1.1, 1.2 Introduction to Stochastic Processes

Definition A <u>stochastic process</u> is a collection of random variables $\{X_t : t \in I\}$

The set I is called the <u>index set</u> of the process and it usually represents time.

<u>Common index sets</u>:

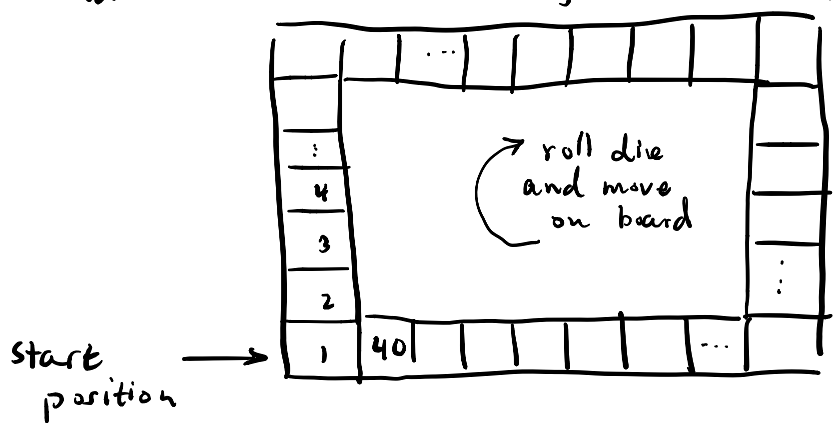
- $I = \{0, 1, 2, \dots\}$ (discrete)

  rounds of a game,
  days of a year

- $I = [0, \infty)$ (continuous time)

The random variables X_t share a common range $S \subseteq \mathbb{R}$ called the <u>state space</u>.

<u>Example</u> Monopoly is a popular board game that can be modeled as a stoch. process. Play by moving position around board according to rolls of a die



State space $S = \{1, \ldots, 40\}$ is the 40 board positions

time index $I = \{0, 1, 2, \ldots\}$ is the rounds of the game.

$X_0, X_1, X_2, \ldots$ denote positions on board

X_k = position after round k.

More examples on worksheet.

Chapter 2 Markov Chains, first steps

2.1 Introduction

Definition Let S be a discrete set.

A Markov chain is a sequence of rv's $\{X_0, X_1, X_2, \dots\}$ taking values in S with the property

$$P(X_{n+1} = j \mid X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = i)$$

$$= P(X_{n+1} = j \mid X_n = i)$$

for all $x_0, \dots, x_{n-1}, i, j \in S$, $n \geq 0$

(this is called the Markov property).

If $P(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i)$

for all $n \geq 0$, the Markov chain is called

time-homogeneous.

(this means one-step probabilities don't

vary with time)

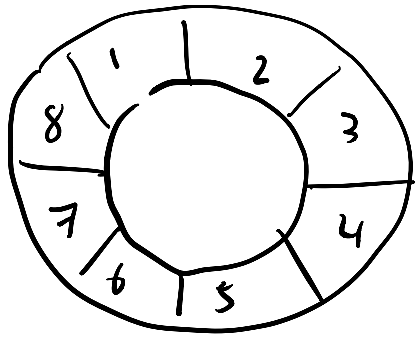
Our focus will be on time-homogeneous chains.

Markov transition matrix P has entries

$$P_{ij} = P(X_1 = j \mid X_0 = j) \text{ called}$$

one-step transition probabilities.

Example (board game)



Start somewhere on board,

roll 6-sided die,

move that many spaces clockwise

repeat

start at 1

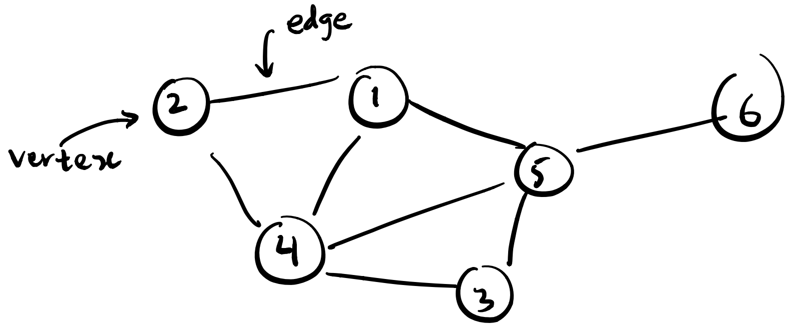
$$P(X_1 = j \mid X_0 = 1) = \begin{cases} \frac{1}{6} & j = 2, \ldots, 7 \\ 0 & j = 1, 8 \end{cases}$$

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \left(\begin{array}{cccccccc} 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 \\ 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ & & & & & & & \\ & & & \vdots & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{array} \right) \end{array}$$

2.2 Markov chain cornucopia

Example (random walk on graph)

Graph is a set of vertices and edges



Two vertices are **neighbors** if connected by edge. eg. 1 and 2 neighbors, write $1 \sim 2$

Degree of vertex is # of neighbors.

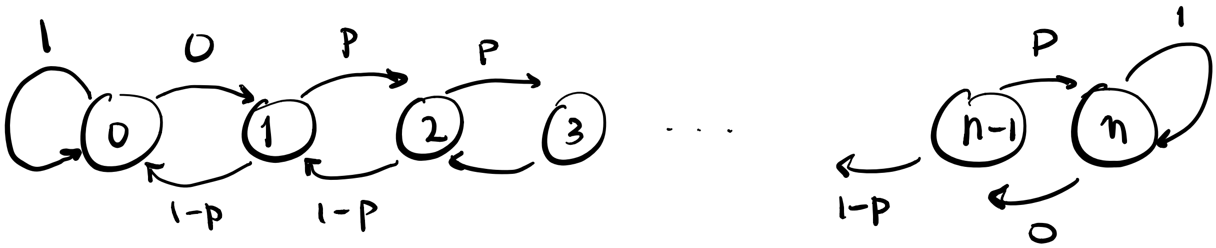
$$\deg(5) = 4$$
$$\deg(6) = 1$$

For the graph above, $P(X_1 = j \mid X_0 = 1)$

$$= \begin{cases} 1/3 & \text{if } j = 2,4,5 \\ 0 & \text{if } j = 1,3,6 \end{cases}$$

Example (Gambler's ruin)

We can visualize transition probabilities with transition state diagram



p = prob. of winning \$1

$1-p$ = prob. of losing \$1

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ n \end{array} \begin{array}{c} \begin{array}{cccccccc} 0 & 1 & 2 & \cdots & & & & n \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & \cdots & & & & 0 \\ 1-p & 0 & p & 0 & \cdots & & & 0 \\ 0 & 1-p & 0 & p & 0 & \cdots & & 0 \\ & & 1-p & 0 & p & \ddots & & \\ & & & \ddots & \ddots & \ddots & \ddots & \\ 0 & \cdot & \cdot & & & & 0 & 1 \end{pmatrix} \end{array}$$

2.3 Basic computations

Definitions $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n)$ is a probability vector if $\sum_{j=1}^{n} \alpha_j = 1$.

If X is a discrete $\{1, \ldots, n\}$-valued random variable and $P(X = j) = \alpha_j$ then $\vec{\alpha}$ is called the distribution of X.

If $X_0, X_1, \ldots$ is a Markov chain and $\vec{\alpha}$ is the distribution of X then $\vec{\alpha}$ is called the initial distribution.

The matrix whose ij-entry is

$$P(X_n = j \mid X_0 = i)$$ is called

the n-step transition matrix.

Lemma If $P = (P_{ij})$ is the Markov transition matrix of a chain with state space S then
$$P^n = \underbrace{P \cdot P \cdots P}_{n\text{-fold product}}$$
is the n-step transition matrix.

Proof

$$P(X_n = j \mid X_0 = i)$$

$$\overset{\text{conditional law of total prob.}}{=} \sum_{k \in S} P(X_n = j \mid X_{n-1} = k, X_0 = i)\, P(X_{n-1} = k \mid X_0 = i)$$

$$= \sum_{k \in S} P(X_n = j \mid X_{n-1} = k)\, P(X_{n-1} = k \mid X_0 = i)$$

$$\overset{\text{induction}}{=} \sum_{k \in S} P_{kj}\, P^{n-1}_{ik}$$

$$= (i\text{th row of } P^{n-1}) \overset{\text{dot product}}{\cdot} (j\text{th col of } P) = P^n_{ij}$$

<u>Question</u> if $X_0 \sim \vec{\alpha}$, then

find $P(X_n = j)$ for $j \in S$.

(the <u>distribution</u> of X_n)

$$P(X_n = j) = \sum_{i \in S} P(X_n = j \mid X_0 = i)\, P(X_0 = i)$$

$$= \sum_{i \in S} P^n_{ij}\, \alpha_i$$ ← dot product of $\vec{\alpha}$ and jth column of P^n

$$= (\underbrace{\alpha P^n}_{\text{row vector - matrix product}})_j$$

jth entry of row vector - matrix product

$$\boxed{X_n \sim \vec{\alpha} P^n \text{ if } X_0 \sim \vec{\alpha}}$$

Joint distribution

Example Suppose $X_0 \sim \alpha$. Find

$$P(X_5 = i, X_6 = j, X_9 = k) \text{ in terms}$$

of α and matrix P powers.

$$P(X_5 = i, X_6 = j, X_9 = k)$$

$$= P(X_9 = k \mid X_5 = i, X_6 = j)\, P(X_5 = i, X_6 = j)$$

$$= P^3_{jk}\, P(X_5 = i, X_6 = j)$$

$$= P^3_{jk}\, P(X_6 = j \mid X_5 = i)\, P(X_5 = i)$$

$$= P^3_{jk}\, P_{ij}\, P(X_5 = i)$$

$$= P^3_{jk}\, P_{ij}\, (\alpha P^5)_i$$

Ch. 3 Markov chains for the long term (limit theory)

3.1, 3.2 Limiting Distribution, Stationary Distributions

Def. A limiting distribution is a probability distribution $\vec{\lambda}$ on the state space such that for all $i, j \in S$

$$\lim_{n\to\infty} P_{ij}^{n} = \lambda_j$$

(notice the RHS doesn't depend on initial state i).

Equivalently ① for any initial distribution $\vec{\alpha}$

$$\lim_{n\to\infty} \vec{\alpha} P^{n} = \vec{\lambda}$$

② $\lim_{n\to\infty} P^{n} = \Lambda$ where Λ is a matrix whose rows are all $\vec{\lambda}$.

Remark In the long term, the distribution of X_n doesn't change step to step when the chain has a limiting distribution

Def A stationary distribution is a probability distribution π which satisfies $\pi P = \pi$.

Remark If π is stationary and $X_0 \sim \pi$

Then $X_n \sim \pi P^n = (\pi P) P^{n-1} = \pi P^{n-1}$

$$= (\pi P) P^{n-2}$$

$$= \pi P^{n-2} = \dots = \pi$$

So $X_n \sim \pi$ for all $n \geq 0$.

Lemma If $\vec{\pi}$ is a limiting distribution then $\vec{\pi} P = \vec{\pi}$. (That is, π is stationary)

Proof Let $\vec{\alpha}$ be any initial distribution.

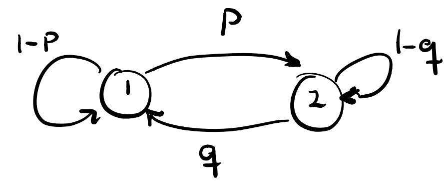
Then
$$\begin{aligned}\pi &= \lim_{n\to\infty} \alpha P^n \\ &= \lim_{n\to\infty} (\alpha P^{n-1}) P \\ &= \left(\lim_{n\to\infty} \alpha P^{n-1}\right) P \\ &= \pi P\end{aligned}$$

Issue How do you know limiting distribution exists? We'll discuss this later

Upside If you know it exists, you find it by solving linear system of equations

$$\vec{x} P = \vec{x}$$

Example $\quad P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$



Find stationary distribution(s). (By solving system of equations)

Let $\quad \vec{x} = (x_1, x_2)$

Then $\quad \vec{x}P = (x_1, x_2)\begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$

$$= \big(x_1(1-p) + x_2 q \,,\; x_1 p + x_2(1-q)\big)$$

and $\quad \vec{x}P = \vec{x} \quad$ means

$$\begin{cases} x_1 = x_1(1-p) + x_2 q \\ x_2 = x_1 p + x_2(1-q) \\ x_1 + x_2 = 1 \end{cases} \Rightarrow \begin{cases} x_1 p = x_2 q \\ x_1 + x_2 = 1 \end{cases}$$

$$\Rightarrow \quad x_1 p = (1-x_1) q$$

$$\boxed{\begin{aligned} x_1 &= \frac{q}{p+q} \\ x_2 &= \frac{p}{p+q} \end{aligned}}$$

Example (making a good guess to find stationary distribution)

Let G be a graph with e edges.

