Continuous Time Markov Chains (CTMCs)

Overview of CTMC Lectures

- Review: Definition of a Continuous-Time Stochastic Process

  Example: frog pond
- Definition of a Continuous-Time Markov Chain (CTMC)
  - The Markov Property
  - Applications of CTMCs
  - Continuous-Time Transition Probability Function
  - The Memoryless Property of Holding Times
  - Digression: The Exponential Distribution
  - Embedded Discrete-Time Markov Chain
  - Transition Rates: Transition Diagrams & Matrix Notation

Examples: barber shop
emergency room
traffic light

Overview of CTMC Lectures, continued

- A Special Class of CTMCs: The Birth and Death Process
  - The Poisson Process
  - The M/M/1 Queue
  - The M/M/∞ Queue
  - The M/M/S Queue
- Characterizing the Long-Run Behavior of CTMCs
  - Classification of States
  - Establishing Positive Recurrence
  - Steady State Distribution
  - Application to Infinite Birth and Death Processes

Examples: barber shop
emergency room
M/M/1 queue
M/M/∞ queue
M/M/S queue

- Costs and Rewards in CTMCs

Examples: barber shop,
traffic light

Continuous-Time Stochastic Processes

In some stochastic systems, it is more natural to model time as continuous than as discrete. This may be true when we are interested in observing the system continuously over time rather than at discrete points in time. We will see that continuous and discrete time Markov Chains are very related.

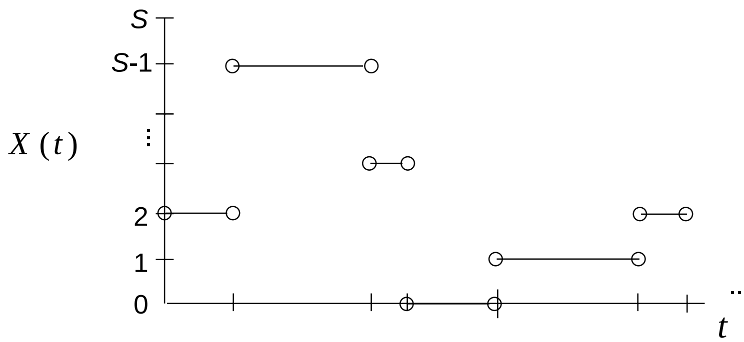
A <u>continuous-time stochastic process</u> $\{ X(t): t \in T \}$ is a collection of random variables, where

$X(t)$ is the <u>state</u> of the process at time *t.*

$T = [0, \infty)$ is the <u>index set</u> of the process

$E = \{0, 1, \ldots, S\}$ is the <u>state space</u> of the stochastic process (the set of possible values $X(t)$ can assume.)

Sample Path of a Continuous-Time Stochastic Process

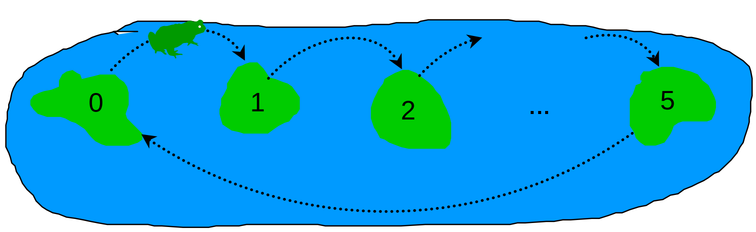


Note: the times at which transitions occur need not be integers.

Example

EXAMPLE 1: Frog Pond

Consider a frog which jumps between lilypads numbered 0 through 5. He starts at lilypad 0 and always jumps to the next highest lilypad in number until he reaches lilypad 5. From lilypad 5 he jumps back to lilypad 0. At each lilypad he waits a number of seconds which is exponentially distributed with parameter $\lambda>0$ (ie., mean $1/\lambda$ seconds) before jumping to the next lilypad.



Example

EXAMPLE 1, continued

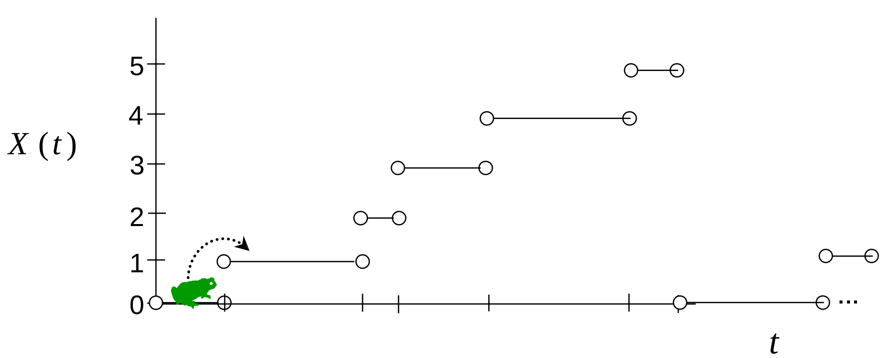
Let *X*(*t*) denote the number of the lilypad that the frog is on at time *t*.

The index set for this process is the time interval $[0,\infty)$.

The state space of the process is {0,1,2,3,4,5}.

X(*t*) is a continuous-time stochastic process.

A Sample Path for Example 1



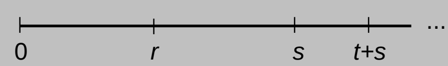
Note: *X*(*t*) is right-continuous.

Continuous-Time Markov Chains

MARKOV PROPERTY

A continuous-time stochastic process $\{ X(t) \colon t \in T \}$ is a Markov Chain if for all $r, s, t \in T$, $t \geq 0$ and $s > r \geq 0$ and for all states i, j, k,

$$P(X(t+s) = j \mid X(s) = i \text{ and } X(r) = k) = P(X(t+s) = j \mid X(s) = i)$$



$\Rightarrow$ the conditional distribution of the future state given the past states and the present state is independent of the past states and depends *only* on the present state.

STATIONARITY ASSUMPTION

$P(X(t+s) = j \mid X(s) = i)$ is independent of *s*

Example of a Continuous-Time Markov Chain

EXAMPLE 1, continued

Reconsider our frog pond. It follows from the memoryless property of the exponential distribution that at time *s*, the distribution of which lilypad the frog will be on at time $s + t$ is independent of where he was at any time before *s*, given that you know where he is at time *s*.

$\Rightarrow$ $X(t)$ is a continuous-time Markov Chain.

Applications of Continuous-Time Markov Chains

Queuing Systems

- number of customers waiting for service at the Stanford post office at each time t
- number of computer jobs (submitted from remote terminals) waiting to be executed by a mainframe at time t
- number of printed circuit boards waiting to be drilled at time t
- number of callers waiting to place an order at L.L.Bean at each time t
- number of cars waiting at the stop light at Page Mill and Peter Coutts at time t
- number of patients waiting in an emergency room at each time t
- number of people waiting to clear a security checkpoint at SFO at time t

Applications of Continuous-Time Markov Chains

Inventory Management

- number of units in stock (or backordered) at each time t

Airline Overbooking

- number of coach seats reserved at each time t

Finance

- price of a given security or option at each time t
- The US GDP at each time t

Epidemiology

- number of infected individuals at each time t

Population Growth

- size of US population at each time t

Workforce Planning

- number of employees in a firm with each level of experience at each time t

Transition Probability Function

CONTINUOUS-TIME TRANSITION PROBABILITY FUNCTION

$$p_{ij}(t) = P(X(t+s) = j \mid X(s) = i) \underset{\text{stationarity assumption}}{=} P(X(t) = j \mid X(0) = i)$$

$p_{ij}(t)$ is called the <u>continuous-time transition probability function.</u> It satisfies

$$p_{ij}(t) \geq 0 \text{ for all } i,j \in \{0,\ldots,S\} \text{ and } t \geq 0$$
$$\sum_{j=0}^{S} p_{ij}(t) = 1 \text{ for all } i \in \{0,\ldots,S\} \text{ and } t \geq 0$$

and

$$p_{ij}(0) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Note:

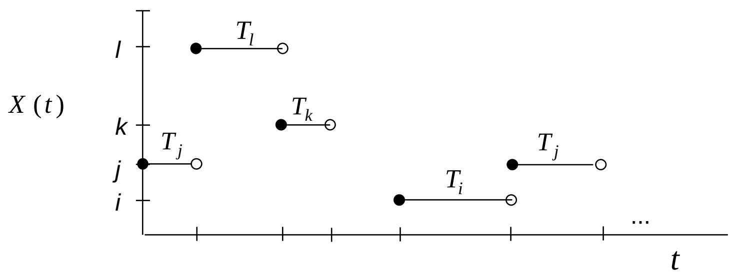
$p_{ij}(t)$ is the continuous time analogy of $p_{ij}^{(n)}$.

