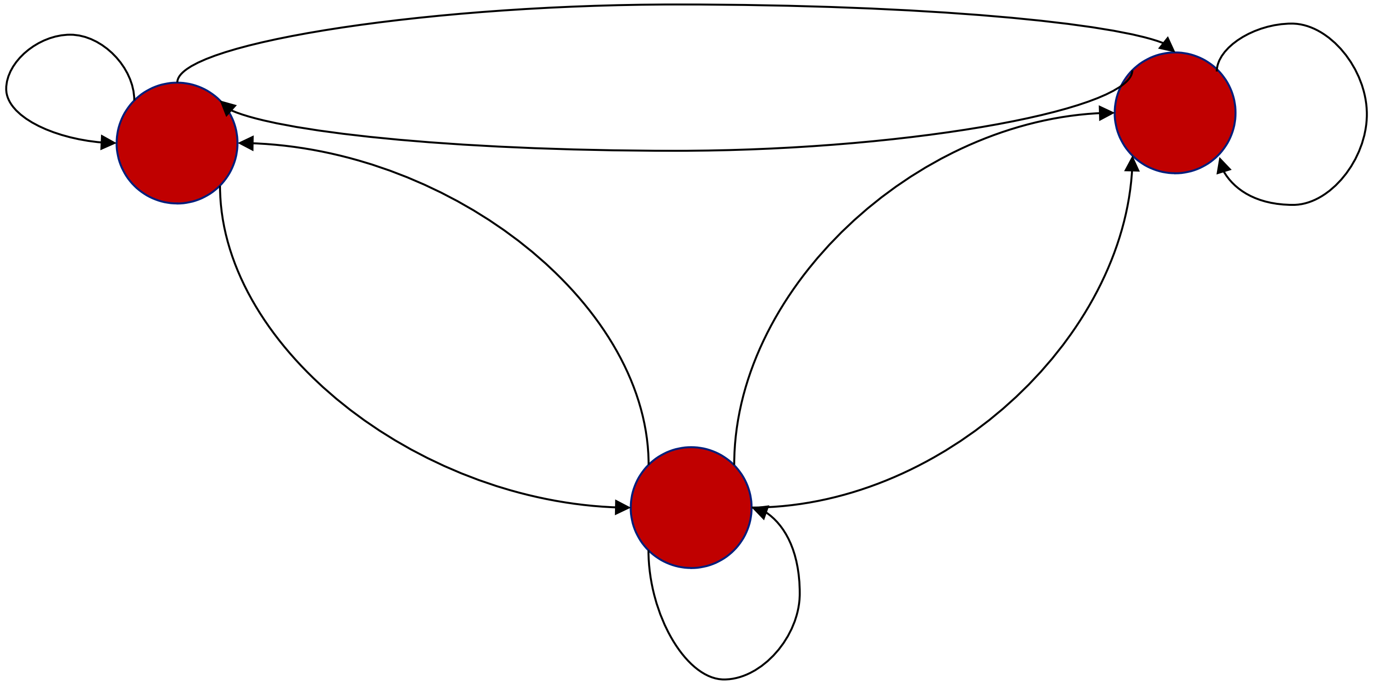
Continuous Time Markov Chains



Markov Chains: Time and State Space Continuity

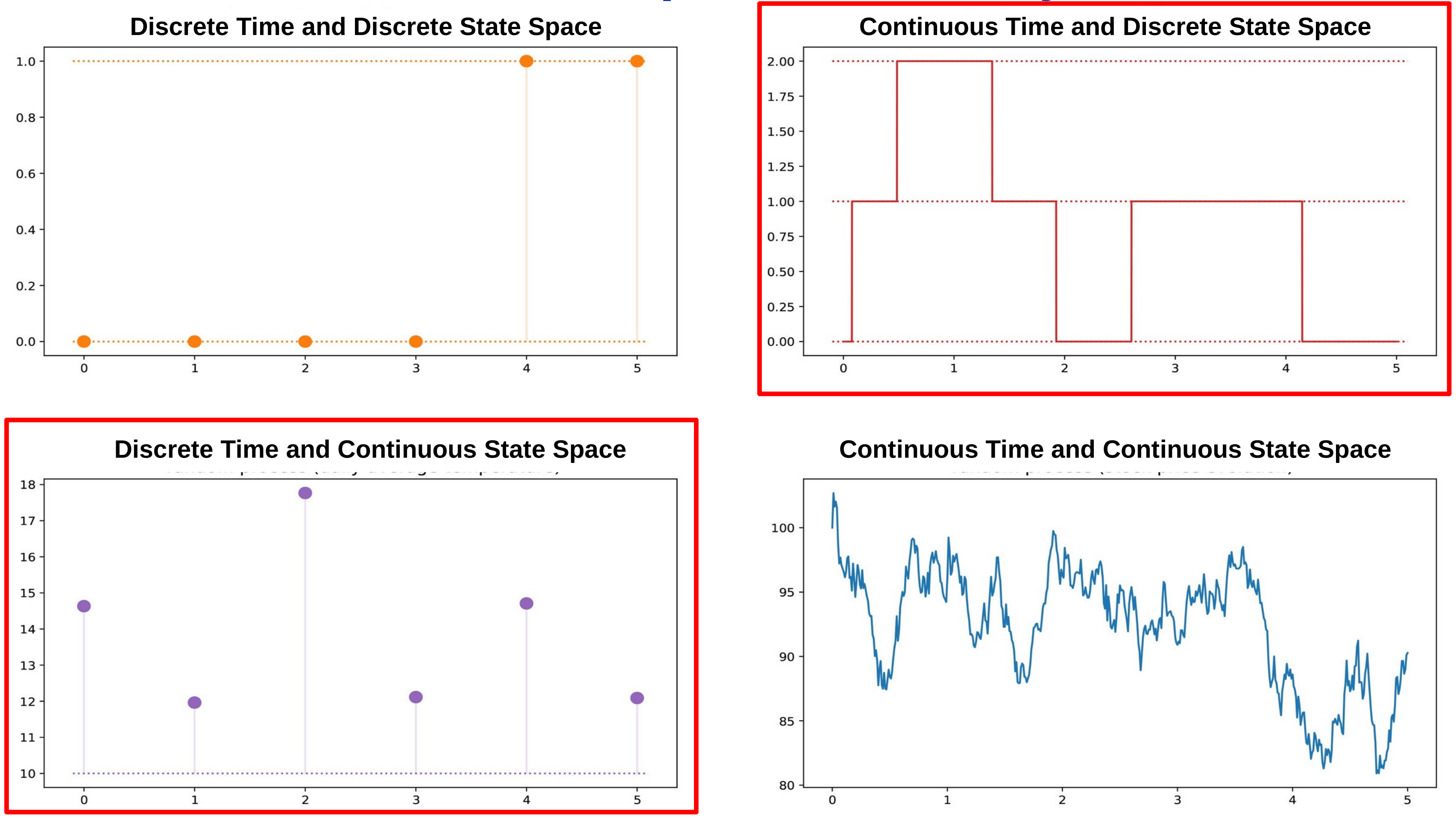


Table of Contents

Some Definitions

- Stochastic process: $\{X(t), t \in [0,\infty)\}$ and discrete State Space: {1,2,3,…}
- Markovian Property (for continuous time):

$$\mathrm{P}\big(X(t+s)=j \mid X(s)=i,\ \{X(u),\ 0\le u\le s\}\big) = P\big(X(t+s)=j \mid X(s)=i\big) =: p_{ij}(s,t+s)$$

If this is fulfilled, then X(t) is called a Continuous Time Markov Chain (CTMC).

- We also assume time stationary transition probabilities ($p_{ij}(t)$ only depends on t):
- This implies: $P_{ij}(t) = P(X(t)=j|X(0)=i)$
- In the discrete setting we had:

$$\mathrm{P}(X_{n+1}=j \mid X_n=i_n, X_{n-1}=i_{n-1},\dots,X_0=i_0) = \mathrm{P}(X_{n+1}=j \mid X_n=i_n)$$

