# Continuous-time Markov Chains

Gonzalo Mateos
Dept. of ECE and Goergen Institute for Data Science
University of Rochester
gmateosb@ece.rochester.edu
http://www.ece.rochester.edu/~gmateosb/

October 31, 2016

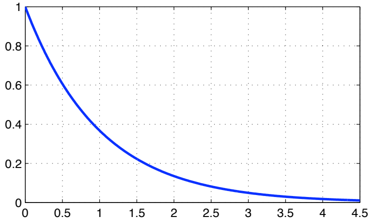
# Exponential distribution

- Exponential RVs often model times at which events occur
  $\Rightarrow$ Or time elapsed between occurrence of random events
- RV $T \sim \exp(\lambda)$ is exponential with parameter $\lambda$ if its pdf is

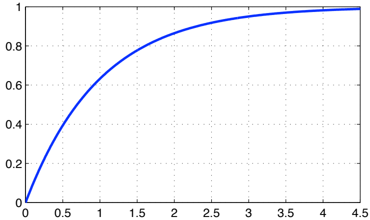
$$f_T(t) = \lambda e^{-\lambda t}, \qquad \text{for all } t \geq 0$$

- Cdf, integral of the pdf, is $\Rightarrow F_T(t) = \mathrm{P}\left(T \leq t\right) = 1 - e^{-\lambda t}$
  $\Rightarrow$ Complementary (c)cdf is $\Rightarrow P(T \geq t) = 1 - F_T(t) = e^{-\lambda t}$

pdf ($\lambda = 1$)




cdf ($\lambda = 1$)



# Expected value

- Expected value of time $T \sim \exp(\lambda)$ is

$$\mathbb{E}[T] = \int_0^\infty t\lambda e^{-\lambda t} dt = -te^{-\lambda t}\Big|_0^\infty + \int_0^\infty e^{-\lambda t} dt = 0 + \frac{1}{\lambda}$$

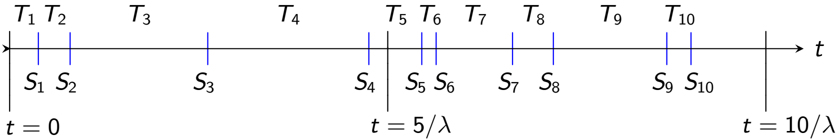
$\Rightarrow$ Integrated by parts with $u = t$, $dv = \lambda e^{-\lambda t} dt$

- Mean time is inverse of parameter $\lambda$

  $\Rightarrow$ $\lambda$ is rate/frequency of events happening at intervals $T$

  $\Rightarrow$ **Interpret:** Average of $\lambda t$ events by time $t$

- Bigger $\lambda$ $\Rightarrow$ smaller expected times, larger frequency of events



# Second moment and variance

- For second moment also integrate by parts ($u = t^2$, $dv = \lambda e^{-\lambda t} dt$)

$$\mathbb{E}\left[T^2\right] = \int_0^\infty t^2 \lambda e^{-\lambda t} dt = -t^2 e^{-\lambda t}\Big|_0^\infty + \int_0^\infty 2te^{-\lambda t} dt$$

- First term is 0, second is $(2/\lambda)\mathbb{E}[T]$

$$\mathbb{E}\left[T^2\right] = \frac{2}{\lambda}\int_0^\infty t\lambda e^{-\lambda t} = \frac{2}{\lambda^2}$$

- The variance is computed from the mean and second moment

$$\text{var}[T] = \mathbb{E}\left[T^2\right] - \mathbb{E}^2[T] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

$\Rightarrow$ Parameter $\lambda$ controls mean and variance of exponential RV

# Memoryless random times

- **Def:** Consider random time $T$. We say time $T$ is memoryless if

$$\mathrm{P}\left(T > s + t \,\middle|\, T > t\right) = \mathrm{P}\left(T > s\right)$$

- Probability of waiting $s$ extra units of time (e.g., seconds) given that we waited $t$ seconds, is just the probability of waiting $s$ seconds
  - $\Rightarrow$ System does not remember it has already waited $t$ seconds
  - $\Rightarrow$ Same probability irrespectively of time already elapsed

Ex: Chemical reaction $A + B \to AB$ occurs when molecules $A$ and $B$ "collide". $A$, $B$ move around randomly. Time $T$ until reaction

# Exponential RVs are memoryless

- Write memoryless property in terms of joint pdf

$$\mathrm{P}\left(T > s + t \,\middle|\, T > t\right) = \frac{\mathrm{P}\left(T > s + t, T > t\right)}{\mathrm{P}\left(T > t\right)} = \mathrm{P}\left(T > s\right)$$

- Notice event $\{T > s + t, T > t\}$ is equivalent to $\{T > s + t\}$
  $\Rightarrow$ Replace $\mathrm{P}\left(T > s + t, T > t\right) = \mathrm{P}\left(T > s + t\right)$ and reorder

$$\mathrm{P}\left(T > s + t\right) = \mathrm{P}\left(T > t\right)\mathrm{P}\left(T > s\right)$$

- If $T \sim \exp(\lambda)$, ccdf is $\mathrm{P}\left(T > t\right) = e^{-\lambda t}$ so that

$$\mathrm{P}\left(T > s + t\right) = e^{-\lambda(s+t)} = e^{-\lambda t}e^{-\lambda s} = \mathrm{P}\left(T > t\right)\mathrm{P}\left(T > s\right)$$

- If random time $T$ is exponential $\Rightarrow$ $T$ is memoryless

# Continuous memoryless RVs are exponential

- Consider a function $g(t)$ with the property $g(t+s) = g(t)g(s)$
- Q: Functional form of $g(t)$? Take logarithms

$$\log g(t+s) = \log g(t) + \log g(s)$$

$\Rightarrow$ Only holds for all $t$ and $s$ if $\log g(t) = ct$ for some constant $c$

$\Rightarrow$ Which in turn, can only hold if $g(t) = e^{ct}$ for some constant $c$

- Compare observation with statement of memoryless property

$$\mathrm{P}\left(T > s+t\right) = \mathrm{P}\left(T > t\right)\mathrm{P}\left(T > s\right)$$

$\Rightarrow$ It must be $\mathrm{P}\left(T > t\right) = e^{ct}$ for some constant $c$

- $T$ continuous: only true for exponential $T \sim \exp(-c)$
- $T$ discrete: only geometric $\mathrm{P}\left(T > t\right) = (1-p)^t$ with $(1-p) = e^c$
- If continuous random time $T$ is memoryless $\Rightarrow$ $T$ is exponential

# Main memoryless property result

Theorem

*A continuous random variable $T$ is memoryless if and only if it is exponentially distributed. That is*

$$\mathrm{P}\left(T > s+t \,\middle|\, T > t\right) = \mathrm{P}\left(T > s\right)$$

*if and only if $f_T(t) = \lambda e^{-\lambda t}$ for some $\lambda > 0$*

- Exponential RVs are memoryless. Do not remember elapsed time
  - $\Rightarrow$ Only type of continuous memoryless RVs
- Discrete RV $T$ is memoryless if and only of it is geometric
  - $\Rightarrow$ Geometrics are discrete approximations of exponentials
  - $\Rightarrow$ Exponentials are continuous limits of geometrics
- Exponential = time until success $\Leftrightarrow$ Geometric = nr. trials until success

# Exponential times example

- First customer's arrival to a store takes $T \sim \exp(1/10)$ minutes
  $\Rightarrow$ Suppose 5 minutes have passed without an arrival
- Q: How likely is it that the customer arrives within the next 3 mins.?
- Use memoryless property of exponential $T$

$$\begin{aligned} \mathrm{P}\left(T \leq 8 \,\middle|\, T > 5\right) &= 1 - \mathrm{P}\left(T > 8 \,\middle|\, T > 5\right) \\ &= 1 - \mathrm{P}\left(T > 3\right) = 1 - e^{-3\lambda} = 1 - e^{-0.3} \end{aligned}$$

- Q: How likely is it that the customer arrives after time $T = 9$?

$$\mathrm{P}\left(T > 9 \,\middle|\, T > 5\right) = \mathrm{P}\left(T > 4\right) = e^{-4\lambda} = e^{-0.4}$$

- Q: What is the expected additional time until the first arrival?
- Additional time is $T - 5$, and $\mathrm{P}\left(T - 5 > t \,\middle|\, T > 5\right) = \mathrm{P}\left(T > t\right)$

$$\mathbb{E}\left[T - 5 \,\middle|\, T > 5\right] = \mathbb{E}\left[T\right] = 1/\lambda = 10$$