Then $\pi = (\pi_v)$ with

$$\pi_v = \frac{\deg(v)}{2e}$$

is stationary for the random walk on G.

$$(\pi P)_v = \sum_w \pi_w P_{wv}$$

$$= \sum_w \frac{\deg(w)}{2e} P_{wv}$$

$$P_{wv} = \begin{cases} \frac{1}{\deg(w)} & \text{if } v \sim w \\ 0 & \text{else} \end{cases}$$

$$= \sum_{w \sim v} \frac{\deg(w)}{2e} \frac{1}{\deg(w)}$$

$$= \frac{1}{2e} \sum_{w \sim v} 1$$

$$= \frac{\deg(v)}{2e} = \pi_v$$

Example (making a good guess to find stationary distribution)

Let G be a graph with e edges.

Then $\pi = (\pi_v)$ with

$$\pi_v = \frac{\deg(v)}{2e}$$

is stationary for the random walk on G.

$$(\pi P)_v = \sum_w \pi_w P_{wv}$$

$$= \sum_w \frac{\deg(w)}{2e} P_{wv}$$

$$P_{wv} = \begin{cases} \frac{1}{\deg(w)} & \text{if } v \sim w \\ 0 & \text{else} \end{cases}$$

$$= \sum_{w \sim v} \frac{\deg(w)}{2e} \frac{1}{\deg(w)}$$

$$= \frac{1}{2e} \sum_{w \sim v} 1$$

$$= \frac{\deg(v)}{2e} = \pi_v$$

Definition A matrix P is positive ($P > 0$) if all its entries are positive. A transition matrix P is regular if $P^n > 0$ for some $n \geq 1$.

Intuitive interpretation: there is a number of steps, n, which guarantees chain can transition between any two states with positive prob.

Theorem (Limit Theorem for Regular Markov Chains)

If a Markov chain has a regular transition matrix, then it has a unique, positive stationary distribution that is a limiting distribution

Examples 1. $\begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$ for $0 < p < 1$, $0 < q < 1$.

2. $P = \begin{pmatrix} 0 & 1-p & p \\ p & 0 & 1-p \\ 1-p & p & 0 \end{pmatrix}$, check $P^2 > 0$. $\pi = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

A reason why this theorem seems reasonable.

two ways we've seen a chain is prevented from having a limiting distribution

① state space is partitioned into chunks. If the process starts in one chunk it stays there

$$P = \begin{pmatrix} 1-p & p & 0 \\ q & 1-q & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

② "periodicity" processes visits certain parts of state space a set intervals of time "even and odd times"

Regularity of P means we can reach any state in N time steps, so both ① and ② are avoided.

3.3 Communication among states

Long-term behavior is related to how often states are visited. Our focus now is how groups of states are accessible to each other.

Defs • j is accessible from i

if $P_{ij}^{n} > 0$ for some $n \geq 0$

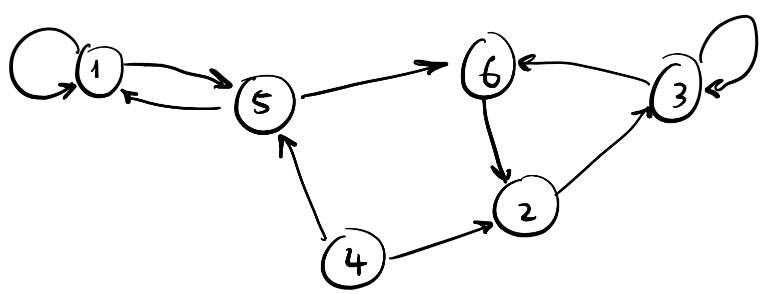
• states i and j communicate if j accessible from i and i accessible from j

We write $i \sim j$ if i and j communicate.

Communication is an equivalence relation, meaning
1) (reflexive) $i \sim i$
2) (symmetric) $i \sim j$ implies $j \sim i$
3) (transitive) $i \sim j$ and $j \sim k$ imply $i \sim k$.

The state space can be partition into equivalence classes called communication classes which are disjoint sets of states that only communicate with each other.

Example Consider the Markov chain whose transition graph is



Find the communication classes.

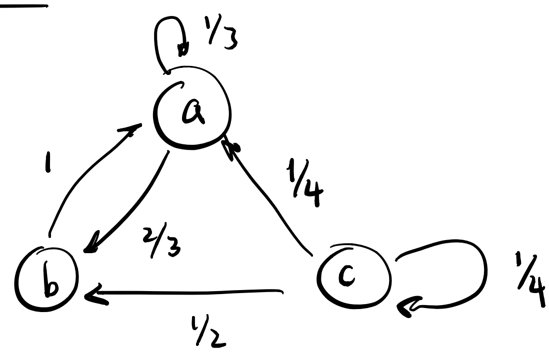
$\{1,5\}, \{2,3,6\}, \{4\}$

Recurrence and transience

Main ideas

- classes contain either all recurr. or all transient states
- recurrent classes are closed.

Example Consider the chain



Question If the chain starts in state a what is the chance it returns to state a? b? c?

a) returns with prob. $\frac{1}{3} + \frac{2}{3} = 1$

b) never returns with prob. $\lim_{n \to \infty} \left(\frac{1}{3}\right)^n = 0$

(ie always stays at a)

c) never returns with prob. $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$

(ie goes to a or b on first step)

Def $T_j = \min\{n > 0 : X_n = j\}$ is called **first passage time** to state j.

Let $f_j = P(T_j < \infty \mid X_0 = j)$ be the prob. that the chain started in j eventually returns to j. $\left(f_a = 1 = f_b, \ f_c = \frac{1}{4} \text{ in last example} \right)$

State j is called **recurrent** if $f_j = 1$.

is called **transient** if $f_j < 1$.

Proposition j is recurrent iff $\sum_{n=0}^{\infty} P_{jj}^{n} = \infty$

j is transient iff $\sum_{n=0}^{\infty} P_{jj}^{n} < \infty$

Proof Let $I_n = \begin{cases} 1 & \text{if } X_n = j \\ 0 & \text{else.} \end{cases}$

Then $\sum_{n=0}^{\infty} I_n = \#$ of visits to j over time.

and $$E\left[\sum_{n=0}^{\infty} I_n \,\middle|\, X_0 = i\right] = \sum_{n=0}^{\infty} E[I_n \mid X_0 = i]$$

$$= \sum_{n=0}^{\infty} P_{ij}^n$$ average number of visits over time.

j <u>recurrent</u> $\Rightarrow$ if we start at j, process eventually returns, starts afresh and returns again, repeats infinitely often

$\Rightarrow$ average number of visits $= \infty$

$$\Rightarrow \sum_{n=0}^{\infty} P_{jj}^n = \infty.$$

j <u>not recurrent</u> $\Rightarrow$ j transient

$\Rightarrow$ possible never to return to j when starting from j.

(happens with probability $1 - f_j$)

$\Rightarrow$ # of visits to j is

Geometric$(1 - f_j)$

$$\Rightarrow \frac{1}{1-f_j} = E\left[\text{Geometric}(1-f_j)\right]$$

$$= E\left[\sum_{n=0}^{\infty} I_n \,\middle|\, X_0 = j\right]$$

$$= \sum_{n=0}^{\infty} P_{jj}^{n} \quad (\text{so} < \infty).$$

<u>Theorem</u> States of a communication class are either all recurrent or all transient.

<u>Proof</u> Let $i \sim j$ be two states in same comm. class. Suppose i is recurrent.

$i \sim j$ means $\exists$ r, s so that

$$P_{ij}^{r} > 0 \quad \text{and} \quad P_{ji}^{s} > 0$$

Then j is recurrent since

$$P_{jj}^{n+r+s} \geq P_{ji}^{s} P_{ii}^{n} P_{ij}^{r} \quad \text{for any } n \geq 0.$$

$$\Rightarrow \sum_{n=0}^{\infty} P_{jj}^{n+r+s} = \sum_{n=0}^{\infty} P_{ji}^{s} P_{ii}^{n} P_{ij}^{r}$$

$$= P_{ji}^{s} P_{ij}^{r} \sum_{n=0}^{\infty} P_{ii}^{n} = \infty$$

$$\Rightarrow \sum_{n=0}^{\infty} P_{jj}^{n} \geq \sum_{n=r+s}^{\infty} P_{jj}^{n}$$

$$= \sum_{n=0}^{\infty} P_{jj}^{n+r+s} = \infty$$

If i is transient then every other state must be too. If there was any recurrent states, that it communicates with i would force i to be recurrent.

Corollary A finite state irreducible Markov chain has all recurrent states.

Proof if there exist a transient state then all are transient since states in the same communicating class are all transient.

If we start at a transient state, say state 1, eventually we don't return. Similarly with state 2, and all other states. But there are only finitely many states.

Lemma If a communication class has all recurrent states, then it's closed (meaning it's impossible to transition to a state in another class).

Canonical Decomposition If S is a finite state space, it can be partitioned as

$$S = \underbrace{T}_{\text{transient states.}} \cup \underbrace{R_1 \cup R_2 \cup \cdots \cup R_m}_{\text{closed, recurrent comm. classes}}$$

We can then reorder states so that

$$P = \begin{array}{cc} & \begin{array}{ccccc} T & R_1 & & \cdots & R_m \end{array} \\ \begin{array}{c} T \\ R_1 \\ \\ \vdots \\ R_m \end{array} & \begin{pmatrix} * & * & & \cdots & * \\ & P_1 & & & \\ & & P_2 & & 0 \\ & 0 & & \ddots & \\ & & & & P_m \end{pmatrix} \end{array}$$

Then

$$P^n = \begin{array}{cc} & \begin{array}{cccc} T & R_1 & \cdots & R_m \end{array} \\ \begin{array}{c} T \\ R_1 \\ \vdots \\ R_m \end{array} & \begin{pmatrix} * & * & \cdots & * \\ & P_1^n & & 0 \\ & & \ddots & \\ & 0 & & P_m^n \end{pmatrix} \end{array}$$

Proposition j is recurrent iff $\sum_{n=0}^{\infty} P_{jj}^{n} = \infty$

j is transient iff $\sum_{n=0}^{\infty} P_{jj}^{n} < \infty$

Proof Let $I_n = \begin{cases} 1 & \text{if } X_n = j \\ 0 & \text{else.} \end{cases}$

Then $\sum_{n=0}^{\infty} I_n =$ # of visits to j over time.

and $E\left[\sum_{n=0}^{\infty} I_n \middle| X_0 = i \right] = \sum_{n=0}^{\infty} E[I_n | X_0 = i]$

$= \sum_{n=0}^{\infty} P_{ij}^{n}$ average number of visits over time.

j recurrent $\Rightarrow$ if we start at j, process eventually returns, starts afresh and returns again, repeats infintely often

$\Rightarrow$ average number of visits $= \infty$

$\Rightarrow \sum_{n=0}^{\infty} P_{jj}^{n} = \infty.$

j <u>not recurrent</u> $\Rightarrow$ j transient

$\Rightarrow$ possible never to return to j when starting from j.

(happens with probability $1 - f_j$)

$\Rightarrow$ # of visits to j is

Geometric$(1 - f_j)$

$$\Rightarrow \frac{1}{1-f_j} = E[\text{Geometric}(1-f_j)]$$

$$= E\left[\sum_{n=0}^{\infty} I_n \,\middle|\, X_0 = j\right]$$

$$= \sum_{n=0}^{\infty} P_{jj}^{n} \quad (\text{so} < \infty).$$

Theorem States of a communication class are either all recurrent or all transient.

Proof Let $i \sim j$ be two states in same comm. class. Suppose i is recurrent.

$i \sim j$ means $\exists$ r, s so that

$$P_{ij}^r > 0 \quad \text{and} \quad P_{ji}^s > 0$$

Then j is recurrent since

$$P_{jj}^{n+r+s} \geq P_{ji}^s P_{ii}^n P_{ij}^r \quad \text{for any } n \geq 0.$$

$$\Rightarrow \sum_{n=0}^{\infty} P_{jj}^{n+r+s} = \sum_{n=0}^{\infty} P_{ji}^s P_{ii}^n P_{ij}^r$$

$$= P_{ji}^s P_{ij}^r \sum_{n=0}^{\infty} P_{ii}^n = \infty$$

$$\Rightarrow \sum_{n=0}^{\infty} P_{jj}^n \geq \sum_{n=r+s}^{\infty} P_{jj}^n$$

$$= \sum_{n=0}^{\infty} P_{jj}^{n+r+s} = \infty$$

If i is transient then every other state must be too. If there was any recurrent states, that it communicates with i would force i to be recurrent.

<u>Corollary</u> A finite state irreducible Markov chain has all recurrent states.

<u>Proof</u> if there exist a transient state then all are transient since states in the same communicating class are all transient.