The continuous-time transition probability function $p_{ij}(t)$ is almost never computed. Instead a CTMC is usually characterized in terms of *transition rates*. We are going to develop the idea of transition rates shortly. In order to do so, we first need to establish some essential facts about CTMCs that follow from the Markov Property, namely:

- When the CTMC visits any state *i*, the distribution of time it remains in state *i* before moving on to another state exhibits the *memoryless property*. This is equivalent to saying that the time spent in state *i* is exponentially distributed.
- The amount of time spent in state *i* is independent of the next state visited.

Holding Times

Suppose our continuous time Markov Chain has just arrived in state *i*. Define the random variable T_i to be the length of time the process spends in state *i* before moving to a different state. We call T_i the <u>holding time</u> in state *i.*



The Memoryless Property

<u>Claim:</u>

The Markov Property implies $P(T_i > t + s \mid T_i > s) = P(T_i > t)$

➡ the distribution of how much longer you'll be in a given state *i* is independent of how long you've already been there.

<u>Proof</u> (by contradiction):

Suppose it is time *s*, you're in state *i*, and

$$P(T_i > t + s \mid T_i > s) \neq P(T_i > t)$$

i.e., the amount of time you've already been in state *i* is relevant in predicting how much longer you'll be there. Then for any time $r < s$, whether or not you were in state *i* at time *r* is relevant in predicting whether you will be in state *i* or a different state *j* at some future time *s + t.* Thus

$$P(X(t+s) = j \mid X(s) = i \text{ and } X(r) = k) \neq P(X(t+s) = j \mid X(s) = i),$$

which violates the Markov Property.

The Memoryless Property

MEMORYLESS PROPERTY

The fact that $P(T_i > t + s \mid T_i > s) = P(T_i > t)$ for each state *i* is called the memoryless property of holding times.

Fact: The only distribution satisfying the memoryless property is the exponential distribution.

EXPONENTIALLY DISTRIBUTED HOLDING TIMES

For a continuous-time Markov Chain, the holding time T_i in state *i* is exponentially distributed with parameter, say, q_i (mean $1/q_i$).

Thus $$F_i(t) = P(T_i \le t) = 1 - e^{-q_i t}$$

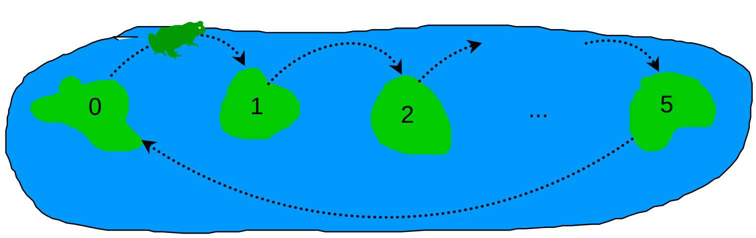
Example of non-Memoryless Holding Time

EXAMPLE 1, revisited

Suppose that whenever the frog lands on lilypad 2, he remains there for a number of seconds that is uniformly distributed between 0 and 10.

$$P(T_2 > 7 \mid T_2 > 5) = \frac{P(T_2 > 7 \text{ and } T_2 > 5)}{P(T_2 > 5)} = \frac{P(T_2 > 7)}{P(T_2 > 5)} = \frac{1 - \frac{7}{10}}{1 - \frac{5}{10}} = \frac{3}{5} \ne \frac{8}{10} = P(T_2 > 2)$$

The uniform distribution does not have the memoryless property. That is, the amount of time the frog has already been on lilypad 2 affects the distribution of how much longer he will stay there. Our example is no longer a Markov Chain.



Digression: The Exponential Distribution

A random variable *X* that is exponentially distributed with parameter λ (mean $1/\lambda$) has probability density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

and cumulative distribution function $F(x) = 1 - e^{-\lambda x}$

USEFUL FACT ABOUT THE EXPONENTIAL DISTRIBUTION

Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables, where Y_i is exponentially distributed with parameter λ_i. Then the random variable

$$Y = \min\{Y_1, Y_2, \ldots, Y_n\}$$

is exponentially distributed with parameter $\lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n$.

Independence of Holding Time and Next State

<u>Claim</u>: The Markov Property also implies that the amount of time spent in state *i* is independent of the next state visited.

<u>Proof</u> (by contradiction): Suppose you're in state *i*, and suppose the next state visited depends on the total length of time T_i that you spend in state *i*. Let's say it is a particular time *s* during your sojourn in state *i.* Then at time *s* , the information as to how long you've already been in state *i* is relevant to the prediction of the next state.

The Markov Property is violated.

$$P(X(t+s) = j | X(s) = i \text{ and } X(r) = k) \neq P(X(t+s) = j | X(s) = i)$$

Alternative Definition of a Continuous-Time Markov Chain

An equivalent definition of a continuous-time Markov Chain is that it is a continuous-time stochastic process with the property that each time it enters state *i*,

- the amount of time T_i the process spends in state *i* before making a transition into another state is exponentially distributed with mean $1/q_i$;
- when the process leaves state *i*, it enters state *j* with some probability, call it p_{ij}, where the p_{ij} satisfy:

$$p_{ii} = 0 \quad \text{and} \quad \sum_j p_{ij} = 1 \text{ for all } i.$$

The transition probabilities $\{p_{ij}\}$ are independent of the time T_i spent in state *i*.

Note: the transition probability p_{ij} is different from (although related to) the transition probability function $p_{ij}(t)$.

Thus, a continuous time Markov Chain is a stochastic process that moves from state to state in accordance with a discrete time Markov Chain, and has the property that the amount of time it spends in each state before proceeding to the next state is exponentially distributed. In addition, the amount of time spent in state *i* is independent of the next state visited.

CTMC = DTMC *(where to move)* + exponential holding times *(when to move)*

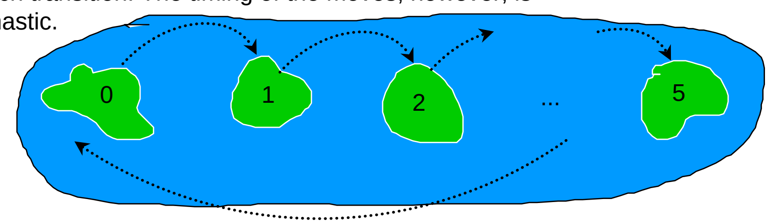
this is called the
Embedded DTMC
of our CTMC

Example

EXAMPLE 1 revisited
Now let's reconsider our original frog pond example with exponential holding times. We can now look at this stochastic process as the combination of:

- a DTMC that determines which lilypad the frog goes to each time he moves, and
- a sequence of exponentially distributed holding times that determine how long he hangs out on each lilypad.

Notice that in this example, the embedded DTMC is actually a deterministic process: we know exactly where the frog will move at each transition. The timing of the moves, however, is stochastic.



Embedded Discrete-Time Markov Chain

The embedded discrete-time Markov Chain of a continuous-time Markov Chain is the process that dictates the sequence of states visited, and is independent of the length of time spent in each state. It is defined by the transition probabilities $\{p_{ij}\}$ which satisfy

$$p_{ij} \geq 0 \quad \text{for all} \quad i, j \in E$$

$$\sum_{j=0}^{S} p_{ij} = 1 \quad \text{for all} \quad i \in E$$

$$p_{ii} = 0 \quad \text{for all} \quad i \in E$$

This condition is the only difference between an embedded DTMC and a normal DTMC. It is needed because we are modeling the holding times and the sequence of states visited seperately.

Embedded Discrete-Time Markov Chain

As before, we can write the transition probabilities for our embedded DTMC in matrix format:

TRANSITION PROBABILITY MATRIX FOR EMBEDDED DTMC

$$P = \begin{bmatrix} 0 & p_{01} & \cdots & p_{0S} \\ p_{10} & 0 & \cdots & p_{1S} \\ \vdots & \cdots & & \cdots \\ p_{S0} & p_{S1} & \cdots & 0 \end{bmatrix}$$

Notice the zeros on the diagonal.

Mean Holding Times & Transition Rates

We showed that for a continuous-time Markov Chain, the holding time T_i in state i is exponentially distributed with mean $1/q_i$.

As an alternative to thinking about mean holding times, we can think about *rates* at which the process moves from state to state.

Think of q_i as the rate at which the process leaves state *i.*

Also, for $j \neq i$ define q_{ij} to be the rate at which the process moves from state *i* to state *j*. Then q_{ij} satisfies

$$q_{ij} \equiv q_i p_{ij}, \text{ for } j \neq i$$

Also define $q_{ii} \equiv -q_i.$

We call the quantities q_i and q_{ij} the <u>transition rates</u> or <u>transition intensities</u> for state *i*.