# Time to first event

- Independent exponential RVs $T_1$, $T_2$ with parameters $\lambda_1$, $\lambda_2$
- Q: Prob. distribution of time to first event, i.e., $T := \min(T_1, T_2)$?
  $\Rightarrow$ To have $T > t$ we need both $T_1 > t$ and $T_2 > t$
- Using independence of $T_1$ and $T_2$ we can write

$$P(T > t) = P(T_1 > t, T_2 > t) = P(T_1 > t)\,P(T_2 > t)$$

- Substituting expressions of exponential ccdfs

$$P(T > t) = e^{-\lambda_1 t} e^{-\lambda_2 t} = e^{-(\lambda_1 + \lambda_2)t}$$

- $T := \min(T_1, T_2)$ is exponentially distributed with parameter $\lambda_1 + \lambda_2$
- In general, for $n$ independent RVs $T_i \sim \exp(\lambda_i)$ define $T := \min_i T_i$
  $\Rightarrow$ $T$ is exponentially distributed with parameter $\sum_{i=1}^{n} \lambda_i$

# First event to happen

- Q: Prob. $\mathrm{P}\left(T_1 < T_2\right)$ of $T_1 \sim \exp(\lambda_1)$ happening before $T_2 \sim \exp(\lambda_2)$?
- Condition on $T_2 = t$, integrate over the pdf of $T_2$, independence

$$\mathrm{P}\left(T_1 < T_2\right) = \int_0^\infty \mathrm{P}\left(T_1 < t \,\middle|\, T_2 = t\right) f_{T_2}(t)\, dt = \int_0^\infty F_{T_1}(t) f_{T_2}(t)\, dt$$

- Substitute expressions for exponential pdf and cdf

$$\mathrm{P}\left(T_1 < T_2\right) = \int_0^\infty (1 - e^{-\lambda_1 t})\lambda_2 e^{-\lambda_2 t}\, dt = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

- Either $T_1$ comes before $T_2$ or vice versa, hence

$$\mathrm{P}\left(T_2 < T_1\right) = 1 - \mathrm{P}\left(T_1 < T_2\right) = \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

$\Rightarrow$ Probabilities are relative values of rates (parameters)

- Larger rate $\Rightarrow$ smaller average $\Rightarrow$ higher prob. happening first

# Additional properties of exponential RVs

- Consider $n$ independent RVs $T_i \sim \exp(\lambda_i)$. $T_i$ time to $i$-th event
- Probability of $j$-th event happening first

$$\mathrm{P}\left(T_j = \min_i T_i\right) = \frac{\lambda_j}{\sum_{i=1}^n \lambda_i}, \ j = 1, \ldots, n$$

- Time to first event and rank ordering of events are independent

$$\mathrm{P}\left(\min_i T_i \geq t, T_{i_1} < \ldots < T_{i_n}\right) = \mathrm{P}\left(\min_i T_i \geq t\right) \mathrm{P}\left(T_{i_1} < \ldots < T_{i_n}\right)$$

- Suppose $T \sim \exp(\lambda)$, independent of non-negative RV $X$
- Strong memoryless property asserts

$$\mathrm{P}\left(T > X + t \,\middle|\, T > X\right) = \mathrm{P}\left(T > t\right)$$

$\Rightarrow$ Also forgets random but independent elapsed times

# Strong memoryless property example

- Independent customer arrival times $T_i \sim \exp(\lambda_i)$, $i = 1, \ldots, 3$
  $\Rightarrow$ Suppose customer 3 arrives first, i.e., $\min(T_1, T_2) > T_3$
- Q: Probability that time between first and second arrival exceeds $t$?
- We want to compute

$$\mathrm{P}\left(\min(T_1, T_2) - T_3 > t \,\middle|\, \min(T_1, T_2) > T_3\right)$$

- Denote $T_{i_2} := \min(T_1, T_2)$ the time to second arrival
  $\Rightarrow$ Recall $T_{i_2} \sim \exp(\lambda_1 + \lambda_2)$, $T_{i_2}$ independent of $T_{i_1} = T_3$
- Apply the strong memoryless property

$$\mathrm{P}\left(T_{i_2} - T_3 > t \,\middle|\, T_{i_2} > T_3\right) = \mathrm{P}\left(T_{i_2} > t\right) = e^{-(\lambda_1+\lambda_2)t}$$

# Probability of event in infinitesimal time

- Q: Probability of an event happening in infinitesimal time $h$?
- Want $\mathrm{P}\left(T < h\right)$ for small $h$

$$\mathrm{P}\left(T < h\right) = \int_0^h \lambda e^{-\lambda t}\, dt \approx \lambda h$$

$\Rightarrow$ Equivalent to $\left.\dfrac{\partial \mathrm{P}\left(T < t\right)}{\partial t}\right|_{t=0} = \lambda$

- Sometimes also write $\mathrm{P}\left(T < h\right) = \lambda h + o(h)$

$\Rightarrow$ $o(h)$ implies $\displaystyle\lim_{h \to 0} \frac{o(h)}{h} = 0$

$\Rightarrow$ Read as "negligible with respect to $h$"

- Q: Two independent events in infinitesimal time $h$?

$$\mathrm{P}\left(T_1 \leq h, T_2 \leq h\right) \approx (\lambda_1 h)(\lambda_2 h) = \lambda_1 \lambda_2 h^2 = o(h)$$

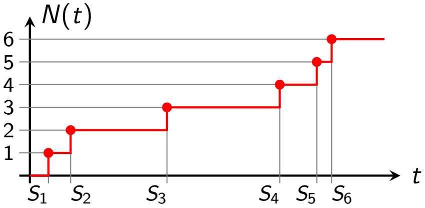
# Counting processes

- Random process $N(t)$ in continuous time $t \in \mathbb{R}_+$
- **Def:** Counting process $N(t)$ counts number of events by time $t$
- Nonnegative integer valued: $N(0) = 0$, $N(t) \in \{0, 1, 2, \ldots\}$
- Nondecreasing: $N(s) \leq N(t)$ for $s < t$
- Event counter: $N(t) - N(s)$ = number of events in interval $(s, t]$
  - $N(t)$ continuous from the right
  - $N(S_4) - N(S_2) = 2$, while $N(S_4) - N(S_2 - \epsilon) = 3$ for small $\epsilon > 0$

Ex.1: # text messages (SMS) typed since beginning of class

Ex.2: # economic crises since 1900

Ex.3: # customers at Wegmans since 8 am this morning



# Independent increments

- Consider times $s_1 < t_1 < s_2 < t_2$ and intervals $(s_1, t_1]$ and $(s_2, t_2]$
  - $\Rightarrow$ $N(t_1) - N(s_1)$ events occur in $(s_1, t_1]$
  - $\Rightarrow$ $N(t_2) - N(s_2)$ events occur in $(s_2, t_2]$
- **Def:** Independent increments implies latter two are independent

$$
\begin{aligned}
\mathrm{P}\left(N(t_1) - N(s_1) = k, N(t_2) - N(s_2) = l\right) \\
= \mathrm{P}\left(N(t_1) - N(s_1) = k\right) \mathrm{P}\left(N(t_2) - N(s_2) = l\right)
\end{aligned}
$$

- Number of events in disjoint time intervals are independent

Ex.1: Likely true for SMS, except for "have to send" messages

Ex.2: Most likely not true for economic crises (business cycle)

Ex.3: Likely true for Wegmans, except for unforeseen events (storms)

- Does not mean $N(t)$ independent of $N(s)$, say for $t > s$
  - $\Rightarrow$ These events are clearly dependent, since $N(t)$ is at least $N(s)$

# Stationary increments

- Consider time intervals $(0, t]$ and $(s, s + t]$

  $\Rightarrow N(t)$ events occur in $(0, t]$

  $\Rightarrow N(s + t) - N(s)$ events in $(s, s + t]$

- **Def:** Stationary increments implies latter two have same prob. dist.

$$\mathrm{P}\left(N(s + t) - N(s) = k\right) = \mathrm{P}\left(N(t) = k\right)$$

- Prob. dist. of number of events depends on length of interval only

Ex.1: Likely true if lecture is good and you keep interest in the class

Ex.2: Maybe true if you do not believe we become better at preventing crises

Ex.3: Most likely not true because of, e.g., rush hours and slow days

# Poisson process

- **Def:** A counting process $N(t)$ is a Poisson process if
  (a) The process has stationary and independent increments
  (b) The number of events in $(0, t]$ has Poisson distribution with mean $\lambda t$

$$\mathrm{P}\left(N(t) = n\right) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

- An equivalent definition is the following
  (i) The process has stationary and independent increments
  (ii) Single event in infinitesimal time $\Rightarrow \mathrm{P}\left(N(h) = 1\right) = \lambda h + o(h)$
  (iii) Multiple events in infinitesimal time $\Rightarrow \mathrm{P}\left(N(h) > 1\right) = o(h)$

  $\Rightarrow$ A more intuitive definition (even hard to grasp now)

- Conditions (i) and (a) are the same
- That (b) implies (ii) and (iii) is obvious
  - Substitute small $h$ in Poisson pmf's expression for $\mathrm{P}\left(N(t) = n\right)$
- To see that (ii) and (iii) imply (b) requires some work

# What is a Poisson process?

- Fundamental defining properties of a Poisson process
  - Events happen in small interval $h$ with probability $\lambda h$ proportional to $h$
  - Whether event happens in an interval has no effect on other intervals
- Modeling questions

  Q1: Expect probability of event proportional to length of interval?