If we start at a transient state, say state 1, eventually we don't return. Similarly with state 2, and all other states. But there are only finitely many states.

Polya's recurrence theorem

Consider random walk on $\mathbb{Z}$ where walker moves left/right with prob. $1-p$ or p.

This chain is irreducible (one comm. class).

Are states transient or recurrent?

Consider state 0.

$$P_{00}^{n} = 0 \quad \text{when } n \text{ is odd}$$

$$P_{00}^{2n} = \binom{2n}{n} p^n (1-p)^n$$

Stirling's approximation $\quad n! \approx n^n e^{-n} \sqrt{2\pi n}$

for large n.

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2} \approx \frac{(2n)^{2n} e^{-2n} \sqrt{4\pi n}}{n^{2n} e^{-2n} \, 2\pi n}$$

$$= \frac{4^n}{\sqrt{\pi n}}$$

$$\sum_{n=0}^{\infty} \binom{2n}{n} p^n (1-p)^n \approx \sum_{n=0}^{\infty} \frac{(4p(1-p))^n}{\sqrt{\pi n}} = \begin{cases} \infty & p = \frac{1}{2} \\ < \infty & p \neq \frac{1}{2} \end{cases}$$

<u>Conclusion</u> RW on $\mathbb{Z}$ is recurrent if

$p = \frac{1}{2}$, transient if $p \neq \frac{1}{2}$.

Same holds for RW on $\mathbb{Z}^2$

(two dimensional walk on plane lattice)

<u>But</u> RW on $\mathbb{Z}^3$ (3d walk)

is transient for all p. even $p = \frac{1}{2}$.

"A drunk person will find their way

home, but a drunk bird may get

lost forever"

3.4 Irreducible Markov Chains

Recall A Markov chain is irreducible if it has one communication class, and it's guaranteed to be recurrent if the state space is finite.

Limit Theorem for finite irreducible Markov chains

Assume $X_0, X_1, \ldots$ is a finite, irreducible chain.

For each j, let $\mu_j = E[T_j \mid X_0 = j]$ be expected return time to j. Then $\mu_j < \infty$ and

$$\pi_j = \frac{1}{\mu_j}$$

defines a unique, positive, stationary distribution π for the chain. Moreover,

$$\pi_j = \lim_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} P_{ij}^m$$ (not quite limiting necessarily).

Remark μ_j = expected excursion time $\left(\frac{\text{steps}}{\text{visit}}\right)$

$\frac{1}{\mu_j}$ $\left(\frac{\text{visits}}{\text{step}}\right)$ = long term proportion of visits to j

eg. 5 steps between visits to j $\Rightarrow$ $\frac{1}{5}$ of visits are to j

$$\frac{1}{\mu_j} = \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{k=0}^{n-1} I_k \,\middle|\, X_0 = i\right]$$

$$= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} P_{ij}^{k} \overset{*}{:=} \pi_j$$

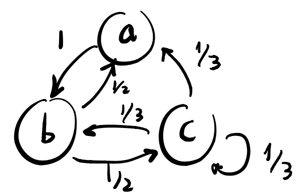
where $I_k = \begin{cases} 1 & \text{if } X_k = j \\ 0 & \text{else} \end{cases}$

* It takes some work to show this limit exists, doesn't depend on i, and defines a stationary measure.

First step analysis (a method for computing $E[T_j \mid X_0 = j]$)

Example

$$P = \begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \end{array}$$



Find $E[T_a \mid X_0 = a]$ (first return time to a)

Let $e_x = E[T_a \mid X_0 = x]$, $\left.\begin{matrix} e_b \\ e_c \end{matrix}\right\}$ first passage time to a.

$$\begin{cases} e_a = \underbrace{1 + e_b}_{\text{time to } b \ + \ \text{time from } b \text{ to } a} \\[2ex] e_b = \frac{1}{2} \cdot \underbrace{1}_{\substack{\text{time directly} \\ \text{to } a}} + \frac{1}{2} \underbrace{(1 + e_c)}_{\substack{\text{time to } c \\ + \text{ time from } c \text{ to } a}} \\[2ex] e_c = \frac{1}{3} \cdot \underbrace{1}_{\substack{\text{time directly} \\ \text{to } a}} + \frac{1}{3} \underbrace{(1 + e_b)}_{\substack{\text{time to } b \\ + \text{ time from } b \text{ to } a}} + \frac{1}{3} \underbrace{(1 + e_c)}_{\substack{\text{time to } c \\ + \text{ time from } c \text{ to } a}} \end{cases}$$

Exercise

Set up system to find $E[T_b \mid X_0 = b]$

Solve this system for e_a. (Get $\underline{\underline{e_a = \frac{10}{3}}}$, $e_b = \frac{7}{3}$, $e_c = \frac{8}{3}$)

This tells us $\pi_a = \frac{3}{10}$.

3.5 Periodicity

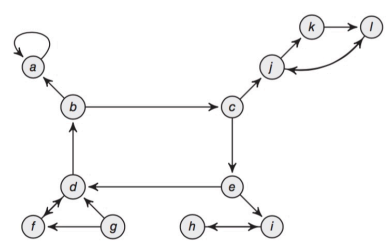
The <u>period</u> of state i, denoted $d(i)$, is

$$d(i) = \underbrace{\gcd\left\{ n > 0 : P^n_{ii} > 0 \right\}}_{\substack{\text{greatest common divisor of} \\ \text{possible return times to } i}}$$

If $d(i) = 1$, then i is called <u>aperiodic</u>.

If set of return times is empty, $d(i) = +\infty$.

<u>Example</u>



Comm. classes are

$$\{a\},\ \{b, c, d, e, f\},\ \{g\},\ \{h, i\},\ \{j, k, l\}$$

<u>Find periods</u>

$d(a) = 1$, $d(b) = 2$, $d(g) = \infty$, $d(h) = 2$, $d(j) \overset{\star}{=} 1$

<u>Lemma</u> The states of a comm. class all have the same period. (periodicity is a class property).

<u>Proof</u> Let $i \sim j$ be in same comm. class.

Then $\exists$ $r, s > 0$ so that $P_{ij}^{r} > 0$, $P_{ji}^{s} > 0$.

$$P_{ii}^{r+s} \geq P_{ij}^{r} P_{ji}^{s} > 0$$

So $r+s$ is a return time to i and

$d(i)$ divides $r+s$

If n is any integer $P_{jj}^{n} > 0$

$$P_{ii}^{r+s+n} \geq P_{ij}^{r} P_{jj}^{n} P_{ji}^{s} > 0$$

so $d(i)$ divides $r+s+n$ too.

Then $d(i)$ divides n. So $d(i)$ is a divisor of $\{n > 0 : P_{jj}^{n} > 0\}$. Hence $d(i) \leq d(j)$

<u>greatest</u> divisor.

3.6 Ergodic chains

A chain is called <u>aperiodic</u> if all states are aperiodic. It's called <u>ergodic</u> if it's irreducible, aperiodic, and all states have finite expected return times (automatic for finite state chains)

Fundamental Limit Theorem for Ergodic Markov chains

Let (X_n) be an ergodic Markov chain. Then it has a unique, positive stationary distribution which is limiting.

Idea of proof

- irreducibility guarantees unique, positive stationary distribution
- aperiodicity $\Rightarrow$ for each state i, $\exists$ N (depending on i) so that $P_{ii}^n > 0$ for all $n \geq N$
- irreducible + aperiodic $\Rightarrow$ P regular
- P regular $\Rightarrow$ can use "coupling argument"

3.8 Absorbing Chains

State i is **absorbing** if $P_{ii} = 1$. A Markov chain is called **absorbing** if it has at least one absorbing state.

Example Gambler's ruin has 0 and n as absorbing states.

Canonical decomposition of absorbing chain

Suppose chain has t transient states and $k-t$ absorbing states

$$P = \begin{array}{c c} & \begin{array}{cc} \text{transient} & \text{absorbing} \end{array} \\ \begin{array}{c} \text{transient} \\ \text{absorbing} \end{array} & \left(\begin{array}{c|c} Q & R \\ \hline O & I \end{array}\right) \end{array}$$

Then $$P^2 = \left(\begin{array}{c|c} Q^2 & QR+R \\ \hline O & I \end{array}\right)$$

$$P^3 = \left(\begin{array}{c|c} Q^3 & Q^2R+QR+R \\ \hline O & I \end{array}\right)$$

$$P^n = \left(\begin{array}{c|c} Q^n & (I+Q+\cdots+Q^{n-1})R \\ \hline O & I \end{array}\right)$$

<u>Remark</u>

$$\lim_{n\to\infty} P^n = \begin{array}{c} \\ t \\ a \end{array}\begin{array}{c} \begin{array}{cc} t & a \end{array} \\ \left(\begin{array}{c|c} \lim\limits_{n\to\infty} Q^n & \lim\limits_{n\to\infty}(I+Q+\cdots+Q^{n-1})R \\ \hline O & I \end{array}\right) \end{array}$$

and the upper right block represents the probabilities of starting in a transient state and eventually being stuck in a certain absorbing state.

<u>Lemma</u> (geometric matrix series)

If A is a square matrix such that $\lim_{n\to\infty} A^n = 0$, then $\sum_{n=0}^{\infty} A^n = (I-A)^{-1}$

<u>Proof</u>

$$(I-A)(I+A+A^2+\cdots+A^n)$$

$$=(I+A+\cdots+A^n)-(A+A^2+\cdots+A^{n+1})$$

$$= I - A^{n+1}$$

If $I-A$ is invertible, then

$$I+A+\cdots+A^n = (I-A)^{-1}(I-A^{n+1})$$

Taking limits as $n\to\infty$ of both sides gives

$$\sum_{n=0}^{\infty} A^n = (I-A)^{-1} \quad \text{since } A^{n+1} \to 0$$

$I-A$ is indeed invertible since $(I-A)x=0$ has only trivial solution $x=0$ since $(I-A)x=0$

$$\Rightarrow \quad x = Ax \Rightarrow x = A(Ax) = A^2 \Rightarrow x = A^n x \Rightarrow$$

$$x = \lim_{n\to\infty} A^n x = 0.$$

Conclusion $\lim_{n\to\infty} Q^n = 0$ by definition of transience, so

$$\lim_{n\to\infty} P^n = \left(\begin{array}{c|c} 0 & (I-Q)^{-1}R \\ \hline 0 & I \end{array}\right)$$

Example Suppose gambler's ruin is played with $p=0.6$ (gambler's win prob.), $n=5$ and $k=2$ (initial stake). Find prob. gambler is eventually ruined.

$$Q = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 0.6 & 0 & 0 \\ 0.4 & 0 & 0.6 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0 & 0.4 & 0 \end{pmatrix} \end{array}$$

$$R = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 5 \end{array} \\ \begin{pmatrix} 0.4 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0.6 \end{pmatrix} \end{array}$$

$(I-Q)^{-1}R$

= solve(I - Q) %*% R

$$= \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 5 \end{array} \\ \begin{pmatrix} * & * \\ 0.36 & 0.64 \\ * & * \\ * & * \end{pmatrix} \end{array}$$

ruin prob = 0.36

Theorem (expected number of visits to transient states)

Consider an absorbing chain with t transient states.

Let F be $t \times t$ matrix (indexed by transient states)

$$F_{ij} = \text{expected number of visits to } j \text{ given chain starts in } i$$

Then $F = (I - Q)^{-1}$ called fundamental matrix.

Proof Let $I_n = \begin{cases} 1 & \text{if } X_n = j \\ 0 & \text{if } X_n \neq j \end{cases}$

$$F_{ij} = E\left[\sum_{n=0}^{\infty} I_n \,\middle|\, X_0 = i\right]$$

$$= \sum_{n=0}^{\infty} P_{ij}^{n}$$

$$= \sum_{n=0}^{\infty} Q_{ij}^{n}$$

$$= \left(\sum_{n=0}^{\infty} Q^n\right)_{ij} = (I - Q)^{-1}_{ij}$$

Theorem (expected time until absorption)

Let i be a transient state and a_i the expected absorption time given that the chain starts at i. Then

$$a_i = \sum_{k \in T} F_{ik}$$

sum of ith row of F

Proof a_i is simply the expected visits to the transient states (eventually the chain stops visiting them and is stuck in an absorbing state.

Summary For chains with only transient and absorbing states, let $F = (I-Q)^{-1}$. Then

1) (absorption prob.) $(FR)_{ij}$ gives prob. of absorption in state j given $X_0 = i$.

2) (absorption time) $(F\mathbf{1})_i$ gives expected time until absorption given $X_0 = i$. Note $\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ col vector of 1's.

5.1 Introduction to Markov Chain Monte Carlo (MCMC)

<u>Goal</u> given a probability distribution π, construct an ergodic chain which has π as its limiting distribution.

