Markov Chains

Alejandro Ribeiro
Dept. of Electrical and Systems Engineering
University of Pennsylvania
aribeiro@seas.upenn.edu
http://www.seas.upenn.edu/users/~aribeiro/

September 15, 2014

Markov chains

- Consider time index $n = 0, 1, 2, \ldots$ & time dependent random state X_n
- State X_n takes values on a countable number of states
  - In general denotes states as $i = 0, 1, 2, \ldots$
  - Might change with problem
- Denote the history of the process $\mathbf{X}_n = [X_n, X_{n-1}, \ldots, X_0]^T$
- Denote stochastic process as $\mathbf{X}_{\mathbb{N}}$
- The stochastic process $X_{\mathbb{N}}$ is a Markov chain (MC) if

$$\mathrm{P}\left[X_{n+1} = j \,\middle|\, X_n = i, \mathbf{X}_{n-1}\right] = \mathrm{P}\left[X_{n+1} = j \,\middle|\, X_n = i\right] = P_{ij}$$

- Future depends only on current state X_n

Observations

- Process's history $\mathbf{X}_{n-1}$ irrelevant for future evolution of the process
- Probabilities P_{ij} are constant for all times (time invariant)
- From the definition we have that for arbitrary m

$$\mathrm{P}\left[X_{n+m} \,\middle|\, X_n, \mathbf{X}_{n-1}\right] = \mathrm{P}\left[X_{n+m} \,\middle|\, X_n\right]$$

- X_{n+m} depends only on X_{n+m-1}, which depends only $on X_{n+m-2}$, ... which depends only on X_n
- Since P_{ij}'s are probabilities they're positive and sum up to 1

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

Matrix representation

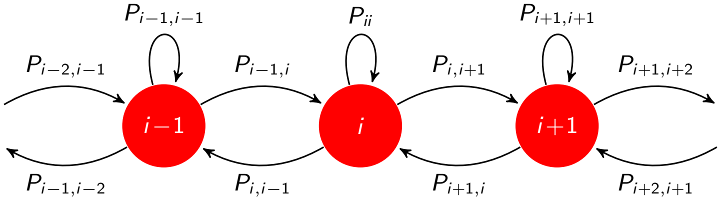
- Group transition probabilities P_{ij} in a "matrix" $\mathbf{P}$

$$\mathbf{P} := \begin{pmatrix} P_{00} & P_{01} & P_{02} & \ldots \\ P_{10} & P_{11} & P_{12} & \ldots \\ \vdots & \vdots & \vdots & \vdots \\ P_{i0} & P_{i1} & P_{i2} & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

- Not really a matrix if number of states is infinite

Graph representation

- A graph representation is also used

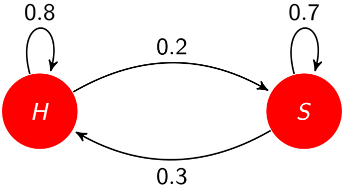


- Useful when number of states is infinite

Example: Happy - Sad

- I can be happy ($X_n = 0$) or sad ($X_n = 1$).
- Happiness tomorrow affected by happiness today only
- Model as Markov chain with transition probabilities

$$\mathbf{P} := \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$$

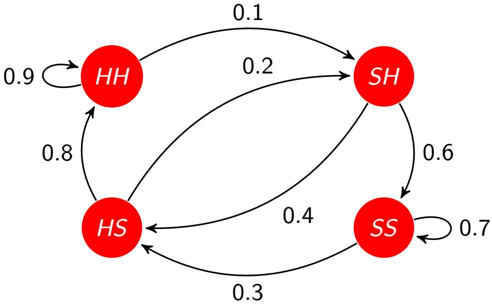


- Inertia $\Rightarrow$ happy or sad today, likely to stay happy or sad tomorrow ($P_{00} = 0.8$, $P_{11} = 0.7$)
- But when sad, a little less likely so ($P_{00} > P_{11}$)