  Q2: Expect subsequent intervals to behave independently?

  $\Rightarrow$ If positive, then a Poisson process model is appropriate
- Typically arise in a large population of agents acting independently

  $\Rightarrow$ Larger interval, larger chance an agent takes an action

  $\Rightarrow$ Action of one agent has no effect on action of other agents

  $\Rightarrow$ Has therefore negligible effect on action of group

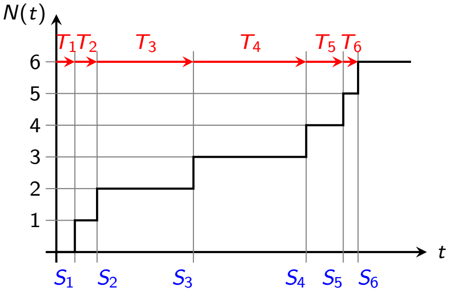
# Examples of Poisson processes

Ex.1: Number of people arriving at subway station. Number of cars arriving at a highway entrance. Number of customers entering a store ... Large number of agents (people, drivers, customers) acting independently

Ex.2: SMS generated by all students in the class. Once you send an SMS you are likely to stay silent for a while. But in a large population this has a minimal effect in the probability of someone generating a SMS

Ex.3: Count of molecule reactions. Molecules are "removed" from pool of reactants once they react. But effect is negligible in large population. Eventually reactants are depleted, but in small time scale process is approximately Poisson

# Arrival times and interarrival times




- Let $S_1, S_2, \ldots$ be the sequence of absolute times of events (arrivals)
- **Def:** $S_i$ is known as the $i$-th arrival time
  
  $\Rightarrow$ Notice that $S_i = \min_t(N(t) \geq i)$
- Let $T_1, T_2, \ldots$ be the sequence of times between events
- **Def:** $T_i$ is known as the $i$-th interarrival time
- Useful identities: $S_i = \sum_{k=1}^{i} T_k$ and $T_i = S_i - S_{i-1}$, where $S_0 = 0$

# Interarrival times are i.i.d. exponential RVs

- Ccdf of $T_1 \Rightarrow \mathrm{P}(T_1 > t) = \mathrm{P}(N(t) = 0) = e^{-\lambda t}$

  $\Rightarrow$ $T_1$ has exponential distribution with parameter $\lambda$

- Since increments are stationary and independent, likely $T_i$ are i.i.d.

Theorem

*Interarrival times $T_i$ of a Poisson process are independent identically distributed exponential random variables with parameter $\lambda$, i.e.,*

$$\mathrm{P}(T_i > t) = e^{-\lambda t}$$

- Have already proved for $T_1$. Let us see the rest

# Interarrival times example

- Let $N_1(t)$ and $N_2(t)$ be Poisson processes with rates $\lambda_1$ and $\lambda_2$
  - $\Rightarrow$ Suppose $N_1(t)$ and $N_2(t)$ are independent
- Q: What is the expected time till the first arrival from either process?
- Denote as $S_1^{(i)}$ the first arrival time from process $i = 1, 2$
  - $\Rightarrow$ We are looking for $\mathbb{E}\left[\min\left(S_1^{(1)}, S_1^{(2)}\right)\right]$
- Note that $S_1^{(1)} = T_1^{(1)}$ and $S_1^{(2)} = T_1^{(2)}$
  - $\Rightarrow$ $S_1^{(1)} \sim \exp(\lambda_1)$ and $S_1^{(2)} \sim \exp(\lambda_2)$
  - $\Rightarrow$ Also, $S_1^{(1)}$ and $S_1^{(2)}$ are independent
- Recall that $\min\left(S_1^{(1)}, S_1^{(2)}\right) \sim \exp(\lambda_1 + \lambda_2)$, then

$$\mathbb{E}\left[\min\left(S_1^{(1)}, S_1^{(2)}\right)\right] = \frac{1}{\lambda_1 + \lambda_2}$$

# Alternative definition of Poisson process

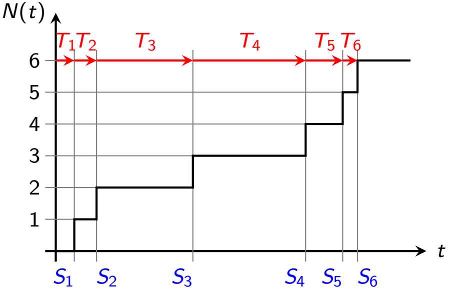
- Start with sequence of independent random times $T_1, T_2, \ldots$
- Times $T_i \sim \exp(\lambda)$ have exponential distribution with parameter $\lambda$
- Define $i$-th arrival time $S_i$

$$S_i = T_1 + T_2 + \ldots + T_i$$

- Define counting process of events occurred by time $t$

$$N(t) = \max_i (S_i \leq t)$$

- $N(t)$ is a Poisson process




- If $N(t)$ is a Poisson process interarrival times $T_i$ are i.i.d. exponential
- To show that definition is equivalent have to show the converse

$\Rightarrow$ If interarrival times are i.i.d. exponential, process is Poisson

# Three definitions of Poisson processes

Def. 1: Prob. of event proportional to interval width. Intervals independent

- Physical model definition
- Can a phenomenon be reasonably modeled as a Poisson process?
- The other two definitions are used for analysis and/or simulation

Def. 2: Prob. distribution of events in $(0, t]$ is Poisson

- Event centric definition. Nr. of events in given time intervals
- Allows analysis and simulation
- Used when information about nr. of events in given time is desired

Def. 3: Prob. distribution of interarrival times is exponential

- Time centric definition. Times at which events happen
- Allows analysis and simulation
- Used when information about event times is of interest

# Number of visitors to a web page example

Ex: Count number of visits to a webpage between 6:00pm to 6:10pm

Def 1: Q: Poisson process? Yes, seems reasonable to have

- Probability of visit proportional to time interval duration
- Independent visits over disjoint time intervals

- Model as Poisson process with rate $\lambda$ visits/second (v/s)

Def 2: Arrivals in $(s, s+t]$ are Poisson with parameter $\lambda t$

- Prob. of exactly 5 visits in 1 sec? $\Rightarrow \mathrm{P}\left(N(1)=5\right) = e^{-\lambda}\lambda^5/5!$
- Expected nr. of visits in 10 minutes? $\Rightarrow \mathbb{E}\left[N(600)\right] = 600\lambda$
- On average, data shows $N$ visits in 10 minutes. Estimate $\hat{\lambda} = N/600$

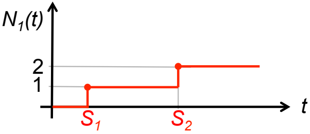
Def 3: Interarrival times are i.i.d. $T_i \sim \exp(\lambda)$

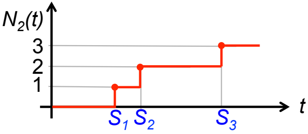
- Expected time between visits? $\Rightarrow \mathbb{E}\left[T_i\right] = 1/\lambda$
- Expected arrival time $S_n$ of $n$-th visitor?

