Reading: Lawler Ch. 3
           Karlin and Taylor Ch. 4

Resuming from last lecture, why is the nontrivial value of
$0 < a < 1$ which satisfies the equation

$$a = P_{\Sigma}(a)$$

the solution to the extinction probability:

$a = P(\cup_{n=1}^{\infty} \{X_n = 0\} | X_0 = 1)$ when the mean offspring $\mathbb{E}Y > 1$?

One needs to consider what is happening at large epoch n by looking at the dynamics.

$a = \lim_{N \to \infty} P(\cup_{n=1}^{N} \{X_n = 0\} | X_0 = 1)$ because probability is continuous under limits of countable sequences.

But $\{X_{n'} = 0\} \subseteq \{X_n = 0\}$ for $n \geq n'$
so $\cup_{n=1}^{N} \{X_n = 0\} = \{X_N = 0\}$

$$a = \lim_{N \to \infty} P(X_N = 0 | X_0 = 1)$$

We can compute finite-horizon statistics involving
branching processes most conveniently using
probability generating functions:

$$P_{X_{n+1}}(s) = P_{X_n}\left(P_{\Sigma}(s)\right)$$

$\underset{\uparrow}{}$ pgf of offspring

Iterating this N times, we get:

$$P_{X_N}(s) = P_{\Sigma}\left(P_{\Sigma} \circ P_{\Sigma} \circ P_{\Sigma} \cdots P_{\Sigma}(s)\right)$$

$$P_{X_N}(s) = P_{X_0}\left(\underbrace{P_X \circ P_X \circ P_X \cdots P_X(s)}_{\text{N-fold composition}}\right)$$

$$P_{X_N}(0) = \sum_{j=0}^{q} p(X_N = j) \, 0^j$$

$$= p(X_N = 0)$$

$$P_{X_0}(s) = \sum_{j=0}^{\infty} p(X_0 = j) s^j$$

$$= 0 s^0 + 1 s^2 + \sum_{j=2}^{\infty} 0 s^j$$

$$= s$$

$$P_{X_N}(0) = \underbrace{P_X \circ P_X \circ \cdots P_X(0)}_{\text{N-fold composition}}$$

$$q = \lim_{N \to \infty} \underbrace{P_X \circ P_X \circ \cdots P_X(0)}_{\text{N-fold}}$$

We can think about this in terms of a discrete dynamical system with the mapping

$$s \to P_X(s)$$

One can show, by making appropriate estimates on the behavior of

$$\rho_{\mathcal{I}}(s)$$

that the nontrivial fixed point $0 < a < 1$ will globally attract all values of $0 \le s < 1$. In particularly it attracts the value starting from $s = 0$, so that establishes that the nontrivial solution of

$$q = \rho_{\mathcal{I}}(a)$$

is the correct answer for the probability of extinction starting from one agent.

Continuous-Time Markov Chains

Now we will try to carry over our theory, simulation, modeling approaches to Markov chains with continuous time (parameter domain); the state space will remain discrete (finite or countable).
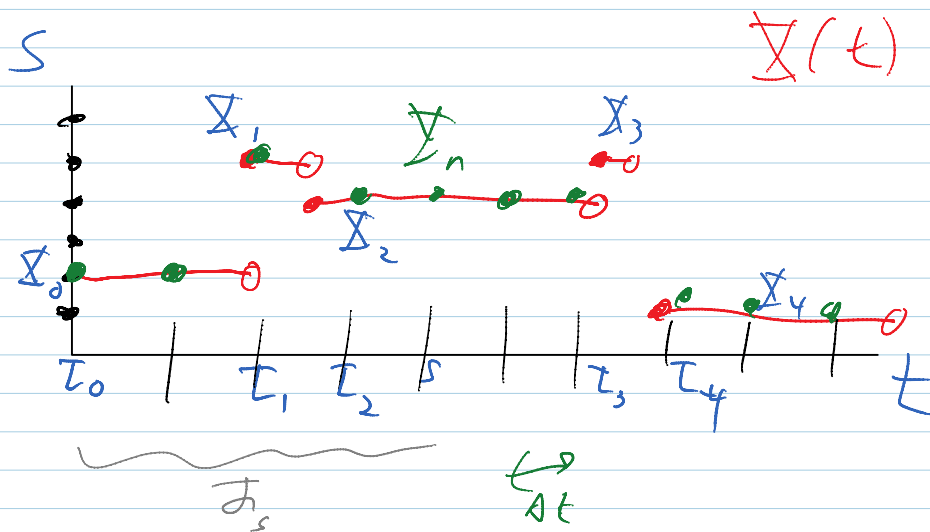
The convention for continuous-time Markov chains (CTMC) is that they be chosen to be right continuous (cadlag).