Example: Happy - Sad, version 2

- Happiness tomorrow affected by today and yesterday
- Define double states HH (happy-happy), HS (happy-sad), SH, SS
- Only some transitions are possible
  - HH and SH can only become HH or HS
  - HS and SS can only become SH or SS

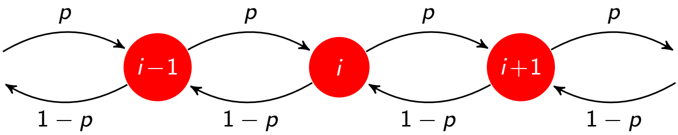
$$\mathbf{P} := \begin{pmatrix} 0.9 & 0.1 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \end{pmatrix}$$



- More time happy or sad increases likelihood of staying happy or sad
- State augmentation $\Rightarrow$ Capture longer time memory

Random (drunkard's) walk

- Step to the right with probability p, to the left with prob. (1-p)



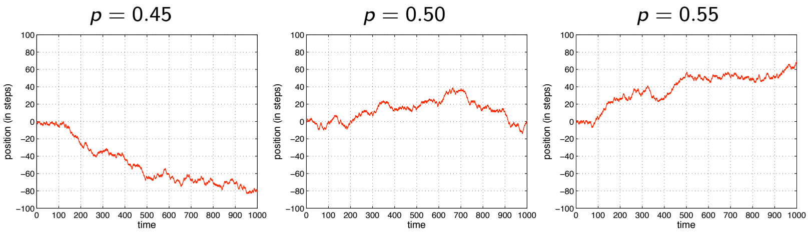
- States are $0, \pm 1, \pm 2, \ldots$, number of states is infinite
- Transition probabilities are

$$P_{i,i+1} = p, \qquad P_{i,i-1} = 1 - p,$$

- $P_{ij} = 0$ for all other transitions

Random (drunkard's) walk - continued

- Random walks behave differently if $p < 1/2$, $p = 1/2$ or $p > 1/2$



- With $p > 1/2$ diverges to the right (grows unbounded almost surely)
- With $p < 1/2$ diverges to the left
- With $p = 1/2$ always come back to visit origin (almost surely)

- Because number of states is infinite we can have all states transient
  - They are not revisited after some time (more later)

Two dimensional random walk

- Take a step in random direction East, West, South or North

  $\Rightarrow$ E, W, S, N chosen with equal probability

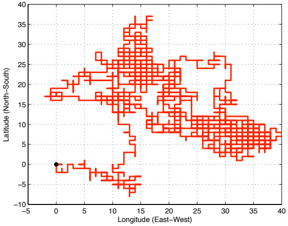
- States are pairs of coordinates (x, y)
  - $x = 0, \pm 1, \pm 2, \ldots$ and $y = 0, \pm 1, \pm 2, \ldots$
- Transiton probabilities are not zero only for points adjacent in the grid

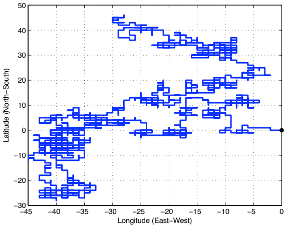
$$\mathrm{P}\left[x(t+1) = i+1, y(t+1) = j \,\middle|\, x(t) = i, y(t) = j\right] = \frac{1}{4}$$

$$\mathrm{P}\left[x(t+1) = i-1, y(t+1) = j \,\middle|\, x(t) = i, y(t) = j\right] = \frac{1}{4}$$

$$\mathrm{P}\left[x(t+1) = i, y(t+1) = j+1 \,\middle|\, x(t) = i, y(t) = j\right] = \frac{1}{4}$$

$$\mathrm{P}\left[x(t+1) = i, y(t+1) = j-1 \,\middle|\, x(t) = i, y(t) = j\right] = \frac{1}{4}$$





More about random walks

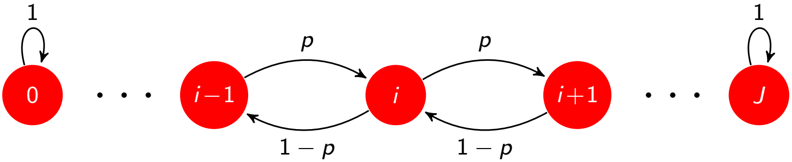
- Some random facts of life for equiprobable random walks

- In one and two dimensions probability of returning to origin is 1
- Will almost surely return home

- In more than two dimensions, probability of returning to origin is less than 1
- In three dimensions probability of returning to origin is 0.34
- Then 0.19, 0.14, 0.10, 0.08, ...