  $\Rightarrow$ Recall $S_n = \sum_{i=1}^n T_i$, hence $\mathbb{E}\left[S_n\right] = \sum_{i=1}^n \mathbb{E}\left[T_i\right] = n/\lambda$

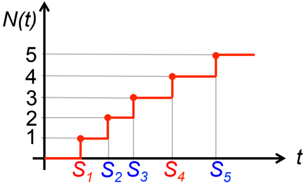
# Superposition of Poisson processes

- Let $N_1(t), N_2(t)$ be Poisson processes with rates $\lambda_1$ and $\lambda_2$
  $\Rightarrow$ Suppose $N_1(t)$ and $N_2(t)$ are independent



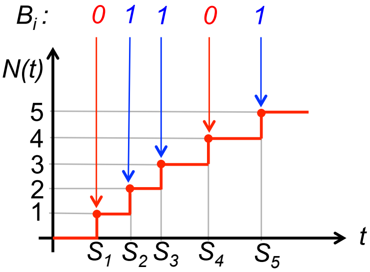




- Then $N(t) := N_1(t) + N_2(t)$ is a Poisson process with rate $\lambda_1 + \lambda_2$



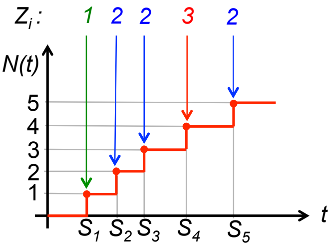
# Thinning of a Poisson process

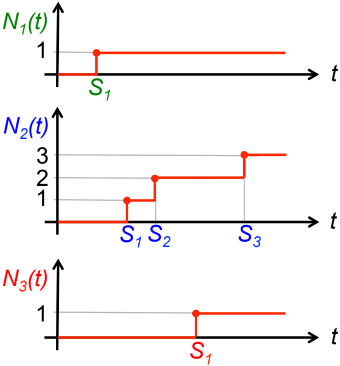
- Let $B_{\mathbb{N}} = B_1, B_2, \ldots$ be an i.i.d. sequence of Bernoulli($p$) RVs
- Let $N(t)$ be a Poisson process with rate $\lambda$, independent of $B_{\mathbb{N}}$
- Then $M(t) := \sum_{i=1}^{N(t)} B_i$ is a Poisson process with rate $\lambda p$



# Splitting of a Poisson process

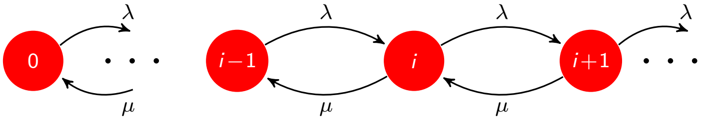
- Let $Z_{\mathbb{N}} = Z_1, Z_2, \ldots$ be an i.i.d. sequence of RVs, $Z_i \in \{1, \ldots, m\}$
- Let $N(t)$ be a Poisson process with rate $\lambda$, independent of $Z_{\mathbb{N}}$
- Define $N_k(t) = \sum_{i=1}^{N(t)} \mathbb{I}\{Z_i = k\}$, for each $k = 1, \ldots, m$
- Then each $N_k(t)$ is a Poisson process with rate $\lambda \mathrm{P}(Z_1 = k)$



# M/M/1 queue example

- An M/M/1 queue is a BD process with $\lambda_i = \lambda$ and $\mu_i = \mu$ constant
- State $Q(t)$ is the number of customers in the system at time $t$
  - $\Rightarrow$ Customers arrive for service at a rate of $\lambda$ per unit time
  - $\Rightarrow$ They are serviced at a rate of $\mu$ customers per unit time




- The M/M is for Markov arrivals/Markov departures
  - $\Rightarrow$ Implies a Poisson arrival process, exponential services times
  - $\Rightarrow$ The 1 is because there is only one server

# Definition

- Continuous-time positive variable $t \in [0, \infty)$
- Time-dependent random state $X(t)$ takes values on a countable set
  - In general denote states as $i = 0, 1, 2, \ldots$, i.e., here the state space is $\mathbb{N}$
  - If $X(t) = i$ we say "the process is in state $i$ at time $t$"
- **Def:** Process $X(t)$ is a continuous-time Markov chain (CTMC) if

$$\begin{aligned} & \mathrm{P}\left(X(t+s) = j \,\middle|\, X(s) = i, X(u) = x(u), u < s\right) \\ = & \mathrm{P}\left(X(t+s) = j \,\middle|\, X(s) = i\right) \end{aligned}$$

- Markov property $\Rightarrow$ Given the present state $X(s)$

  $\Rightarrow$ Future $X(t+s)$ is independent of the past $X(u) = x(u), u < s$
- In principle need to specify functions $\mathrm{P}\left(X(t+s) = j \,\middle|\, X(s) = i\right)$

  $\Rightarrow$ For all times $t$ and $s$, for all pairs of states $(i, j)$

# Notation and homogeneity

- **Notation**
  - $X[s:t]$ state values for all times $s \leq u \leq t$, includes borders
  - $X(s:t)$ values for all times $s < u < t$, borders excluded
  - $X(s:t]$ values for all times $s < u \leq t$, exclude left, include right
  - $X[s:t)$ values for all times $s \leq u < t$, include left, exclude right
- Homogeneous CTMC if $\mathrm{P}\left(X(t+s)=j \,\middle|\, X(s)=i\right)$ invariant for all $s$

  $\Rightarrow$ We restrict consideration to homogeneous CTMCs
- Still need $P_{ij}(t) := \mathrm{P}\left(X(t+s)=j \,\middle|\, X(s)=i\right)$ for all $t$ and pairs $(i,j)$

  $\Rightarrow$ $P_{ij}(t)$ is known as the transition probability function. More later
- Markov property and homogeneity make description somewhat simpler

# Transition times

- $T_i$ = time until transition out of state $i$ into any other state $j$
- **Def:** $T_i$ is a random variable called transition time with ccdf

$$\mathrm{P}\left(T_i > t\right) = \mathrm{P}\left(X(0:t] = i \,\middle|\, X(0) = i\right)$$

- Probability of $T_i > t + s$ given that $T_i > s$? Use cdf expression

$$\begin{aligned}
\mathrm{P}\left(T_i > t + s \,\middle|\, T_i > s\right) &= \mathrm{P}\left(X(0:t+s] = i \,\middle|\, X[0:s] = i\right)\\
&= \mathrm{P}\left(X(s:t+s] = i \,\middle|\, X[0:s] = i\right)\\
&= \mathrm{P}\left(X(s:t+s] = i \,\middle|\, X(s) = i\right)\\
&= \mathrm{P}\left(X(0:t] = i \,\middle|\, X(0) = i\right)
\end{aligned}$$

- Used that $X[0:s] = i$ given, Markov property, and homogeneity
- From definition of $T_i$ $\Rightarrow \mathrm{P}\left(T_i > t + s \,\middle|\, T_i > s\right) = \mathrm{P}\left(T_i > t\right)$

  $\Rightarrow$ Transition times are exponential random variables

# Alternative definition

- Exponential transition times is a fundamental property of CTMCs

  ⇒ Can be used as "algorithmic" definition of CTMCs

- Continuous-time random process $X(t)$ is a CTMC if

  (a) Transition times $T_i$ are exponential random variables with mean $1/\nu_i$

  (b) When they occur, transition from state $i$ to $j$ with probability $P_{ij}$

$$\sum_{j=1}^{\infty} P_{ij} = 1, \qquad P_{ii} = 0$$

  (c) Transition times $T_i$ and transitioned state $j$ are independent

- Define matrix $\mathbf{P}$ grouping transition probabilities $P_{ij}$

- CTMC states evolve as in a discrete-time Markov chain

  ⇒ State transitions occur at exponential intervals $T_i \sim \exp(\nu_i)$

  ⇒ As opposed to occurring at fixed intervals

# Embedded discrete-time Markov chain

- Consider a CTMC with transition matrix $\mathbf{P}$ and rates $\nu_i$
- **Def:** CTMC's embedded discrete-time MC has transition matrix $\mathbf{P}$
- Transition probabilities $\mathbf{P}$ describe a discrete-time MC
  - $\Rightarrow$ No self-transitions ($P_{ii} = 0$, $\mathbf{P}$'s diagonal null)
  - $\Rightarrow$ Can use underlying discrete-time MCs to study CTMCs
- **Def:** State $j$ accessible from $i$ if accessible in the embedded MC
- **Def:** States $i$ and $j$ communicate if they do so in the embedded MC
  - $\Rightarrow$ Communication is a class property
- Recurrence, transience, ergodicity. Class properties . . . More later

# Transition rates

- Expected value of transition time $T_i$ is $\mathbb{E}[T_i] = 1/\nu_i$
  - $\Rightarrow$ Can interpret $\nu_i$ as the rate of transition out of state $i$
  - $\Rightarrow$ Of these transitions, a fraction $P_{ij}$ are into state $j$
- **Def:** Transition rate from $i$ to $j$ is $q_{ij} := \nu_i P_{ij}$
- Transition rates offer yet another specification of CTMCs
- If $q_{ij}$ are given can recover $\nu_i$ as

$$\nu_i = \nu_i \sum_{j=1}^{\infty} P_{ij} = \sum_{j=1}^{\infty} \nu_i P_{ij} = \sum_{j=1}^{\infty} q_{ij}$$