<u>Uses</u>

(1) If $r : S \to \mathbb{R}$ is a function, we may be interested in $E[r(X)]$ when $X \sim \pi$.

<u>SLLN for Ergodic</u> Chains (aka the Ergodic Theorem)

If X_n is ergodic with limiting distribution π then for any $\alpha \sim X_0$,

$$\lim_{n \to \infty} \frac{r(X_0) + \cdots + r(X_{n-1})}{n} = E[r(X)]$$

with probability 1.

<u>Moral</u> you can run one long simulation to approximate an integral/expected value.

(2) You may be interested in finding the minimum of a function $E : S \longrightarrow [0,\infty)$ on a finite but massive set S.

Searching uniformly at random for it is hopeless. Instead you construct a probability measure

$$\pi_i = e^{-E(i)/kT} \quad \text{for } i \in S.$$

(for fixed choice of parameters $k, T > 0$)

- lots of probability mass where E is small
- little mass where E is big.

<u>Moral</u>

Construct a MC with π as its limiting distribution. Long term: chain spends largest fraction of time where E is minimized.

keywords: stochastic optimization, simulated annealing.

(3) Lots of other uses including

- simulating large scale particle systems a la statistical mechanics (original use)
- simulating samples from high-dimensional distributions (Gibbs sampler, important in Bayesian statistics)

5.2 **Metropolis-Hastings Algorithm**

Let S be a finite state space and let $\pi = (\pi_1, \dots, \pi_{|S|})$ be a given probability distribution on S.

Goal Construct a Markov chain X_n which is ergodic and has π as its stationary distribution.

Let T_{ij} be transition matrix for any irreducible chain with state space S. The chain associated to T_{ij} is called the proposal chain

(apologies for the overlap in notation. This T_{ij} is unrelated to temperature)

In traveling salesman problem, states are ω, ω'

and $$T_{\omega\omega'} = \begin{cases} \frac{1}{n} \cdot \frac{1}{n-1} & \text{if } \omega, \omega' \text{ are different except for one pair of cities swapped} \\ 0 & \text{else} \end{cases}$$

Note $T_{\omega\omega'} = T_{\omega'\omega}$

Algorithm (Metropolis-Hastings)

In TSP

$\pi_j T_{ji} \geq \pi_i T_{ij}$

$\iff \pi_j \geq \pi_i$

$\iff e^{-E(j)/kT} \geq e^{-E(i)/kT}$

$\iff E(j) \leq E(i)$

If $X_n = i$ then choose X_{n+1} by proposing a jump to a new state j

- Choose j with probability T_{ij}
- If $\pi_j T_{ji} \geq \pi_i T_{ij}$ then $X_{n+1} = j$ (accept j)
- If $\pi_j T_{ji} < \pi_i T_{ij}$ then flip a coin with head prob.

$$\frac{T_{ji}}{T_{ij}} \cdot \frac{\pi_j}{\pi_i} \qquad \left(= \frac{T_{ji}}{T_{ij}} e^{-\frac{\Delta E}{kT}} = e^{-\frac{\Delta E}{kT}} \text{ in TSP} \right).$$

note head prob is less than 1

- if it's heads, then $X_{n+1} = j$ (accept j)
- if it's tails, stay at current state (reject j) so $X_{n+1} = i$.

Transition matrix for X_n:

$$\text{for } j \neq i \qquad P_{ij} = \begin{cases} T_{ij} & \text{if } \pi_j T_{ji} \geq \pi_i T_{ij} \\ \cancel{T_{ij}} \cdot \dfrac{T_{ji}}{\cancel{T_{ij}}} \dfrac{\pi_j}{\pi_i} & \text{if } \pi_j T_{ji} < \pi_i T_{ij} \end{cases}$$

$$\text{and} \qquad P_{ii} = 1 - \sum_{j \neq i} P_{ij}$$

Proporition If π is a probability vector which satisfies

$$\pi_i P_{ij} = \pi_j P_{ji} \quad \text{for all } i,j$$

(called detailed balance equations)

then π is stationary. (Chains that have this property are called time reversible)

Proof $(\pi P)_i = \sum_j \pi_j P_{ji} = \sum_j \pi_i P_{ij}$

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

Theorem π_i is stationary for P defined in the MH-algorithm. (and π is limiting)

Proof Suffices to show detailed balance. (Clear that P is ergodic since T is irreducible and there are self-loops)

Case I $\pi_j T_{ij} > \pi_i T_{ij}$

then $\pi_i P_{ij} = \pi_i T_{ij}$ and $\pi_j P_{ji} = \pi_j \underbrace{T_{ij} \frac{\pi_i}{\pi_j}}_{P_{ji}} = \pi_i T_{ij}$

Case II $\pi_j T_{ij} \leq \pi_i T_{ij}$

then $\pi_i P_{ij} = \pi_i T_{ji} \frac{\pi_j}{\pi_i}$

$$= \pi_j T_{ji} = \pi_j P_{ji}$$

Example Find P with $\pi = \left(\frac{1}{8}, \frac{3}{8}, \frac{1}{2}\right)$ as stationary by MH-algorithm using uniform proposals

$$T = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

Note $T_{ij} = T_{ji}$ for all i,j and

$$\pi_3 > \pi_2 > \pi_1$$

for $i \neq j$ $$P_{ij} = \begin{cases} T_{ij} & \text{if } \pi_j \geq \pi_i \\ T_{ji} \dfrac{\pi_j}{\pi_i} & \text{if } \pi_j < \pi_i \end{cases}$$

$$= \begin{cases} 1/3 & \text{if } \pi_j \geq \pi_i \\ 1/3 \dfrac{\pi_j}{\pi_i} & \text{if } \pi_j < \pi_i \end{cases}$$

$P_{12} = 1/3$ $\quad P_{13} = 1/3$, $\quad P_{11} = 1 - (1/3 + 1/3) = 1/3$

$P_{21} = 1/3 \cdot \frac{\pi_1}{\pi_2} = \frac{1}{3} \cdot \frac{1/8}{3/8} = 1/9$ $\quad P_{23} = 1/3$ $\quad P_{22} = 5/9$

$P_{31} = 1/3 \cdot \pi_1/\pi_3 = 1/3 \cdot \frac{1/8}{1/2} = 1/12$ $\quad P_{32} = 1/3 \cdot \frac{3/8}{1/2} = 1/4$, $P_{33} = \frac{5}{12}$

$$P = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/9 & 5/9 & 1/3 \\ 1/12 & 1/4 & 2/3 \end{pmatrix}$$

Using R, one sees

$$\lim_{n \to \infty} P^n = \begin{pmatrix} 1/8 & 3/8 & 1/2 \\ 1/8 & 3/8 & 1/2 \\ 1/8 & 3/8 & 1/2 \end{pmatrix}.$$

Ch4 Branching Processes

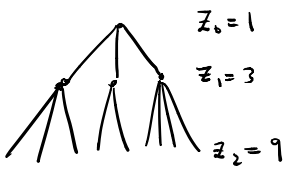
4.1 Introduction

Model of population generation by generation

$$Z_n = \text{size of } n^{th} \text{ generation}$$

Given $Z_n = k$, Z_{n+1} is determined as follows

- $X_1, \ldots, X_k$ are the number of offspring of the k individuals



- $X_1, \ldots, X_k$ i.i.d. with a common prob. distribution $a = (a_0, a_1, a_2, \ldots)$
  (offspring distribution)
  prob. of 0 offspring
  prob. of 1 offspring
  ...

- then $Z_{n+1} = X_1 + \cdots + X_k$ or

- More generally, $Z_{n+1} = \sum_{i=1}^{Z_n} X_i$

Assumptions

- $0 < a_0 < 1$ (if $a_0 = 0$, process always just grows, if $a_0 = 1$ it always becomes extinct)

- $a_0 + a_1 < 1$ (there's the possibility of individuals having more than one offspring)

<u>Lemma</u> In a branching process, all nonzero states are transient. (0 is absorbing)

<u>Proof</u> Consider state $i>0$ (i individuals in population. To show transience, we consider

$$f_i = P\left(Z_n = i \text{ for some } n \geq 1 \,\middle|\, Z_0 = i\right)$$

The event $Z_n = i > 0$ for some $n \geq 1$ implies $Z_1 > 0$.

$$\leq P\left(Z_1 > 0 \,\middle|\, Z_0 = i\right)$$

$$= 1 - P\left(Z_1 = 0 \,\middle|\, Z_0 = i\right)$$

$$= 1 - (a_0)^i$$

all i individuals have 0 offspring.

< 1, so i is transient. ///

<u>Conclusion</u> Since all positive states are transient, either process gets absorbed at 0 or grows without bound (depending on offspring dist)

4.2 Mean generation size

$X_i \sim \vec{a}$ are i.i.d. offspring rv's

Suppose mean number offspring for an individual is $\mu = E[X_i] = \sum_{k=0}^{\infty} k a_k$

Let $Z_n = \sum_{i=1}^{Z_{n-1}} X_i$ be n^{th} generation size.

Then
$$E[Z_n] = \sum_{k=0}^{\infty} E[Z_n \mid Z_{n-1}=k]\, P(Z_{n-1}=k)$$
$$= \sum_{k=0}^{\infty} E\left[\sum_{i=1}^{k} X_i\right] P(Z_{n-1}=k)$$
$$= \sum_{k=0}^{\infty} k\mu \cdot P(Z_{n-1}=k)$$
$$= \mu \sum_{k=0}^{\infty} k \cdot P(Z_{n-1}=k) = \mu E[Z_{n-1}]$$

Thus
$$E[Z_n] = \mu E[Z_{n-1}] = \mu(\mu E[Z_{n-2}])$$
$$= \mu^2(\mu E[Z_{n-3}])$$
$$= \mu^n E[Z_0]$$

Assuming $Z_0 = 1$ for simplicity, $E[Z_n] = \mu^n$

Three cases

$$\lim_{n\to\infty} E(Z_n) = \begin{cases} 0 & \text{if } \mu < 1 \text{ (subcritical)} \\ 1 & \text{if } \mu = 1 \text{ (critical)} \\ \infty & \text{if } \mu > 1 \text{ (supercritical)} \end{cases}$$

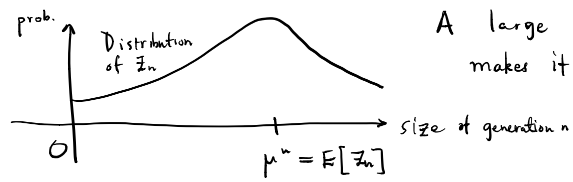
Main question what can you say about extinction probability in each case?

Example $a = \left(\frac{1}{8}, \frac{1}{2}, \frac{3}{8}\right)$. Find long term mean generation size.

$$\mu = \frac{1}{8}\cdot 0 + \frac{1}{2}\cdot 1 + \frac{3}{8}\cdot 2 = 1.25$$

$$\lim_{n\to\infty} E[Z_n] = \lim_{n\to\infty} (1.25)^n = \infty.$$

Does this guarantee the process doesn't go extinct? No! (See simulation results in book)



A large variance $V(Z_n)$ makes it possible to go extinct.

Variance of Z_n

Let $\sigma^2 = V(X_i) = \sum_{k=0}^{\infty} k^2 a_k$

Then $V(Z_n \mid Z_{n-1} = k) = V\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} V(X_i) = \sigma^2 k$

and $V(Z_n \mid Z_{n-1}) = \sigma^2 Z_{n-1}$. Law of Total Variance:

$$\begin{aligned} V(Z_n) &= V(E(Z_n \mid Z_{n-1})) + E(V(Z_n \mid Z_{n-1})) \\ &= V(\mu Z_{n-1}) + E(\sigma^2 Z_{n-1}) \\ &= \mu^2 V(Z_{n-1}) + \sigma^2 \mu^{n-1} \end{aligned}$$

$V(Z_1) = \sigma^2$, $V(Z_2) = \mu^2\sigma^2 + \mu\sigma^2$ $\quad V(Z_3) = \mu^2\sigma^2\mu(\mu+1) + \sigma^2\mu^2 = \mu^2\sigma^2(\mu^2+\mu+1)$

$= \sigma^2\mu(1+\mu)$

$V(Z_4) = \sigma^2\mu^4(\mu^2+\mu+1) + \sigma^2\mu^3 = \sigma^2\mu^3(\mu^3+\mu^2+\mu+1)$

$$V(Z_n) = \sigma^2\mu^{n-1}(\mu^{n-1} + \dots + \mu + 1) = \begin{cases} \sigma^2 n & \mu = 1 \\ \sigma^2\mu^{n-1}\dfrac{\mu^n - 1}{\mu - 1} & \mu \neq 1 \end{cases}$$

grows when $\mu > 1$!