$$p_{ij}(n) := \mathrm{P}(X_{n+1}=j \mid X_n=i) \longrightarrow p_{ij} := \mathrm{P}(X_{n+1}=j \mid X_n=i)$$

The P Matrix

- $\mathbf{P}(t) = \big(p_{ij}(t)\big) \in \mathbb{R}^{K\times K}$ in the discrete setting we had: $\mathbf{P}(n) = \big(p_{ij}(n)\big) \in \mathbb{R}^{K\times K}$
- We also have: $\mathbf{P}(0) = \mathbf{I}$
- The row vectors are: $\boldsymbol{p}(t) = \big(p_1(t), \ldots, p_K(t)\big)$
- Reasonable assumption: $\sum_j p_{ij}(t) = 1$,so no $p_{ij}(0) = \delta_{ij}$
- Chapman-Kolmogorov: $P(s+t) = P(s)P(t).$ I ii in the discrete set $\mathbf{P}^n = \mathbf{P}^k\mathbf{P}^{n-k}$

$$P(t) = \begin{bmatrix} p_{11}(t) & p_{12}(t) & \ldots & p_{1r}(t) \\ p_{21}(t) & p_{22}(t) & \ldots & p_{2r}(t) \\ \vdots & \vdots & \vdots & \vdots \\ p_{r1}(t) & p_{r2}(t) & \ldots & p_{rr}(t) \end{bmatrix}$$

The P Matrix

- Let {X(t), t ≥ 0} be a CTMC, then we have:

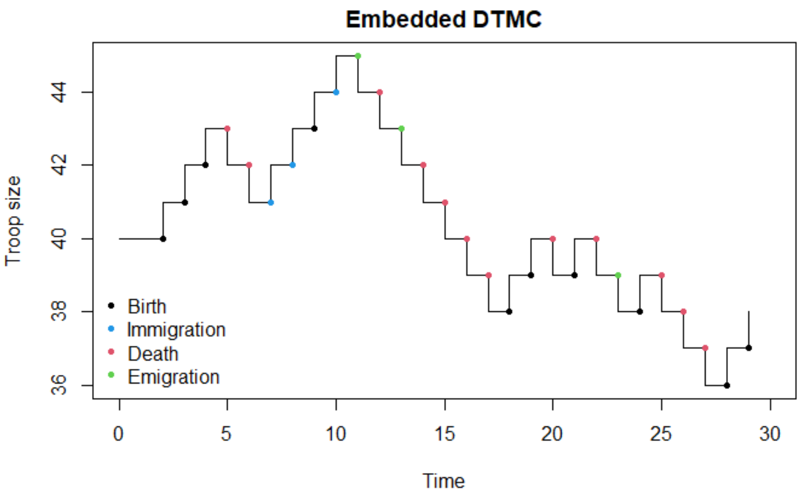
$p_i(t) = \mathrm{P}\big(X(t) = i\big)$ is called the absolute state probability at time t.

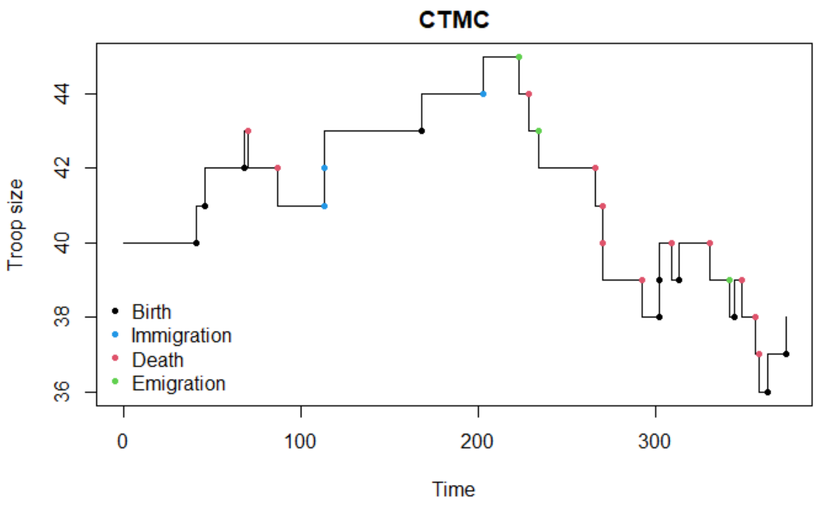
$\{p_i(t), i \in S\}$ is called the absolute probability distribution of the Markov chain at time t.

$\{p_i(0), i \in S\}$ is the initial probability distribution of the Markov chain. $\diamondsuit$

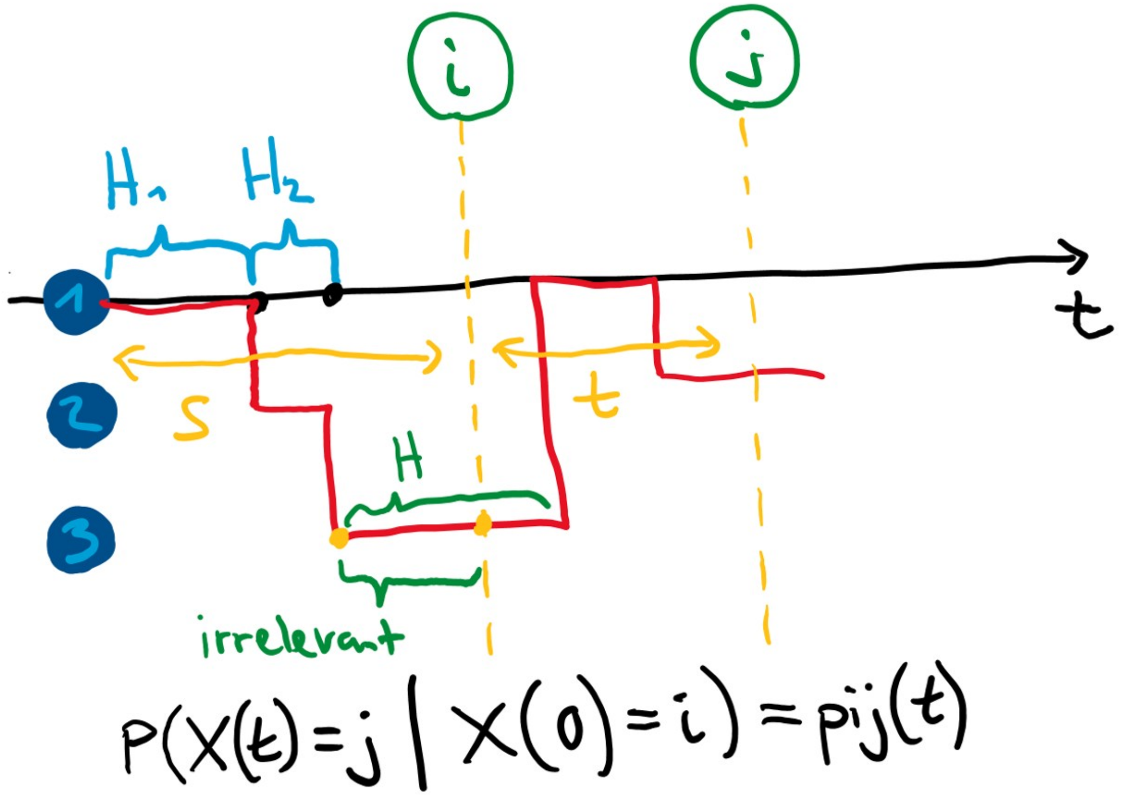
Discrete vs Continuous Time Markov Processes