Random walk with borders (gambling)

- As a random walk, but stop moving when $i = 0$ or $i = J$
  - Models a gambler that stops playing when ruined, $X_n = 0$
  - Or when reaches target gains $X_n = J$



- States are $0, 1, \ldots, J$. Finite number of states (J). Transition probs.

$$P_{i,i+1} = p, \quad P_{i,i-1} = 1 - p, \qquad P_{00} = 1, \quad P_{JJ} = 1$$

- $P_{ij} = 0$ for all other transitions

- States 0 and J are called absorbing. Once there stay there forever
- The rest are transient states. Visits stop almost surely

Multiple step transition probabilities

- What can be said about multiple transitions ?
- Transition probabilities between two time slots

$$P_{ij}^{2} := \mathrm{P}\left[X_{m+2} = j \,\middle|\, X_m = i\right]$$

- Probabilities of X_{m+n} given X_n $\Rightarrow$ n-step transition probabilities

$$P_{ij}^{n} := \mathrm{P}\left[X_{m+n} = j \,\middle|\, X_m = i\right]$$

- Relation between n-step, m-step and $(m+n)$-step transition probs.
  - Write P_{ij}^{m+n} in terms of P_{ij}^{m} and P_{ij}^{n}
- All questions answered by Chapman-Kolmogorov's equations

2-step transition probabilities

- Start considering transition probs. between two time slots

$$P_{ij}^2 = \mathrm{P}\left[X_{n+2} = j \,\middle|\, X_n = i\right]$$

- Using the theorem of total probability

$$P_{ij}^2 = \sum_{k=1}^{\infty} \mathrm{P}\left[X_{n+2} = j \,\middle|\, X_{n+1} = k, X_n = i\right] \mathrm{P}\left[X_{n+1} = k \,\middle|\, X_n = i\right]$$

- In the first probability, conditioning on $X_n = i$ is unnecessary. Thus

$$P_{ij}^2 = \sum_{k=1}^{\infty} \mathrm{P}\left[X_{n+2} = j \,\middle|\, X_{n+1} = k\right] \mathrm{P}\left[X_{n+1} = k \,\middle|\, X_n = i\right]$$

- Which by definition yields

$$P_{ij}^2 = \sum_{k=1}^{\infty} P_{kj} P_{ik}$$

n-step, m-step and $(m+n)$-step

- Identical argument can be made (condition on X_0 to simplify notation, possible because of time invariance)

$$P_{ij}^{m+n} = \mathrm{P}\left[X_{n+m} = j \,\middle|\, X_0 = i\right]$$

- Use theorem of total probability, remove unnecessary conditioning and use definitions of n-step and m-step transition probabilities

$$P_{ij}^{m+n} = \sum_{k=1}^{\infty} \mathrm{P}\left[X_{m+n} = j \,\middle|\, X_m = k, X_0 = i\right] \mathrm{P}\left[X_m = k \,\middle|\, X_0 = i\right]$$

$$P_{ij}^{m+n} = \sum_{k=1}^{\infty} \mathrm{P}\left[X_{m+n} = j \,\middle|\, X_m = k\right] \mathrm{P}\left[X_m = k \,\middle|\, X_0 = i\right]$$

$$P_{ij}^{m+n} = \sum_{k=1}^{\infty} P_{kj}^{n} P_{ik}^{m}$$

Interpretation

- Chapman Kolmogorov is intuitive. Recall

$$P_{ij}^{m+n} = \sum_{k=1}^{\infty} P_{kj}^{n} P_{ik}^{m}$$

- Between times 0 and $m+n$ time m occurred
- At time m, the chain is in some state $X_m = k$
  - $\Rightarrow$ P_{ik}^{m} is the probability of going from $X_0 = i$ to $X_m = k$
  - $\Rightarrow$ P_{kj}^{n} is the probability of going from $X_m = k$ to $X_{m+n} = j$
  - $\Rightarrow$ Product $P_{ik}^{m} P_{kj}^{n}$ is then the probability of going from $X_0 = i$ to $X_{m+n} = j$ passing through $X_m = k$ at time m
- Since any k might have occurred sum over all k