Intuition Behind Transition Rates

Think of the transition rate q_i out of state *i* as:

- the expected number of times the process leaves state *i* per unit of time spent in state *i,* or equivalently
- the reciprocal of the expected time that the process spends in state *i* per visit to state *i,* $1/E(T_i)$

Similarly, think of the transition rate from *i* to *j*, q_{ij} , as the expected number of times the process transits from *i* to *j* per unit of time spent in *i.*

From their definition, it should be clear that $q_i = \sum_{j \neq i} q_{ij}$. Since we have defined $q_{ii} \equiv -q_i$, the above equation is equivalent to

$$\sum_{j=0}^{S} q_{ij} = 0$$

More Intuition Behind Transition Rates: Competing Exponentials

Another way to think about transition rates is as follows. Suppose you're in state *i*. Now set *S* independent kitchen timers --one for each state $j \neq i$. The random time before the *j*th kitchen timer goes off, which we will denote by T_{ij}, is exponentially distributed with parameter q_{ij} . The holding time T_i in state *i* is simply the time that the first timer goes off, which can be expressed as

$$T_i = \min_{j \neq i} \{T_{ij}\}$$

Now, if we recall our useful fact about the exponential distribution, we know that T_i is exponentially distributed with parameter

$$q_i = \sum_{j \neq i} q_{ij}$$

Absorbing States

ABSORBING STATE

State *j* is **absorbing** if once in state *j* the process will never leave state *j*, i.e., $q_{jk} = 0$ for all $k \neq j$. Naturally, for an absorbing state *j*, $q_j = -q_{jj} = 0$.

Until further notice, we will consider only CTMCs with no absorbing states (except for the following example.)

Absorbing States: Example

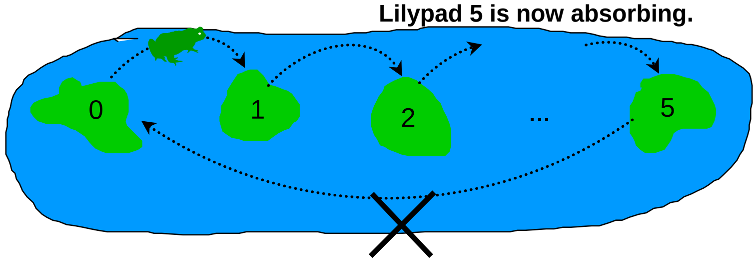
EXAMPLE 1: Frog Pond again

Suppose we are back to the original frog pond example (exponential holding times) but we now assume that once the frog lands on lilypad 5 he stays there forever. Then

$$q_{50} = q_{51} = q_{52} = q_{53} = q_{54} = 0$$

$$q_5 = -q_{55} = \sum_{j \neq 5} q_{5j} = 0$$

Lilypad 5 is now absorbing.



Matrix Notation for Transition Rates

TRANSITION RATE MATRIX

$$Q = \begin{bmatrix} q_{00} & q_{01} & \cdots & q_{0S} \\ q_{10} & q_{11} & \cdots & q_{1S} \\ \vdots & \vdots & & \vdots \\ q_{S0} & q_{S1} & \cdots & q_{SS} \end{bmatrix}$$

Transition Rates must satisfy

$$q_{ij} \geq 0 \text{ for all } i,j \in \{0,\ldots,S\},\ i \neq j$$
$$q_{ii} = -q_i \leq 0 \quad \text{for all } i \in \{0,\ldots,S\}$$
$$\sum_{j=0}^{S} q_{ij} = 0 \text{ for all } i \in \{0,\ldots,S\}$$

Q is sometimes called the infinitesimal generator of the CTMC.

Note: This *Q* is <u>not related</u> to the initial probability distribution vector *Q* in DTMCs.

Example of a Continuous-Time Markov Chain

EXAMPLE 2: Barber Shop

Consider a barber shop with 2 barbers (each with a chair) and 2 additional waiting chairs. Customers arriving to a fully occupied shop leave without being served (they *balk*). When the shop opens at 8am, there are already 2 waiting customers. Assume that the time between arrivals of consecutive customers is exponentially distributed with a mean of 12 minutes. Each barber's haircut times are also exponentially distributed with a mean of 30 minutes. Also, the customer arrival process is independent of the haircut times.

Example: Formulation as a CTMC

EXAMPLE 2, continued.

Let $X(t)$ be the number of customers in the barber shop at time for $t \geq 0$.

The index set for this process is the time interval $[0,\infty)$.

The state space of the process is {0,1,2,3,4}.

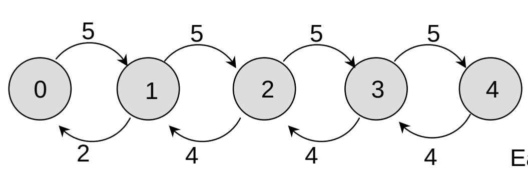
At time *s*, the distribution of the number of customers in the barber shop at some future time $s + t$ is independent of the number of customers in the shop at any time before *s*, given that you know the number of customers at time *s*.

➡ $X(t)$ is a continuous-time Markov Chain.

Transition Rate Diagram

Just as we used transition diagrams for discrete-time Markov Chains, they are useful for continuous-time Markov Chains. Here, we use it to depict the transition rates. In doing so, we show the rates q_{ij} for $j \neq i$ and omit the rates $q_{ii} = -q_i$ since the latter are easily computed from the former.

TRANSITION RATE DIAGRAM FOR EXAMPLE 2



5 people arrive per hour as long as the number in the shop is less than 4.

Each barber gives haircuts at a rate of 2 per hour. The rate of completion of haircuts is
2 x (# of busy barbers)

We assume that at most one arrival or departure can occur at any one point in time.

Relationship between Transition Rates and Transition Probabilities

Note that

$$q_{ij} \equiv q_i p_{ij}, \text{ for } j \neq i$$

Thus, for any non-absorbing state *i* we can write the transition probabilities of the embedded discrete-time Markov Chain as

$$p_{ij} = \begin{cases} q_{ij} / q_i & \text{for } j \neq i \\ 0 \text{ for } j = i \end{cases}$$

$$= \begin{cases} -q_{ij} / q_{ii} & \text{for } j \neq i \\ 0 \text{ for } j = i \end{cases}$$

This shows that transition probabilities of the embedded DTMC can be derived directly from the transition rates.

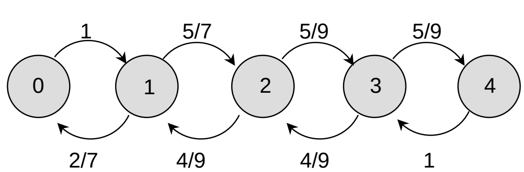
Transition Probability Diagram

You can also use the transition diagram to depict the transition probabilities of the embedded DTMC. To do this, recall:

$$p_{ij} = \begin{cases} q_{ij} / q_i & \text{for } j \neq i \\ 0 \text{ for } j = i \end{cases} = \begin{cases} -q_{ij} / q_{ii} & \text{for } j \neq i \\ 0 \text{ for } j = i \end{cases}$$

Furthermore, $\sum_{j=0}^{S} q_{ij} = 0$ so $q_{ii} = -\sum_{j \neq i} q_{ij}$.

TRANSITION PROBABILITY DIAGRAM FOR EXAMPLE 2



Example: Transition Rate and Probability Matrices

TRANSITION RATE MATRIX FOR EXAMPLE 2

$$Q = \begin{bmatrix} -5 & 5 & 0 & 0 & 0 \\ 2 & -7 & 5 & 0 & 0 \\ 0 & 4 & -9 & 5 & 0 \\ 0 & 0 & 4 & -9 & 5 \\ 0 & 0 & 0 & 4 & -4 \end{bmatrix}$$

TRANSITION PROBABILITY MATRIX FOR EXAMPLE 2

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 2/7 & 0 & 5/7 & 0 & 0 \\ 0 & 4/9 & 0 & 5/9 & 0 \\ 0 & 0 & 4/9 & 0 & 5/9 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Example

EXAMPLE 3: Emergency Room

Consider a small emergency room that has 2 operating tables and 2 beds for holding waiting patients. The interarrival times of ambulances follow an exponential distribution with mean of 30 minutes. A given ambulance may bring into the emergency room 1, 2 or 3 patients with probabilities $p_1 = 0.7,\ p_2 = 0.2,\ p_3 = 0.1$

Patients who arrive and find no bed to wait in are taken to a neighboring hospital.

A patient's length of stay on an operating table follows an exponential distribution with mean 2.5 hours. Operation durations are independent of each other and the arrival process.

Example

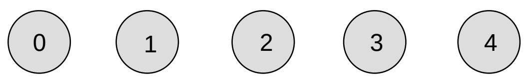
EXAMPLE 3, continued.

Let $X(t)$ be the number of patients in the emergency room at time $t \geq 0$.

The index set for this process is the time interval $[0,\infty)$.

The state space of the process is {0,1,2,3,4}.

Clearly the process is a continuous-time Markov Chain. What are the transition rates?