- DTMC: Initial Distribution and Transition Matrix (where we go)
- CTMC: Initial Distribution and Transition Matrix (where we go and **when we change state**)
- To describe this when we use **Holding Times**, but how can we describe them?





Holding times



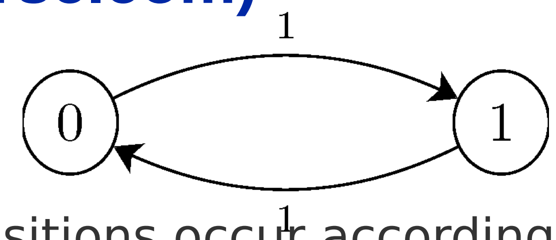
Holding times

- H_i has to fulfill the memoryless property: P(H > a + b | h > a) = P (h > b), where a >b are ∈ t∈[0,∞),

→ Only continuous distribution that fulfills the memoryless property is the exponential distribution

→ **In a CTMC, the holding times H_i are exponentially distributed $H_i \sim Exp(q_i)$.**

P(t) matrix: A simple example (from: probabilitycourse.com)



We have a Transition Matrix (Jump Chain)

We have holding time parameters = $\lambda_0 = \lambda_1 > 0$ and since $H_i \sim \exp(\lambda)$ the transitions occur according to poisson process:

$$P_{00}(t) = P(X(t) = 0 | X(0) = 0) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^{2n}}{(2n)!} = \frac{1}{2} + \frac{1}{2} e^{-2\lambda t}$$

$$P_{01}(t) = 1 - P_{00}(t) = \frac{1}{2} - \frac{1}{2} e^{-2\lambda t}$$

Because of symmetry we have:

$$P_{11}(t) = P_{00}(t) \quad \text{and} \quad P_{10}(t) = P_{01}(t)$$

$$P(t) = \begin{bmatrix} \frac{1}{2} + \frac{1}{2} e^{-2\lambda t} & \frac{1}{2} - \frac{1}{2} e^{-2\lambda t} \\ \frac{1}{2} - \frac{1}{2} e^{-2\lambda t} & \frac{1}{2} + \frac{1}{2} e^{-2\lambda t} \end{bmatrix}$$

The Generator Matrix

- Often it is very complicated to calculate P(t)….
- Other way to describe a CTMC using the Generator Matrix, based on exponential holding times: $H_i \sim \exp(q_i)$
- We define:

$$q_i = \lim_{h \to 0} \frac{1 - p_{ii}(h)}{h}$$

$$g_{ij} = \lim_{h \to 0} \frac{p_{ij}(h)}{h}$$

The parameters q_i are the unconditional transition rates of leaving state i for any other state and the parameters g_{ij} are the conditional transition rates of making a transition from state i to state j.

These **transition rates** define the Generator matrix.

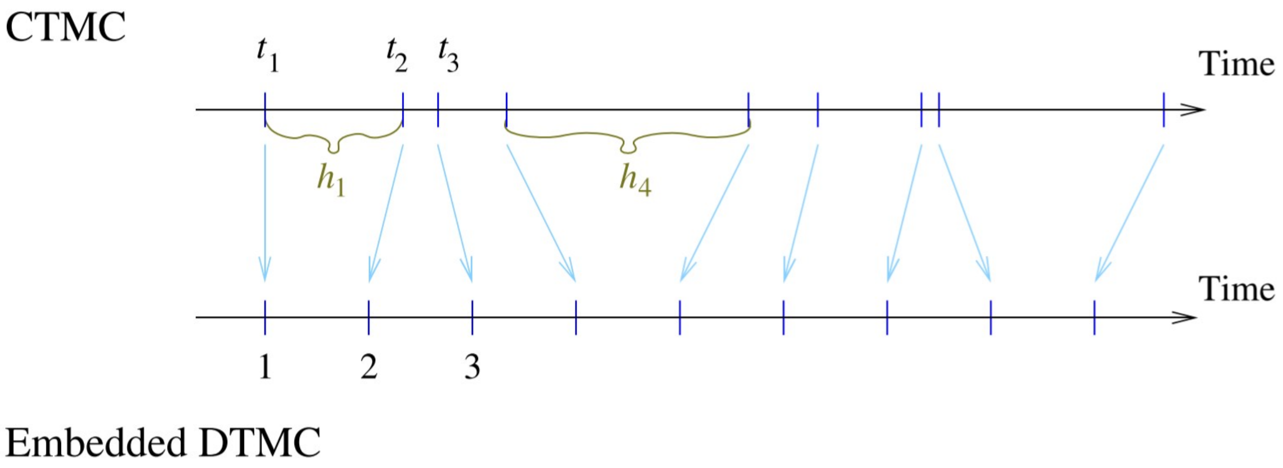
Some more Generator Matrix properties (all from last week)