Extinction in subcritical case

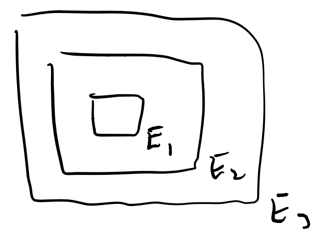
Suppose branching process $Z_0, Z_1, \ldots$ is subcritical ($\mu < 1$).

Let $E_n = \{ Z_n = 0 \}$ event that extinct in n^{th} generation

let E = event that extinction eventually happens

$$= \{ Z_n = 0 \text{ for some } n \geq 1 \}$$

$$= \bigcup_{n=0}^{\infty} E_n$$



and $E_1 \subseteq E_2 \subseteq E_3 \subseteq \ldots$ so

$$P(E) = P\left(\bigcup_{n=1}^{\infty} E_n \right) = \lim_{n \to \infty} P(E_n)$$

$$= \lim_{n \to \infty} P(Z_n = 0)$$

$$P(Z_n = 0) = 1 - P(Z_n \geq 1)$$

$$= 1 - \sum_{k=1}^{\infty} P(Z_n = k) \qquad P(Z_n = k) \leq k P(Z_n = k)$$

$$\geq 1 - \sum_{k=1}^{\infty} k P(Z_n = k) = 1 - E[Z_n]$$

$$\longrightarrow 1 \text{ as } n \to \infty$$

Thus $P(E) = 1$, guaranteed to eventually go extinct.

4.3 Probability generating functions (PGF's)

If X takes values in $\{0,1,2,\dots\}$

the <u>pgf</u> of X is the function

$$G(s) = E[s^X] = \sum_{k=0}^{\infty} s^k P(X=k)$$

$$= P(X=0) + s\,P(X=1) + s^2 P(X=2) + \dots$$

<u>Some notes</u>

* $G(1) = 1$
* if $|s| \leq 1$, then $G(s) = \sum_{k=0}^{\infty} s^k \underbrace{P(X=k)}_{\leq 1}$

$$\leq \sum_{k=0}^{\infty} s^k < \infty \text{ exists.}$$

* $G'(s) = P(X=1) + 2s\,P(X=2) + 3s^2 P(X=3) + \dots$

so $G'(0) = P(X=1)$

* Similarly, $\frac{G''(0)}{2} = P(X=2)$ and $P(X=k) = \frac{G^{(k)}(0)}{k!}$

G is a function which lets you represent the prob. distribution of X as coefficients of a power series.

<u>Example</u> Let $X \sim \text{Geom}(p)$. Find $G(s)$

$$G(s) = E[s^X] = \sum_{k=1}^{\infty} s^k P(X=k)$$

$$= \sum_{k=1}^{\infty} s^k (1-p)^{k-1} p$$

$$= sp \sum_{k=1}^{\infty} (s(1-p))^{k-1}$$

$$= sp \cdot \frac{1}{1-s(1-p)} \quad \text{for } |s|<1$$

<u>Example</u> Suppose $G(s) = (1-p+sp)^n$, find X.

$$G'(s) = np(1-p+sp)^{n-1}$$

$$G''(s) = n(n-1)p^2(1-p+sp)^{n-2}$$

$$G^{(k)}(s) = n(n-1)\cdots(n-k+1)p^k(1-p+sp)^{n-k} \quad k \le n$$

So
$$P(X=k) = \frac{G^{(k)}(0)}{k!}$$

$$= \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

Thus, $X \sim \text{Binomial}(n,p)$

Moments

Note
$$G'(s) = \frac{d}{ds} E[s^X]$$
$$= E\left[\frac{d}{ds}(s^X)\right]$$
$$= E\left[X s^{X-1}\right]$$

So
$$G'(1) = E[X]$$

Also
$$G''(s) = E\left[\frac{d}{ds}(X s^{X-1})\right]$$
$$= E\left[X(X-1) s^{X-2}\right]$$
$$G''(1) = E[X(X-1)] = E[X^2] - E[X]$$

So
$$Var(X) = E[X^2] - E[X]^2$$
$$= G''(1) + G'(1) - G'(1)^2$$

Example Find mean and variance for $X \sim Geom(p)$

$$G(s) = \frac{sp}{1-(1-p)s} \Rightarrow G'(s) = \frac{p}{(1-s(1-p))^2}, \quad G''(s) = \frac{2(1-p)p}{(1-s(1-p))^3}$$

$$E[X] = G'(1) = \frac{1}{p}$$

$$Var(X) = G''(1) + G'(1) - G'(1)^2$$
$$= \frac{2(1-p)p}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

Sums

Suppose X, Y independent

Then $$G_{X+Y}(s) = E\left[s^{X+Y}\right]$$

$$= E\left[s^X s^Y\right] = E\left[s^X\right] E\left[s^Y\right]$$

$$= G_X(s)\, G_Y(s)$$

Example Let $X_1, \dots, X_n \sim Ber(p)$ be i.i.d.

$$G_{X_1}(s) = E\left[s^{X_1}\right]$$

$$= s^1 \cdot p + s^0 \cdot (1-p)$$

$$= 1 - p + sp$$

$$G_{X_1 + \dots + X_n}(s) = G_{X_1}(s) \cdots G_{X_n}(s)$$

$$= (1 - p + sp)^n$$

Same as Binom(n, p) PGF !

4.4 Extinction

Goal understand probability e of eventual extinction

Theorem Given a branching process, let $G(s)$ be the pgf of the offspring distribution. Then the probability of eventual extinction is the smallest positive root of $s = G(s)$.

In the critical and subcritical cases ($\mu \le 1$) the probability is 1.

Proof to come.

Example Suppose $a = (1/6, 1/2, 1/3)$. Find extinction probability.

$$\mu = 1/6 \cdot 0 + 1/2 \cdot 1 + 1/3 \cdot 2 = 7/6 > 1. \quad (\text{super-critical})$$

$$G(s) = \sum_{k=0}^{\infty} s^k a_k = s^0 \cdot 1/6 + s^1 \cdot 1/2 + s^2 \cdot 1/3$$
$$= 1/6 + 1/2\, s + 1/3\, s^2$$

Solve $s = G(s)$ $\qquad 0 = \frac{1}{6} - \frac{1}{2} s + \frac{1}{3} s^2$

$$s = \frac{1/2 \pm \sqrt{1/4 - 4(1/6)(1/3)}}{2/3}$$

$$= \frac{1/2 \pm \sqrt{1/4 - 2/9}}{2/3} = (1/2 \pm 1/6)\left(\frac{3}{2}\right)$$

$$= \left(\frac{2}{3}\right)\left(\frac{3}{2}\right) \text{ or } \left(\frac{1}{3}\right)\left(\frac{3}{2}\right)$$

$$= 1 \text{ or } \frac{1}{2}$$

extinction probability $e = \min\left(1, \frac{1}{2}\right) = \frac{1}{2}$.

<u>Example</u> Branching process has offspring distribution

$$a_k = (1-p)^k p \quad , \quad k=0,1,\dots$$

(not quite geometric) Find extinction probability e.
in the supercritical case.

$$G(s) = \sum_{k=0}^{\infty} s^k a_k = \sum_{k=0}^{\infty} \left(s(1-p)\right)^k \cdot p$$

$$= \frac{p}{1-s(1-p)} \quad \text{when } |s(1-p)|<1$$

$$\mu = G'(1) = \left.\frac{p(1-p)}{(1-s(1-p))^2}\right|_{s=1} = \frac{p(1-p)}{p^2} = \frac{1-p}{p}$$

$\mu > 1$ when $\frac{1-p}{p} > 1 \iff p < 1-p$
$\iff 2p<1 \iff p<\frac{1}{2}.$

e is smallest root of $s = \dfrac{p}{1-s(1-p)}$

$$\Rightarrow 0 = s(1-s(1-p)) - p$$
$$= (1-p)s^2 - s + p$$

roots are $s = \dfrac{1 \pm \sqrt{1-4p(1-p)}}{2(1-p)} = \dfrac{1 \pm \sqrt{1-4p+4p^2}}{2(1-p)}$

$$= \frac{1 \pm (1-2p)}{2(1-p)}$$

$$e = \frac{1-(1-2p)}{2(1-p)} = \frac{p}{1-p} = \frac{1}{\mu}$$

4.4 Proof of main theorem

Background: For $n \geq 0$, let $G_n(s)$ be PGF of Z_n

so $$G_n(s) = \sum_{k=0}^{\infty} s^k \cdot P(Z_n = k)$$

and $G(s)$ the PGF of the offspring distribution

$$G(s) = \sum_{k=0}^{\infty} s^k a_k$$

Expression for $G_n(s)$

$$G_n(s) = E\left(s^{Z_n}\right) = E\left(s^{\sum_{k=1}^{Z_{n-1}} X_k}\right)$$

$$= E\left(E\left(s^{\sum_{k=1}^{Z_{n-1}} X_k} \,\middle|\, Z_{n-1}\right)\right)$$

$$E\left(s^{\sum_{k=1}^{Z_{n-1}} X_k} \,\middle|\, Z_{n-1} = m\right) = E\left(s^{\sum_{k=1}^{m} X_k}\right) = \left[E\left(s^{X_k}\right)\right]^m = G(s)^m$$

So $$E\left(s^{\sum_{k=1}^{Z_{n-1}} X_k} \,\middle|\, Z_{n-1}\right) = G(s)^{Z_{n-1}}$$

and $$G_n(s) = \underbrace{E\left[G(s)^{Z_{n-1}}\right]}_{\substack{\text{this is PGF of } Z_{n-1} \\ \text{evaluated at } G(s)}} = G_{n-1}\left(G(s)\right)$$

Iterating recursive expression

$$G_0(s) = E\left[s^{Z_0}\right] = E[s] = s.$$

$$G_1(s) = G_0\left(G(s)\right) = G(s)$$

$$G_2(s) = G_1\left(G(s)\right) = G\left(G(s)\right)$$

$$G_n(s) = G_{n-1}(G(s)) = \underbrace{G \circ G \circ \cdots \circ G(s)}_{n\text{-fold composition/iterate}} = G\left(G_{n-1}(s)\right)$$

Proof of Theorem Part I Let $e_n = P(Z_n = 0)$. Then

$$e_n = P(Z_n = 0) = G_n(0) = G(G_{n-1}(0))$$
$$= G(P(Z_{n-1} = 0))$$
$$= G(e_{n-1})$$

Then $e = P(\text{eventual extinction})$

$$= \lim_{n\to\infty} e_n = \lim_{n\to\infty} G(e_{n-1})$$

$$= G\left(\lim_{n\to\infty} e_{n-1}\right) = G(e).$$

So e is a solution (aka root) of $s = G(s)$.

Part II Remains to show e is smallest positive root of $s = G(s)$.

Let x be any positive root of $s = G(s)$.

$\star$ Want to show $e \leq x$. $\star$

Claim $e_n \leq x$ for all n.

Pf. of claim $n=1$ case $e_1 = P(Z_1 = 0) = G_1(0) = G(0) \leq G(x) = x$

($\text{note}\quad G(s) = \sum_{k=0}^{\infty} s^k a_k$ increases as s increases)

Induction step Assume claim is true for $n=k$.

$$e_{k+1} = G_{k+1}(0) = G(G_k(0)) = G(e_k) \leq G(x) = x$$

Conclusion Take limit of both sides of $e_n \leq x$

$$\lim_{n\to\infty} e_n \leq \lim_{n\to\infty} x \implies e \leq x$$

Part III

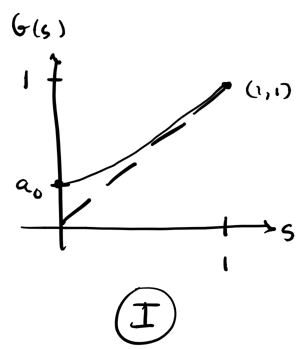
Why is $e = 1$ when $\mu \le 1$?

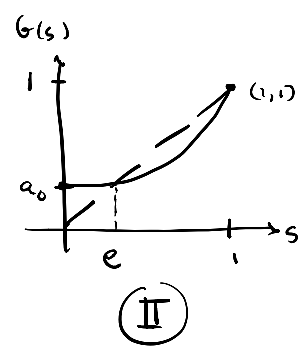
$$G(s) = \sum_{k=0}^{\infty} s^k a_k$$

- $G(0) = a_0$
- $G(1) = 1$
- $G''(s) = \sum_{k=2}^{\infty} k(k-1)s^{k-2}a_k > 0$

so G is concave up.

Two possibilities for graph of $G(s)$:





either intersects $y = s$ at $s = 1$ only or at $s = 1$ and some $s < 1$.

(I) Suppose $\mu \le 1$. Then $G'(1) = \mu \le 1$.

Let $h(s) = G(s) - s$.