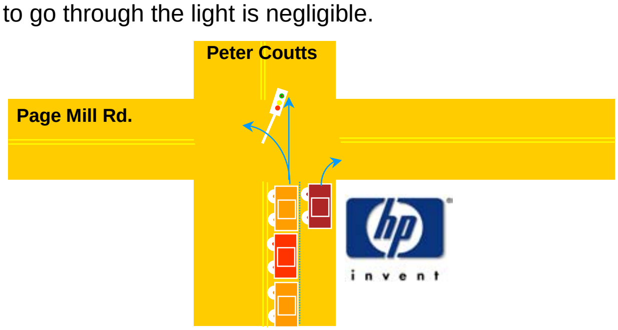


In this example (unlike example 2), there can be up to 3 people arriving simultaneously. However, at most one person departs in any instant.

Example

EXAMPLE 4: Traffic Light

Cars line up to leave the HP Labs parking lot at the intersection of Peter Coutts and Page Mill Roads, with exponentially distributed interarrival times having mean 1 minute. An arriving car will queue up in the left lane with probability p_L = 0.6; it will get in the right turn lane instead with probability p_R = 0.4. The times between the light turning green are exponentially distributed with mean 5 minutes. When the light turns green, it stays green until 5 cars go through the intersection (if there are that many waiting.) Assume the time required to go through the light is negligible.



Example

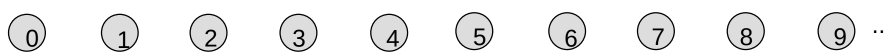
EXAMPLE 4, continued.

Let *X*(*t*) be the number of cars waiting *in the left lane* to leave the parking lot at time $t \geq 0$.

The index set for this process is the time interval $[0,\infty)$.

The state space of the process is {0,1,2,….}.

This process is a continuous-time Markov Chain. What are the transition rates?



Special Case: Birth and Death Processes

BIRTH AND DEATH PROCESS

A continuous-time Markov Chain $X(t),\ t \geq 0$ is called a birth and death process if the state space *E* = {0,1,2,…,*N*} or {0,1,2,….} and

$$q_{ij} = 0 \quad \text{if} \quad |j - i| > 1$$

A birth and death process is a special case of a CTMC in which transitions from state *i* can only go to either state *i* - 1 or state *i* + 1.

For the infinite state space case we can write the transition rates of a birth and death process as:

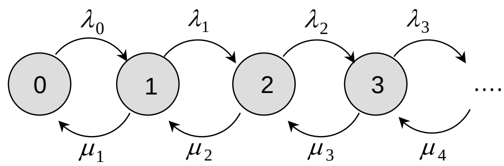
$q_{i,i+1} = \lambda_i \ , \ \ q_{i,i-1} = \mu_i \ , \ \ q_i = -q_{ii} = \mu_i + \lambda_i$ and $q_{ij} = 0 \ \text{ if } \ |j - i| > 1$

The parameters $\{\lambda_i\}$ are called birth rates and $\{\mu_i\}$ are called death rates.

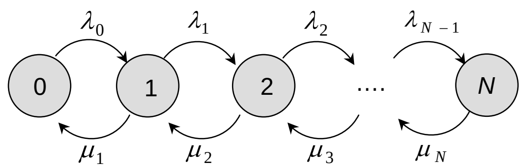
Birth and Death Processes

TRANSITION RATE DIAGRAM FOR A BIRTH AND DEATH PROCESS

(infinite State Space)



(Finite State Space)



Birth and Death Processes

From the transition rates we can compute the transition probabilities of the embedded DTMC of a birth and death process using the formulas:

$$p_{ij} = \begin{cases} q_{ij}/q_i & \text{for } j \neq i \\ 0 \text{ for } j = i \end{cases} = \begin{cases} -q_{ij}/q_{ii} & \text{for } j \neq i \\ 0 \text{ for } j = i \end{cases}$$

We obtain the following transition probabilities for 1<*i*<*N*:

$$p_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_\mathrm{i}},$$

$$p_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_\mathrm{i}},$$ and $$p_{ij} = 0 \quad \text{for } j \notin \{i-1, i+1\}$$

plus the end-state conditions: $p_{01} = 1$ and $p_{N,N-1} = 1 \quad (\text{if } N < \infty)$

Examples of Birth and Death Processes: The Poisson Process

EXAMPLE 5: The Poisson Process

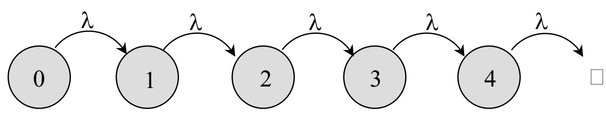
A birth and death process is a Poisson process with rate λ if it has transition rates

$$q_{i,i+1} = \lambda_i = \lambda \quad \text{for all } i \geq 0$$ (birth rates are independent of *i*)

$$q_{i,i-1} = \mu_i = 0 \quad \text{for all } i \geq 1$$ (no deaths occur)

$$q_i = -q_{ii} = \lambda \quad \text{for all } i \geq 0$$

$$q_{ij} = 0 \quad \text{if } j \notin \{i, i+1\}$$



The Poisson process is an example of a *pure birth process*. It is also often called a counting process.

Examples of Birth and Death Processes: The Poisson Process

EXAMPLE 5, continued.

Why is this CTMC called a Poisson process? Recall: A discrete random variable N has a Poisson distribution with α if, for n = 0,1,2,…..

$$P(N = n) = f(n) = \frac{\alpha^n e^{-\alpha}}{n!}$$

RELATIONSHIP BETWEEN THE EXPONENTIAL AND POISSON DISTRIBUTIONS

Suppose we observe a sequence of events occuring at random times starting at time 0. Then the length of time between consecutive events is exponentially distributed with parameter λ if and only if the number $N(t)$ of events that occur in the interval $[0, t]$ has a Poisson distribution with parameter λt.

Examples of Birth and Death Processes: The M/M/1 Queue

EXAMPLE 6: M/M/1 QUEUING SYSTEM

In an M/M/1 queuing system, we have a single server in a service system with an unlimited number of spaces for people to wait. Customers arrive according to a Poisson process with rate λ (or equivalently, their interarrival times are independent exponential random variables with parameter λ). Upon arrival, each customer goes directly to the server if the queue is empty, or joins the end of the queue otherwise. Service times are independent and exponentially distributed with rate μ. Once the server is finished serving a customer, that customer leaves the system and the next customer in line (if any) enters service.

The number of people in the system (in queue plus being served) can be modeled as a birth and death process.

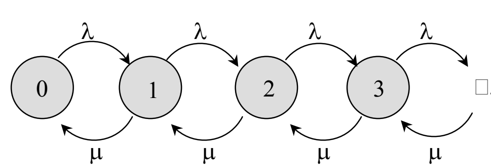
Note: In the notation M/M/1, the first M refers to the fact that the arrival process is Markovian, the second M refers to the service process being Markovian, and the 1 refers to the number of servers.

The M/M/1 Queue, continued

$$\lambda_i = \lambda \quad \text{for all } i \geq 0$$

$$\mu_i = \mu \quad \text{for all } i \geq 1$$

TRANSITION RATE DIAGRAM:



TRANSITION PROBABILITIES FOR THE M/M/1 QUEUE:

$$p_{i,i+1} = \frac{\lambda}{\lambda + \mu}, \quad p_{i,i-1} = \frac{\mu}{\lambda + \mu} \qquad p_{ij} = 0 \quad \text{for } j \notin \{i-1, i+1\}$$

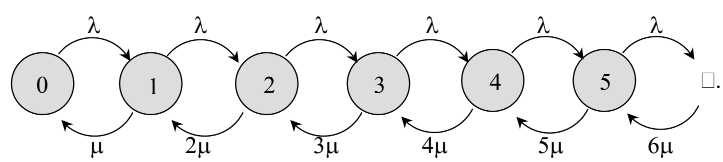
except for the i = 0 case, in which $p_{01} = 1,\ p_{0j} = 0 \text{ for } j \neq 1$

Examples of Birth and Death Processes: The M/M/∞ Queue

EXAMPLE 7: M/M/∞ QUEUING SYSTEM

Now consider a service system with *infinitely many* servers. Assume again that arrivals follow a Poisson process with rate λ and that all servers' service times are independent and exponentially distributed with rate μ.

TRANSITION RATE DIAGRAM

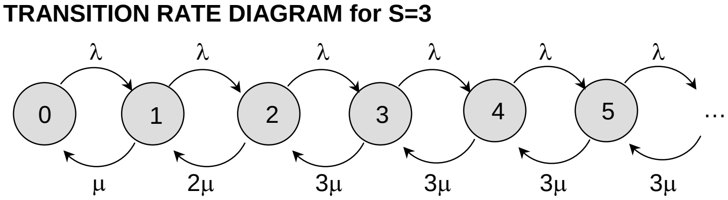


Examples of Birth and Death Processes: The M/M/S Queue

EXAMPLE 8: M/M/S QUEUING SYSTEM

Now suppose instead that we have S servers. Assume again that arrivals follow a Poisson process with rate λ and that all servers' service times are independent and exponentially distributed with rate μ. An entering customer waits in line for the first available server.