Matrix form

- Define matrices $\mathbf{P}^{(m)}$ with elements P_{ij}^{m}, $\mathbf{P}^{(n)}$ with elements P_{ij}^{n} and $\mathbf{P}^{(m+n)}$ with elements P_{ij}^{m+n}
- $\sum_{k=1}^{\infty} P_{kj}^{n} P_{ik}^{m}$ is the (i,j)-th element of matrix product $\mathbf{P}^{(m)}\mathbf{P}^{(n)}$
- Chapman Kolmogorov in matrix form

$$\mathbf{P}^{(m+n)} = \mathbf{P}^{(m)}\mathbf{P}^{(n)}$$

- Matrix of $(n+m)$-step transitions is product of n-step and m-step

- For $m = n = 1$ (2-step transition probabilities) matrix form is

$$\mathbf{P}^{(2)} = \mathbf{P}\mathbf{P} = \mathbf{P}^2$$

- Proceed recursively backwards from n

$$\mathbf{P}^{(n)} = \mathbf{P}^{(n-1)}\mathbf{P} = \mathbf{P}^{(n-2)}\mathbf{P}\mathbf{P} = \ldots = \mathbf{P}^n$$

- Have proved the following

Theorem

The matrix of n-step transition probabilities $P^{(n)}$ is given by the n-th power of the transition probability matrix $\mathbf{P}$. i.e.,

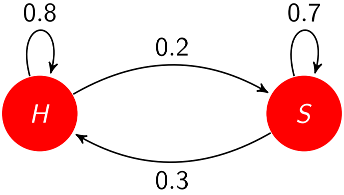
$$\mathbf{P}^{(n)} = \mathbf{P}^n$$

Henceforth we write $\mathbf{P}^n$

Example: Happy-Sad

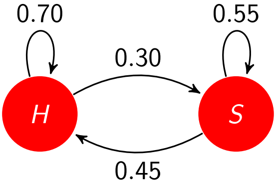
- Happiness transitions in one day (not the same as earlier example)

$$\mathbf{P} := \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$$



- Transition probabilities between today and the day after tomorrow?

$$\mathbf{P}^2 := \begin{pmatrix} 0.70 & 0.30 \\ 0.45 & 0.55 \end{pmatrix}$$



Example: Happy-Sad (continued)

- ... After a week and after a month

$$\mathbf{P}^7 := \begin{pmatrix} 0.6031 & 0.3969 \\ 0.5953 & 0.4047 \end{pmatrix} \qquad \mathbf{P}^{30} := \begin{pmatrix} 0.6000 & 0.4000 \\ 0.6000 & 0.4000 \end{pmatrix}$$

- Matrices $\mathbf{P}^7$ and $\mathbf{P}^{30}$ almost identical $\Rightarrow \lim_{n\to\infty} \mathbf{P}^n$ exists
  - Note that this is a regular limit
- After a month transition from H to H with prob. 0.6 and from S to H also 0.6
- State becomes independent of initial condition
- Rationale: 1-step memory $\Rightarrow$ initial condition eventually forgotten

Unconditional probabilities

- All probabilities so far are conditional, i.e., $\mathrm{P}\left[X_n = j \,\middle|\, X_0 = i\right]$
- Want unconditional probabilities $p_j(n) := \mathrm{P}\left[X_n = j\right]$
- Requires specification of initial conditions $p_i(0) := \mathrm{P}\left[X_0 = i\right]$