- Can also recover $P_{ij}$ as $\Rightarrow P_{ij} = q_{ij}/\nu_i = q_{ij}\left(\sum_{j=1}^{\infty} q_{ij}\right)^{-1}$

# Birth and death process example

- State $X(t) = 0, 1, \ldots$ Interpret as number of individuals
- Birth and deaths occur at state-dependent rates. When $X(t) = i$
- Births $\Rightarrow$ Individuals added at exponential times with mean $1/\lambda_i$
  $\Rightarrow$ Birth or arrival rate $= \lambda_i$ births per unit of time
- Deaths $\Rightarrow$ Individuals removed at exponential times with rate $1/\mu_i$
  $\Rightarrow$ Death or departure rate $= \mu_i$ deaths per unit of time
- Birth and death times are independent
- Birth and death (BD) processes are then CTMCs

# Transition times and probabilities

- Q: Transition times $T_i$? Leave state $i \neq 0$ when birth or death occur
- If $T_B$ and $T_D$ are times to next birth and death, $T_i = \min(T_B, T_D)$

  $\Rightarrow$ Since $T_B$ and $T_D$ are exponential, so is $T_i$ with rate

$$\nu_i = \lambda_i + \mu_i$$

- When leaving state $i$ can go to $i+1$ (birth first) or $i-1$ (death first)

  $\Rightarrow$ Birth occurs before death with probability $\dfrac{\lambda_i}{\lambda_i + \mu_i} = P_{i,i+1}$

  $\Rightarrow$ Death occurs before birth with probability $\dfrac{\mu_i}{\lambda_i + \mu_i} = P_{i,i-1}$

- Leave state 0 only if a birth occurs, then

$$\nu_0 = \lambda_0, \qquad P_{01} = 1$$

  $\Rightarrow$ If CTMC leaves 0, goes to 1 with probability 1

  $\Rightarrow$ Might not leave 0 if $\lambda_0 = 0$ (e.g., to model extinction)

# Transition rates

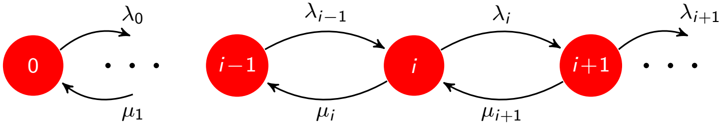
- Rate of transition from $i$ to $i+1$ is (recall definition $q_{ij} = \nu_i P_{ij}$)

$$q_{i,i+1} = \nu_i P_{i,i+1} = (\lambda_i + \mu_i)\frac{\lambda_i}{\lambda_i + \mu_i} = \lambda_i$$

- Likewise, rate of transition from $i$ to $i-1$ is

$$q_{i,i-1} = \nu_i P_{i,i-1} = (\lambda_i + \mu_i)\frac{\mu_i}{\lambda_i + \mu_i} = \mu_i$$

- For $i = 0$ $\Rightarrow q_{01} = \nu_0 P_{01} = \lambda_0$



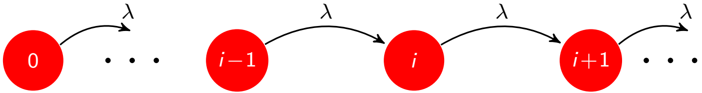

- Somewhat more natural representation. Similar to discrete-time MCs

# Poisson process example

- A Poisson process is a BD process with $\lambda_i = \lambda$ and $\mu_i = 0$ constant
- State $N(t)$ counts the total number of events (arrivals) by time $t$

  $\Rightarrow$ Arrivals occur a rate of $\lambda$ per unit time

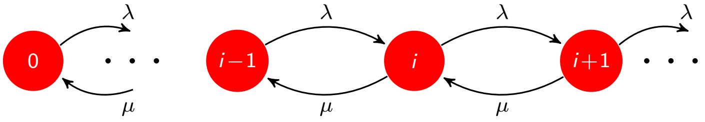
  $\Rightarrow$ Transition times are the i.i.d. exponential interarrival times




- The Poisson process is a CTMC

# M/M/1 queue example

- An M/M/1 queue is a BD process with $\lambda_i = \lambda$ and $\mu_i = \mu$ constant
- State $Q(t)$ is the number of customers in the system at time $t$
  - $\Rightarrow$ Customers arrive for service at a rate of $\lambda$ per unit time
  - $\Rightarrow$ They are serviced at a rate of $\mu$ customers per unit time




- The M/M is for Markov arrivals/Markov departures
  - $\Rightarrow$ Implies a Poisson arrival process, exponential services times
  - $\Rightarrow$ The 1 is because there is only one server

# Transition probability function

- Two equivalent ways of specifying a CTMC

1) Transition time averages $1/\nu_i$ + transition probabilities $P_{ij}$
   - $\Rightarrow$ Easier description
   - $\Rightarrow$ Typical starting point for CTMC modeling

2) Transition probability function $P_{ij}(t) := \mathrm{P}\left(X(t+s) = j \,\middle|\, X(s) = i\right)$
   - $\Rightarrow$ More complete description for all $t \geq 0$
   - $\Rightarrow$ Similar in spirit to $P_{ij}^n$ for discrete-time Markov chains

- Goal: compute $P_{ij}(t)$ from transition times and probabilities
   - $\Rightarrow$ Notice two obvious properties $P_{ij}(0) = 0$, $P_{ii}(0) = 1$

# Roadmap to determine $P_{ij}(t)$

- Goal is to obtain a differential equation whose solution is $P_{ij}(t)$

  $\Rightarrow$ Study change in $P_{ij}(t)$ when time changes slightly

- Separate in two subproblems (divide and conquer)

  $\Rightarrow$ Transition probabilities for small time $h$, $P_{ij}(h)$

  $\Rightarrow$ Transition probabilities in $t + h$ as function of those in $t$ and $h$

- We can combine both results in two different ways

1) Jump from 0 to $t$ then to $t + h$ $\Rightarrow$ Process runs a little longer

   $\Rightarrow$ Changes where the process is going to $\Rightarrow$ Forward equations

2) Jump from 0 to $h$ then to $t + h$ $\Rightarrow$ Process starts a little later

   $\Rightarrow$ Changes where the process comes from $\Rightarrow$ Backward equations

# Transition probability in infinitesimal time

**Theorem**

*The transition probability functions $P_{ii}(t)$ and $P_{ij}(t)$ satisfy the following limits as $t$ approaches 0*

$$\lim_{t\to 0} \frac{P_{ij}(t)}{t} = q_{ij}, \qquad \lim_{t\to 0} \frac{1 - P_{ii}(t)}{t} = \nu_i$$

- Since $P_{ij}(0) = 0$, $P_{ii}(0) = 1$ above limits are derivatives at $t = 0$

$$\left.\frac{\partial P_{ij}(t)}{\partial t}\right|_{t=0} = q_{ij}, \qquad \left.\frac{\partial P_{ii}(t)}{\partial t}\right|_{t=0} = -\nu_i$$

- Limits also imply that for small $h$ (recall Taylor series)

$$P_{ij}(h) = q_{ij}h + o(h), \qquad P_{ii}(h) = 1 - \nu_i h + o(h)$$

- Transition rates $q_{ij}$ are "instantaneous transition probabilities"

  $\Rightarrow$ Transition probability coefficient for small time $h$

# Chapman-Kolmogorov equations

**Theorem**

*For all times $s$ and $t$ the transition probability functions $P_{ij}(t+s)$ are obtained from $P_{ik}(t)$ and $P_{kj}(s)$ as*

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t) P_{kj}(s)$$

- As for discrete-time MCs, to go from $i$ to $j$ in time $t+s$
  - $\Rightarrow$ Go from $i$ to some state $k$ in time $t$ $\Rightarrow P_{ik}(t)$
  - $\Rightarrow$ In the remaining time $s$ go from $k$ to $j$ $\Rightarrow P_{kj}(s)$
  - $\Rightarrow$ Sum over all possible intermediate states $k$

# Chapman-Kolmogorov equations (proof)

Proof.

$$
\begin{aligned}
&P_{ij}(t+s) \\
&\quad = \mathrm{P}\left(X(t+s)=j \,\middle|\, X(0)=i\right) && \text{Definition of } P_{ij}(t+s) \\
&\quad = \sum_{k=0}^{\infty} \mathrm{P}\left(X(t+s)=j \,\middle|\, X(t)=k, X(0)=i\right)\mathrm{P}\left(X(t)=k \,\middle|\, X(0)=i\right) \\
& && \text{Law of total probability} \\
&\quad = \sum_{k=0}^{\infty} \mathrm{P}\left(X(t+s)=j \,\middle|\, X(t)=k\right) P_{ik}(t) && \text{Markov property of CTMC} \\
& && \text{and definition of } P_{ik}(t) \\
&\quad = \sum_{k=0}^{\infty} P_{kj}(s) P_{ik}(t) && \text{Definition of } P_{kj}(s)
\end{aligned}
$$