TRANSITION RATE DIAGRAM for S=3



Long Run Behavior of CTMCs

As with DTMCs, we are particularly interested in knowing how a CTMC behaves in the long run. In the discrete time case, we know that under certain conditions

$$\pi_j = \lim_{n \to \infty} p_{ij}^{(n)} \quad \text{exists and is independent of } i.$$

Thus, in the long run, the probability of being in state *j* is independent of which state the process started out in.

It turns out that for continuous-time Markov Chains as well, under certain conditions

$$\omega_j = \lim_{t \to \infty} p_{ij}(t) \quad \text{exists and is independent of } i.$$

Analogous to π_j for DTMCs, in a CTMC ω_j represents the probability of being in state *j* in the long run. We will establish conditions under which the values $\{\omega_j\}$ exist and how to compute them in the continuous time case. To do so, we will first discuss how to classify states of a continuous-time Markov Chain.

Classification of States: Overview

It turns out that much of the classification of states of a CTMC depends on the classification of states in its embedded DTMC. Therefore, we can use what we have already learned about classifying states of a DTMC. The only concepts that do not port directly from DTMC state classification are:

- periodicity: there is no concept of periodicity in CTMCs.
- positive and null recurrence: we develop a method to determine whether a state is positive or null recurrent

Classification of States: Summary

The communicating classes of a CTMC are the communicating classes of its embedded DTMC.

A CTMC is irreducible if and only if its embedded DTMC is irreducible.

Periodicity does not arise in CTMCs.

A state in a CTMC is recurrent (transient) if and only if it is recurrent (transient) in the embedded DTMC.

Facts about Positive and Null Recurrence: Same as DTMCs

A finite-state CTMC cannot have null recurrent states. Thus all recurrent states in a finite state CTMC are positive recurrent.

States in the same communicating class are either all transient, all positive recurrent or all null recurrent.

All states are positive recurrent in an irreducible finite state Markov Chain.

We establish positive and null recurrence differently in CTMCs and DTMCs.

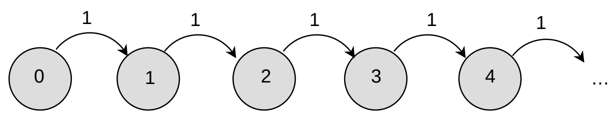
Classification of States: Examples

EXAMPLE 5: THE POISSON PROCESS

$$p_{i,i+1} = 1 \text{ for all } i \geq 0$$

$$p_{ij} = 0 \quad \text{if} \quad j \neq i+1$$

TRANSITION PROBABILITY DIAGRAM FOR THE POISSON PROCESS:



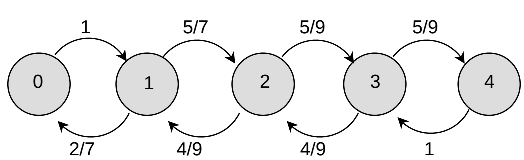
Communicating classes: {0}, {1}, {2}, {3}, …..
Thus, the chain is not irreducible. Moreover, each state is transient.

Note: you can use either the transition rate diagram or the transition probability diagram to depict positive transition probabilities. Wherever there is an arc in the transition probability diagram, there will be one in the transition rate diagram.

Classification of States: Examples

EXAMPLE 2: THE BARBERSHOP

TRANSITION PROBABILITY DIAGRAM FOR EXAMPLE 2

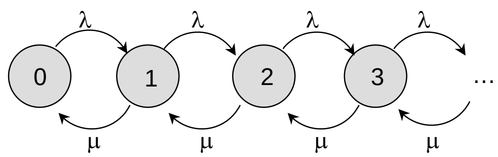


All states communicate. The chain is irreducible. Since it is finite, all states are positive recurrent.

Classification of States: Examples

EXAMPLE 6: THE M/M/1 QUEUE

TRANSITION RATE DIAGRAM:



TRANSITION PROBABILITIES FOR THE M/M/1 QUEUE:

$$p_{i,i+1} = \frac{\lambda}{\lambda + \mu}, \quad p_{i,i-1} = \frac{\mu}{\lambda + \mu} \qquad p_{ij} = 0 \quad \text{for } j \notin \{i-1, i+1\}$$

except for the *i* = 0 case, in which $p_{01} = 1,\ p_{0j} = 0 \text{ for } j \neq 1$

All states communicate if and only if λ>0 and μ>0. In that case, the chain is irreducible. Therefore, either all states are recurrent or all states are transient. Since the chain is not finite, we must check to see whether it is recurrent, and if so whether it's positive recurrent.

Establishing Positive Recurrence: Continuous-Time Case

Unfortunately, in the case of infinite-state continuous time Markov Chains it is not easy to see whether recurrence and positive recurrence holds. (Recall this was true for DTMCs as well.) The following theorem gives us a way to test for positive recurrence in irreducible continuous-time chains. Naturally if we establish positive recurrence, then recurrence follows. Once we have used this theorem to establish positive recurrence, it turns out we have already solved for the stationary probability distribution.

Establishing Positive Recurrence: Theorem for CTMCs

Theorem 1

Suppose $X(t),\ t \geq 0$ is an irreducible continuous-time Markov Chain. Then $X(t)$ is positive recurrent if and only if there exists a unique nonnegative solution $\omega = (\omega_0, \omega_1, \ldots, \omega_S)$

to the system:

$$\sum_{i=0}^{S} \omega_i q_{ij} = 0 \quad \text{for } j = 0, \ldots, S$$

$$\sum_{i=0}^{S} \omega_i = 1$$

Note: if a nonnegative solution to the above system exists, it must be unique.

Remember that if you're testing for positive recurrence in a finite chain, you don't need this theorem. All recurrent states are automatically positive recurrent in that case.

Establishing Positive Recurrence: Example

EXAMPLE 6: THE M/M/1 QUEUE, continued
We apply the theorem to determine whether the M/M/1 queue is positive recurrent. The transition rate matrix looks like

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & . & . & . \\ \mu & -(\lambda+\mu) & \lambda & . & . & . \\ 0 & \mu & -(\lambda+\mu) & \lambda & . & . \\ 0 & 0 & \mu & -(\lambda+\mu) & \lambda & . \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

We seek a nonnegative solution $\omega = (\omega_0, \omega_1, \omega_2, \ldots.)$ to the system

$$\sum_{i=0}^{\infty} \omega_i q_{ij} = 0 \quad \text{for } j = 0, 1, 2, \ldots..$$

$$\sum_{i=0}^{\infty} \omega_i = 1$$

Establishing Positive Recurrence: Example

EXAMPLE 6, continued

For this example, the first system of equations can be written as

$$-\lambda\omega_0 + \mu\omega_1 = 0$$
$$\lambda\omega_0 - (\lambda+\mu)\omega_1 + \mu\omega_2 = 0$$
$$\lambda\omega_1 - (\lambda+\mu)\omega_2 + \mu\omega_3 = 0$$
$$\vdots$$
$$\lambda\omega_{j-1} - (\lambda+\mu)\omega_j + \mu\omega_{j+1} = 0 \text{ for } j = 1,2,3,\ldots.$$

By adding each equation to the one above it, we obtain

$$\lambda\omega_0 = \mu\omega_1$$
$$\lambda\omega_1 = \mu\omega_2$$
$$\lambda\omega_2 = \mu\omega_3$$
$$\vdots$$
$$\lambda\omega_j = \mu\omega_{j+1} \text{ for } j = 0,1,2,3,\ldots.$$

Establishing Positive Recurrence: Example

EXAMPLE 6, continued

Solving each in terms of ω_0 gives us: $\omega_1 = \frac{\lambda}{\mu}\omega_0$

$$\omega_2 = \frac{\lambda}{\mu}\omega_1 = \left(\frac{\lambda}{\mu}\right)^2\omega_0$$
$$\vdots$$
$$\omega_{j+1} = \frac{\lambda}{\mu}\omega_j = \left(\frac{\lambda}{\mu}\right)^{j+1}\omega_0$$

Now, using the equation $\sum_{i=0}^{\infty}\omega_i = 1$ we have $\sum_{i=0}^{\infty}\omega_i = \omega_0\sum_{i=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^i = 1$

case 1: $\lambda < \mu$

If $\lambda < \mu$, the sum $\sum_{i=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^i$ converges to $\frac{1}{1-\frac{\lambda}{\mu}} = \frac{\mu}{\mu-\lambda}$. In this case,

$$\omega_0 = \frac{1}{\sum_{i=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^i} = \frac{\mu-\lambda}{\mu} = 1 - \frac{\lambda}{\mu} \quad \text{and} \quad \omega_j = \left(\frac{\lambda}{\mu}\right)^j\omega_0 = \left(\frac{\lambda}{\mu}\right)^j - \left(\frac{\lambda}{\mu}\right)^{j+1}$$

We have found a nonegative solution, so the queue is positive recurrent.