- $$\mathbf{P}(t) = \exp(-t\mathbf{G}) = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathbf{G}^k$$
- $$q_i = \sum_{j=1, j\neq i}^{K} g_{ij} \text{ and thus } q_i = -g_{ii}$$
- $$\pi G = 0$$

The Embedded Markov Chain (DTMC) (Jump Chain)

$$\mathrm{P}(\text{ transition from } i \text{ to } j \text{ in } [t, t+h] \mid \text{ transition occurred }) = \frac{p_{ij}(h)}{1 - p_{ii}} \approx \frac{g_{ij}}{q_i}$$

- For a CTMC with generator matrix G, we define $p_{ij} = g_{ij}/q_i$. The DTMC with transition matrix $P = (p_{ij})$ is the associated embedded DTMC
- we directly have $p_{ii} = 0$, no self-transition



$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

Classification of CTMC

- Very similar to the DTMC for the CTMC we use the embedded DTMC and we can say:
- The accessibility relation divides states into different classes
- A CTMC is irreducible if and only if its embedded chain is irreducible
- A state is recurrent/transient for a CTMC if and only if it is recurrent/transient for its embedded chain
- CTMC with transition probabilities $p_{ij}(t)$, a state i is recurrent or transient if:

$$\int_0^\infty p_{ii}(t)\,\mathrm{d}t = \infty \qquad \text{or} \qquad \int_0^\infty p_{ii}(t)\,\mathrm{d}t < \infty,$$

- Compared to the discrete case: a recurrent state $\sum_{n=1}^{\infty}(\mathbf{P}^n)_{ii} = \infty$

Classification of CTMC (Ergodic Chains)

- **Definition**: An **irreducible aperiodic** positive chain is called **ergodic**.
- We well see later that these have some handy properties later….

Stationarity of a CMTC

Let $X(t)$ be a continuous-time Markov chain with transition matrix $P(t)$ and state space $S = \{0, 1, 2, \cdots\}$. A probability distribution π on S, i.e, a vector $\pi = [\pi_0, \pi_1, \pi_2, \cdots]$, where $\pi_i \in [0, 1]$ and

$$\sum_{i \in S} \pi_i = 1,$$

is said to be a **stationary distribution** for $X(t)$ if

$$\pi = \pi P(t), \quad \text{for all } t \geq 0.$$

Stationarity of a CTMC the limiting Distribution

The probability distribution $\pi = [\pi_0, \pi_1, \pi_2, \cdots]$ is called the **limiting distribution** of the continuous-time Markov chain $X(t)$ if

$$\pi_j = \lim_{t \to \infty} P(X(t) = j | X(0) = i)$$

for all $i, j \in S$, and we have

$$\sum_{j \in S} \pi_j = 1.$$

- for ergodic (aperiodic, irreducible) MC the limiting and the stationary distribution is the same.
- And we also have $0 = \pi G$

Stationarity of a CTMC: Using the embedded Chain

Let $\{X(t), t \geq 0\}$ be a continuous-time Markov chain with an irreducible positive recurrent jump chain. Suppose that the unique stationary distribution of the jump chain is given by

$$\tilde{\pi} = \left[\tilde{\pi}_0, \tilde{\pi}_1, \tilde{\pi}_2, \cdots\right].$$

Further assume that

$$0 < \sum_{k \in S} \frac{\tilde{\pi}_k}{\lambda_k} < \infty.$$

Then,

$$\pi_j = \lim_{t \to \infty} P(X(t) = j | X(0) = i) = \frac{\frac{\tilde{\pi}_j}{\lambda_j}}{\sum_{k \in S} \frac{\tilde{\pi}_k}{\lambda_k}}.$$

for all $i, j \in S$. That is, $\pi = [\pi_0, \pi_1, \pi_2, \cdots]$ is the limiting distribution of $X(t)$.

