As we discussed last time, for a time-homogenous, continuous-time Markov chain, all the statistics can be encoded in terms of a probability transition (matrix) function:

\[ P_{ij}(t) = P(\mathbb{X}(t, t') = j | \mathbb{X}(t') = i) \]

Actually this probability transition function is itself completely determined by its behavior over infinitesimal time intervals due to the Markov property, which implies the Chapman-Kolmogorov equation:

\[ P_{ij}(t, t_2) = \sum_{k \in S} P_{ik}(t_1) P_{kj}(t_2) \]

This allows the probability transition function to be determined by its behavior on arbitrarily small intervals.

So consider the probability transition function over a very short time interval

\[ P(t) = \{ P_{ij}(t) \}_{ij} \]

**Taylor expansion**

\[ P(t) = P(0) + P'(0) t + o(t) \]

\[ \lim_{t \to 0} \frac{o(t)}{t} = 0 \]

\[ P(0) = I \]

\[ P' = \begin{pmatrix} P_{ij}(0) \end{pmatrix} \]

\[ P'(0) = A \]

plays the fundamental role in the formulation of a CTMC

It's called transition rate matrix or infinitesimal generator.

What are the entries of this matrix \( A \)?
\[ A = \lim_{\Delta t \downarrow 0} \frac{P(\Delta t) - I}{\Delta t} \]

Off-diagonal terms:
\[ A_{ij} = \lim_{\Delta t \downarrow 0} \frac{P_{ij}(\Delta t)}{\Delta t} = \lim_{\Delta t \downarrow 0} \frac{P(X(t+\Delta t) = j | X(t) = i)}{\Delta t} \]

Note that this describes a rate at which the probability to jump from state \( i \) to \( j \) increases. It is not a probability!

What about the diagonal terms?
Recall that
\[ \sum_{j \in S} P_{ij}(t) = 1 \]
\[ P_{ij}(t) = \delta_{ij} + \frac{A_{ij} \Delta t + o(\Delta t)}{\Delta t} \]
\[ \sum_{j \in S} P_{ij}(t) = \sum_{j \in S} \left( \delta_{ij} A_{ij} \Delta t + o(\Delta t) \right) \]
\[ 1 = 1 + \left( \sum_{j \in S} A_{ij} \right) \Delta t + o(\Delta t) \]
\[ \implies 0 = \sum_{j \in S} A_{ij} \quad \text{for all } i \]
\[ \implies A_{ii} = - \sum_{j \in S} A_{ij} = - \lambda_i \]
\[ \lambda_i = \text{total rate at which the MC leaves state } i. \]

So what is the intuitive meaning behind all these quantities?
- \( A_{ij} \) denotes the inverse of the average amount of time to wait to see a transition from \( i \) to \( j \) if all other transitions were disabled
- \( \lambda_i \) is the inverse the average amount of time spent in state \( i \) before some transition happens.

So to define a continuous time Markov chain, one need only:
- For each possible transition \( i \rightarrow j \) compute or model the rate at which this transition would happen if no other process intervened (in other words, one over the average amount of time required for the process to happen) and set \( A_{ij} \) equal to this for \( i \neq j \)
- Set the diagonal terms according to:
\[ A_{ii} = - \lambda_i = - \sum_{j \neq i} A_{ij} \]
Define the initial probability distribution for the state of the system:

\[ \mathbf{\theta}_0 = \mathbf{P}(X(t) = j) \]

Note there is a technical condition on infinite state CTMCs that you might want to check to be sure you have a well defined CTMC (Karlin and Taylor Ch. 4) -- the concern is the Markov chain "blowing up" in finite time.

Examples

1) Birth-death processes

\[
\begin{align*}
A_{i,i+1} &= \lambda_i & \text{"aggregate birth rates"} \\
A_{i,i-1} &= \mu_i & \text{"aggregate death rates"} \\
A_{i,i} &= -\lambda_i - \mu_i \\
A_{i,j} &= 0 \quad \text{for} \quad |i-j| \geq 2
\end{align*}
\]

Many useful models fall under this framework

a) Poisson counting process (arrivals, number of occurrences of random events)

\[
N(t)
\]

\[
\lambda_i = \lambda, \quad \mu_i > 0
\]

b) Queueing models (k servers, queue is allowed to be unbounded)

\[
X(t) = \text{# requests in or awaiting service}
\]

\[
\lambda_i = \lambda \quad \text{(arrival rate of demand)}
\]
c) Population models

Various possibilities depending on assumptions of reproduction, competition for resources, etc.

\[
\begin{align*}
\dot{N}_i &= \lambda_i \\
M_i &= M_i
\end{align*}
\]

Compare with logistic growth deterministic model:

\[
\frac{dX}{dt} = \lambda X \left(1 - \frac{X}{K}\right)
\]

By the way the choice

\[
\dot{N}_i = \lambda_i^i \quad M_i = \frac{M_i^i}{K}
\]

violates the technical condition for a well-defined CTMC but could regularize the model by having, for example, the birth rate slow down to linear growth at a large enough population.

