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.

$X(t) \ : t \geq 0$

Semigroup for Continuous time Markov Chain

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

t_{n-1} = present time

impose the Markov property for all times.

$$P(X(t_n) = j \mid X(t_1) = i_1, \ldots, X(t_{n-1}) = i_{n-1}) = P(X(t_n) = j \mid X(t_{n-1}) = i_{n-1})$$

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

t_n = future

The chain is homogeneous if the transition probability

$$P(X(t) = j \mid X(s) = i)$$

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

$t_1 \ t_2 \ \cdots \ t_{n-1} \ t_n$

could involve multiple jumps

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

$i \to j$, s, t

The chain is described by the family of matrices $\{P(t) \mid t \geq 0\}$, 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

$$P(0)_{ii} = 1 \quad \text{for all } i \in \Omega \tag{2}$$

$$P(0)_{ij} = 0 \quad \text{for all } i \neq j \in \Omega \tag{3}$$

Analogy: Discrete | Continuous

	Discrete	Continuous
one-step	P_{ij}	$P(t)_{ij}$
2-steps	$P_{ij}(2)$	$t=2 \Rightarrow P(2)_{ij}$
3-steps	$P_{ij}(3)$	$t=3 \Rightarrow P(3)_{ij}$
$\vdots$	$\vdots$	

$$P_{ij}(2) = \mathbb{P}(X_2 = j \mid X_0 = i)$$

$$= \sum_k \mathbb{P}(X_2 = j \mid X_1 = k)\ \mathbb{P}(X_1 = k \mid X_0 = i)$$

$$= \sum_k P_{kj} P_{ik}$$

$$= \sum_k P_{ik} P_{kj}$$

$$= (P \cdot P)_{ij}$$

$$P(s+t)_{ij} = \mathbb{P}(X(s+t) = j \mid X(0) = i)$$

$$= \sum_k \mathbb{P}(X(s+t) = j \mid X(s) = k).$$

$$\mathbb{P}(X(s) = k \mid X(0) = i)$$

(homgeneous) $$= \sum_k \mathbb{P}(X(t) = j \mid X(0) = k).$$

$$\mathbb{P}(X(s) = k \mid X(0) = i)$$

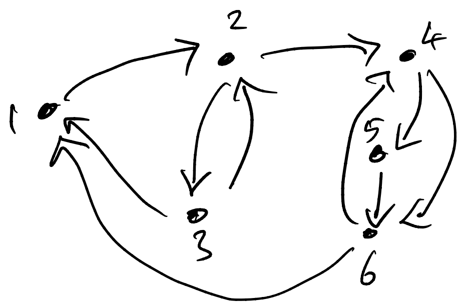
$$= \sum_k P(t)_{kj} \; P(s)_{ik}.$$

$$= (P(s) P(t))_{ij}$$

$\boxed{\textbf{Generator for Continuous time Markov Chain}}$

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.



P_{ij} : no time dependence

P_{ij} = prob. to jump next to state j, from state i

$P_{45} = q$

$P_{46} = 1-q$

Discrete chain with all diagonal entries are zero.

T_i = holding time at state i

$\sim$ exponential r.v. rate λ_i.

pdf: $f_{T_i}(t) = \lambda_i e^{-\lambda_i t}$ $(t \geq 0)$

$\mathbb{P}(T_i \geq t) = e^{-\lambda_i t}$

Mean time before jumps out of

state i $= \dfrac{1}{\lambda_i} = \mathbb{E}[T_i]$

Holding times are independent at

each state.

$$\lambda_{ij} = \lambda_i P_{ij} = \text{rate to jump } i \to j.$$

$\downarrow$

For each state j we have a "clock"

T_{ij} with exponential waiting time,

rate λ_{ij}. These clocks are independent.

$$T_i = \min_j T_{ij}$$

Independent exponentials $\{T_{ij}\}$

$\Rightarrow$

$$\Rightarrow \min_j T_{ij} \sim \text{exponential}$$

$$\text{rate } \lambda_i = \sum_j \lambda_{ij}$$

$$= \sum_j \lambda_i P_{ij}$$

$$= \lambda_i .$$

FACT: The process with exponential holding times is a Markov chain.

This requires that the holding times have exponential distribution.

b/c "memoryless property" of exponential.