Establishing Positive Recurrence: Example

EXAMPLE 6, continued

case 2: $\lambda \geq \mu$

In this case, the sum $\sum_{i=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^i$ diverges and therefore

$$\omega_0 = 0 \quad \text{and} \quad \omega_j = 0 \text{ for all } j = 1,2,\ldots$$

But then $\sum_{i=0}^{\infty} \omega_i \neq 1$ so our system does not have a solution.

Thus in case 2 the queue is not positive recurrent.

We have shown that the M/M/1 queue is positive recurrent if and only if $\lambda < \mu$.

Intuition: the service rate must exceed the arrival rate of customers… otherwise the queue will become infinitely long in the long run. In this case it is not positive recurrent, and there is no steady state distribution!

It turns out that the conditions in Theorem 1 to characterize positive recurrence also provide a way to compute the steady state distribution for a CTMC. That is, if have an irreducible CTMC and we find a nonegative solution $\omega = (\omega_0, \omega_1, \ldots, \omega_S)$ to the system

$$\sum_{i=0}^{S} \omega_i q_{ij} = 0 \quad \text{for} \quad j = 0,\ldots,S$$

$$\sum_{i=0}^{S} \omega_i = 1$$

then ω is the steady state distribution for the CTMC. This is the point of the next theorem. It is analogous to our big theorem for DTMCs.

Big Theorem for Continuous-Time Markov Chains

Theorem 2

If $X(t),\ t \geq 0$ is an irreducible, positive recurrent continuous-time Markov Chain, then there exists a unique nonnegative solution $\omega = (\omega_0, \omega_1, \ldots, \omega_S)$ to the system:

$$\sum_{i=0}^{S} \omega_i q_{ij} = 0 \quad \text{for} \quad j = 0, \ldots, S$$

$$\sum_{i=0}^{S} \omega_i = 1$$

Moreover, for $j = 0,\ldots,S$, $\omega_j = \lim_{t \to \infty} p_{ij}(t)$

Steady State Probabilities

$\omega = (\omega_0, \omega_1, \ldots, \omega_S)$ is called the vector of **steady state probabilities.**

Intuition for Big Theorem

$$\lim_{t \to \infty} P(t) = \begin{bmatrix} \omega_0 & \omega_1 & \cdots & \omega_S \\ \omega_0 & \omega_1 & \cdots & \omega_S \\ \vdots & \vdots & & \vdots \\ \omega_0 & \omega_1 & \cdots & \omega_S \end{bmatrix}$$

The big theorem is analogous to our big theorem for DTMCs. The intuition behind it is that if the CTMC is irreducible and positive recurrent, the probability of being in state *j* after a long time is independent of which state you started out in.

Thus ω_j is the probability of being in state *j* in the long run. It can also be interpreted as the long run proportion of time that the Markov Chain will be in state *j*.

Steady State Equations

Steady State Equations

The S+2 equations $\sum_{i=0}^{S} \omega_i q_{ij} = 0 \quad \text{for} \quad j = 0,\ldots,S$

$$\sum_{i=0}^{S} \omega_i = 1$$

are called the **steady state equations.**

Steady State Equations: Rate In = Rate Out

To understand the steady-state equations for CTMCs, let's take a closer look:

$$\sum_{i=0}^{S} \omega_i q_{ij} = 0 \quad \text{for} \quad j = 0,\ldots,S$$

$$\sum_{i=0}^{S} \omega_i = 1$$

Recall that for every state i, $q_{ii} = -q_i$. Furthermore, we know $q_{ij} = q_i p_{ij}$.

Using these facts, we can obtain an equivalent statement of the steady-state equations:

$$\underset{\uparrow}{q_j}\omega_j = \sum_{i \neq j} \omega_i q_{ij} = \sum_{i \neq j} \underline{q_i \omega_i p_{ij}} \quad \text{for} \quad j = 0,\ldots,S \quad \text{and} \quad \sum_{i=0}^{S} \omega_i = 1$$

average rate of leaving state *j*

average rate of entering state *j* from state *i*

Thus, the steady state equations require that for each state j

the average rate out of state j = the average total rate into state j

Example: Steady State Equations

EXAMPLE 2: THE BARBERSHOP, revisited

Our barbershop example was finite and positive recurrent. Therefore we know (from the big theorem) that a steady state distribution

$$\omega = (\omega_0, \omega_1, \ldots, \omega_S)$$

exists. To find it we need to solve the steady state equations:

$$\sum_{i=0}^{S} \omega_i q_{ij} = 0 \quad \text{for} \quad j = 0, \ldots, S$$

$$\sum_{i=0}^{S} \omega_i = 1$$

Recall the transition rate matrix for this example was:

$$Q = \begin{bmatrix} -5 & 5 & 0 & 0 & 0 \\ 2 & -7 & 5 & 0 & 0 \\ 0 & 4 & -9 & 5 & 0 \\ 0 & 0 & 4 & -9 & 5 \\ 0 & 0 & 0 & 4 & -4 \end{bmatrix}$$

Example: Steady State Equations

EXAMPLE 2, continued

The steady state equations for this example are

$$\begin{aligned} -5\omega_0 + 2\omega_1 &= 0 \\ 5\omega_0 - 7\omega_1 + 4\omega_2 &= 0 \\ 5\omega_1 - 9\omega_2 + 4\omega_3 &= 0 \\ 5\omega_2 - 9\omega_3 + 4\omega_4 &= 0 \\ 5\omega_3 - 4\omega_4 &= 0 \end{aligned} \qquad \text{and} \qquad \sum_{i=0}^{S} \omega_i = 1$$

The first equation gives us: $\omega_1 = \frac{5}{2}\omega_0$ Then adding the first two equations and solving for ω_2, we get: $\omega_2 = \frac{5}{4}\omega_1$ Adding the first three equations and solving for ω_3 gives us: $\omega_3 = \frac{5}{4}\omega_2$ Adding the first four equations and solving for ω_4, we get $\omega_4 = \frac{5}{4}\omega_3$. (The fifth equation is redundant.) Now using $\sum_{i=0}^{S} \omega_i = 1$ to solve these equations gives us:

$$\omega = \left(\frac{128}{1973}, \frac{320}{1973}, \frac{400}{1973}, \frac{500}{1973}, \frac{625}{1973}\right)$$

Example: Steady State Equations

EXAMPLE 3: THE EMERGENCY ROOM

Recall the emergency room example. Its transition rate matrix is

$$Q = \begin{bmatrix} -2 & 1.4 & 0.4 & 0.2 & 0 \\ 0.4 & -2.4 & 1.4 & 0.4 & 0.2 \\ 0 & 0.8 & -2.8 & 1.4 & 0.6 \\ 0 & 0 & 0.8 & -2.8 & 2 \\ 0 & 0 & 0 & 0.8 & -0.8 \end{bmatrix}$$

Since this example is also finite and irreducible, it is positive recurrent. Thus, again a steady state distribution

$$\omega = (\omega_0, \omega_1, ..., \omega_S)$$

exists and is obtained from the steady state equations, given by

$$\sum_{i=0}^{S} \omega_i = 1$$

and

$$\begin{aligned} -2\omega_0 + 0.4\omega_1 &= 0 \\ 1.4\omega_0 - 2.4\omega_1 + 0.8\omega_2 &= 0 \\ 0.4\omega_0 + 1.4\omega_1 - 2.8\omega_2 + 0.8\omega_3 &= 0 \\ 0.2\omega_0 + 0.4\omega_1 + 1.4\omega_2 - 2.8\omega_3 + 0.8\omega_4 &= 0 \\ 0.2\omega_1 + 0.6\omega_2 + 2\omega_3 - 0.8\omega_4 &= 0 \end{aligned}$$

Verify that the solution is: $\omega = (\frac{4}{638}, \frac{20}{638}, \frac{58}{638}, \frac{145}{638}, \frac{411}{638})$

Steady State Distribution for Infinite Birth and Death Processes

We already found conditions under which an M/M/1 queue is positive recurrent and solved for its stationary distribution:

$$\omega_0 = \frac{1}{\sum_{i=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^i} = \frac{\mu - \lambda}{\mu} = 1 - \frac{\lambda}{\mu} \qquad \omega_j = \left(\frac{\lambda}{\mu}\right)^j \omega_0 = \left(\frac{\lambda}{\mu}\right)^j - \left(\frac{\lambda}{\mu}\right)^{j+1}$$

We now show how to verify whether a general infinite birth and death process (of which an M/M/1 queue is an example) is positive recurrent, and in doing so, solve for the steady state distribution (if one exists.)