Other examples of CTMCs:

2) Finite-state processes without linear ordering
   ○ atomic state transitions in atomic physics
   ○ conformational changes in biomolecules
   ○ climactic transitions

3) Biochemical and genetic regulatory networks

Typical framework is to identify the types of biomolecules of interest, for brevity, call them

\[S_1, S_2, \ldots, S_8\]

State space will be vectors of nonnegative integers which provide the current count of the number of molecule of each type \(X_1, X_2, \ldots, X_8\)

This is a huge (but countable!) state space and the transition rate matrix \(A\) is of course hard to write down directly (but somewhat sparse)....

Instead, one describes it implicitly by identifying each possible process that could cause a state change, indicate how the state changes due to that process, and the rate at which that change occurs.
For the heat shock model, the table provided by Tom Kurtz quoted in the last lecture identifies all the relevant processes, now let’s consider how one of those processes can induce changes in the state of the Markov chain and determine its rate:

\[ \text{State } i = (x_1, x_3, \ldots, x_8) \text{ makes a transition to state } \]
\[ (x_1, x_2 + 1, x_3, x_4, x_5 + 1, x_6, x_7, x_8) \text{ with rate } 6.3 \times 10^{-3} \]

provided \( x_3 \geq 1 \) otherwise this process cannot change the state.

This is fine for giving a well-defined CTMC model with a sparse infinitesimal generator matrix -- but how to compute statistics or even simulate efficiently -- topics of current research.

**Statistical computations for Continuous-Time Markov Chains**

**Finite-time statistics**

For time-homogenous CTMC (as we will generally assume), this amounts to the question of how to compute \( P(t) \) given the model \((A, \phi)\)?

We will simply mimic the argument for the discrete-time case using the Chapman-Kolmogorov equation:

\[
\begin{align*}
P(t) & = \{ P_{ij}(t) \} \\
\dot{P}_{ij}(t) & = P(\tau(t + \Delta t) \rightarrow \tau(t) = s) \\
\end{align*}
\]

\[
P_{ij}(t + \Delta t) = \sum_{k \in S} P_{ik}(\Delta t) P_{kj}(t)
\]

Consider this for small \( \Delta t \):

\[
P_{ik}(\Delta t) = S_{ik} + \Delta t A_{ik} + o(\Delta t)
\]

\[
P_{ij}(t + \Delta t) = \sum_{k \in S} \left( S_{ik} + \Delta t A_{ik} + o(\Delta t) \right) P_{kj}(t)
\]
For technical precision, one has to be a bit careful about whether the error term really can be neglected in this limit because the sum could be infinite -- however, it can be shown with sufficient effort that the error term can be rigorously neglected (Karlin and Taylor, Vol I and II).

Then we have:

\[
\begin{align*}
P_{ij}(t + \Delta t) &= \sum_{k \in S} \left( \delta_{ik} + A_{ik} \Delta t + o_{ik}(\Delta t) \right) P_{kj}(t) \\
&= P_{ij}(t) + \sum_{k \in S} A_{ik} P_{kj}(t) \Delta t \\
&\quad + \sum_{k \in S} o_{ik}(\Delta t) P_{kj}(t)
\end{align*}
\]

\[
\frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} = \sum_{k \in S} A_{ik} P_{kj}(t) + \sum_{k \in S} \frac{o_{ik}(\Delta t)}{\Delta t} P_{kj}(t)
\]

\[
\text{At } \Delta t \to 0 \; \text{limit } \rightarrow \text{error term?}
\]

For technical precision, one has to be a bit careful about whether the error term really can be neglected in this limit because the sum could be infinite -- however, it can be shown with sufficient effort that the error term can be rigorously neglected (Karlin and Taylor, Vol I and II).

Then we have:

\[
\frac{dP_{ij}(t)}{dt} = \sum_{k \in S} A_{ik} P_{kj}(t)
\]

We can write this in matrix form:

\[
\frac{d\rho(t)}{dt} = A \rho(t)
\]

By instead using Chapman-Kolmogorov equation in this form:

\[
P_{ij}(t + \Delta t) = \sum_{k \in S} P_{ik}(t) P_{kj}(\Delta t)
\]

then by similar formal arguments we would achieve the

**Kolmogorov forward equation**

\[
\frac{dP_{ij}(t)}{dt} = \sum_{k \in S} P_{ik}(t) A_{kj}
\]
Kolmogorov forward equation

\[
\frac{d p_{ij}(t)}{dt} = \sum_{k \in S} p_{ik}(t) A_{kj}
\]

\[p_{ij}(t=0) = \delta_{ij}\]

\[
\frac{d}{dt} p(t) = p(t) A
\]

\[p(t=0) = J\]