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.

$$\frac{dP(t)}{dt} = G\,P(t) = P(t)\,G$$

where G is a constant matrix.

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*

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

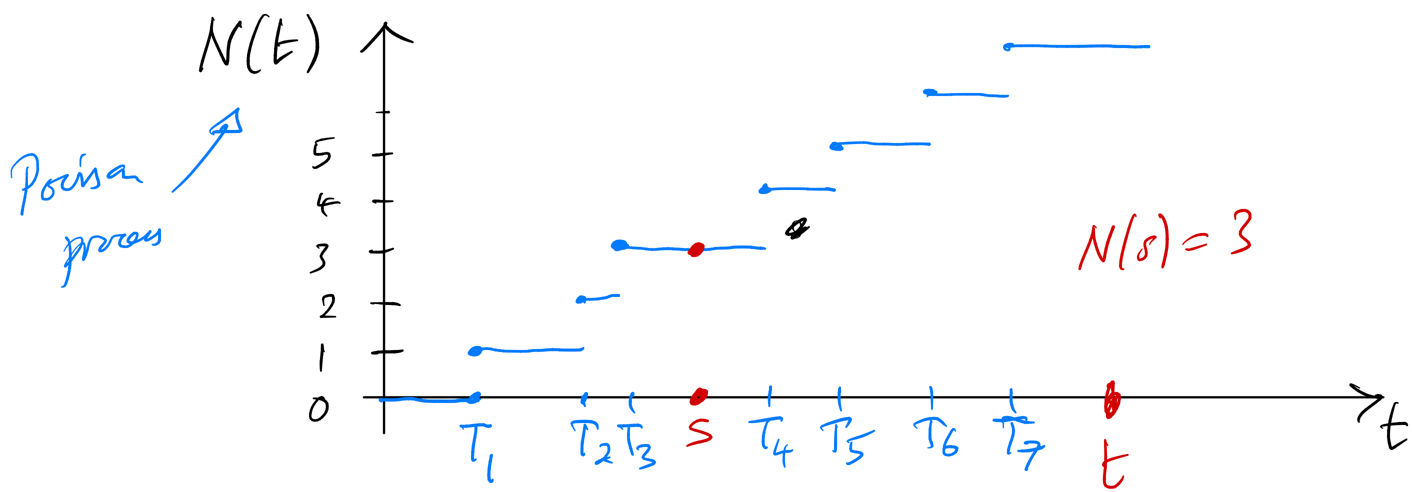
- *The arrival times $0 = T_0 < T_1 < T_2 < \dots$ have gamma distribution.*
- *If $s < t$ then $N(t) - N(s)$ is Poisson with rate $\lambda(t - s)$, and is independent of the arrival times during the interval $[0, s]$.*

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.

States of chain: $0 \to 1 \to 2 \to 3 \to 4 \to \dots$

Jumps $k \to k+1$, $k = 0, 1, 2, \dots$

Holding times are IID exponential with same rate λ.



T_1 = first arrival time

T_2 = second arrival time

$\vdots$ $\qquad\qquad\vdots$

T_k = k^{th} arrival time.

$\vdots$

$N(t)$ = number of arrivals up to time t

= number of arrivals in $[0, t]$.

$N(0) = 0$.

If $s < t \Rightarrow N(s) \leq N(t)$

$\Rightarrow N(t)$ is a counting process.

$N(t)$ is a Poisson process with rate λ if

i) $N(t)$ is a counting process

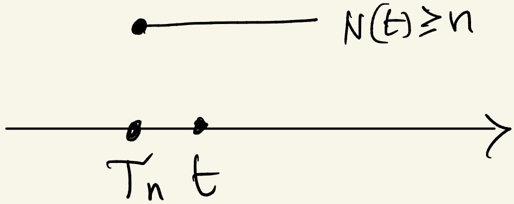
ii) all interarrival times $X_k = T_k - T_{k-1}$

are IID exponential rate λ.

