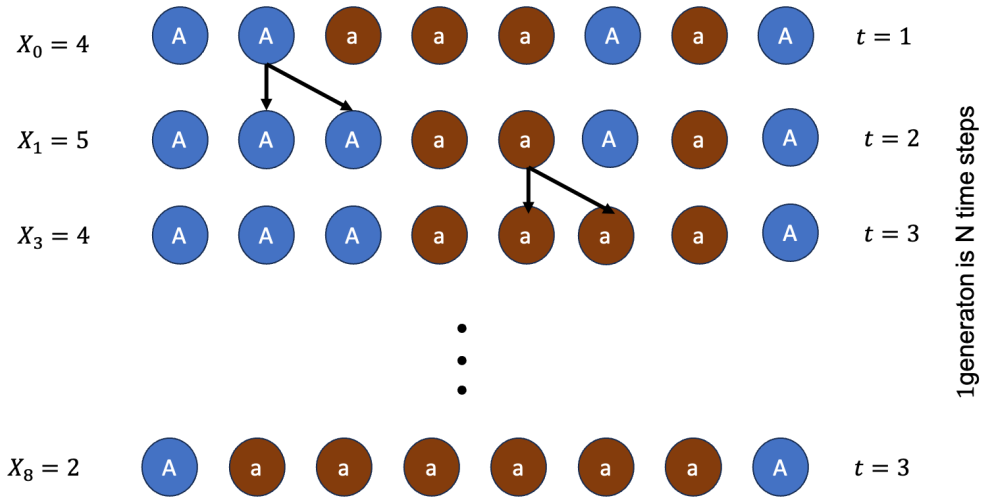
Lecture 2.4 Simulating a Stochastic Process

Python: Lecture2_4.ipynb

Lecture 2.5 The Moran Process

Lecture 2.2 introduced the Wright-Fisher Model of genetic drift. Recall that in this model the population is characterized by **discrete generations**, many populations however do not display synchrony in their reproduction and hence must be modelled with **overlapping generations**. The **Moran Model** is the natural model for genetic drift in such populations.

Like the WF model, we begin by assuming that there are N haploid individuals/gene copies in the population. Each time step one random individual reproduces and another random individual (can be the same as the one who gave birth) dies. Like the WF model, we can show this process with a diagram.



To build an analogy to the WF model, we define **1 generation** as the time it takes for N individuals to be replaced or in other words N time steps.

The transition rates in this model are much simpler than in the WF model. If there are currently i 'A' alleles the probability that there are j 'a' alleles in the next time step is:

$$p_{i,j} = \begin{cases} \underbrace{\frac{i}{N}}_{\text{'A' born}} * \underbrace{\left(1 - \frac{i}{N}\right)}_{\text{'a' dies}} & \text{if } j = i+1 \\ 1 - 2 * \frac{i}{N} * \left(1 - \frac{i}{N}\right) = \left(\frac{i}{N}\right)^2 + \left(1 - \frac{i}{N}\right)^2 & \text{if } j = i \\ \underbrace{\left(1 - \frac{i}{N}\right)}_{\text{'a' born}} * \underbrace{\frac{i}{N}}_{\text{'A' dies}} & \text{if } j = i-1 \end{cases}$$

These can be derived by considering who is born and who dies.

Example 2.12: Moran Model
Python: Lecture2_5.ipynb

Consider a population with $N = 6$ with an initial frequency of $p = 0.5$

1. What does the transition matrix look like?

2. Characterize the states of this process.

Discussion: How does this compare to the WF model?

3. What is the probability that the population will have a frequency of $p = 0.5$ in the next time step? How about in the next generation? How does this result compare to the WF model?

4. Simulate 100 trajectories in both the WF and Moran models. Compare them. For each trajectory, calculate the dynamics of *expected heterozygosity* $H(t) = 2p(t)(1 - p(t))$, how do these dynamics compare?

As $N \to \infty$ 1 generation in the Moran model is many many tiny time steps and hence this model is effectively in **continuous time** even though it is technically a **discrete-time stochastic process**.

Just like in the WF model, we can introduce mutation to this model. Suppose that once again focal 'A' individuals mutate to be 'a' with probability u and that 'a' mutate to 'A' with probability v.

