MTH 7241 Fall 2022: Prof. C. King

Notes 9: Continuous time Markov chains

The state space Ω is discrete, and the chain $\{X(t)\}$ is indexed by the half line $\{t \geq 0\}$. There are two useful ways to characterize continuous time Markov Chains.

Semigroup for Continuous time Markov Chain

The first is the obvious generalization of the discrete time Markov property:

$$\begin{aligned} P(X(t_n) = j \mid X(t_1) = i_1, \dots, X(t_{n-1}) = i_{n-1}) = \\ P(X(t_n) = j \,|\, X(t_{n-1}) = i_{n-1}) \end{aligned}$$

for all states $j, i_1, \dots, i_n \in \Omega$ and *any sequence of times* $t_1 < t_2 < \cdots < t_n$.

The chain is homogeneous if the transition probability

$$P(X(t) = j \,|\, X(s) = i)$$

is a function of $t - s$ for all $i, j \in \Omega$ and all $s \leq t$. In this case we write

$$P(t)_{ij} = P(X(t) = j \,|\, X(0) = i)$$

The chain is described by the family of matrices $\{P(t) \,|\, t \geq 0\}$, known as a stochastic semigroup. They satisfy the property

$$P(s)P(t) = P(s+t) \quad \text{for all } s, t \geq 0 \tag{1}$$

They also satisfy

$$\begin{aligned} P(0)_{ii} &= 1 \quad \text{for all } i \in \Omega & (2) \\ P(0)_{ij} &= 0 \quad \text{for all } i \neq j \in \Omega & (3) \end{aligned}$$

Generator for Continuous time Markov Chain

The second way to characterize the continuous time Markov Chain $\{X(t)\}$ is by a discrete Markov chain on Ω, together with independent exponential holding times $\{T_i\}$. After arrival at site i, the chain waits for a random time T_i, then jumps to site j with probability p_{ij}. So the chain is described by the transition matrix $P = (p_{ij})$ of the underlying discrete Markov chain, together with the rates of the holding times λ_i at each state.

Note that we require $p_{ii} = 0$ for all states i, so it must jump somewhere after the holding time.

Semigroup and Generator

How are these descriptions related? Starting with the second description, define the matrix

$$G_{ij} = \begin{cases} -\lambda_i & \text{for } i = j \\ \lambda_i \, p_{ij} & \text{for } i \neq j \end{cases}$$

Then G is called the *generator* of the chain. The exponential of G gives the transition matrices of the first description, that is

$$P(t) = \exp(tG) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \, G^n$$

So the first description can be derived from the second one in this way.

The other direction does not always hold, and needs some additional regularity assumptions on the family $P(t)$.

Definition 1 *The semigroup* $P(t)$ *is called* uniform *if* $P(t) \to \mathbf{1}$ *uniformly as* $t \downarrow 0$.

The symbol $\mathbf{1}$ denotes the identity matrix or identity operator:

$$\mathbf{1}_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \tag{4}$$

We always have convergence in the sense that $P(t)_{ij} \to \mathbf{1}_{ij}$ for all states $i, j \in \Omega$. Concretely, this means that for all $\epsilon > 0$ and all $i, j \in \Omega$, there is $\delta > 0$ such that

$$0 \leq t < \delta \Rightarrow |P(t)_{ij} - \mathbf{1}_{ij}| < \epsilon \tag{5}$$

Uniform convergence is the additional statement that for every ϵ the same value of δ can be used for every i, j. This is always true for a finite dimensional matrix, so for a finite state space Ω the convergence is always uniform. For infinite state space it may happen that the convergence is not uniform.

Theorem 1 *If* $P(t)$ *is uniform, then there is a matrix* G *such that* $P(t) = \exp(tG)$ *for all* $t \geq 0$, *and the matrix* G *defines a chain in the second sense above (as a generator).*

Henceforth we assume that the chain is uniform, and we go back and forth between the two descriptions. Here is a simple example to illustrate the correspondence.