- Using theorem of total probability and definitions of P_{ij}^n and $p_j(n)$

$$p_j(n) := \mathrm{P}\left[X_n = j\right] = \sum_{i=1}^{\infty} \mathrm{P}\left[X_n = j \,\middle|\, X_0 = i\right] \mathrm{P}\left[X_0 = i\right]$$

$$= \sum_{i=1}^{\infty} P_{ij}^n p_i(0)$$

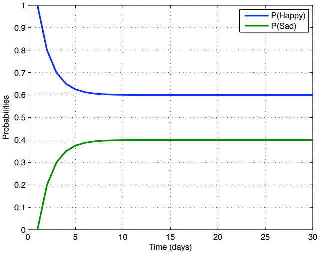
- Or in matrix form (define vector $\mathbf{p}(n) := [p_1(n), p_2(n), \ldots]^T$)

$$\mathbf{p}(n) = \mathbf{P}^{n\,T} \mathbf{p}(0)$$

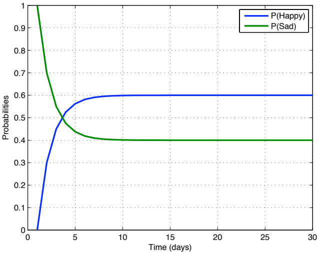
Example: Happy-Sad

- Transition probability matrix $\Rightarrow$ $\mathbf{P} := \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$

$\mathbf{p}(0) = [1, 0]$



$\mathbf{p}(0) = [0, 1]$



- For large n probabilities $\mathbf{p}(t)$ are independent of initial state $\mathbf{p}(0)$

Gambler's ruin problem

Gambler's ruin problem

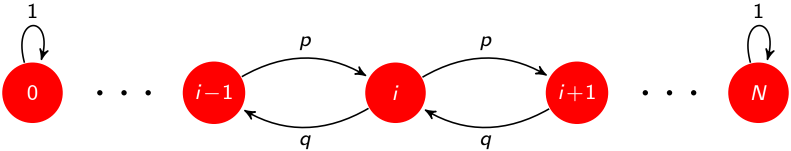
- You place $1 bets,
  (a) With probability p you gain \$1 and
  (b) With probability $q = (1 - p)$ you loose your \$1 bet
- Start with an initial wealth of $\$i_0$
- Define bias factor $\alpha := q/p$
  - If $\alpha > 1$ more likely to loose than win (biased against gambler)
  - $\alpha < 1$ favors gambler (more likely to win than loose)
  - $\alpha = 1/2$ game is fair
- You keep playing until
  (a) You go broke (loose all your money)
  (b) You reach a wealth of $\$N$
- Prob. S_i of reaching $\$N$ before going broke for initial wealth $\$i$?
  - S stands for success

Gambler's Markov chain

- Model as Markov chain $X_{\mathbb{N}}$. Transition probabilities

$$P_{i,i+1} = p, \quad P_{i,i-1} = q, \quad P_{00} = P_{NN} = 1$$

- Realizations $x_{\mathbb{N}}$. Initial state = initial wealth = i_0



- Sates 0 and N said absorbing. Eventually end up in one of them
- Remaining states said transient (visits eventually stop)

- Being absorbing states says something about the limit wealth

$$\lim_{n\to\infty} x_n = 0, \text{ or } \lim_{n\to\infty} x_n = N, \quad \Rightarrow \quad S_i := \mathrm{P}\left(\lim_{n\to\infty} X_n = N \,\middle|\, X_0 = i\right)$$

Recursive relations

- Prob. S_i of successful betting run depends on current state i only
- We can relate probabilities of SBR from adjacent states

$$S_i = S_{i+1}P_{i,i+1} + S_{i-1}P_{i,i-1} = S_{i+1}p + S_{i-1}q$$

- Recall $p + q = 1$. Reorder terms

$$p(S_{i+1} - S_i) = q(S_i - S_{i-1})$$

- Recall definition of bias $\alpha = q/p$

$$S_{i+1} - S_i = \alpha(S_i - S_{i-1})$$

Recursive relations (continued)

