We'll characterize the nature (transient/null recurrent/positive recurrent) for this Markov chain based on the values of the birth and death parameters. We'll focus on the irreducible case where:

- $p_i, q_i > 0$ for $i \geq 1$
- $p_0 > 0$ (otherwise, if $p_0 = 0$ then 0 is an absorbing state, and therefore the rest of the states would be transient)

Irreducible because can simply go between any two desired states by one step at a time.

For the irreducible case, need some analysis, not just topology to sort out the long-time properties.

Following the script from last lecture, we attempt track 2A, which is to look for an invariant measure.

Write down the equations for the invariant measure:

$$
\rho \cdot r = \rho \\
\rho_j \geq 0
$$

$\rho_0 = 1$

$$
\rho = \\
\begin{pmatrix}
0 & r_0 & p_0 \\
q_0 & q_1 & r_1 \\
q_2 & q_2 & r_2 \\
& & \\
& & \\
& &
\end{pmatrix}
$$

$$
0 \quad \rho_{i,j} = \begin{cases} 
\rho_i & \text{if } j = i + 1 \\
q_i & \text{if } j > i \\
r_i & \text{if } j < i \\
0 & \text{else }
\end{cases}
$$
This is an infinite hierarchy of linear equations. Could solve by recursion, but there's a cleaner way because of the special structure of the probability transition matrix. Recall tricks for stationary distribution!

Look for a detailed balance solution. These are special cases of stationary distributions/invariant measures -- might not exist, but if they do, they serve as stationary distributions/invariant measures.

The only pairs of states we need to consider are those for which $|i - j| = 1$.

WLOG, take $j = i + 1$.

This invariant measure is valid and satisfies the detailed balance conditions, so we do have an invariant measure.

Next question: Can I normalize the invariant measure to get a stationary distribution?
Next question: Can I normalize the invariant measure to get a stationary distribution?

If this sum is finite, then we have a stationary distribution, and the MC is positive recurrent.

- necessary and sufficient; the only question is whether one can find a different invariant measure that is normalizable. But the coupling argument shows that if an invariant measure exists for an irreducible MC, it is unique up to constant.

But if this sum is infinite, then MC is null recurrent or transient...have to follow track b to sort it.

Now we implement the decisive test for transience. Choose a reference state \( i_c = 0 \). Form the matrix \( Q \) by deleting the corresponding row and column:

\[
Q = \begin{pmatrix}
1 & r_1 & r_2 & \cdots \\
r_1 & q_1 & q_2 & \cdots \\
r_2 & q_2 & q_3 & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\]

or

\[
Q_{ij} = \begin{cases}
p_i & \text{if } j = i + 1 \\
q_i & \text{if } j = i - 1 \\
r_i & \text{if } j = i \\
0 & \text{else}
\end{cases}
\]

Look for solutions of \( Qx = x \) or equivalently:

\[
\sum_{j \geq 1} Q_{ij} x_j = x_i \quad i \geq 1
\]
Can solve either by recursion, or by applying the same technique we used to solve for the absorption probability in a finite state birth-death chain. Remember that the $Qx = x$ equation arose from an absorption probability calculation for $i_e = 0$. Implementing this idea, we arrive at the solution:

$$X_j = X_1 \left( \sum_{k=1}^{j} \delta_k \right)$$

where

$$\delta_k = \prod_{k'=1}^{k} q_{k'} = \frac{p_0}{\prod_{k'=1}^{k} p_{k'}}$$

Transience is equivalent to the existence of bounded nontrivial nonnegative solutions to $Qx = x$.

$$\iff \sum_{k=1}^{\infty} \delta_k < \infty$$

Putting these two tests (Tracks 2a and 2b together), we can classify irreducible birth-death chains based on the values of their birth and death parameters as follows.