Discussion: What do you think the transition probabilities are in this model?

$$p_{i,j} = \begin{cases} \underbrace{\left[\left(1-\frac{i}{N}\right)v + (1-u)\frac{i}{N}\right]}_{\text{'A' born}} * \underbrace{\left(1-\frac{i}{N}\right)}_{\text{'a' dies}} & \text{if } j = i+1 \\ 1 - p_{i,i-1} - p_{i,i+1} & \text{if } j = i \\ \underbrace{\left[\left(1-\frac{i}{N}\right)(1-v) + u\frac{i}{N}\right]}_{\text{'a' born}} * \underbrace{\frac{i}{N}}_{\text{'A' dies}} & \text{if } j = i-1 \end{cases}$$

Example 2.13: Moran Model With Mutation

Python: Lecture2_5.ipynb

1. Consider the same population with $N = 6$, and calculate the stationary distribution of the Moran process.

Discussion: How does this compare to the WF model?

2. Calculate the time until fixation/loss of an allele in this model. How does it compare to the WF model?

Lecture 2.6 Branching Processes

Definition

The "Galton-Watson" **branching process**, considers the number of individuals present in generation n is, X_n:

1. Each individual in generation n gives birth to $Y \in \mathbb{Z}^+$. In other words, the **offspring distribution** in generation n is given by:

$$\Pr(Y = k) = p_k$$

2. Each individual gives birth independently and iid number of offspring (independent of all other individuals).
3. The same offspring distribution applies to all n generations (i.e. there is no density dependence)!

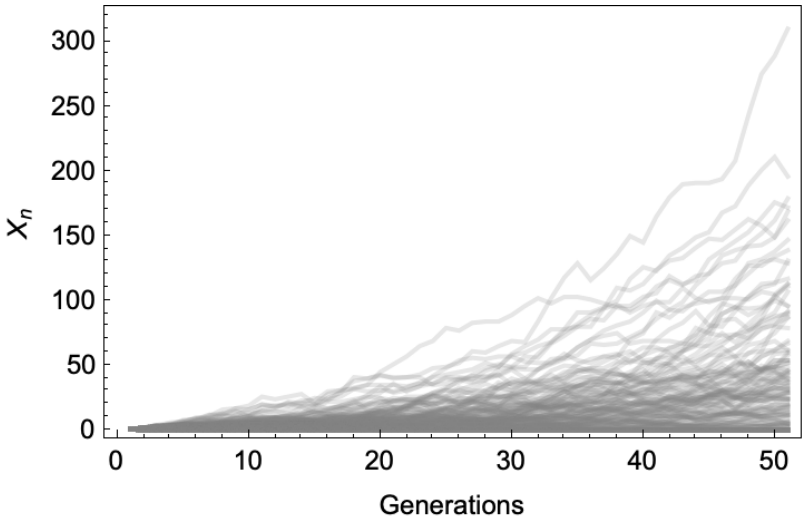
Example 2.14: Birth/Death

Suppose that each generation an individual can give birth with probability b giving rise to two individuals (itself and its child), or die with probability d. With probability $1 - b - d$ neither occurs.

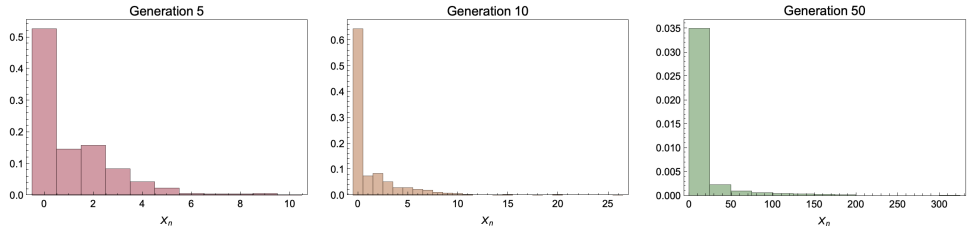
1. What is the offspring distribution?

$$Y_k = \begin{cases} b & k = 2 \\ 1 - b - d & k = 1 \\ d & k = 0 \end{cases}$$

2. Simulate 500 Sample trajectories of this process with $b = 0.25$ and $d = 0.2$ given that $X_0 = 1$



3. What is the distribution of X_5, X_{10}, and X_{50}?



4. Characterize the states of the branching process.

This process like all branching processes has one absorbing state $X = 0$ and all other states are transient.

As we can see in a branching process, many trajectories go extinct and a handful that do not. How do we calculate this probability of extinction? Given that there is only one absorbing state the probability that the process goes extinct is after all 100%. But, if enough births happen it may take a very very long time for this to occur. What then is the probability that the branching process goes extinct in **finite time**?

Once again, the first step analysis comes to the rescue.

Define g as the probability that a single individual is alive in generation n and all of its subsequent descendants eventually go extinct. By noting that the offspring distribution of each individual is independent we have:

$$g = \sum_{j=0}^{k_{max}} p_k g^k = p_0 + \sum_{j=1}^{k_{max}} p_k g^k$$

This polynomial will have k roots. We can show that one of these roots will be $\hat{g} = 1$.

$$1 = \sum_{j=0}^{k_{max}} p_k 1^k = \sum_{j=0}^{k_{max}} p_k$$

Noting that p_k is a probability distribution we have $1 = 1$.