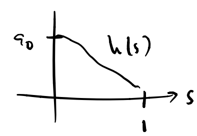
we know $h(1) = 0$ since $G(1) = 1$

and $h(0) = G(0) = a_0 > 0$

Next, note $h'(s) = G'(s) - 1 < G'(1) - 1 = 0$

when $s<1$ since $G'(s) < G'(1)$ since $G''(s) > 0$ means G' is increasing

So h is decreasing and never 0 before $s=1$.



(II) Suppose $\mu > 1$. Then $G'(1) = \mu > 1$

so $h'(1) = G'(1) - 1 > 0$ (h increasing at 1)

$h(0) = G(0) = a_0 > 0$

$h(1) = G(1) - 1 = 0$

- $h(s)$ starts above s-axis at a_0, when $s=0$
- $h(s)$ ends at s-axis, when $s=1$
  - was increasing at $s=1$
- so must have been below s-axis at some $s<1$
- so must have crossed s-axis at some $e<1$

$$h(e) = 0 \Rightarrow G(e) = e.$$

6.1 Intro to Poisson process

Recall Poisson rv counts number of occurrences of event that happens with mean rate λ over fixed time interval.

Ex's
- phone calls in a day
- earthquakes in a year
- customers at store in a day

count "arrivals" of events

If on average you get λ customers per day and $X \sim \text{Poisson}(\lambda)$ counts how many customers you get today $P(X=k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$ $k=0,1,\dots$

Poisson process models arrivals over variable time interval $[0,t]$ where $t>0$.

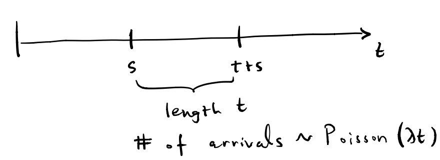
Def A Poisson process with parameter λ is a counting process $(N_t)_{t\geq 0}$ (means N_t is integer valued and increases with time t) such that

1) $N_0 = 0$

2) For all $t>0$, $N_t \sim \text{Poisson}(\lambda t)$

3) (stationary increments) For all $s,t>0$

$$N_{t+s} - N_s \sim N_t \sim \text{Poisson}(\lambda t)$$



4) (independent increments)

For $0 \leq q < r \leq s < t$,

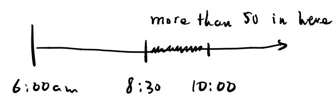
$N_t - N_s$ and $N_r - N_q$ are independent

# of arrivals in these two intervals of time are indep

Remark This is our first continuous time process!

Example Suppose customers arrive at a bakery at a rate of 30 customers per hour according to a Poisson process, starting at 6 a.m. today.

① Find prob. that more than 50 arrive between 8:30 and 10:00 am.



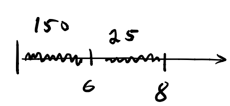
$$P(N_4 - N_{2.5} > 50)$$

$$\overset{\text{Stationary incr.}}{=} P(N_{1.5} > 50) \qquad N_{1.5} \sim \text{Pois}(1.5 \cdot 30)$$

$$= 1 - \underbrace{\text{ppois}(50, 1.5 * 30)}_{\text{R command for computing } P(X \le 50) \text{ when } X \sim \text{Pois}(1.5 \cdot 30)}$$

② Find prob. of getting 150 customers by 12:00pm (noon) and 175 by 2:00 pm.

$$P(N_6 = 150,\ N_8 = 175)$$



$$= P(N_6 = 150,\ N_8 - N_6 = 25)$$

$$\overset{\text{indep. increm.}}{=} P(N_6 = 150)\, P(N_2 = 25)$$

$$= \text{dpois}(150, 6 * 30) * \text{dpois}(25, 2 * 30)$$

③ Given that 60 people arrive by 8:00 am, find prob. that more than 100 total have arrived by 9:00am.

$$P(N_3 > 100 \mid N_2 = 60) = P(N_1 > 40) = 1 - \text{ppois}(40, 30)$$

6.2 Arrival, inter-arrival times

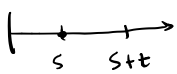
Another way of thinking about Poisson process is through the times when arrivals occur or the time gaps between arrivals (inter-arrival times).

Let X_1 be time of first arrival of Poisson process (N_t) with rate λ.

$$P(X_1 > t) = P(N_t = 0) = e^{-\lambda t}, \quad t > 0$$

So $X_1 \sim Exp(\lambda)$!

Let X_2 be time between 1st and 2nd arrivals.



$$\begin{aligned} P(X_2 > t \mid X_1 = s) &= P(N_{s+t} - N_s = 0) \\ &= P(N_t = 0) \\ &= e^{-\lambda t} \quad \text{for any } s, t > 0 \end{aligned}$$

So X_2 is independent of X_1 !

and $X_2 \sim Exp(\lambda)$!

Theorem The interarrival times $X_1, X_2, \ldots$ of a Poissin process are i.i.d. $Exp(\lambda)$

Remarks

1) If $X_1, X_2, \ldots$ are i.i.d. $Exp(\lambda)$, then

$$N_t = \max\{n : X_1 + \cdots + X_n \le t\}, \quad N_0 = 0$$

is a Poisson process. (This simply says that any counting process with i.i.d. $Exp(\lambda)$ inter-arrival times is a Poisson process).

2) The arrival times $S_1, S_2, \ldots$ given by

$$S_k = X_1 + \cdots + X_k$$

are the times when arrivals occur.

$$S_k \sim Gamma(k, \lambda)$$

$$E(S_k) = \frac{k}{\lambda} \quad \text{and} \quad V(S_k) = \frac{k}{\lambda^2}$$

3) Memoryless property if $X \sim Exp(\lambda)$

$$\underbrace{P(X > t+s \mid X > t)}_{\text{given you've already waited } t \text{ time units, probability of waiting additional } s \text{ time units}} = \underbrace{P(X > s)}_{\text{prob. of waiting } s \text{ time units.}}$$

Two important results Let $X_1, \ldots, X_n$ be independent exponential random variables with parameters $\lambda_1, \ldots, \lambda_n$. Let $M = \min\{X_1, \ldots, X_n\}$.

① For $t > 0$, $P(M > t) = e^{-t(\lambda_1 + \cdots + \lambda_n)}$

Therefore, $M \sim \text{Exp}(\lambda_1 + \cdots + \lambda_n)$.

Proof
$$\begin{aligned} P(M > t) &= P(X_1 > t, \ldots, X_n > t) \\ &= P(X_1 > t) \cdots P(X_n > t) \\ &= e^{-t\lambda_1} \cdots e^{-t\lambda_n} = e^{-t(\lambda_1 + \cdots + \lambda_n)} \end{aligned}$$

② For $k = 1, \ldots, n$, $P(M = X_k) = \dfrac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}$

Proof $P(M = X_k) = P(X_1 \geq X_k, \ldots, X_n \geq X_k)$
$$\begin{aligned} &= \int_0^\infty P(X_1 \geq t, \ldots, X_n \geq t \mid X_k = t)\, f_{X_k}(t)\, dt \\ &= \int_0^\infty P(X_1 \geq t) \cdots P(X_{k-1} \geq t) P(X_{k+1} \geq t) \cdots P(X_n \geq t)\, \lambda_k e^{-\lambda_k t}\, dt \\ &= \lambda_k \int_0^\infty e^{-(\lambda_1 + \cdots + \lambda_n)t}\, dt = \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n} \end{aligned}$$

6.4 Thinning and superposition

Vehicles arrive at a toll booth according to a Poisson process (N_t) with rate λ.

The chance that an arriving vehicle is a bus is p and the chance it's a car is $1-p$, independent of other arrivals.

Claim The processes (B_t) and (C_t) corresponding to busses and cars are independent Poisson process with parameters λp and $\lambda(1-p)$.

B_t and C_t are called thinned Poisson processes

Sketch of proof

$$P(B_t=b, C_t=c)$$

$$= P(B_t=b, C_t=c, N_t=b+c)$$

$$= P(B_t=b, C_t=c \mid N_t=b+c)\, P(N_t=b+c)$$

$$= P(B_t=b \mid N_t=b+c)\, P(N_t=b+c)$$

$$= \binom{b+c}{b} p^b (1-p)^c \times e^{-\lambda t} \frac{(\lambda t)^{b+c}}{(b+c)!}$$

$$= \frac{1}{b!\,c!} p^b (1-p)^c \underbrace{e^{-\lambda}}_{e^{-\lambda p} e^{-\lambda(1-p)}} (\lambda t)^{b+c}$$

$$= \underbrace{e^{-\lambda p} \frac{(\lambda p t)^b}{b!}}_{\text{Poisson}(\lambda p t)} \underbrace{e^{-\lambda(1-p)} \frac{(\lambda(1-p)t)^c}{c!}}_{\text{Poisson}(\lambda(1-p)t)}$$

On last week's worksheet on PGF's we saw if $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\mu)$ are independent, then $X + Y \sim \text{Poisson}(\lambda + \mu)$.

this implies...

If (N_t) and (M_t) are independendent Poisson processes with parameters λ_1 and λ_2 then $(N_t + M_t)$ is a Poisson process with parameter $\lambda_1 + \lambda_2$ called a <u>superposition</u> process.

Ch 7 Continuous time Markov chains (CTMC)

7.1 Introduction

Let's consider a process with discrete state space S with a continuous time index $[0, \infty)$.

If $X_0 = i$, then X_t stays at i for a random amount of time T_1 (where T_1 is a continuous rv) and then jumps to a new state j. The process then waits again for a random amount of time T_2 before jumping to a new state and this repeats.

Markov property A continuous-time stochastic process $(X_t)_{t \geq 0}$ is a CTMC if

$$P(X_{t+s} = j \mid X_s = i, X_u = x_u \quad 0 \leq u < s)$$
$$= P(X_{t+s} = j \mid X_s = i) \quad \text{for all } i, j, x_u \in S, \quad s, t \geq 0$$

we'll assume our chains are time-homogeneous, meaning $P(X_{t+s} = j \mid X_s = i) = P(X_t = j \mid X_0 = i)$

The transition probabilities can be represented as a matrix-valued function $P(t)$ called the transition function where

$$P_{ij}(t) = P(X_t = j \mid X_0 = i).$$

Example Poisson process

$$P_{ij}(t) = P(N_{t+s}=j \mid N_s=i)$$
$$= P(N_t = j-i)$$
$$= e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}$$

Chapman-Kolmogorov Equations

$$P(t+s) = P(t)P(s)$$

This says that the transition probabilities over $t+s$ time units can be broken down in terms of t time units and s time units.

Proof

$$P_{ij}(t+s) = P(X_{t+s}=j \mid X_0=i)$$
$$= \sum_k P(X_{t+s}=j \mid X_0=i, X_s=k)\,P(X_s=k \mid X_0=i)$$
$$= \sum_k P_{kj}(t)\,P_{ik}(s)$$
$$= (P(s)P(t))_{ij}$$

How CTMC's work, given $X_0=i$

- wait a random amount of time (distribution depends on state i)
- jump to a random state j (with prob. $\tilde{P}_{ij}$)

different from $P(t)$

Worksheet introduces holding times, $\tilde{P}$-matrix.

7.3 Infinitesimal generator

We introduced the transition rates q_{ij} last time as the alarm clock parameter for state j when our current state is i.

Here's how they arise:

q_{ij} = mean instantaneous rate of transition from i to j

$$= \lim_{h \to 0^+} \frac{E[\# \text{ of transitions to } j \text{ in } (t, t+h] \mid X_t = i]}{h}$$

$$= \lim_{h \to 0^+} \frac{P(X_{t+h} = j \mid X_t = i)}{h}$$

$$= \lim_{h \to 0^+} \frac{P_{ij}(h)}{h} = \lim_{h \to 0^+} \frac{P_{ij}(h) - P_{ij}(0)}{h}$$

$$= P_{ij}'(0)$$

Let $Q = (q_{ij}) = P'(0)$

Note $q_{ii} = \lim_{h \to 0^+} \frac{P_{ii}(h) - P_{ii}(0)}{h}$

$$= \lim_{h \to 0^+} \frac{P_{ii}(h) - 1}{h}$$

$$= \lim_{h \to 0^+} \frac{(1 - \sum_{j \neq i} P_{ij}(h)) - 1}{h}$$

$$= -\sum_{j \neq i} \lim_{h \to 0^+} \frac{P_{ij}(h)}{h} = -\sum_{j \neq i} P_{ij}'(0)$$

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

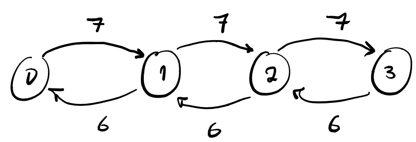
<u>Conclusion</u> Q (called the <u>infinitesimal generator</u>) consists of a matrix with

- rows sum to 0
- diagonal entry i is $-q_i = -\sum_{j \neq i} q_{ij}$ where q_i is hold time parameter
- q_{ij} alarm clock parameter.