□

# Combining both results

- Let us combine the last two results to express $P_{ij}(t+h)$
- Use Chapman-Kolmogorov's equations for $0 \to t \to h$

$$P_{ij}(t+h) = \sum_{k=0}^{\infty} P_{ik}(t)P_{kj}(h) = P_{ij}(t)P_{jj}(h) + \sum_{k=0, k\neq j}^{\infty} P_{ik}(t)P_{kj}(h)$$

- Substitute infinitesimal time expressions for $P_{jj}(h)$ and $P_{kj}(h)$

$$P_{ij}(t+h) = P_{ij}(t)(1-\nu_j h) + \sum_{k=0, k\neq j}^{\infty} P_{ik}(t)q_{kj}h + o(h)$$

- Subtract $P_{ij}(t)$ from both sides and divide by $h$

$$\frac{P_{ij}(t+h) - P_{ij}(t)}{h} = -\nu_j P_{ij}(t) + \sum_{k=0, k\neq j}^{\infty} P_{ik}(t)q_{kj} + \frac{o(h)}{h}$$

- Right-hand side equals a "derivative" ratio. Let $h \to 0$ to prove ...

# Kolmogorov's forward equations

**Theorem**

*The transition probability functions $P_{ij}(t)$ of a CTMC satisfy the system of differential equations (for all pairs $i, j$)*

$$\frac{\partial P_{ij}(t)}{\partial t} = \sum_{k=0, k\neq j}^{\infty} q_{kj} P_{ik}(t) - \nu_j P_{ij}(t)$$

- Interpret each summand in Kolmogorov's forward equations
  - $\partial P_{ij}(t)/\partial t$ = rate of change of $P_{ij}(t)$
  - $q_{kj} P_{ik}(t)$ = (transition into $k$ in $0 \to t$) $\times$ (rate of moving into $j$ in next instant)
  - $\nu_j P_{ij}(t)$ = (transition into $j$ in $0 \to t$) $\times$ (rate of leaving $j$ in next instant)
- Change in $P_{ij}(t) = \sum_k$ (moving into $j$ from $k$) $-$ (leaving $j$)
- Kolmogorov's forward equations valid in most cases, but not always

# Kolmogorov's backward equations

- For forward equations used Chapman-Kolmogorov's for $0 \to t \to h$
- For backward equations we use $0 \to h \to t$ to express $P_{ij}(t+h)$ as

$$P_{ij}(t+h) = \sum_{k=0}^{\infty} P_{ik}(h)P_{kj}(t) = P_{ii}(h)P_{ij}(t) + \sum_{k=0, k\neq i}^{\infty} P_{ik}(h)P_{kj}(t)$$

- Substitute infinitesimal time expression for $P_{ii}(h)$ and $P_{ik}(h)$

$$P_{ij}(t+h) = (1-\nu_i h)P_{ij}(t) + \sum_{k=0, k\neq i}^{\infty} q_{ik} h P_{kj}(t) + o(h)$$

- Subtract $P_{ij}(t)$ from both sides and divide by $h$

$$\frac{P_{ij}(t+h) - P_{ij}(t)}{h} = -\nu_i P_{ij}(t) + \sum_{k=0, k\neq i}^{\infty} q_{ik} P_{kj}(t) + \frac{o(h)}{h}$$

- Right-hand side equals a "derivative" ratio. Let $h \to 0$ to prove ...

# Kolmogorov's backward equations

**Theorem**

*The transition probability functions $P_{ij}(t)$ of a CTMC satisfy the system of differential equations (for all pairs $i, j$)*

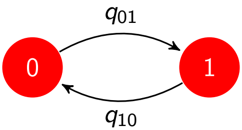
$$\frac{\partial P_{ij}(t)}{\partial t} = \sum_{k=0,k\neq i}^{\infty} q_{ik} P_{kj}(t) - \nu_i P_{ij}(t)$$

- Interpret each summand in Kolmogorov's backward equations
  - $\partial P_{ij}(t)/\partial t =$ rate of change of $P_{ij}(t)$
  - $q_{ik} P_{kj}(t) =$ (transition into $j$ in $h \to t$) $\times$ (rate of transition into $k$ in initial instant)
  - $\nu_i P_{ij}(t) =$ (transition into $j$ in $h \to t$) $\times$ (rate of leaving $i$ in initial instant)
- Forward equations $\Rightarrow$ change in $P_{ij}(t)$ if finish $h$ later
- Backward equations $\Rightarrow$ change in $P_{ij}(t)$ if start $h$ earlier
- Where process goes (forward) vs. where process comes from (backward)

# A CTMC with two states

Ex: Simplest possible CTMC has only two states. Say 0 and 1

- Transition rates are $q_{01}$ and $q_{10}$
- Given $q_{01}$ and $q_{10}$ can find rates of transitions out of $\{0, 1\}$




$$\nu_0 = \sum_j q_{0j} = q_{01}, \qquad \nu_1 = \sum_j q_{1j} = q_{10}$$

- Use Kolmogorov's equations to find transition probability functions

$$P_{00}(t), \quad P_{01}(t), \quad P_{10}(t), \quad P_{11}(t)$$

- Transition probabilities out of each state sum up to one

$$P_{00}(t) + P_{01}(t) = 1, \qquad P_{10}(t) + P_{11}(t) = 1$$

# Kolmogorov's forward equations

- Kolmogorov's forward equations (process runs a little longer)

$$P'_{ij}(t) = \sum_{k=0, k\neq j}^{\infty} q_{kj} P_{ik}(t) - \nu_j P_{ij}(t)$$

- For the two state CTMC

$$P'_{00}(t) = q_{10} P_{01}(t) - \nu_0 P_{00}(t), \qquad P'_{01}(t) = q_{01} P_{00}(t) - \nu_1 P_{01}(t)$$
$$P'_{10}(t) = q_{10} P_{11}(t) - \nu_0 P_{10}(t), \qquad P'_{11}(t) = q_{01} P_{10}(t) - \nu_1 P_{11}(t)$$

- Probabilities out of 0 sum up to 1 $\Rightarrow$ eqs. in first row are equivalent
- Probabilities out of 1 sum up to 1 $\Rightarrow$ eqs. in second row are equivalent

  $\Rightarrow$ Pick the equations for $P'_{00}(t)$ and $P'_{11}(t)$

# Solution of forward equations

- Use $\Rightarrow$ Relation between transition rates: $\nu_0 = q_{01}$ and $\nu_1 = q_{10}$
  $\Rightarrow$ Probs. sum 1: $P_{01}(t) = 1 - P_{00}(t)$ and $P_{10}(t) = 1 - P_{11}(t)$

$$P'_{00}(t) = q_{10}\big[1 - P_{00}(t)\big] - q_{01}P_{00}(t) = q_{10} - (q_{10} + q_{01})P_{00}(t)$$
$$P'_{11}(t) = q_{01}\big[1 - P_{11}(t)\big] - q_{10}P_{11}(t) = q_{01} - (q_{10} + q_{01})P_{11}(t)$$

- Can obtain exact same pair of equations from backward equations
- First-order linear differential equations $\Rightarrow$ Solutions are exponential
- For $P_{00}(t)$ propose candidate solution (just differentiate to check)

$$P_{00}(t) = \frac{q_{10}}{q_{10} + q_{01}} + ce^{-(q_{10}+q_{01})t}$$

$\Rightarrow$ To determine $c$ use initial condition $P_{00}(0) = 1$

# Solution of forward equations (continued)

- Evaluation of candidate solution at initial condition $P_{00}(0) = 1$ yields

$$1 = \frac{q_{10}}{q_{10} + q_{01}} + c \Rightarrow c = \frac{q_{01}}{q_{10} + q_{01}}$$

- Finally transition probability function $P_{00}(t)$

$$P_{00}(t) = \frac{q_{10}}{q_{10} + q_{01}} + \frac{q_{01}}{q_{10} + q_{01}} e^{-(q_{10}+q_{01})t}$$

- Repeat for $P_{11}(t)$. Same exponent, different constants

$$P_{11}(t) = \frac{q_{01}}{q_{10} + q_{01}} + \frac{q_{10}}{q_{10} + q_{01}} e^{-(q_{10}+q_{01})t}$$

- As time goes to infinity, $P_{00}(t)$ and $P_{11}(t)$ converge exponentially
  - $\Rightarrow$ Convergence rate depends on magnitude of $q_{10} + q_{01}$