One of the roots may also be real and $0 \leq \hat{g} \leq 1$, in which case this gives the probability that the process goes extinct in finite time.

Example 2.15: Birth/Death Continued

1. Calculate the probability of extinction in this branching process

$$g = d + (1 - b - d)g + bg^2$$

Solving we have:

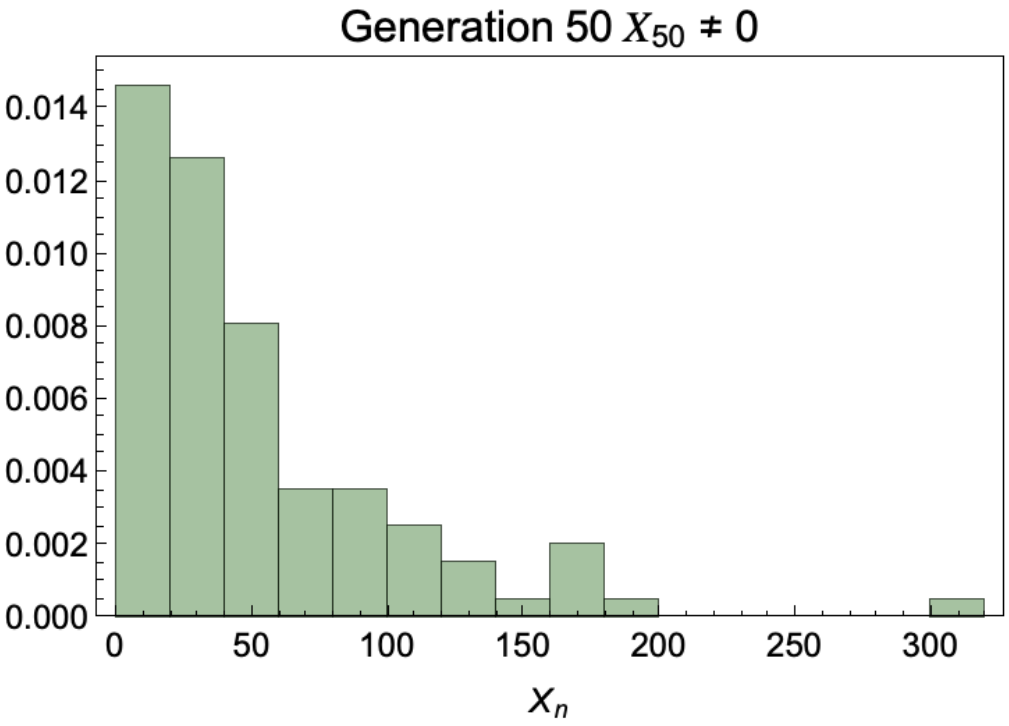
$$\hat{g} = 1 \quad \& \quad \hat{g} = \frac{d}{b}$$

This tells us that if the probability of death is less than birth $b > d$ there is a non-zero chance that the population will not go extinct $(1 - \hat{g})$ in finite time.

2. How does this analytical result compare to your simulations?

In the simulations shown above 401 of 500 simulations have gone extinct by generation 50 or 80.2.

The analytical prediction is $\hat{g} = \frac{0.2}{0.25} = 0.8 = 80$. Note that the distribution of lineage that have not gone extinct looks like this:



So the number of individuals present is usually large and hence none of these are likely to go extinct anytime soon.

While framed in a discrete-time context the concept of a branching process can be used to conclude events occurring continuously through time by changing the meaning of the time index n.

Specifically, let's redefine n as the n^{th} event.

Example 2.16: Birth-death process

Suppose that each individual gives birth at *rate* λ and dies at *rate* μ.

1. If there are currently n individuals what is the probability that a birth event occurs before a death event?

We will derive this more formally in the next lecture but:

$$\Pr(\text{birth}) = b = \frac{\lambda * n}{\lambda * n + \mu * n} = \frac{\lambda}{\lambda + \mu}$$

2. What is the probability that a death event occurs before a birth event?

$$\Pr(\text{death}) = d = \frac{\mu * n}{\lambda * n + \mu * n} = \frac{\mu}{\lambda + \mu} = 1 - b$$

3. Define a branching process describing the number of individuals in the population.

Let X_n be the number of individuals present in the population after the n^{th} event. Then the offspring distribution:

$$\Pr(Y=k) = \begin{cases} b = \frac{\lambda}{\lambda+\mu} & k=2 \\ 1-b & k=0 \\ 0 & \text{otherwise} \end{cases}$$

Note that, unlike the stochastic process above we do not have a $k=1$ option because we are counting events.