<u>Example</u> 3 machines running independently each break on average 7 times per year and the repair time is on average 2 months. Let X_t be the number of machines broken at time t. Write Q, the mean hold times, and the embedded chain matrix $\tilde{P}$.

7 breaks per year

6 repairs per year



$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} -7 & 7 & 0 & 0 \\ 6 & -13 & 7 & 0 \\ 0 & 6 & -13 & 7 \\ 0 & 0 & 6 & -6 \end{pmatrix} \end{array}$$

Mean hold times $\frac{1}{7}, \frac{1}{13}, \frac{1}{13}, \frac{1}{6}$ (units of years)

for states $0, 1, 2, 3$

if $i \neq j$ $\tilde{P}_{ij} = \dfrac{q_{ij}}{q_i} \Rightarrow \tilde{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{6}{13} & 0 & \frac{7}{13} & 0 \\ 0 & \frac{6}{13} & 0 & \frac{7}{13} \\ 0 & 0 & 1 & 0 \end{pmatrix}$

Forward and Backward Equations

From Q, it's possible to derive $P(t)$:

$P(t)$ satisfies the following differential equation

forward equation $$P'(t) = P(t)\,Q$$

because $$P'(t) = \lim_{h \to 0^+} \frac{P(t+h) - P(t)}{h}$$

$P(t+h) = P(t)P(h)$ is Chapman-Kolmogorov equation

$$= \lim_{h \to 0^+} \frac{P(t)P(h) - P(t)}{h}$$

$$= P(t) \lim_{h \to 0^+} \frac{P(h) - I}{h}$$

$$= P(t)\,P'(0) = P(t)\,Q$$

this says that if you know initial distribution and infinitesimal transition rates (Q), then you can find probabilities forward in time by solving eq.

backward equation $$P'(t) = Q\,P(t)$$

this says if you know distribution at time T then you can find initial distribution at time 0 by solving equation

Example

$$\begin{tikzcd} \textcircled{1} \arrow[r, bend left, "\lambda"] & \textcircled{2} \arrow[l, bend left, "\mu"] \end{tikzcd}$$

transition rate diagram

$$Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

forward equations $\quad P'(t) = P(t)Q = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$

$$\begin{cases} P_{11}' = -\lambda P_{11} + \mu P_{12} = -\lambda P_{11} + \mu(1-P_{11}) = \mu - (\lambda+\mu)P_{11} \\ P_{22}' = \lambda P_{21} - \mu P_{22} = \lambda(1-P_{22}) - \mu P_{22} = \lambda - (\lambda+\mu)P_{22} \\ P_{12}' = \dots \\ P_{21}' = \dots \end{cases}$$

$$\Rightarrow \begin{cases} P_{11}' = \mu - (\lambda+\mu)P_{11} & \qquad P_{12} = 1 - P_{11} \\ P_{22}' = \lambda - (\lambda+\mu)P_{22} & \qquad P_{21} = 1 - P_{22} \end{cases}$$

These are decoupled differential equations (independent of each other) that can be solved using a standard technique like separation of variables.

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

Matrix Exponential

It's possible to solve the forward, backward equations using a special tool called the matrix exponential.

$$e^{tQ} := I + tQ + \frac{t^2}{2!}Q^2 + \frac{t^3}{3!}Q^3 + \cdots$$

(this comes from the power series $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$)

<u>Theorem</u> $P(t) = e^{tQ}$ solves the equation $P' = PQ$ (and $P' = QP$)

<u>Proof</u> we plug e^{tQ} into both sides for P

$$\begin{aligned}
LHS = P' &= \frac{d}{dt}\left(e^{tQ}\right) \\
&= \frac{d}{dt}\left(I + tQ + \frac{t^2}{2!}Q^2 + \frac{t^3}{3!}Q^3 + \cdots\right. \\
&= Q + tQ^2 + \frac{t^2}{2!}Q^3 + \cdots
\end{aligned}$$

$$\begin{aligned}
RHS &= PQ \\
&= \left(I + tQ + \frac{t^2}{2!}Q^2 + \cdots\right)Q = LHS,
\end{aligned}$$

<u>Example</u> $Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$ $\lambda = 3$, $\mu = 7$

Wolfram Alpha command for e^{tQ}

```
MatrixExp( t * { {-3,3}, {7,-7} } )
```

(symbolic result)

R command

do this once →

```
install.packages("expm")
```

do this once per R session →

```
library(expm)

Q <- matrix( c(-3, 3, 7, -7), nrow = 2,
             byrow = TRUE)

P <- function(t) expm(t * Q)
```

P(2.5) gives transition function a time 2.5

Example

$$\begin{tikzcd} 1 \arrow[r, bend left, "3"] & 2 \arrow[l, bend left, "7"] \end{tikzcd}$$

$$Q = \begin{pmatrix} -3 & 3 \\ 7 & -7 \end{pmatrix}$$

$$0 = \det\left(Q - \lambda I\right) = \lambda^2 + 10\lambda = \lambda(\lambda + 10)$$

$\lambda = 0, -10$ are eigenvalues

eigenvectors for $\lambda = 0$ solve $(Q - \lambda I)v = 0$ for v

$$Q - \lambda I = Q = \begin{pmatrix} -3 & 3 \\ 7 & -7 \end{pmatrix}$$

$$\xrightarrow[\text{elimination}]{\text{Gauss}} \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \qquad \begin{array}{l} x = y \\ y \text{ free} \end{array}$$

eigenvectors are the set

$$\left\{ \begin{bmatrix} y \\ y \end{bmatrix} : y \in \mathbb{R} \right\} = \operatorname{span}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$$

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

eigenvectors for $\lambda = -10$ $\quad Q + 10 I = \begin{pmatrix} 7 & 3 \\ 7 & 3 \end{pmatrix} \to \begin{pmatrix} 7 & 3 \\ 0 & 0 \end{pmatrix}$

$$v_2 = \begin{bmatrix} -3 \\ 7 \end{bmatrix} \qquad \left\{ \begin{bmatrix} -3/7\, y \\ y \end{bmatrix} : y \in \mathbb{R} \right\} \qquad \begin{array}{l} 7x = -3y \\ x = \frac{-3}{7} y \end{array}$$

Fact from linearly algebra

1) eigenvectors from different eigenvalues are linearly independent.

2) if an $n \times n$ matrix Q has n linearly independent eigenvectors $v_1, \dots, v_n$, then

$$Q = \underbrace{\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}^{-1}}_{S^{-1}}$$

3) $$Q^k = (SDS^{-1})^k = SDS^{-1}SDS^{-1} \cdots SDS^{-1}$$
$$= SD^kS^{-1}$$

Therefore,

$$e^{tQ} = e^{tSDS^{-1}} = I + (tSDS^{-1}) + \frac{(tSDS^{-1})^2}{2!} + \frac{(tSDS^{-1})^3}{3!} + \cdots$$

$$= S\left(I + tD + \frac{t^2D^2}{2!} + \frac{t^3D^3}{3!} + \cdots\right)S^{-1}$$

$$= S\begin{bmatrix} e^{t\lambda_1} & & \\ & \ddots & \\ & & e^{t\lambda_n} \end{bmatrix}S^{-1}$$

Back to example

$$Q = \begin{bmatrix} 1 & -3 \\ 1 & 7 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -10 \end{bmatrix}\begin{bmatrix} 1 & -3 \\ 1 & 7 \end{bmatrix}^{-1}$$

$$P(t) = e^{tQ} = \begin{bmatrix} 1 & -3 \\ 1 & 7 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & e^{-10t} \end{bmatrix}\begin{bmatrix} 1 & -3 \\ 1 & 7 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 1 & -3e^{-10t} \\ 1 & 7e^{-10t} \end{bmatrix}\begin{bmatrix} 7/10 & 3/10 \\ -1/10 & 1/10 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{7}{10} + \frac{3}{10}e^{-10t} & \frac{3}{10} - \frac{3}{10}e^{-10t} \\ \frac{7}{10} - \frac{7}{10}e^{-10t} & \frac{3}{10} + \frac{7}{10}e^{-10t} \end{bmatrix}$$

Example Diagonalize Q in order to find closed form of e^{tQ}, where $Q = \begin{pmatrix} -3 & 1 & 2 \\ 2 & -5 & 3 \\ 1 & 1 & -2 \end{pmatrix}$

eigenvalues $0 = \det(Q - \lambda I)$

$$= \det \begin{pmatrix} -3-\lambda & 1 & 2 \\ 2 & -5 & 3 \\ 1 & 1 & -2 \end{pmatrix}$$

$$= -\lambda^3 - 10\lambda^2 - 24\lambda$$

$$= -\lambda(\lambda^2 + 10\lambda + 24) = -\lambda(\lambda+4)(\lambda+6)$$

$\lambda_1 = 0, \ \lambda_2 = -4, \ \lambda_3 = -6.$

eigenvectors for $\lambda_1 = 0$ $\quad Q - \lambda_1 I = Q = \begin{pmatrix} -3 & 1 & 2 \\ 2 & -5 & 3 \\ 1 & 1 & -2 \end{pmatrix}$

$$\longrightarrow \begin{pmatrix} 1 & 1 & -2 \\ 0 & -7 & 7 \\ 0 & 4 & -4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & -2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{array}{l} x = z \\ y = z \\ z \text{ free} \end{array}$$

eigenvector for $\lambda_1 = 0 = \left\{ \begin{bmatrix} z \\ z \\ z \end{bmatrix} : z \in \mathbb{R} \right\} = \text{span}\Big(\underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{= v_1} \Big)$

eigenvectors for $\lambda_2 = -4, \ \lambda_3 = -6$

$$v_2 = \begin{bmatrix} -5 \\ -1 \\ 3 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1 \\ -5 \\ 1 \end{bmatrix}$$

The command Eigenvalues({{-3,1,2}, {2,-5,3}, {1,1,-2}})

can be used in Wolfram Alpha to find these.

Then $e^{tQ} = \begin{bmatrix} 1 & -5 & 1 \\ 1 & -1 & -5 \\ 1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-4t} & 0 \\ 0 & 0 & e^{-6t} \end{bmatrix} \begin{bmatrix} 1 & -5 & 1 \\ 1 & -1 & -5 \\ 1 & 3 & 1 \end{bmatrix}^{-1}$

7.4 Long-term behavior

For CTMC's, we can also talk about

- limiting distribution $\quad \lim_{t \to \infty} P_{ij}(t) = \pi_j$

- stationary distribution $\quad \pi P(t) = \pi$ for all $t \geq 0$

Let's start with stationary distributions like we did in discrete case.

$$\pi P(t) = \pi \iff \pi \left(P(t) - I \right) = 0$$

$$\iff \pi \lim_{t \to 0^+} \frac{P(t) - I}{t} = 0$$

$$\iff \pi P'(0) = 0$$

$$\iff \pi Q = 0$$

Conclusion The stationary distribution π must satisfy $\boxed{\pi Q = 0}$

<u>Example</u>

$$Q = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & & & \\ 2 & & & \\ 3 & & & \end{array} \quad \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{pmatrix} -1 & 1 & 0 \\ 4 & -5 & 1 \\ 0 & 4 & -4 \end{pmatrix}$$

$$(\pi_1, \pi_2, \pi_3) \; Q = (0, 0, 0)$$

$\Longleftrightarrow$ solve $Q^T \pi^T = 0$

$$\Longleftrightarrow \begin{pmatrix} -1 & 4 & 0 \\ 1 & -5 & 4 \\ 0 & 1 & -4 \end{pmatrix}$$

$$\xrightarrow{\text{Row reduce}} \begin{pmatrix} 1 & -4 & 0 \\ 0 & -1 & 4 \\ 0 & 1 & -4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -16 \\ 0 & 1 & -4 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\pi_1 = 16\,\pi_3, \quad \pi_2 = 4\pi_3$$

$$\pi_1 + \pi_2 + \pi_3 = 1$$

$$(16\pi_3, 4\pi_3, \pi_3), \quad 21\pi_3 = 1 \Rightarrow \pi_3 = \frac{1}{21}$$

$$\Rightarrow \quad \pi = \left(\frac{16}{21}, \frac{4}{21}, \frac{1}{21} \right)$$

Limiting distribution

We don't have to worry about aperiodicity in continuous time case because of randomness of jump times.

irreducibility for any states i, j we must be able to get from i to j in a finite number of jumps (ie. $\tilde{P}$ must be irreducible)

a finite state CTMC is ergodic if it's irreducible

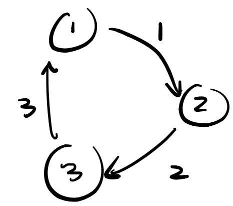
Theorem If $(X_t)_{t\geq 0}$ is a finite state, irreducible CTMC, then $\exists$ a unique, stationary, limiting distribution π.