# Convergence of transition probabilities

- Recall $P_{01}(t) = 1 - P_{00}(t)$ and $P_{10}(t) = 1 - P_{11}(t)$
- Limiting (steady-state) probabilities are

$$\lim_{t\to\infty} P_{00}(t) = \frac{q_{10}}{q_{10}+q_{01}}, \qquad \lim_{t\to\infty} P_{01}(t) = \frac{q_{01}}{q_{10}+q_{01}}$$
$$\lim_{t\to\infty} P_{11}(t) = \frac{q_{01}}{q_{10}+q_{01}}, \qquad \lim_{t\to\infty} P_{10}(t) = \frac{q_{10}}{q_{10}+q_{01}}$$

- Limit distribution exists and is independent of initial condition
  - $\Rightarrow$ Compare across diagonals

# Kolmogorov's forward equations in matrix form

- Restrict attention to finite CTMCs with $N$ states

  $\Rightarrow$ Define matrix $\mathbf{R} \in \mathbb{R}^{N \times N}$ with elements $r_{ij} = q_{ij}$, $r_{ii} = -\nu_i$

- Rewrite Kolmogorov's forward eqs. as (process runs a little longer)

$$P'_{ij}(t) = \sum_{k=1, k \neq j}^{N} q_{kj} P_{ik}(t) - \nu_j P_{ij}(t) = \sum_{k=1}^{N} r_{kj} P_{ik}(t)$$

- Right-hand side defines elements of a matrix product

$$r_{1j}P_{i1}(t) \qquad r_{kj}P_{ik}(t) \qquad r_{Nj}P_{iN}(t)$$

$$\begin{pmatrix} r_{11} & \cdot & r_{1j} & \cdot & r_{1N} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{k1} & \cdot & r_{kj} & \cdot & r_{kN} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{N1} & \cdot & r_{Nj} & \cdot & r_{NN} \end{pmatrix} = \mathbf{R}$$

$$\mathbf{P}(t) = \begin{pmatrix} P_{11}(t) & \cdot & P_{1k}(t) & \cdot & P_{1N}(t) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ P_{i1}(t) & \cdot & P_{ik}(t) & \cdot & P_{iN}(t) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ P_{N1}(t) & \cdot & P_{Nk}(t) & \cdot & P_{JN}(t) \end{pmatrix} \qquad \begin{pmatrix} s_{11} & \cdot & s_{1j} & \cdot & s_{1N} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ s_{i1} & \cdot & s_{ij} & \cdot & s_{iN} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ s_{N1} & \cdot & s_{Nk} & \cdot & s_{NN} \end{pmatrix} = \mathbf{P}(t)\mathbf{R} = \mathbf{P}'(t)$$

# Kolmogorov's backward equations in matrix form

- Similarly, Kolmogorov's backward eqs. (process starts a little later)

$$P_{ij}^{'}(t) = \sum_{k=1,k\neq i}^{N} q_{ik}P_{kj}(t) - \nu_i P_{ij}(t) = \sum_{k=1}^{N} r_{ik}P_{kj}(t)$$

- Right-hand side also defines a matrix product

$r_{i1}P_{1j}(t)$ $\quad r_{ik}P_{kj}(t)$ $\quad r_{iN}P_{Nj}(t)$

$$\begin{pmatrix} P_{11}(t) & \cdot & P_{1j}(t) & \cdot & P_{1N}(t) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ P_{k1}(t) & \cdot & P_{kj}(t) & \cdot & P_{kN}(t) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ P_{N1}(t) & \cdot & P_{Nj}(t) & \cdot & P_{NN}(t) \end{pmatrix} = \mathbf{P}(t)$$

$$\mathbf{R} = \begin{pmatrix} r_{11} & \cdot & r_{1k} & \cdot & r_{1N} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{i1} & \cdot & r_{ik} & \cdot & r_{iN} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{N1} & \cdot & r_{Nk} & \cdot & r_{JN} \end{pmatrix} \quad \begin{pmatrix} s_{11} & \cdot & s_{1j} & \cdot & s_{1N} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ s_{i1} & \cdot & s_{ij} & \cdot & s_{iN} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ s_{N1} & \cdot & s_{Nk} & \cdot & s_{NN} \end{pmatrix} = \mathbf{RP}(t) = \mathbf{P}^{'}(t)$$

# Kolmogorov's equations in matrix form

- Matrix form of Kolmogorov's forward equation $\Rightarrow \mathbf{P}'(t) = \mathbf{P}(t)\mathbf{R}$
- Matrix form of Kolmogorov's backward equation $\Rightarrow \mathbf{P}'(t) = \mathbf{R}\mathbf{P}(t)$
  - $\Rightarrow$ More similar than apparent
  - $\Rightarrow$ But not equivalent because matrix product not commutative
- Notwithstanding both equations have to accept the same solution

# Matrix exponential

- Kolmogorov's equations are first-order linear differential equations
  - $\Rightarrow$ They are coupled, $P'_{ij}(t)$ depends on $P_{kj}(t)$ for all $k$
  - $\Rightarrow$ Accepts exponential solution $\Rightarrow$ Define matrix exponential
- **Def:** The matrix exponential $e^{\mathbf{A}t}$ of matrix $\mathbf{A}t$ is the series

$$e^{\mathbf{A}t} = \sum_{n=0}^{\infty} \frac{(\mathbf{A}t)^n}{n!} = \mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2} + \frac{(\mathbf{A}t)^3}{2 \times 3} + \ldots$$

- Derivative of matrix exponential with respect to $t$

$$\frac{\partial e^{\mathbf{A}t}}{\partial t} = \mathbf{0} + \mathbf{A} + \mathbf{A}^2 t + \frac{\mathbf{A}^3 t^2}{2} + \ldots = \mathbf{A}\left(\mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2} + \ldots\right) = \mathbf{A}e^{\mathbf{A}t}$$

- Putting $\mathbf{A}$ on right side of product shows that $\Rightarrow \dfrac{\partial e^{\mathbf{A}t}}{\partial t} = e^{\mathbf{A}t}\mathbf{A}$

# Solution of Kolmogorov's equations

- Propose solution of the form $\mathbf{P}(t) = e^{\mathbf{R}t}$
- $\mathbf{P}(t)$ solves backward equations, since derivative is

$$\frac{\partial \mathbf{P}(t)}{\partial t} = \frac{\partial e^{\mathbf{R}t}}{\partial t} = \mathbf{R}e^{\mathbf{R}t} = \mathbf{R}\mathbf{P}(t)$$

- It also solves forward equations

$$\frac{\partial \mathbf{P}(t)}{\partial t} = \frac{\partial e^{\mathbf{R}t}}{\partial t} = e^{\mathbf{R}t}\mathbf{R} = \mathbf{P}(t)\mathbf{R}$$

- Notice that $\mathbf{P}(0) = \mathbf{I}$, as it should ($P_{ii}(0) = 1$, and $P_{ij}(0) = 0$)

# Computing the matrix exponential

- Suppose $\mathbf{A} \in \mathbb{R}^{n \times n}$ is diagonalizable, i.e., $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}$
  - $\Rightarrow$ Diagonal matrix $\mathbf{D} = \text{diag}(\lambda_1, \ldots, \lambda_n)$ collects eigenvalues $\lambda_i$
  - $\Rightarrow$ Matrix $\mathbf{U}$ has the corresponding eigenvectors as columns
- We have the following neat identity

$$e^{\mathbf{A}t} = \sum_{n=0}^{\infty} \frac{(\mathbf{U}\mathbf{D}\mathbf{U}^{-1}t)^n}{n!} = \mathbf{U}\left(\sum_{n=0}^{\infty} \frac{(\mathbf{D}t)^n}{n!}\right)\mathbf{U}^{-1} = \mathbf{U}e^{\mathbf{D}t}\mathbf{U}^{-1}$$

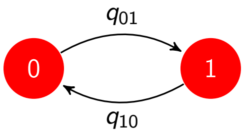
- But since $\mathbf{D}$ is diagonal, then

$$e^{\mathbf{D}t} = \sum_{n=0}^{\infty} \frac{(\mathbf{D}t)^n}{n!} = \begin{pmatrix} e^{\lambda_1 t} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & e^{\lambda_n t} \end{pmatrix}$$

# Two state CTMC example

Ex: Simplest CTMC with two states 0 and 1

- Transition rates are $q_{01} = 3$ and $q_{10} = 1$




- Recall transition time rates are $\nu_0 = q_{01} = 3$, $\nu_1 = q_{10} = 1$, hence

$$\mathbf{R} = \begin{pmatrix} -\nu_0 & q_{01} \\ q_{10} & -\nu_1 \end{pmatrix} = \begin{pmatrix} -3 & 3 \\ 1 & -1 \end{pmatrix}$$

