Long-Time Properties of CTMCs

Homework 4 will be posted by tomorrow; due Wednesday, May 6 at 5 PM.

Special math colloquia by Michael Brenner:
- Monday, 4 PM Sage 5101: Linear algebra and bird beaks
- Tuesday, 12:30 PM Eaton 216: Droplet splashing

How do we modify the long-time formulas for DTMCs to apply to CTMCs.

Classification of Markov chains (positive recurrence, null recurrence, transience)

- **Recurrence vs. transience** is decided by simply looking at whether the embedded DTMC is recurrent or transient. (The timing spent in a state is irrelevant to this question).
- **Communication classes** are determined by same topological considerations.
- Positive recurrence cannot be entirely decided by embedded DTMC because timing is important.
  - However, positive recurrence can again be decided (for an irreducible Markov chain or closed communication class) by the presence of a stationary distribution. But the stationary distribution for a CTMC is not quite the stationary distribution for the embedded DTMC. (That's why the CTMC and embedded DTMC can have different positive/null recurrence properties).

Stationary Distribution for CTMC

\[ \pi_i = P(X(t) = i) \]

should be a stationary solution of the forward Kolmogorov equation for the probability distribution of the state of the MC.

\[ \frac{d\pi}{dt} = \pi \cdot A \]

\[ \pi = \pi \cdot A \]

\[ \pi_i \geq 0 \]

\[ \sum \pi_i = 1 \]
This will in general give different answer than the stationary distribution for the embedded DTMC because the stationary distribution for a CTMC is sensitive to the amount of time spent in the various states.

The trick of trying to look for detailed balance solutions for stationary distributions is also a good idea for CTMCs:

\[ \pi_i A_{ij} = \pi_j A_{ji}. \]

So in summary, to classify a CTMC:

- Construct the embedded DTMC
- Apply the classification procedure for embedded DTMC to determine transience vs. recurrence
- But to determine positive recurrence, need to check for the existence of the stationary distribution for the CTMC (not for the embedded DTMC)

Once the Markov chain or its communication classes have been classified, long-time behavior is analyzed in a similar way to DTMCs, with the following modifications:

1. For Markov chains starting within a positive recurrent class, the long-time behavior is governed by the stationary distribution associated to that positive recurrent class, and the LLN for CTMCs reads:

\[
\lim_{T \to \infty} \frac{1}{T} \int_0^T f(X(t)) dt = \sum_{j \in S} \pi_j f(j).
\]

And there is no need to worry about periodicity in continuous-time in stating that the stationary distribution is a limit distribution:

\[ \lim_{t \to \infty} P(X(t) = j) = \pi_j. \]
2. **Null recurrent classes** behave roughly similarly to how they did for DTMCs, and in particular, there's no particularly useful formulas for extracting finer detail about the long-time behavior in general.

3. For Markov chains starting in a **transient class**, one can do similar calculations as for DTMCs regarding absorption probabilities (which, if any, recurrent class is eventually visited by the MC) and the accumulation of cost/reward (additive functional) while moving through transient states.
   a. **Absorption probabilities**: Can simply compute the absorption probability for the embedded DTMC, and will give the correct result because it doesn't rely on timing.  
   - Alternatively, there is a way to compute absorption probabilities directly on the CTMC. A proper derivation is given in
     - Goutsias and Jenkinson, "Markovian dynamics on complex reaction networks," Physics Reports 529 (2013), 199-264
     - Here is a heuristic derivation of the formula from DTMC formula. Think about the DTMC $X_n = X(n \Delta t)$.
     - If we apply the absorption probability calculation for this associated DTMC, it is:

\[
\begin{align*}
    U_{ij} &= P_{ij} + \sum_{k \in T} P_{ik} U_{kj} \\
    \text{for } i \in I^+, j \in T^c
    \\
    U_{ij} &= P(X_T = j | X_0 = i) \\
    I^c &= \min\{n > 0 : X_n \notin T\}
\end{align*}
\]

Approximate the probability transition matrix for the associated DTMC in terms of the transition rate matrix for the CTMC:

\[
\begin{align*}
    \rho &= e^{A \Delta t} \\
    P_{ij} \text{ in } U_{ij} &= \delta_{ij} + A_{ij} \Delta t + o(\Delta t) \\
    U_{ij} &= A_{ij} \Delta t + A_{ij} \rho \Delta t + o(\Delta t) \\
    U_{ij} &= A_{ij} \rho \Delta t + A_{ij} \sum_{k \in T} A_{ik} U_{kj} + o(\Delta t)
\end{align*}
\]
This can be expressed in matrix form by doing a canonical decomposition on $A$.

b. **Accumulated cost/reward (additive functional)** while in transient states

$$W_\xi = \mathbb{E}\left( \int_0^\infty f(x(t)) \, dt \mid x(0) = \xi \right)$$

$T := \min \{ t > 0 : t \not\in T \}$

Just as for stationary distribution, the embedded DTMC does not have enough information to compute this quantity so we will present a different CTMC formula. Note also that the function $f(x)$ is the rate at which cost/reward is accumulated while the CTMC is in state $x$.