- If current state is 0 then $S_i = S_0 = 0$. Can write

  $$S_2 - S_1 = \alpha(S_1 - S_0) = \alpha S_1$$

- Substitute this in the expression for $S_3 - S_2$

  $$S_3 - S_2 = \alpha(S_2 - S_1) = \alpha^2 S_1$$

- Apply recursively backwards from $S_i - S_{i-1}$

  $$S_i - S_{i-1} = \alpha(S_{i-1} - S_{i-2}) = \ldots = \alpha^{i-1} S_1$$

- Sum up all of the former to obtain

$$S_i - S_1 = S_1 \left(\alpha + \alpha^2 + \ldots + \alpha^{i-1} \right)$$

- The latter can be written as a geometric series

$$S_i = S_1 \left(1 + \alpha + \alpha^2 + \ldots + \alpha^{i-1} \right)$$

Geometric series

- Geometric series can be summed. Assuming $\alpha \neq 1$

$$S_i = \frac{1 - \alpha^i}{1 - \alpha} S_1$$

- Write for $i = 1$. When in state N, $S_N = 1$

$$1 = S_N = \frac{1 - \alpha^N}{1 - \alpha} S_1$$

- Compute S_1 from latter and substitute into expression for S_i

$$S_i = \frac{1 - \alpha^i}{1 - \alpha^N}$$

- For $\alpha = 1 \;\Rightarrow S_i = iS_1$, $1 = S_N = NS_1$, $\;\Rightarrow S_i = (i/N)$

For large N

- Consider exit bound N arbitrarily large.
- For $\alpha \geq 1$, $S_i \approx (\alpha^i - 1)/\alpha^N \to 0$
- If win prob. does not exceed loose probability will almost surely loose all money
- For $\alpha < 1$, $P_i \to 1 - \alpha^i$
- If win prob. exceeds loose probability might win
- If initial wealth i sufficiently high, will most likely win

$\Rightarrow$ Which explains what we saw on first lecture and homework

Queues in communication systems

Queues in communication systems

- Communication systems goal

  $\Rightarrow$ Move packets from generating sources to intended destinations

- Between arrival and departure we hold packets in a memory buffer
- Want to design buffers appropriately

Non concurrent queue

- Time slotted in intervals of duration Δt
- n-th slot between times $n\Delta t$ and $(n+1)\Delta t$
- Average arrival rate is $\bar{\lambda}$ packets per unit time
- During slot of duration Δt probability of packet arrival is $\lambda = \bar{\lambda}\Delta t$
- Packets are transmitted (depart) at a rate of $\bar{\mu}$ packets per unit time
- During interval Δt probability of packet departure is $\mu = \bar{\mu}\Delta t$
- Assume no simultaneous arrival and departure (no concurrence)
  - Reasonable for small Δt (μ and λ are likely small)

Queue evolution equations

- q_n denotes number of packets in queue in n-th time slot
- $\mathbb{A}_n =$ nr. of packet arrivals, $\mathbb{D}_n =$ nr. of departures (during n-th slot)

- If there are no packets in queue $q_n = 0$ then there are no departures
- Queue length at time $n+1$ can be written as

$$q_{n+1} = q_n + \mathbb{A}_n, \quad \text{if } q_n = 0$$

- If $q_n > 0$, departures and arrivals may happen

$$q_{n+1} = \left[q_n + \mathbb{A}_n - \mathbb{D}_n\right]^+, \quad \text{if } q_n > 0$$

- $\mathbb{A}_n \in \{0, 1\}$, $\mathbb{D}_n \in \{0, 1\}$ and either $\mathbb{A}_n = 1$ or $\mathbb{D}_n = 1$ but not both
- Arrival and departure probabilities are

$$\mathrm{P}\left[\mathbb{A}_n = 1\right] = \lambda, \qquad \mathrm{P}\left[\mathbb{D}_n = 1\right] = \mu$$