Steady State Distribution for Infinite Birth and Death Processes

The transition matrix for the infinite birth and death process is:

$$Q = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & . & . & . \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & . & . & . \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & . & . \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \lambda_3 & . \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

We seek a nonnegative solution $\omega = (\omega_0, \omega_1, ..., \omega_S)$ to the system

$$\sum_{i=0}^{S} \omega_i q_{ij} = 0 \quad \text{for} \quad j = 0,\ldots,S$$

$$\sum_{i=0}^{S} \omega_i = 1$$

Assume nonzero birth and death rates.

Steady State Distribution for Infinite Birth and Death Processes

For the birth and death process, the first system of equations can be written as

$$\begin{aligned} &\mu_1\omega_1 - \lambda_0\omega_0 = 0 \\ &\lambda_0\omega_0 - (\lambda_1+\mu_1)\omega_1 + \mu_2\omega_2 = 0 \\ &\lambda_1\omega_1 - (\lambda_2+\mu_2)\omega_2 + \mu_3\omega_3 = 0 \\ &\vdots \\ &\lambda_{j-1}\omega_{j-1} - (\lambda_j+\mu_j)\omega_j + \mu_{j+1}\omega_{j+1} = 0 \quad \text{for} \quad j = 1,2,3,\ldots. \end{aligned}$$

By adding each equation to the one above it, we obtain

$$\begin{aligned} &\lambda_0\omega_0 = \mu_1\omega_1 \\ &\lambda_1\omega_1 = \mu_2\omega_2 \\ &\lambda_2\omega_2 = \mu_3\omega_3 \\ &\vdots \\ &\lambda_j\omega_j = \mu_{j+1}\omega_{j+1} \quad \text{for} \quad j = 0,1,2,3,\ldots. \end{aligned}$$

Steady State Distribution for Infinite Birth and Death Processes

Solving each in terms of ω_0 gives us:

$$\omega_1 = \frac{\lambda_0}{\mu_1}\omega_0$$
$$\omega_2 = \frac{\lambda_1}{\mu_2}\omega_1 = \frac{\lambda_0\lambda_1}{\mu_1\mu_2}\omega_0$$
$$\vdots$$
$$\omega_j = \frac{\lambda_{j-1}}{\mu_j}\omega_{j-1} = \frac{\lambda_0\lambda_1\cdots\lambda_{j-1}}{\mu_1\mu_2\cdots\mu_j}\omega_0 \quad \text{for} \quad j = 1, 2, \ldots$$

Now, using the equation $\sum_{i=0}^{\infty}\omega_i = 1$ we have

$$\sum_{i=0}^{\infty}\omega_i = \omega_0 + \omega_0\sum_{i=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i} = 1$$

Steady State Distribution for Infinite Birth and Death Processes

Thus, w must satisfy

$$\omega_0 = \left(1 + \sum_{i=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i}\right)^{-1}$$

$$\omega_j = \frac{\lambda_0\lambda_1\cdots\lambda_{j-1}}{\mu_1\mu_2\cdots\mu_j}\left(1 + \sum_{i=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i}\right)^{-1}$$

From these expressions, it is clear that the birth and death process is positive recurrent if and only if

$$\sum_{i=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i} < \infty$$

and its stationary distribution ω is given by the equations above.

Verify that the results we obtained for the M/M/1 queue follow from the above results.

Theorem: Steady State Distribution for Infinite Birth and Death Processes

What we have just shown is:

Theorem 3

An infinite birth and death process with birth and death rates $\{\lambda_i\}$ and $\{\mu_i\}$ is positive recurrent and has a steady state distribution given by

$$\omega_0 = \left(1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}\right)^{-1}$$

$$\omega_j = \frac{\lambda_0 \lambda_1 \cdots \lambda_{j-1}}{\mu_1 \mu_2 \cdots \mu_j}\left(1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}\right)^{-1} \quad \text{for } j = 1, 2, \ldots$$

if and only if $\displaystyle\sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} < \infty$

Note: this theorem is not needed for finite birth and death processes, since they are automatically positive recurrent if the birth and death rates are positive.

Implications for the M/M/∞ Queue

Let's apply this result to determine when the M/M/∞ and M/M/S queues have steady state distributions.

EXAMPLE 7: M/M/∞ Queue

The M/M/∞ queue is positive recurrent (and therefore has a steady state distribution) if and only if

$$\sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} < \infty$$

Substituting in the birth and death rates and λ_i and μ_i for the M/M/∞ queue, we get:

$$\sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} = \sum_{i=1}^{\infty} \frac{\lambda^i}{\mu(2\mu)(3\mu)\ldots(i\mu)} = \sum_{i=1}^{\infty} \frac{\lambda^i}{i!\,\mu^i}$$

But recall that

$$\sum_{i=0}^{\infty} \frac{\lambda^i}{i!\,\mu^i} = e^{\lambda/\mu}$$

which is clearly finite. So the M/M/∞ queue is *always* positive recurrent!

Implications for the M/M/∞ Queue

EXAMPLE 7: M/M/∞ Queue, continued

The stationary distribution ω is given by:

$$\omega_0 = \left(1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}\right)^{-1} \qquad \omega_j = \frac{\lambda_0 \lambda_1 \cdots \lambda_{j-1}}{\mu_1 \mu_2 \cdots \mu_j}\left(1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}\right)^{-1}$$

Using the previous results gives us:

$$\omega_0 = \left(e^{\lambda/\mu}\right)^{-1} = e^{-\lambda/\mu} \qquad \omega_j = \frac{\lambda^j}{j!\,\mu^j} e^{-\lambda/\mu} \quad \text{for } j = 1,2,\ldots..$$

Implications for the M/M/S Queue

EXAMPLE 8: M/M/S Queue

The M/M/S queue is positive recurrent and has a steady state distribution if and only if

$$\sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} < \infty$$

where $\sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} = \sum_{i=1}^{S-1} \frac{\lambda^i}{\mu(2\mu)(3\mu)\ldots(i\mu)} + \sum_{i=S}^{\infty} \frac{\lambda^i}{S!\,S^{i-S}\mu^i}$

$$= \sum_{i=1}^{S-1} \frac{\lambda^i}{i!\,\mu^i} + \frac{S^S}{S!}\sum_{i=S}^{\infty} \frac{\lambda^i}{(S\mu)^i}$$

If and only if $\lambda < S\mu$, the sum $\sum_{i=S}^{\infty} \frac{\lambda^i}{(S\mu)^i}$ is finite and equals

$$\sum_{i=S}^{\infty} \frac{\lambda^i}{(S\mu)^i} = \sum_{j=0}^{\infty} \frac{\lambda^{j+S}}{(S\mu)^{j+S}} = \frac{\lambda^S}{(S\mu)^S}\sum_{j=0}^{\infty} \frac{\lambda^j}{(S\mu)^j} = \frac{\lambda^S}{(S\mu)^S}\left(1 - \frac{\lambda}{S\mu}\right)^{-1}$$

So if $\lambda < S\mu$, we can write $\sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} = \sum_{i=1}^{S-1} \frac{\lambda^i}{i!\,\mu^i} + \frac{\lambda^S}{S!\,\mu^S}\left(1 - \frac{\lambda}{S\mu}\right)^{-1}$

Implications for the M/M/S Queue

EXAMPLE 8, continued

Let $$K \equiv \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} = \sum_{i=1}^{S-1} \frac{\lambda^i}{i!\,\mu^i} + \frac{\lambda^S}{S!\,\mu^S}\left(1 - \frac{\lambda}{S\mu}\right)^{-1}$$

Then the steady state distribution for the M/M/S queue exists if and only if $\lambda < S\mu$, and in that case is equal to

$$\omega_0 = \left(1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}\right)^{-1} = (1 + K)^{-1}$$

$$\omega_j = \frac{\lambda_0 \lambda_1 \cdots \lambda_{j-1}}{\mu_1 \mu_2 \cdots \mu_j}\left(1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}\right)^{-1} = \begin{cases} \dfrac{\lambda^j}{j!\,\mu^j}(1+K)^{-1} & \text{for } j \le S \\ \dfrac{\lambda^j}{S!\,S^{j-S}\mu^j}(1+K)^{-1} & \text{for } j > S \end{cases}$$

Vector Notation for Steady State Equations

$$0 = \sum_{i=0}^{S} \omega_i q_{ij} = (\omega_0, \omega_1, \ldots, \omega_S)\begin{pmatrix} q_{0j} \\ q_{1j} \\ \vdots \\ q_{Sj} \end{pmatrix} \leftarrow j\text{th column of } Q$$

$$\sum_{i=0}^{S} \omega_i = (\omega_0, \omega_1, \ldots, \omega_S)\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \leftarrow \text{vector of ones is denoted by } e$$

So the steady state equations can be rewritten in vector form as

$$\omega Q = 0$$
$$\omega e = 1$$

Solving for the Steady State Probabilities with MATLAB

For a finite CTMC, we can solve the system $\boxed{\begin{matrix} \omega Q = 0 \\ \omega e = 1 \end{matrix}}$ as follows. We know that the first set of S+1 equations has one redundant equation. Therefore we can delete one of those equations. To solve the remaining S equations along with the last one, we create the following matrix which we call B:

$$B = \begin{bmatrix} q_{00} & q_{01} & \cdots & q_{0,S-1} & 1 \\ q_{10} & q_{11} & \cdots & q_{1,S-1} & 1 \\ \vdots & & \ddots & & \vdots \\ q_{S0} & q_{S1} & \cdots & q_{S,S-1} & 1 \end{bmatrix}$$

B is simply Q with the last column replaced by a vector of 1's.