Construct the vectors

$$\vec{f} := \begin{pmatrix} f(1) \\ f(2) \\ \vdots \end{pmatrix} \quad \vec{w} := \begin{pmatrix} w_1 \\ w_2 \\ \vdots \end{pmatrix}$$

$$A \cdot \vec{w} = \vec{f}$$

$$w_i = f_i = 0 \text{ for } i \in T.$$  

This is morally a boundary value problem.

This can be heuristically derived from the DTMC formula by the same kind of discretization of the CTMC into time steps of size $\Delta t$ and taking $\Delta t \downarrow 0$ as we did above. A more proper derivation (from scratch) can be found in Lawler Sec. 3.3, and Karlin and Taylor Ch. 4.

Some extensions of these basic ideas:

1. One can compute not only expected time to be absorbed by the probability density for the time to be absorbed by employing the trick of relating time statistics to state statistics.
\[ \tau = \inf \{ t > 0 : X(t) \in T^c \} \]

\[ P(T > t) = P(X(t) \in T) = \sum_{j \in T} \varphi_j(t) \]

which can be obtained from the forward Kolmogorov equation.

And then

\[ p_t(t) = -\frac{d}{dt} P(T > t) \]

2. All these formulas can be extended to stochastic differential equations by replacing the transition rate matrix \( A \) by the appropriate infinitesimal generator of the stochastic differential equation, which is typically a second order partial differential operator (elliptic).

3. The formulas we have developed often are written in an equivalent but different form that is useful for complex systems that are typical of reaction networks (including chemical, neuronal, social, epidemiological) where the state space is typically something like a high-dimensional lattice. In these settings it is often impractical to directly represent the transition rate matrix \( A_{ij} \). (See Goutsias and Jenkinson, "Markovian dynamics on complex reaction networks," Physics Reports 529 (2013), 199-264).

Instead one uses a kind of "sparse matrix" representation that is well adapted to many complex stochastic models of modern interest.

For reaction networks, this is done by listing a collection \( M \) of possible reactions; typically this a finite list. This is encoded through a paired list:

- change induced by reaction \( m \): \( s_m \) (stochiometric term)
- rate of reaction \( m \): \( \alpha_m(x) \) as a function of the state \( x \in S \).

This list implicitly describes the transition rate matrix as a sparse matrix. But typically one doesn't actually construct the matrix itself. Instead, one rewrites the equations involving the matrix \( A \) to directly refer to this sparse description.

For example, the forward Kolmogorov equation (which is often called the "master equation" when it is written in this modified way) can be written:

\[ \Omega(t) = (\varphi_1(t), \varphi_2(t), \ldots) \]

\[ \varphi_j(t) = P(X(t) = j) \]

In these more complex state spaces, it is often useful to parameterize the state space by a vector \( x \) with each component representing the number of species of a certain type, or the state of a given node in a network.

\[ \varphi_{\tilde{x}}(t) = P(\tilde{X}(t) = \tilde{x}) \]

\[ \gamma \approx \cdot (t) < \gamma \cdot \gamma \cdot \gamma (\tilde{x}) \cdot \gamma \cdot \gamma \cdot \gamma \]
This is just the sparse matrix representation of the forward Kolmogorov equation

\[
\frac{\partial \Psi_x(t)}{\partial t} = \sum_{m \in M} \left( \alpha_m(\mathbf{x}^- - \mathbf{s}_m) \Psi_{\mathbf{x}^- - \mathbf{s}_m}(t) - \alpha_m(\mathbf{x}^+ \Psi_{\mathbf{x}^+}(t) \right)
\]

One has to be careful still about getting the terms right, regarding how the transition rate matrix is multiplying the desired vector of variables. Forward Kolmogorov equation converts, in sparse matrix representation, to equations of the above type which involve rates of moving in from precursor states, but just total rates of moving out.

On the other hand, with the backward Kolmogorov equation (when the vector of the desired variable is multiplied on the right of the transition rate matrix), one has instead that the sparse matrix/reaction network representation will refer to successor states (\(\phi_{x^+ s_m}\)) of the reaction from state \(x\), but not precursor states (\(\phi_{x^- s_m}\)).

We now turn attention to stochastic point processes.

These are stochastic models that describe how to distribute points randomly over some domain. They are not Markov chains. We’ll start with thinking about point processes in one dimension, where they are most commonly seen.

State space is actually hard to formulate (something like \(\mathbb{R}^\infty\) but not terribly useful). One instead tends to restrict attention to point processes that can be generated by a simpler probability space.

- **Poisson point processes** (special, widespread, and work well in any finite dimension)
- **Renewal processes** (more general one-dimensional point processes that have a certain kind of lack of memory that is reminiscent of Markov processes)