Queue evolution probabilities

- Future queue lengths depend on current length only
- Probability of queue length increasing

$$\mathrm{P}\left[q_{n+1} = i+1 \,\middle|\, q_n = i\right] = \mathrm{P}\left[\mathbb{A}_n = 1\right] = \lambda, \qquad \text{for all } i$$

- Queue length might decrease only if $q_n > 0$. Probability is

$$\mathrm{P}\left[q_{n+1} = i-1 \,\middle|\, q_n = i\right] = \mathrm{P}\left[\mathbb{D}_n = 1\right] = \mu, \qquad \text{for all } i > 0$$

- Queue length stays the same if it neither increases nor decreases

$$\mathrm{P}\left[q_{n+1} = i \,\middle|\, q_n = i\right] = 1 - \lambda - \mu, \qquad \text{for all } i > 0$$
$$\mathrm{P}\left[q_{n+1} = 0 \,\middle|\, q_n = 0\right] = 1 - \lambda$$

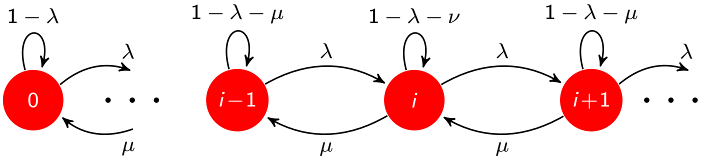
- No departures when $q_n = 0$ explain second equation

Queue as a Markov chain

- MC with states $0, 1, 2, \ldots$. Identify states with queue lengths
- Transition probabilities for $i \neq 0$ are

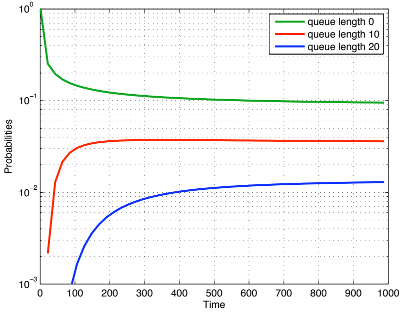
$$P_{i,i-1} = \lambda, \qquad P_{i,i} = 1 - \lambda - \mu, \qquad P_{i,i+1} = \mu$$

- For $i = 0$ $P_{0,0} = 1 - \lambda$ and $P_{01} = \lambda$



Numerical example: Probability propagation

- Build matrix $\mathbf{P}$ truncating at maximum queue length $L = 100$
- Arrival rate $\lambda = 0.3$. Departure rate $\mu = 0.33$
- Initial probability distribution $\mathbf{p}(0) = [1, 0, 0, \ldots]^T$ (queue empty)



- Propagate probabilities with product $\mathbf{P}^n\mathbf{p}(0)$
- Probabilities obtained are

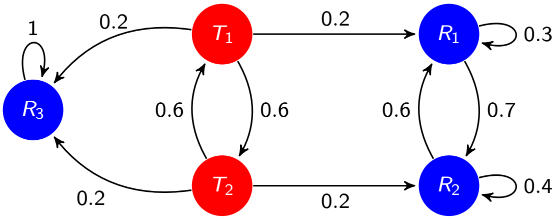
$$\mathrm{P}\left[q_n = i \,\middle|\, q_0 = 0\right] = p_i(n)$$

- A few i's (0, 10, 20) shown
- Probability of empty queue ≈ 0.1.
- Occupancy decrease with index

Classes of states

Transient and recurrent states

- States of a MC can be recurrent or transient
- Transient states might be visited at the beginning but eventually visits stop
- Almost surely, $X_n \neq i$ for n sufficiently large (qualifications needed)
- Visits to recurrent states keep happening forever
- Fix arbitrary m
- Almost surely, $X_n = i$ for some $n \geq m$ (qualifications needed)



Definitions

- Let f_i be the probability that starting at i, MC ever reenters state i

$$f_i := \mathrm{P}\left[\bigcup_{n=1}^{\infty} X_n = i \,\Big|\, X_0 = i\right] = \mathrm{P}\left[\bigcup_{n=m+1}^{\infty} X_n = i \,\Big|\, X_m = i\right]$$