Example with $p_i = p_i * P(t)$ (using P(t) from before)

$$P(t) = \begin{bmatrix} \frac{1}{2} + \frac{1}{2}e^{-2\lambda t} & \frac{1}{2} - \frac{1}{2}e^{-2\lambda t} \\ \frac{1}{2} - \frac{1}{2}e^{-2\lambda t} & \frac{1}{2} + \frac{1}{2}e^{-2\lambda t} \end{bmatrix}$$

For $\pi = [\pi_0, \pi_1]$, we obtain

$$\pi P(t) = [\pi_0, \pi_1] \begin{bmatrix} \frac{1}{2} + \frac{1}{2}e^{-2\lambda t} & \frac{1}{2} - \frac{1}{2}e^{-2\lambda t} \\ \frac{1}{2} - \frac{1}{2}e^{-2\lambda t} & \frac{1}{2} + \frac{1}{2}e^{-2\lambda t} \end{bmatrix} = [\pi_0, \pi_1].$$

We also need

$$\pi_0 + \pi_1 = 1.$$

Solving the above equations, we obtain

$$\pi_0 = \pi_1 = \frac{1}{2}.$$

Example using the embedded chain

holding time parameters are given by $\lambda_1 = 2,\ \lambda_2 = 1$, and $\lambda_3 = 3$

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

Using $\tilde{\pi} P = \tilde{\pi}$, we arrive at $\tilde{\pi} = \frac{1}{5}[1,\ 2,\ 2]$

$$\pi_j = \frac{\frac{\tilde{\pi}_j}{\lambda_j}}{\sum_{k \in S} \frac{\tilde{\pi}_k}{\lambda_k}}.$$

– By plugging in λ and :

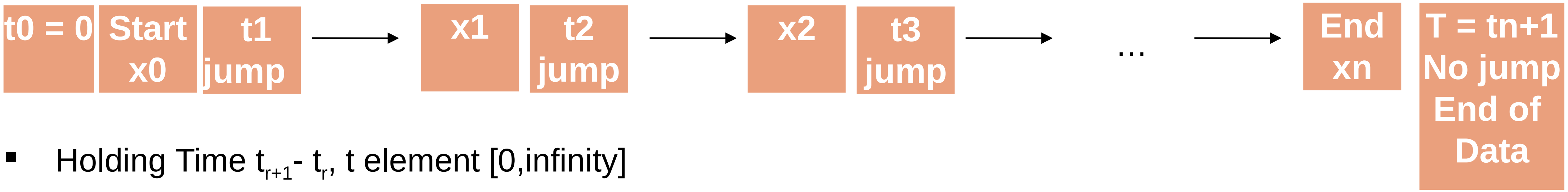
we conclude that $\pi = \frac{1}{19}[3, 12, 4]$ is the limiting distribution of $X(t)$.

Estimating the Generator Matrix Theory

- Often very difficult to find P(t)
- Using the Generator matrix (G)
- Estimate with real world data
- To find G we have to estimate q_i and g_{ij}, but how…?
- Since the holding times are exponential, it is possible to setup a likelihood function.
- Likelihood function based on the number of jumps, the holding times and the jump chain (this information provides a sufficient statistic)
- v = {n, x0, t1, x1, t2, x2, . . . , tn, xn} represent the observed chain

Estimating the Generator Matrix: Likelihood Function Intuition

- Likelihood Function of Form:



- Holding Time t_{r+1}- t_r, t element [0,infinity]
- t_0 = 0 and t_{n+1} = T
- We assume Fisher Regularity Condition

Estimating the Generator Matrix: Likelihood Function

$$L(\mathbf{G}; v) = p_{x_0}(0) \times q_{x_0} \exp(-q_{x_0} t_1) \times \prod_{j=1}^{n-1} \frac{g_{x_{j-1},x_j}}{q_{x_{j-1}}} \cdot q_{x_j} \exp\big(-q_{x_j}(t_{j+1} - t_j)\big) \times \frac{g_{x_{n-1},x_n}}{q_{x_{n-1}}} \cdot \exp\big(-q_{x_j}(T - t_n)\big)$$

$$\text{P(transition from } i \text{ to } j \text{ in } [t, t+h] \mid \text{ transition occurred)} = \frac{p_{ij}(h)}{1 - p_{ii}} \approx \frac{g_{ij}}{q_i}$$

Estimating the Generator Matrix: Likelihood Function and Results

$$= p_{x_0}(0) \times \left(\prod_{j=1}^{n} g_{x_{j-1},x_j}\right) \times \prod_{j=0}^{n} \exp\left(-q_{x_j}(t_{j+1} - t_j)\right)$$

$$= p_{x_0}(0) \left(\prod_{\substack{i,j \\ i \neq j}}^{K} g_{i,j}^{N(i,j)}\right) \exp\left(-\sum_{i=1}^{K} A(i) q_i\right), \quad \text{where } N(i,j) \text{ is the number of transitions from } i \text{ to } j \text{ and } A(i) \text{ is the total time spend in state } i.$$