Example For $Q = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} -1 & 1 & 0 \\ 4 & -5 & 1 \\ 0 & 4 & -4 \end{pmatrix} \end{array}$,

we found $\pi = \left(\frac{16}{21}, \frac{4}{21}, \frac{1}{21}\right)$ is stationary.

Thus $\lim_{t\to\infty} P(t) = \begin{pmatrix} - \pi - \\ - \pi - \\ - \pi - \end{pmatrix}$

Example $\quad Q = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -2 & 2 \\ 3 & 0 & -3 \end{pmatrix}$



$$\tilde{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{(periodic chain with period 3)}$$

$$Q^T \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} -1 & 0 & 3 \\ 1 & -2 & 0 \\ 0 & 2 & -3 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 1 & 0 & -3 \\ 0 & -2 & 3 \\ 0 & 2 & -3 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & -3/2 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\pi_1 = 3\pi_3, \quad \pi_2 = \frac{3}{2}\pi_3, \quad \pi_1 + \pi_2 + \pi_3 = 1$$

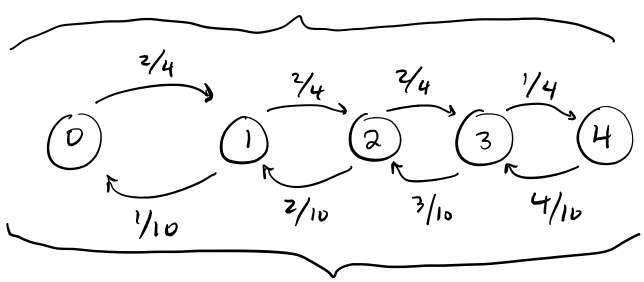
$$\left(3\pi_3, \ \frac{3}{2}\pi_3, \ \pi_3\right) \qquad \frac{11}{2}\pi_3 = 1 \Rightarrow \pi_3 = \frac{2}{11}$$

$$\pi = \left(\frac{6}{11}, \frac{3}{11}, \frac{2}{11}\right)$$

<u>Example</u> A facility has 4 machines, with 2 repair workers to maintain them. Individual machines fail every 10 hours. It takes an individual repair worker 4 hours to fix a machine. Repair and failure times are independent, exponentially dist.

X_t = # of operational machines

these are the rates at which machines are fixed. when more than one machine is broken they're fixed at a rate of $\frac{2 \text{ machines}}{4 \text{ hours}}$



these are rates at which machines fail. when n machines are operational, they fail at rate $\frac{n \text{ machines}}{10 \text{ hours}}$

$$Q = \begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \\ \hline 0 & -\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 1 & \frac{1}{10} & -\frac{3}{5} & \frac{1}{2} & 0 & 0 \\ 2 & 0 & \frac{1}{5} & -\frac{7}{10} & \frac{1}{2} & 0 \\ 3 & 0 & 0 & \frac{3}{10} & -\frac{11}{20} & \frac{1}{4} \\ 4 & 0 & 0 & 0 & \frac{2}{5} & -\frac{2}{5} \end{array}$$

$$\pi Q = 0 \Rightarrow \pi = (0.0191, 0.0955, 0.2388, 0.3979, 0.2487)$$

expected number of working machines in long-term

$$0 \cdot \pi_0 + 1 \cdot \pi_1 + 2 \cdot \pi_2 + 3\pi_3 + 4\pi_4$$

$$= 2.76$$

$P_{42}(5)$ = prob. of 2 working machines after 5 hours, given 4 working machines initially

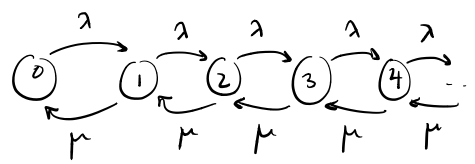
$$= 0.188$$

7.6 Queuing Theory

Example 1

Cars arrive according to a Poisson process at a car wash at average rate λ cars per hour and the wash can wash μ cars per hour, one at a time.

Let X_t count the number of cars at the wash at time t.



$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \cdots & \end{array} \\ \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & & \\ \mu & -(\lambda+\mu) & \lambda & 0 & \cdots & \\ 0 & \mu & -(\lambda+\mu) & \lambda & \cdots & \\ & & & & & \\ & & & & & \end{pmatrix} \end{array}$$

This is called an M/M/1 queue

1st M = memoryless arrivals (exponentially dist'd inter-arrival times)

2nd M = memoryless service times (exponentially distributed service times)

1 = one server

X_t is clearly irreducible.

What is its stationary distribution?

Detailed balance like in discrete time

if there exists distribution π such that

$$\pi_i q_{ij} = \pi_j q_{ji} \quad \text{for all } i, j$$

then π is stationary and X_t is time reversible.

Example 1 (M/M/1 queue)

Note
$$\pi_1 = \pi_0 \frac{q_{01}}{q_{10}} = \pi_0 \frac{\lambda}{\mu}$$

$$\pi_2 = \pi_1 \frac{q_{12}}{q_{21}} = \pi_1 \frac{\lambda}{\mu} = \pi_0 \left(\frac{\lambda}{\mu}\right)^2$$

$$\vdots$$

$$\pi_k = \pi_0 \left(\frac{\lambda}{\mu}\right)^k$$

we also need
$$1 = \sum_{k=0}^{\infty} \pi_k = \pi_0 \sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^k$$

So there is a stationary distribution

if $\frac{\lambda}{\mu} < 1 \iff \lambda < \mu$

and
$$\pi_0 = \frac{1}{\sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^k} = \frac{1}{\frac{1}{1-\frac{\lambda}{\mu}}} = 1 - \frac{\lambda}{\mu}$$

$$\pi_k = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^k \quad k = 0, 1, \ldots$$

Shifted geometric distribution, $p = 1 - \frac{\lambda}{\mu}$

Question in the long term how many cars are there at the car wash on average?

Answer Long-term (ie limiting distribution) of X_t is π. Let $X_\infty \sim \pi$

so $\lim_{t\to\infty} E[X_t] = E[X_\infty]$

$$= \frac{1-p}{p} = \frac{\lambda/\mu}{1-\lambda/\mu} = \frac{\lambda}{\mu-\lambda}$$

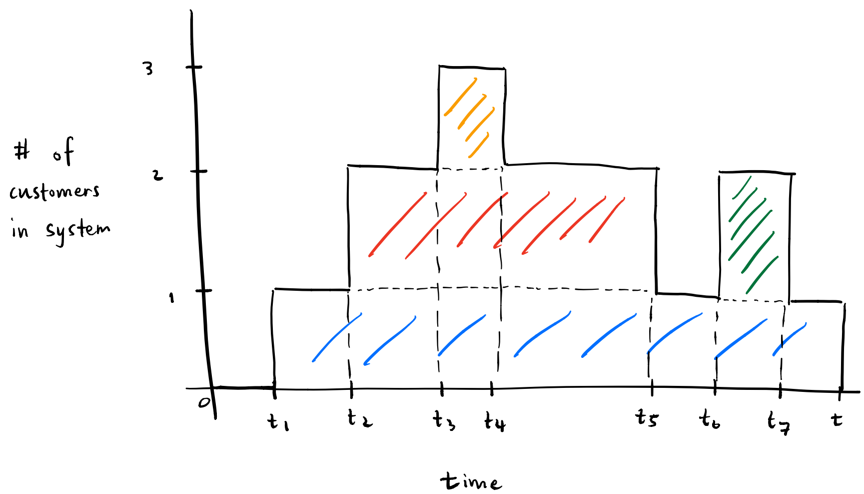
Little's Formula in a queueing system let

- L be the long-term average number of customers in the system
- λ the rate of arrivals
- W the long-term average time a customer spends in the system (in line and being served)

Then $$L = \lambda W$$

Intuition for why the formula works Suppose we have the following data

customer	arrival time	departure time
1	t_1	t_5
2	t_2	t_4
3	t_3	t_7
4	t_6	t



Let A be area under the curve above and N_t be number of customer arrivals in $[0,t]$

$$W = \text{average time customer spends in system}$$

$$= \frac{\sum_{i=1}^{4} \text{time customer } i \text{ spends in system}}{4}$$

$$= \frac{(t_5 - t_1) + (t_4 - t_2) + (t_7 - t_3) + (t - t_6)}{4}$$

$$= \frac{(t - t_1) + (t_5 - 2) + (t_7 - t_6) + (t_4 - t_3)}{N_t}$$

$$= \frac{A}{N_t}$$

$$N = \text{average number of customers in system}$$

$$= 1 \cdot \left[\frac{t_2 - t_1}{t} + \frac{t_6 - t_5}{t} + \frac{t - t_7}{t} \right] + 2 \cdot \left[\frac{t_3 - t_2}{t} + \frac{t_5 - t_4}{t} + \frac{t_7 - t_6}{2} \right]$$

$$+ 3 \cdot \left[\frac{t_4 - t_3}{t} \right]$$

$$= \frac{A}{t}$$

SLLN for Poisson process

$$\text{So } L = \frac{A}{t} = \frac{W N_t}{t} = W \cdot \frac{N_t}{t} \longrightarrow W\lambda \quad \text{as } t \to \infty$$

<u>Example</u> (M/M/1 queue) for car wash with

$\lambda = 9$ customers arriving per hour

$\mu = 12$ customers served per hour

(1) How many customers on average are at car wash?

$$L = \frac{\lambda}{\mu - \lambda} = \frac{9}{12-9} = 3$$

(2) How long on average is the customer at the wash?

$$W = \frac{L}{\lambda} = \frac{3}{9} \text{ hours} = 20 \text{ minutes}$$

(3) How long on average does the customer wait before being served?

$$W = \text{waiting time in queue} + \text{waiting time in service}$$
$$= W_q + W_s$$

$$W_q = W - W_s = \frac{1}{3} - \frac{1}{12} = \frac{1}{4} \text{ hour} = 15 \text{ mins}$$

(4) What is expected number of cars waiting to be served?

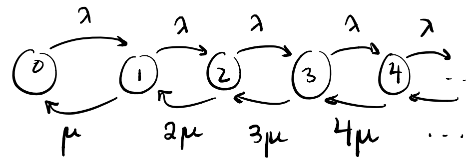
We could consider the process that counts just the cars in line (and not include those being serviced) as its own queue

Then $L_q = \lambda W_q = 9 \cdot \frac{1}{4} = 2.25$

customers in line but not being washed.

Example 2 Suppose the wash has c washing stations instead of 1 and each station can wash on average μ cars per hour like before.

Let X_t count the number of cars at the wash at time t.



if μ_i = rate from i to $i-1$ then

$$\mu_i = \begin{cases} i\mu & i = 1, \dots, c \\ c\mu & i = c+1, \dots \end{cases}$$

Example 2 (M/M/c queue stationary dist.)

$$\pi_1 q_{10} = \pi_0 q_{01} \Rightarrow \quad \pi_1 = \frac{q_{01}}{q_{10}} \pi_0 = \frac{\lambda}{\mu} \pi_0$$

$$\pi_2 q_{21} = \pi_1 q_{12} \Rightarrow \quad \pi_2 = \frac{q_{12}}{q_{21}} \pi_1 = \frac{\lambda}{2\mu} \pi_1 = \frac{\lambda^2}{2\mu^2} \pi_0$$

$$\pi_3 = \frac{q_{23}}{q_{32}} \pi_2 = \frac{\lambda}{3\mu} \pi_2 = \frac{\lambda^3}{3!\, \mu^3} \pi_0$$

$$\pi_k = \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k \pi_0 \qquad k \le c$$

$$\pi_{c+1} = \frac{q_{c,c+1}}{q_{c+1,c}} \pi_c = \frac{\lambda}{c\mu} \frac{1}{c!} \left(\frac{\lambda}{\mu}\right)^c \pi_0$$

$$= \frac{1}{c \cdot c!} \left(\frac{\lambda}{\mu}\right)^{c+1} \pi_0$$

$$\pi_{c+2} = \frac{1}{c^2 \cdot c!} \left(\frac{\lambda}{\mu}\right)^{c+2} \pi_0$$

$$\vdots$$

$$1 = \sum_{k=0}^{\infty} \pi_0 = \sum_{k=0}^{c} \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k \pi_0 + \sum_{k=1}^{\infty} \frac{1}{c^k\, c!} \left(\frac{\lambda}{\mu}\right)^{c+k} \pi_0$$

$$= \sum_{k=0}^{c} \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k \pi_0 + \left(\frac{\lambda}{\mu}\right)^c \cdot \frac{\pi_0}{c!} \sum_{k=1}^{\infty} \left(\frac{\lambda}{c\mu}\right)^k$$

$$= \pi_0 \left(\sum_{k=0}^{c} \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k + \left(\frac{\lambda}{\mu}\right)^c \frac{1}{c!} \cdot \frac{\lambda/c\mu}{1 - \lambda/c\mu} \right) \quad \text{for } \frac{\lambda}{c\mu} < 1.$$