Poisson distribution:

$$\mathbb{P}(N(t)=n) = \mathbb{P}(N(t)\geq n) - \mathbb{P}(N(t)\geq n+1)$$

$$= \mathbb{P}(T_n \leq t) - \mathbb{P}(T_{n+1}\leq t)$$



pdf of T_n : $\quad f_{T_n}(u) \quad$ gamma dist.

$$\Rightarrow \quad \mathbb{P}(T_n \leq t) = \int_0^t f_{T_n}(u)\,du.$$

FACT (follows from the gamma dist.)

$$\int_0^t f_{T_n}(u)\,du = \frac{1}{\lambda} f_{T_{n+1}}(t) + \int_0^t f_{T_{n+1}}(u)\,du.$$

$$\Rightarrow \mathbb{P}(N(t)=n) = \int_0^t f_{T_n}(u)\,du - \int_0^t f_{T_{n+1}}(u)\,du$$

FACT $= \frac{1}{\lambda} f_{T_{n+1}}(t)$ ← gamma distribution

$$= \frac{1}{\lambda} \lambda \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

$$= \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (n \geq 0)$$

$\Rightarrow N(t)$ is a Poisson r.v. with mean λt.

Example (Ross, p. 336, #9).

Model of cell division mistakes occur as a Poisson process with rate $\lambda = 2.5 \text{ year}^{-1}$

Organism dies when 196 errors occur.

$N(t) =$ # errors occurred up to time t.

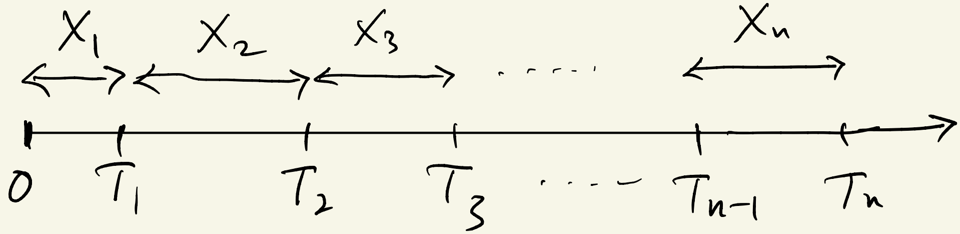
$N(t) \sim$ Poisson process rate $\lambda = 2.5 \text{ year}^{-1}$

a) Find mean lifetime of organism.

L = lifetime = T_{196}

196$^{\text{th}}$ arrival time

$$\mathbb{E}[L] = \mathbb{E}[T_{196}]$$



X_k = interarrival time IID exponential rate λ.

$$T_n = X_1 + X_2 + X_3 + \cdots + X_n.$$

$$\Rightarrow \mathbb{E}[T_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]$$

$$= n\,\mathbb{E}[X]$$

$X \sim$ exponential $\quad f_X(u) = \lambda e^{-\lambda u} \quad (u>0)$

$$\Rightarrow \mathbb{E}[X] = \int_0^\infty u \lambda e^{-\lambda u}\, du = \frac{1}{\lambda}.$$

$$\Rightarrow \mathbb{E}[T_n] = \frac{n}{\lambda}.$$

$$\Rightarrow \mathbb{E}[L] = \mathbb{E}[T_{196}] = \frac{196}{2.5}$$

$$= 78 \text{ years.}$$

b) Standard deviation of L.

$$VAR[L] = VAR[T_{196}]$$

$$= VAR[X_1 + X_2 + \cdots + X_{196}]$$

$$= 196 . VAR[X] \quad (\text{b/c } \underline{\underline{\text{IID}}})$$

$$= 196 . \frac{1}{\lambda^2}$$

$$= \frac{196}{(2.5)^2} = \left(\frac{14}{2.5}\right)^2$$

$\Rightarrow STD[L] = \sqrt{VAR[L]} = \frac{14}{2.5} = 5.6$ year.

c) Find prob. that L exceeds 90 years.

$$\mathbb{P}(L \geq 90) = \mathbb{P}(T_{196} \geq 90)$$

$$= \mathbb{P}(N(90) \leq 196)$$

(in many cases one of these is easier to calculate than the other.)

e.g. $\mathbb{P}(T_3 \geq 1.5) = \mathbb{P}(N(1.5) \leq 3)$

$$\int_{1.5}^{\infty} \lambda \frac{(\lambda u)^2}{2!} e^{-\lambda u} \, du$$

$$\mathbb{P}(N(1.5)=0) + \mathbb{P}(N(1.5)=1)$$

$$+ \mathbb{P}(N(1.5)=2)$$

$$\lambda = 2.5$$
$$t = 1.5$$

$$+ \mathbb{P}(N(1.5)=3)$$

$$= e^{-\lambda t} + \lambda t e^{-\lambda t} + \frac{(\lambda t)^2}{2!} e^{-\lambda t}$$

$$+ \frac{(\lambda t)^3}{3!} e^{-\lambda t}$$

$$\mathbb{P}(T_{196} \geq 90)$$

Since $T_{196} = \sum_{k=1}^{196} X_k$ (IID)

$\Rightarrow$ use CLT.