Define: $v_j = \prod_{i=0}^{j-1} \left( \frac{p_i}{q_{i+1}} \right)$

Then the Markov chain is:
- Positive recurrent if $\sum_{j=1}^{\infty} v_j < \infty$
- Transient if $\sum_{j=1}^{\infty} (p_j v_j) < \infty$
- Null recurrent otherwise.

In practice, then, one can plug in the birth/death parameters for a given model into these equations, and examine their convergence properties; examples in the books.

**Probability Generating Functions**
Just as Fourier transforms/series can be used to solve problems like partial differential equations by representing the given functions by suitable invertible transformations, various kinds of generating functions play a similar role in analytical probability theory. Laplace transform is also another similar idea, and generating functions in probability theory are just analogues of these kinds of transforms in signal analysis and PDE.

One particular generating function, which is particularly well-suited to discrete probability spaces is the probability generating function:

$$P_X(s) = \mathbb{E} s^X = \sum_{j=0}^{\infty} s^j \mathbb{P}(X = j)$$

Here $s$ plays the role of a dummy variable like in Laplace transform. Actually this transformation is more like what is known as the z transform. There are analogs to Laplace and Fourier transforms in probability theory; these are known as moment generating functions and characteristic functions. They play a similar role as the probability generating function for continuous state spaces, which we’re not focusing on.

The probability generating function is an invertible encoding of the probability distribution of the random variable $X$. To recover the probability distribution from the probability generating function, simply observe that the probabilities are just Taylor coefficients of the probability generating function evaluated at $s=0$.

$$\mathbb{P}(X = j) = \left. \frac{1}{j!} \left( \frac{d}{ds} \right)^j P_X(s) \right|_{s=0}$$

But how do I know the probability generating function is actually analytic at $s=0$? Because the probability distribution adds up to 1, so the radius of convergence of the power series is at least 1.

So why would we introduce this gadget? Some calculations are easier to do in terms of probability generating functions.

For example, moments of a probability distribution can be computed from the probability generating function as follows:

$$\mathbb{E} X^m = \left. \mathbb{E} \left( s \frac{d}{ds} \right)^m (sX) \right|_{s=1} = \left. \left( s \frac{d}{ds} \right)^m \mathbb{E} sX \right|_{s=1}$$
So let's see how this works for computing the mean and second moment of the binomial distribution.

\[
X \sim \text{Binomial}(N, p) \\
P(X = j) = \binom{N}{j} p^j (1-p)^{N-j} \quad \text{for} \quad j = 0, \ldots, N \\
\mathbb{E}X = \sum_{j=0}^{N} j \cdot P(X = j) = \text{beware mess}
\]

Probability generating function is easy to compute...

\[
P_X(s) = \mathbb{E}s^X = \sum_{j=0}^{N} s^j P(X = j) \\
= \sum_{j=0}^{N} s^j \binom{N}{j} p^j (1-p)^{N-j} \\
= \sum_{j=0}^{N} \binom{N}{j} (ps)^j (1-p)^{N-j}\binom{N}{j} (ps)^j (1-p)^{N-j} \\
P_X(s) = (ps + 1-p)^N \\
\mathbb{E}X = \left. \left( s \frac{d}{ds} P_X(s) \right) \right|_{s = 1} \\
= Np s (ps + 1-p)^{N-1}\bigg|_{s = 1} = Np
Where generating functions play a particularly important role is with sums of independent random variables.

The basic reason is as follows:

Consider $Y = \sum_{j=1}^{N} X_j$ where the $X_j$ are independent (not necessarily identical) random variables with known probability distributions and let’s suppose known probability generating functions $P_{X_j}(s)$. 

$$P_Y(s) = \left[ \prod_{j=1}^{N} P_{X_j}(s) \right]^{\frac{1}{N}}$$

$$= \left[ N \prod_{j=1}^{N} \left( \frac{1}{N!} \right)^{X_j} \right]^{\frac{1}{N}}$$

$$= N^{\frac{1}{N}} \prod_{j=1}^{N} P_{X_j}(s)$$

Where generating functions play a particularly important role is with sums of independent random variables.