The Markov property for continuous-time stochastic processes can be written in two more-or-less equivalent ways:

Abstract:

$$P\left(X(t) = j \mid \mathcal{F}_s\right) = P\left(X(t) = j \mid X(s)\right)$$

$$\mathcal{J}_s^i$$

for $s < t$

is a filtration generated by the stochastic process X, and intuitively it means the information that is known by viewing the stochastic process up to and including time s.

Concrete: $P(X(t) = j | X(t_1) = i_1, X(t_2) = i_2, \ldots, X(t_n) = i_n) = P(X(t) = j | X(t_n) = i_n)$

for any finite sequence of times $0 \le t_1 < t_2 < t_3 \ldots < t_n < t$.

These descriptions are equivalent provided the stochastic process is nice enough (separable).

Note that a realization of a CTMC (in fact any separable discrete-state continuous-time process) can be encoded by a countable sequence:
$\{(X_0, \tau_0), (X_1, \tau_1), \ldots, (X_n, \tau_n), \ldots\}$
where
- $\tau_0 = 0, X_0 = X(\tau_0) = X(0)$
- $\tau_n = \min\{t > \tau_{n-1} : X(t) \ne X(\tau_{n-1}) = X_{n-1}\}, X_n = X(\tau_n)$ for $n \ge 1$.

For a CTMC, this encoding has a discrete-time Markov property w.r.t. each epoch corresponding to one transition of the CTMC
- except actually the τ_n have continuous state space
- also the sequence $\{X_0, X_1, \ldots, X_n, \ldots\}$ forms a Markov chain in its own right, the so-called embedded DTMC
  - this relies on the strong Markov property (apply the Markov property at the random times τ_n) and that the $\{\tau_n\}$ are Markov times meaning they are determined only by the stochastic process up to the time they occur (don't look into the future). More on Markov times later…

Another way to associate a DTMC to a CTMC:
- regularly sample the CTMC: $Y_n = X(n\Delta t)$; this is a DTMC.

So, what is a good way in practice to deal with the three modeling frameworks we've discussed for continuous-time Markov processes
- CTMC, embedded DTMC, regularly sampled DTMC
- Regularly sampled DTMC avoids all technicalities of continuous time, so the mathematics is "simpler" but:
  - you can miss transitions, particularly if the system is multiscale/stiff

- the model involves a nonphysical parameter Δt; the same issue in "simplifying" differential equations to difference equations
- often it's easier to formulate models involving rates of change in continuous time than changes over finite time intervals
- Embedded DTMC misses timing information, which can be relevant. Add back the timing information, i.e., look at the whole sequence
  - $\{(X_0, \tau_0), (X_1, \tau_1), \dots, (X_n, \tau_n), \dots\}$

  then this is a complete description of a CTMC. This representation is very useful for simulating CTMCs by Monte Carlo (Gillespie method, kinetic Monte Carlo). But it's not the most natural representation for analyzing the CTMC over specific times.

Common applications for CTMC (master equation):
- chemical/biomolecular reactions
- network dynamics
- pharmokinetics, epidemiology, ecology, neuroscience, evolutionary game theory

There are two useful ways to look at a CTMC:
- Embedded DTMC with timing
- Infinitesmal evolution

The second version is usually the one that is modeled, because rates of change enter naturally. And then we will later discuss how to relate this infinitesmal evolution framework to the embedded DTMC with timing framework.

Infinitesmal Modeling Framework for CTMCs

Recall from DTMC that the Markov property implies that all statistics about a DTMC can be related to the statistics of the Markov chain at any given epoch $\phi_j^{(n)} = P(X_n = j)$ and the probability of transition from one epoch to another $P(X_{n'} = j | X_n = i)$ for $n < n'$. The same arguments imply that the key quantities of interest for computing statistics of CTMCs are:

$\phi_j(t) = P(X(t) = j)$
$P_{ij}(s, t) = P(X(t) = j | X(s) = i)$
for $s < t$

Going forward we will assume time-homogeneity of the CTMC, which will simplify the analysis just as for time-homogenous DTMCs. It implies:
$P(X(t) = j | X(s) = i) = P_{ij}(t - s)$

By discreteness of state we can as we did for DTMCs represent the transition probabilities in terms of a matrix, which is now a matrix function which we call the probability transition function

$$P(t) = \{P_{ij}(t)\}_{i,j \in S}.$$