4. What is the probability of extinction in this model?

Again let g be the probability that a single individual and its descendants go extinct. Then:

$$g = bg^2 + (1-b)$$

Solving we have $\hat{g} = 1$ and $\hat{g} = \frac{1-b}{b}$. Since b is always > 0 (assuming $\lambda > 0$) then we always have a non-zero probability that the process does *not* go extinct.

To understand how this procedure can help us conclude the biological world. Let's consider the SIR epidemiological model represented by the set of differential equations:

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta}{\kappa} S(t) I(t) \\ \frac{dI}{dt} &= \frac{\beta}{\kappa} S(t) I(t) - \gamma I(t) \\ \frac{dR}{dt} &= \gamma I(t) \end{aligned}$$

Here $\kappa = S(t) + I(t)$ is the the total population size.

We can derive the **Basic Reproductive Ratio** in this model by considering when $\frac{dI}{dt} > 0$.

$$\frac{\beta}{\kappa} S(t) \cancel{I(t)} > \gamma \cancel{I(t)} \qquad \frac{\frac{\beta}{\kappa} S(t)}{\gamma} > 1 \tag{1}$$

The basic reproductive ratio apply to a **emergent** infectious disease where $S(t) \approx \kappa$

$$R_0 = \frac{\beta}{\gamma} > 1$$

This tells us that in the **deterministic** model, the disease will go extinct if $\beta < \gamma$ and will spread when $\beta > \gamma$. Of course, the world is not deterministic, particularly when the infection is rare.

Example 2.17: R_0

**Consider the classic SIR as defined above where there is a single initial infection $I(0) = 1$. **

1. Propose a branching process that describes the early spread of the disease.

When the number of infections is small we can *approximate* the dynamics of the infection as a birth-death process where "births" (new infections) occur at rate $\frac{\beta}{\kappa}\kappa = \beta$ and "deaths" (recovery) occurs at rate. Using the same logic as above, let the number of infections after the n^{th} event be X_n and the offspring distribution:

$$\Pr(Y=k) = \begin{cases} b = \frac{\beta}{\beta+\gamma} & k=2 \\ 1-b & k=0 \\ 0 & \text{otherwise} \end{cases}$$

2. What is the probability that the emergent disease will go extinct?

From above we have that the probability of extinction is

$$\hat{g} = \frac{1-b}{b} = \frac{\frac{\gamma}{\beta+\gamma}}{\frac{\beta}{\beta+\gamma}} = \frac{\gamma}{\beta} = \frac{1}{R_0}$$

So the probability of extinction is inversely proportional to the R_0.

3. What is the probability that the emergent disease with $R_0 = 3$ (similar to COVID-19 in early 2020) goes extinct?

$\hat{g} = 0.33$ or a 33% chance of extinction or a 66% chance of emergence.

The basic SIR model above makes a critical assumption though that all individuals are equally likely to transmit the infectious disease. This is certainly not the case. As with COVID, some events are known as **super-spreading events** where there is no pair-wise infection but instead where a single infected host infects a large number of other individuals. This begs the question, how does the presence of super-spreading events change the probability of epidemic emergence?

Example 2.18: Superspreaders

Consider a population in which 5% of events are super-spreading events where 3 individuals are infected and 2% of events lead to 4 infections.

1. Express the early dynamics of the disease as a branching process.

Defining $b = \frac{\beta}{\beta+\gamma}$ we have

$$\Pr(Y = k) = \begin{cases} b * 0.02 & k = 4 \\ b * 0.05 & k = 3 \\ b * 0.93 & k = 2 \\ 1 - b & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

2. What is the probability of extinction given $\beta = 3$ and $\gamma = 1$?

Defining g as before we have:

$$g = 1 * (1-b) + 0.97bg^2 + 0.02bg^3 + 0.01bg^4 \quad b = \frac{3}{3+1} = 4$$

Solving numerically, the one real root between 0 and 1 is $\hat{g} = 0.325$. In other words, the probability of emergence is $1 - \hat{g} = 0.675$.

So introducing super spreading events increases the probability of emergence by 1.5%.

3. What if 99% of events are pairwise transmission but 1% are huge such that 10 individuals are infected?

$$g = 1 * (1-b) + 0.99bg^2 + 0.01bg^10 \quad b = 4$$

Solving numerically we have that $\hat{g} = 0.331$ so a 0.1% increase in the probability of emergence.