- Finding The ML Estimators through the normal likelihood approach:

$$\widehat{g}_{ij} = \frac{N(i,j)}{A(i)} \qquad \widehat{q}_i = \frac{\sum_{j \neq i} N(i,j)}{A(i)}$$

- Also it can be shown that:

$$\sqrt{A(i)\widehat{g}_{ij}}\,(\widehat{g}_{ij} - g_{ij}) \xrightarrow{t \to \infty} \mathcal{N}(0,1) \qquad \text{in distribution.}$$

Estimating the Generator Matrix: Ergodic Chains

If we have an ergodic chain, it is possible to show that

$$\frac{A(i)}{t} = \frac{A_t(i)}{t} \xrightarrow{t\to\infty} \pi_i \qquad \text{in probability,}$$

$$\frac{N(i,j)}{t} = \frac{N_t(i,j)}{t} \xrightarrow{t\to\infty} \pi_i g_{ij} \qquad \text{in probability.}$$

- A(i) = Total time spent in state i
- N(i,j) Number of transitions from state i to j

Example: Baboons in R

- Using Real World Data
- Overview
- Find Generator (G) matrix using Likelihood approach
- Stationary distribtion (= limiting Distribution)
- Find the embedded markov chain
- Try to find the P(t) Matrix

Picture from: www.smithsonianmag.com

Example: Baboon discussion

- What are some of the problems with the approach that we did?

DTMC on continuous state space: Theory (based on Quan Lin Li)

- A Markov chain is a sequence of random variables $X_1, X_2, \ldots, X_n$, taking values in state space Ω
- Markovian Property.
- We have discrete time: T = (1,2,3,…) and state space is Ω continuous with values from (a,b)
- transition kernel K(x,A)

$$P\{X_{k+1} \in A \mid X_k, X_{k-1}, \ldots, X_0\} = K(X_k, A).$$

It is clear that $K(x, A)$ denotes the probability to move in one step from the state x into the state set A. The transition kernel $K(x, A)$ has two main properties as follows:

(1) $K(x, \cdot)$ is a probability measure for each $x \in \Omega$, and
(2) $K(\cdot, A)$ is measurable for each $A \subset \Omega$.

More Theory

If there exists a function $K(x, y)$ such that for all $x \in \Omega$ and $A \subset \Omega$,

$$K(x, A) = \int_A K(x, y)\mathrm{d}y$$

then $K(x, y)$ is said to be a density of the transition kernel $K(x, A)$. We write

Stationarity

If the kernel $K(x, A)$ is well behaved, then the Markov chain will have a stationary distribution $\pi(x)$ such that

$$\pi(y) = \int_{\Omega} \pi(x) K(x, y) \mathrm{d}x, \quad y \in \Omega, \tag{5.1}$$

or

$$\pi(A) = \int_{\Omega} \pi(x) K(x, A) \mathrm{d}x, \quad A \in \Omega. \tag{5.2}$$

It is clear that $\pi(y)$ is a density of the stationary probability $\pi(A)$.

Stationarity

For the DTMC we still have that:

- Every aperiodic, irreducible stochastic matrix has a stationary distribution [Perron-Frobenius Theorem]
- **Irreducibility** definition:

  for any $x, y \in \Omega$, there always exists a positive integer n such that $K^n(x, y) > 0$
- **Aperiodic** definition:

  aperiodicity means that there exist no subsets of the state space Ω that can only be periodically