Example 1 *Suppose $\Omega = \{0,1\}$. In the second description we have*

$$G = \begin{pmatrix} -1 & 1 \\ 0.5 & -0.5 \end{pmatrix}$$

Thus the chain oscillates between the states; the mean holding time at state 0 is $(\lambda_0)^{-1} = 1$, the mean holding time at state 1 is $(\lambda_1)^{-1} = 2$. After the holding time at state 0 expires, the chain must jump to state 1, and $p_{01} = 1$. Similarly $p_{10} = 1$.

In the first description, the semigroup is

$$P_t = e^{tG} = \sum_{n=0}^{\infty} \frac{t^n}{n!} G^n$$

Since G is rank one, can check that

$$G^n = \left(-\frac{3}{2}\right)^{n-1} G$$

Thus

$$P_t = \frac{1}{3} \begin{pmatrix} 1 + 2e^{-3t/2} & 2 - 2e^{-3t/2} \\ 1 - e^{-3t/2} & 2 + e^{-3t/2} \end{pmatrix}$$

Note that as $t \to \infty$ this converges to

$$\lim_{t \to \infty} P_t = \frac{1}{3} \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$$

0.1 Stationary distributions and convergence

There are convergence theorems analogous to, but simpler than, the ones for discrete time chains. First we note the following fact: for any pair of states i, j, either $P(t)_{ij} = 0$ for all $t > 0$, or else $P(t)_{ij} > 0$ for all $t > 0$. Notice that this implies all chains are aperiodic. This also gives the definition of irreducibility.

Definition 2 *The CTMC is* irreducible *if $P(t)_{ij} > 0$ for every pair of states i, j, for all $t > 0$.*

As usual, a distribution π_i is stationary if

$$\pi_j = \sum_i \pi_i \, P(t)_{ij}$$

for all $t \geq 0$. By taking the limit $t \downarrow 0$ and using $P(t) = e^{tG}$ this is equivalent to the condition

$$0 = \sum_i \pi_i \, G_{ij}$$

Theorem 2 *Let X be a uniform irreducible CTMC. If there is a stationary distribution π_i then it is unique and*

$$P(t)_{ij} \to \pi_j$$

as $t \to \infty$, for all i, j. If there is no stationary distribution then

$$P(t)_{ij} \to 0$$

as $t \to \infty$, for all i, j.

Finally note that the relation between $P(t)$ and G can also be expressed as a differential equation: the forward equations

$$P(t)'_{ij} = \sum_k P(t)_{ik} G_{kj}$$

and the backward equations

$$P(t)'_{ij} = \sum_k G_{ik} P(t)_{kj}$$

0.2 The Poisson process

This is the following special case, given in the second description: the process is $\{N(t)\}$ on $\Omega = \{0, 1, 2, \dots\}$, with $N(0) = 0$, $\lambda_i = \lambda$ and

$$p_{ij} = \begin{cases} 1 & \text{for } j = i + 1 \\ 0 & \text{else} \end{cases}$$

So the chain starts at 0, and increases by jumps which occur after holding times which are IID exponential with rate λ. There are several immediate consequences.

Theorem 3
- *For every t, $N(t)$ has the Poisson distribution with parameter λt, that is*

$$P(N(t) = j) = \frac{(\lambda t)^j}{j!} e^{-\lambda t}$$

- *The arrival times $0 = T_0 < T_1 < T_2 < \dots$ have gamma distribution.*
- *If $s < t$ then $N(t) - N(s)$ is Poisson with rate $\lambda(t - s)$, and is independent of the arrival times during the interval $[0, s]$.*

The Poisson process is widely used as the model for an arrival process, both because it does a good job in many cases, and also because it allows exact computations in many ways.

The Poisson process has many remarkable properties not shared by other CTMC's, here are some examples.

Theorem 4 (The superposition theorem) *Let $N_1(t)$ and $N_2(t)$ be independent Poisson processes with rates λ_1 and λ_2 respectively. Then $N_1 + N_2$ is a Poisson process with rate $\lambda_1 + \lambda_2$.*

There is also a converse.