Also create an $S+1$ vector called b: $b = (\underbrace{0, 0, \ldots, 0}_{(S \text{ zeros})}, 1)$

Then the system $\boxed{\begin{matrix} \omega Q = 0 \\ \omega e = 1 \end{matrix}}$ can also be expressed as $\boxed{\omega B = b}$

Solving for the Steady State Probabilities with MATLAB

The matrix B has an inverse, B^{-1}. Solve the system $\boxed{\omega B = b}$ using:

$$\boxed{\omega = b B^{-1}}$$

The specific steps in MATLAB are:

1. type in the matrix B
2. type in the vector b
3. type w=b*inv(B)

Example: Solving Steady State Equations using MATLAB

EXAMPLE 2 using MATLAB

To use matlab for example 2, we would create the matrix *B* from Q by replacing the last column of *Q* with a column of 1's:

$$B = \begin{bmatrix} -5 & 5 & 0 & 0 & 1 \\ 2 & -7 & 5 & 0 & 1 \\ 0 & 4 & -9 & 5 & 1 \\ 0 & 0 & 4 & -9 & 1 \\ 0 & 0 & 0 & 4 & 1 \end{bmatrix}$$

We'd also create the vector *b*: $b = (0, 0, 0, 0, 1)$

At the MATLAB prompt, type ω = b*inv(B)

We get: $\omega = (0.0649, 0.1622, 0.2027, 0.2534, 0.3138)$

Distinction Between ω and π

Recall that we used the notation π for the steady state distribution in a DTMC. In fact, Hillier and Lieberman use π for the steady state distribution in CTMCs as well. We choose to use the notation ω in order to distinguish between the stationary distribution of a CTMC and that of its its embedded DTMC.

Suppose we have a CTMC for which the steady state probabilities exist both in the CTMC and its embedded DTMC. Let ω be the steady state probabilities of a
CTMC and π be the steady state probabilities of its embedded DTMC. In this context,

$\omega_j =$ the long run proportion of time that the CTMC spends in state *j*.

$\pi_j =$ the long run proportion of transitions that take the process into state *j*.

Then it is possible to show that $$\omega_i = \frac{\frac{\pi_i}{q_i}}{\sum_{j=0}^{S} \frac{\pi_j}{q_j}} \quad \text{for } i = 0, \ldots, S$$

Example: Distinction Between ω and π

EXAMPLE 2, continued

In the barbershop example, the transition probability matrix for the embedded DTMC is:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 2/7 & 0 & 5/7 & 0 & 0 \\ 0 & 4/9 & 0 & 5/9 & 0 \\ 0 & 0 & 4/9 & 0 & 5/9 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

The **discrete-time** Markov Chain steady state equations $\boxed{\begin{aligned} \pi &= \pi P \\ \pi e &= 1 \end{aligned}}$ yield:

$$\pi = (0.0475,\ 0.1662,\ 0.2671,\ 0.3338,\ 0.1885)$$

Verify on your own that $\omega_i = \dfrac{\pi_i}{q_i} \Big/ \sum_{j=0}^{S} \dfrac{\pi_j}{q_j} \text{ for } i = 0, \ldots, S$ for this example.

Costs and Rewards in CTMCs

The steady-state distribution is perhaps most useful in computing the expected costs or rewards that are incurred as a continuous-time Markov Chain evolves. Specifically, we may be interested in computing the long run expected cost (or reward) per unit time to be used as the basis for decision making. We have a result for computing this long run expected cost per unit time in CTMCs that is the analog of the result we had for DTMCs.

Long Run Expected Average Cost

Suppose there is a cost (or reward) *C(j)* incurred *per unit time* that the process is in state *j*.

Let $C = (C(0), C(1), \ldots, C(S))$ be the vector of cost (or reward) per unit time in each state.

The expected average cost (or reward) incurred by the process over [0,*t*) per unit time is

$$E\left[\frac{1}{t}\int_0^t C(X(s))\,ds\right]$$

Theorem 4

If $X(t),\ t \geq 0$ is an irreducible, positive recurrent Markov Chain then the **long run expected average cost per unit time** is given by

$$\lim_{t\to\infty} E\left[\frac{1}{t}\int_0^t C(X(s))\,ds\right] = \sum_{j=0}^{S} \omega_j C(j) = \omega C$$

Long Run Expected Average Cost: Example

EXAMPLE 2 revisited

Reconsider our barbershop example. Suppose that the barbers charge according to how long a haircut takes; they each charge $1 per minute. If a person has to wait for a haircut, he or she drinks $0.10 worth of coffee per minute while waiting. What is the expected income of the shop per minute?

Let $C = (C(0), C(1), \ldots, C(S)) = (0, 1, 2, 2-.1, 2-.2) = (0, 1, 2, 1.9, 1.8)$

represent the reward vector (in dollars per minute). Since the chain is irreducible and finite, it is positive recurrent. So we can compute the long run expected average reward per minute using theorem 4:

$$\lim_{t\to\infty} E\left[\frac{1}{t}\int_0^t C(X(s))\,ds\right] = \sum_{j=0}^{S} \omega_j C(j) = \omega C$$

Long Run Expected Average Cost: Example

EXAMPLE 2, continued

We have already computed the steady state probability vector ω for this example:

$$\omega = (0.0649,\ 0.1622,\ 0.2027,\ 0.2534,\ 0.3138)$$

To compute the long run expected average reward per minute, we take the scalar product of the vector w with the reward vector

$$C = (C(0), C(1), \ldots, C(4)) = (0, 1, 2, 1.9, 1.8)$$

The long run expected average reward per minute is:

$$\sum_{j=0}^{S} \omega_j C(j) = 0 \cdot (0.0649) + 1 \cdot (0.1622) + 2 \cdot (0.2027)$$
$$+ 1.9 \cdot (0.2534) + 1.8 \cdot (0.3138)$$
$$= \$1.6625$$

Long Run Expected Average Cost: Example

EXAMPLE 4 revisited

Let's go back to our traffic light example, and now assume that there can be at most 7 cars waiting (cars that arrive and find 7 waiting use a different exit). Suppose that HP is actually paying its employees for the time they spend commuting. Each employees' time is valued at $0.50 per minute. In the long run, what is the expected average cost of delays caused by this traffic light to HP per minute?

Let $C = (C(0), C(1), \ldots\ldots C(7)) = \$0.50 \cdot (0, 1, 2, \ldots, 7)$

represent the cost vector (in dollars per minute). The chain is finite (now that we've bounded the maximum queue length) and irreducible, so it it necessarily positive recurrent. We must first compute its steady state distribution.

Long Run Expected Average Cost: Example

EXAMPLE 4 continued

First we need the transition rate matrix:

$$Q = \begin{bmatrix} -.6 & .6 & & & & & & \\ .2 & -.8 & .6 & & & & & \\ .2 & & -.8 & .6 & & & & \\ .2 & & & -.8 & .6 & & & \\ .2 & & & & -.8 & .6 & & \\ .2 & & & & & -.8 & .6 & \\ & .2 & & & & & -.8 & .6 \\ & & .2 & & & & & -.2 \end{bmatrix}$$

Using Matlab to solve for the steady state distribution, we get:

ω = (0.0141, 0.0221, 0.1461, 0.1096, 0.0822, 0.0616, 0.0462, 0.5181)

Long Run Expected Average Cost: Example

EXAMPLE 4, continued

We compute the long run expected cost per minute using theorem 4:

$$\lim_{t \to \infty} E\left[\frac{1}{t}\int_0^t C(X(s))\,ds\right] = \sum_{j=0}^{S} \omega_j C(j) = \omega C$$

$$= (0.0141, 0.0221, 0.1461, 0.1096, 0.0822, 0.0616, 0.0462, 0.5181) \begin{pmatrix} 0 \\ 0.5 \\ 1.0 \\ 1.5 \\ 2.0 \\ 2.5 \\ 3.0 \\ 3.5 \end{pmatrix}$$

This traffic lights costs HP an average of $2.59 per minute (in expectation) in the long run.

= 2.5921