  visited by the Markov chain

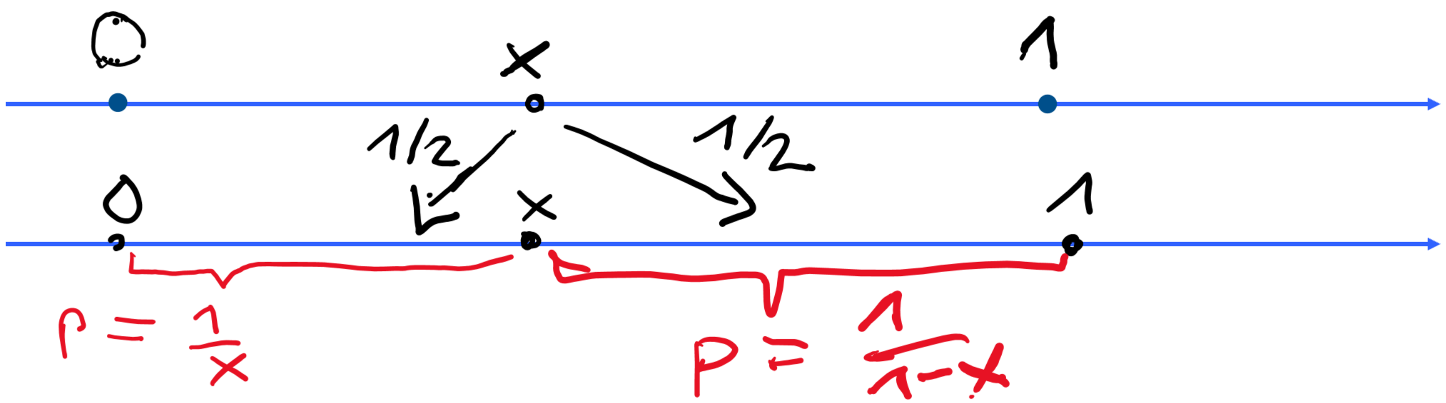
Example

- Consider a DTMC with continuous State Space (0,1) and continuous Time = [0,1,2,…]
- If the chain is at x, it picks one of the two intervals (0, x) or (x, 1) with equal probability 1/2, and then moves to a random y in the chosen interval.

(a) Show that the transition kernel is

$$k(x,y) = \left(\frac{\mathbf{1}_{(0,x)}(y)}{x} + \frac{\mathbf{1}_{(x,1)}(y)}{1-x} \right) / \, 2.$$

Example: Transition density derivation



- Hence the kernel is :

$$k(x,y) = \frac{1}{2}\frac{1}{x}1_{(0,x)}(y) + \frac{1}{2}\frac{1}{1-x}1_{(x,1)}(y)$$

Example stationary distribution derivation

(b) Show that $f(y) = 1/\left(\pi\sqrt{y(1-y)}\right)$ satisfies $f(y) = \int_0^1 k(x,y)f(x)\,\mathrm{d}x$ and hence is a stationary distribution of the process.

- Use the stationarity definition from before: $\pi(y) = \int_\Omega \pi(x)K(x,y)\mathrm{d}x, \quad y \in \Omega,$

$$k(x,y) = \left(\frac{\mathbf{1}_{(0,x)}(y)}{x} + \frac{\mathbf{1}_{(x,1)}(y)}{1-x}\right)/\,2.$$

Use $\mathbf{1}_{(0,x)}(y) = \mathbf{1}_{(y,1)}(x)$ to get

$$\begin{aligned} f(y) = \int_0^1 k(x,y)f(x)dx &= \frac{1}{2}\int_0^1 \mathbf{1}_{(0,x)}(y)\frac{f(x)}{x} + \mathbf{1}_{(x,1)}(y)\frac{f(x)}{1-x}dx \\ &= \frac{1}{2}\left[\int_y^1 \frac{f(x)}{x}dx + \int_0^y \frac{f(x)}{1-x}dx\right] \end{aligned}$$

Example stationary distribution derivation

Fundamental Theorem of Calculus:

from a to x

$$F(x) = \int_a^x f(t)\,dt \;\rightarrow\; \frac{dF}{dx} = f(x)$$

$$= \frac{1}{2}\left[\int_y^1 \frac{f(x)}{x}dx + \int_0^y \frac{f(x)}{1-x}dx\right] \longrightarrow f'(y) = \frac{1}{2}\left[-\frac{f(y)}{y} + \frac{f(y)}{1-y}\right]$$

Example stationary distribution derivation

- For ***first order homogeneous linear differential equation*** we can do the following:

$$\begin{aligned} y' &= -p(t)y \\ \int \frac{1}{y}\,dy &= \int -p(t)\,dt \\ \ln|y| &= P(t) + C \\ y &= \pm\, e^{P(t)+C} \\ y &= Ae^{P(t)}, \end{aligned}$$

use this for our func

$$\begin{aligned} y(x) &= c\exp\left\{\frac{1}{2}\int^x -\frac{1}{\zeta} + \frac{1}{1-\zeta}\,d\zeta\right\} \\ &= c\exp\left\{-\frac{1}{2}\left(\log|x| + \log|1-x|\right)\right\} \\ &= c\left(|x|\,|1-x|\right)^{-\frac{1}{2}} \end{aligned}$$

c = 1/π

– We arrive at:

$$f(y) = 1/\left(\pi\sqrt{y(1-y)}\right)$$

Example task c) : Arcsine Distribution

(c) Use R to simulate from the stationary density of the process, once using an iterative approach and once using $F(z) = \int_0^z f(y)\,\mathrm{d}y = \frac{1}{\pi} \arcsin\sqrt{z}$. Verify empirically that the distributions are equal.

- CDF: $$F(x) = \frac{2}{\pi}\arcsin(\sqrt{x})$$
- PDF: $$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$$
- Derivation: using the substitution $u = \sqrt{t}, t = u^2, dt = 2u\,du$:

$$G(x) = \int_0^x \frac{1}{\pi\sqrt{t(1-t)}} dt = \int_0^{\sqrt{x}} \frac{2}{\pi\sqrt{1-u^2}} du = \frac{2}{\pi}\arcsin(t)\Big|_0^{\sqrt{x}} = \frac{2}{\pi}\arcsin(\sqrt{x})$$

Discussion and Questions