- Eigenvalues of $\mathbf{R}$ are $0, -4$, eigenvectors $[1, 1]^T$ and $[-3, 1]^T$. Thus

$$\mathbf{U} = \begin{pmatrix} 1 & -3 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{U}^{-1} = \begin{pmatrix} 1/4 & 3/4 \\ -1/4 & 1/1 \end{pmatrix}, \quad e^{\mathbf{D}t} = \begin{pmatrix} 1 & 0 \\ 0 & e^{-4t} \end{pmatrix}$$

- The solution to the forward equations is

$$\mathbf{P}(t) = e^{\mathbf{R}t} = \mathbf{U}e^{\mathbf{D}t}\mathbf{U}^{-1} = \begin{pmatrix} 1/4 + (3/4)e^{-4t} & 3/4 - (3/4)e^{-4t} \\ 1/4 - (1/4)e^{-4t} & 3/4 + (1/4)e^{-4t} \end{pmatrix}$$

# Recurrent and transient states

- Recall the embedded discrete-time MC associated with any CTMC
  - $\Rightarrow$ Transition probs. of MC form the matrix $\mathbf{P}$ of the CTMC
  - $\Rightarrow$ No self transitions ($P_{ii} = 0$, $\mathbf{P}$'s diagonal null)
- States $i \leftrightarrow j$ communicate in the CTMC if $i \leftrightarrow j$ in the MC
  - $\Rightarrow$ Communication partitions MC in classes
  - $\Rightarrow$ Induces CTMC partition as well
- **Def:** CTMC is irreducible if embedded MC contains a single class
- State $i$ is recurrent if it is recurrent in the embedded MC
  - $\Rightarrow$ Likewise, define transience and positive recurrence for CTMCs
- Transience and recurrence shared by elements of a MC class
  - $\Rightarrow$ Transience and recurrence are class properties of CTMCs
- Periodicity not possible in CTMCs

# Limiting probabilities

Theorem

*Consider irreducible, positive recurrent CTMC with transition rates* $\nu_i$ *and* $q_{ij}$. *Then,* $\lim_{t\to\infty} P_{ij}(t)$ *exists and is independent of the initial state* $i$, *i.e.,*

$$P_j = \lim_{t\to\infty} P_{ij}(t) \quad \textit{exists for all } (i,j)$$

*Furthermore, steady-state probabilities* $P_j \geq 0$ *are the unique nonnegative solution of the system of linear equations*

$$\nu_j P_j = \sum_{k=0, k\neq j}^{\infty} q_{kj} P_k, \qquad \sum_{j=0}^{\infty} P_j = 1$$

- Limit distribution exists and is independent of initial condition
  - $\Rightarrow$ Obtained as solution of system of linear equations
  - $\Rightarrow$ Like discrete-time MCs, but equations slightly different

# Algebraic relation to determine limit probabilities

- As with MCs difficult part is to prove that $P_j = \lim_{t\to\infty} P_{ij}(t)$ exists

- Algebraic relations obtained from Kolmogorov's forward equations

$$\frac{\partial P_{ij}(t)}{\partial t} = \sum_{k=0,k\neq j}^{\infty} q_{kj} P_{ik}(t) - \nu_j P_{ij}(t)$$

- If limit distribution exists we have, independent of initial state $i$

$$\lim_{t\to\infty} \frac{\partial P_{ij}(t)}{\partial t} = 0, \qquad \lim_{t\to\infty} P_{ij}(t) = P_j$$

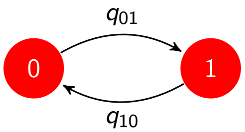
- Considering the limit of Kolomogorov's forward equations yields

$$0 = \sum_{k=0,k\neq j}^{\infty} q_{kj} P_k - \nu_j P_j$$

- Reordering terms the limit distribution equations follow

# Two state CTMC example

Ex: Simplest CTMC with two states 0 and 1




- Transition rates are $q_{01}$ and $q_{10}$
- From transition rates find mean transition times $\nu_0 = q_{01}$, $\nu_1 = q_{10}$
- Stationary distribution equations

$$\nu_0 P_0 = q_{10} P_1, \qquad \nu_1 P_1 = q_{01} P_0, \qquad P_0 + P_1 = 1,$$
$$q_{01} P_0 = q_{10} P_1, \qquad q_{10} P_1 = q_{01} P_0$$

- Solution yields $\Rightarrow P_0 = \dfrac{q_{10}}{q_{10} + q_{01}}, \qquad P_1 = \dfrac{q_{01}}{q_{10} + q_{01}}$
- Larger rate $q_{10}$ of entering 0 $\Rightarrow$ Larger prob. $P_0$ of being at 0
- Larger rate $q_{01}$ of entering 1 $\Rightarrow$ Larger prob. $P_1$ of being at 1

# Ergodicity

- **Def:** Fraction of time $T_i(t)$ spent in state $i$ by time $t$

$$T_i(t) := \frac{1}{t}\int_0^t \mathbb{I}\left\{X(\tau) = i\right\} d\tau$$

$\Rightarrow$ $T_i(t)$ a time/ergodic average, $\lim\limits_{t\to\infty} T_i(t)$ is an ergodic limit

- If CTMC is irreducible, positive recurrent, the ergodic theorem holds

$$P_i = \lim_{t\to\infty} T_i(t) = \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{I}\left\{X(\tau) = i\right\} d\tau \qquad \text{a.s.}$$

- Ergodic limit coincides with limit probabilities (almost surely)

# Function's ergodic limit

- Consider function $f(i)$ associated with state $i$. Can write $f\big(X(t)\big)$ as

$$f\big(X(t)\big) = \sum_{i=1}^{\infty} f(i)\mathbb{I}\left\{X(t) = i\right\}$$

- Consider the time average of $f\big(X(t)\big)$

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f\big(X(\tau)\big)d\tau = \lim_{t\to\infty} \frac{1}{t}\int_0^t \sum_{i=1}^{\infty} f(i)\mathbb{I}\left\{X(\tau) = i\right\}d\tau$$

- Interchange summation with integral and limit to say

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f\big(X(\tau)\big)d\tau = \sum_{i=1}^{\infty} f(i)\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{I}\left\{X(\tau) = i\right\}d\tau = \sum_{i=1}^{\infty} f(i)P_i$$

- Function's ergodic limit = Function's expectation under limiting dist.

# Limit distribution equations as balance equations

- Recall limit distribution equations $\Rightarrow \nu_j P_j = \sum_{k=0, k\neq j}^{\infty} q_{kj} P_k$
- $P_j$ = fraction of time spent in state $j$
- $\nu_j$ = rate of transition out of state $j$ given CTMC is in state $j$

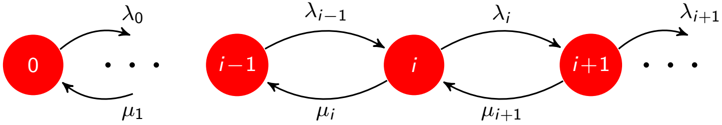
  $\Rightarrow \nu_j P_j$ = rate of transition out of state $j$ (unconditional)
- $q_{kj}$ = rate of transition from $k$ to $j$ given CTMC is in state $k$

  $\Rightarrow q_{kj} P_k$ = rate of transition from $k$ to $j$ (unconditional)

  $\Rightarrow \sum_{k=0, k\neq j}^{\infty} q_{kj} P_k$ = rate of transition into $j$, from all states
- Rate of transition out of state $j$ = Rate of transition into state $j$
- Balance equations $\Rightarrow$ Balance nr. of transitions in and out of state $j$

# Limit distribution for birth and death process

- Birth/deaths occur at state-dependent rates. When $X(t) = i$
- Births $\Rightarrow$ Individuals added at exponential times with mean $1/\lambda_i$
  $\Rightarrow$ Birth rate = upward transition rate $= q_{i,i+1} = \lambda_i$
- Deaths $\Rightarrow$ Individuals removed at exponential times with mean $1/\mu_i$
  $\Rightarrow$ Death rate = downward transition rate $= q_{i,i-1} = \mu_i$
- Transition time rates $\Rightarrow \nu_i = \lambda_i + \mu_i,\ i > 0$ and $\nu_0 = \lambda_0$




- Limit distribution/balance equations: Rate out of $j$ = Rate into $j$

$$(\lambda_i + \mu_i)P_i = \lambda_{i-1}P_{i-1} + \mu_{i+1}P_{i+1}$$
$$\lambda_0 P_0 = \mu_1 P_1$$