Pretend T_{196} is normal:

mean $\mu = \mathbb{E}[T_{196}] = \frac{196}{\lambda} = 78$

std. dev. $\sigma = \mathrm{STD}[T_{196}] = \frac{14}{2.5} = 5.6$.

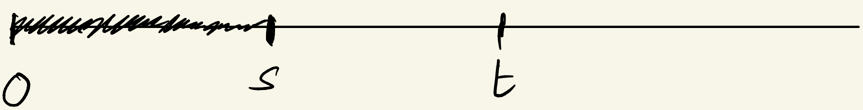
$$\Rightarrow \frac{T_{196} - \mu}{\sigma} = Z$$ approx. standard normal.

$$\Rightarrow \mathbb{P}(T_{196} \geq 90) = \mathbb{P}\left(Z \geq \frac{90-78}{5.6}\right)$$

$$> \mathbb{P}(Z \geq 2)$$

normal tables

$$= 0.01$$

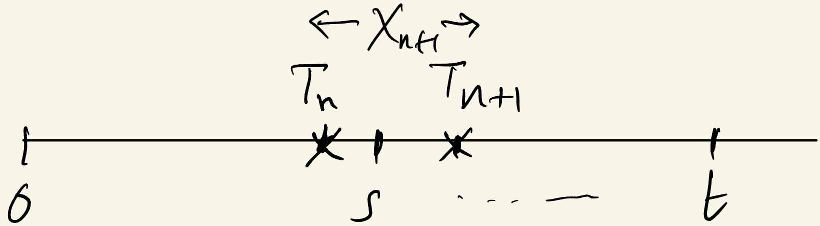


$0 < s < t$: $N(t) - N(s)$ = increment of process between s and t

= # arrivals in $[s, t]$.

$$\mathbb{P}(N(t) = n+m \mid N(s) = m)$$

$$= \mathbb{P}(N(t) - N(s) = n \mid N(s) = m)$$



$$\mathbb{P}(T_{n+1} - s \geq u \mid N(s) = n)$$

$$= \mathbb{P}(X \geq u)$$ (memoryless property of exponential)

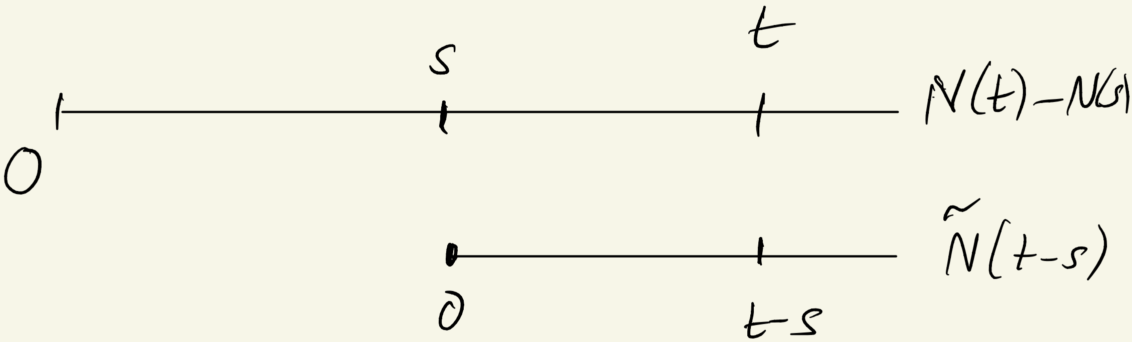
$$= e^{-\lambda u}.$$

Holding time starts over again at time s

So process starts over at time s.

$N(t) - N(s)$ = increment

$$\sim N(t-s)$$



s
t
0
N(t)−N(s)
0
t−s
$\tilde{N}(t-s)$