Theorem 5 (The thinning theorem) *Suppose that $N(t)$ is a Poisson process with rate λ, and each arrival is independently classified as one of two types: Type 1 with probability p, or Type 2 with probability $1 - p$. Let $N_1(t)$ and $N_2(t)$ be the number of arrivals of type 1 and 2 respectively. Then N_1 and N_2 are independent Poisson processes with rates λp and $\lambda(1 - p)$ respectively.*

0.3 The birth-death process

This can be viewed as a generalization of the Poisson process where there are both arrivals and departures. For each state j there is a departure rate μ_j and an arrival rate λ_j, and the holding time at state j is exponential with rate $\lambda_j + \mu_j$. The transition matrix is

$$p_{ij} = \begin{cases} \frac{\lambda_i}{\lambda_i+\mu_i} & \text{for } j = i+1 \\ \frac{\mu_i}{\lambda_i+\mu_i} & \text{for } j = i-1 \\ 0 & \text{else} \end{cases}$$

where $\mu_0 = 0$. So the generator is

$$G_{ij} = \begin{cases} -(\lambda_i + \mu_i) & \text{for } i = j \\ \lambda_i & \text{for } j = i+1 \\ \mu_i & \text{for } j = i-1 \\ 0 & \text{else} \end{cases}$$

This is used as a model for population growth where arrivals are births and departures are deaths. It also arises in queuing systems, as we will see shortly. For the moment we just determine the stationary distribution. This satisfies

$$\sum_i \pi_i \, G_{ij} = 0$$

for all j. The equations are

$$\begin{aligned} \lambda_0\pi_0 + \mu_1\pi_1 &= 0 \\ \lambda_{n-1}\pi_{n-1} - (\lambda_n + \mu_n)\pi_n + \mu_{n+1}\pi_{n+1} &= 0 \end{aligned}$$

Rather than solving directly, we note that similarly to the discrete Markov chain, the condition for time reversibility is

$$\pi_i p_{ij}(t) = \pi_j p_{ji}(t)$$

for all $t > 0$. Taking the limit $t \to 0$ this gives

$$\pi_i G_{ij} = \pi_j G_{ji}$$

Applying this to the birth-death model we get the equation

$$\pi_i \lambda_i = \pi_{i+1} \mu_{i+1}$$

which immediately gives the recursion relation

$$\pi_{i+1} = \frac{\lambda_i}{\mu_{i+1}} \pi_i$$

Hence we get the solution for $n \geq 1$

$$\pi_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} \pi_0$$

So the condition for existence of the stationary distribution is

$$\sum_n \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < \infty$$

Example 2 *The Yule process is a pure birth process in which each individual gives birth at rate λ, independent of all others. Also known as the preferential attachment process. So each starting individual $N_i(0) = 1$ gives rise to a family $N_i(t)$, where $N_i(t)$ is a CTMC with*

$$\lambda_n = n\lambda, \quad n \geq 1$$

The total population at time t is $\sum_i N_i(t)$.

The process $N_i(t)$ has arrival times $T_1 < T_2 < \cdots$, so let $X_1, X_2, \ldots$ be the interarrival times. These are independent exponential r.v.'s with increasing rates $\lambda, 2\lambda, \ldots$.

Now

$$N_i(t) > n \Leftrightarrow T_n < t \Leftrightarrow X_1 + \cdots + X_n < t$$

Thus

$$P(N_i(t) > n) = P(X_1 + \cdots + X_n < t)$$

This probability is given by

$$P(X_1 + \cdots + X_n < t) = \left(1 - e^{-\lambda t}\right)^n$$

This can be proved by induction on n. Thus

$$\begin{aligned} P(N_i(t) > n) &= P(X_1 + \cdots + X_n < t) \\ &= \left(1 - e^{-\lambda t}\right)^n \end{aligned}$$

and so

$$P(N_i(t) = n) = \left(1 - e^{-\lambda t}\right)^{n-1} e^{-\lambda t}$$

This has a useful interpretation; let $p = e^{-\lambda t}$, and consider a biased coin which comes up Heads with probability p. Then $N_i(t)$ is the number of tosses needed until the first Heads appears. Thus $\sum_{i=1}^m N_i(t)$ is the number of tosses needed until we get m Heads. This has a negative binomial distribution:

$$P\left(\sum_{i=1}^m N_i(t) = k\right) = \binom{k-1}{m-1} e^{-m\lambda t} \left(1 - e^{-\lambda t}\right)^{k-1}$$