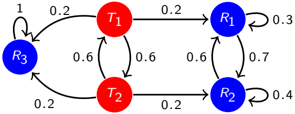
- State i is recurrent if $f_i = 1$
- Process reenters i again and again (almost surely). Infinitely often

- State i is transient if $f_i < 1$
- Positive probability $(1 - f_i)$ of never coming back to i

Recurrent states example

- State R_3 is recurrent because $\mathrm{P}\left[X_1 = R_3 \,\middle|\, X_0 = R_3\right] = 1$

- State R_1 is recurrent because



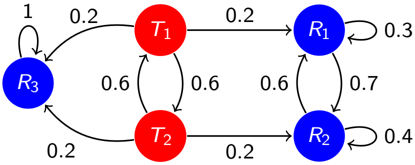
$$\mathrm{P}\left[X_1 = R_1 \,\middle|\, X_0 = R_1\right] = 0.3$$
$$\mathrm{P}\left[X_2 = R_1, X_1 \neq R_1 \,\middle|\, X_0 = R_1\right] = (0.7)(0.6)$$
$$\mathrm{P}\left[X_3 = R_1, X_2 \neq R_1, X_1 \neq R_1 \,\middle|\, X_0 = R_1\right] = (0.7)(0.4)(0.6)$$
$$\vdots$$
$$\mathrm{P}\left[X_n = R_1, X_{n-1} \neq R_1, \ldots, X_1 \neq R_1 \,\middle|\, X_0 = R_1\right] = (0.7)(0.4)^{n-1}(0.6)$$

- Sum up: $f_i = \sum_{n=1}^{\infty} \mathrm{P}\left[X_n = R_1, X_{n-1} \neq R_1, \ldots, X_1 \neq R_1 \,\middle|\, X_0 = R_1\right]$

$$= 0.3 + 0.7\left(\sum_{n=1}^{\infty} 0.4^{n-1}\right)0.6 = 0.3 + 0.7\left(\frac{1}{1-0.4}\right)0.6 = 1$$

Transient state example

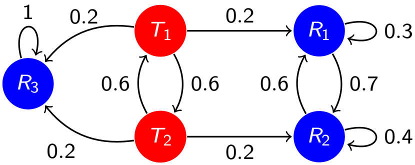
- States T_1 and T_2 are transient
- Probability of returning to T_1 is $f_{T_1} = (0.6)^2 = 0.36$
- Might come back to T_1 only if it goes to T_2 (with prob. 0.6)
- Will come back only if it moves back from T_2 to T_1 (with prob. 0.6)



- Likewise, $f_{T_2} = (0.6)^2 = 0.36$

Accessibility

- State j is accessible from state i if $P_{ij}^n > 0$ for some $n \geq 0$
- It is possible to enter j if MC initialized at $X_0 = i$
- Since $P_{ii}^0 = \text{P}\left[X_0 = 1 \,\middle|\, X_0 = i\right] = 1$, state i is accessible from itself



- All states accessible from T_1 and T_2
- Only R_1 and R_2 accessible from R_1 or R_2
- None other than itself accessible from R_3

Communication

- States i and j are said to communicate ($i \leftrightarrow j$) if
  $\Rightarrow$ i is accessible from j, $P_{ij}^n > 0$ for some n; and
  $\Rightarrow$ j is accessible from i, $P_{ji}^m > 0$ for some m

- Communication is an equivalence relation
- Reflexivity: $i \leftrightarrow i$
  - true because $P_{ii}^0 = 1$
- Symmetry: If $i \leftrightarrow j$ then $j \leftrightarrow i$
  - If $i \leftrightarrow j$ then $P_{ij}^n > 0$ and $P_{ji}^m > 0$ from where $j \leftrightarrow i$
- Transitivity: If $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$
  - Just notice that $P_{ik}^{n+m} \geq P_{ij}^n P_{jk}^m > 0$

- Partitions set of states into disjoint classes (as all equivalences do)
- What are these classes? (start with a brief detour)

Expected number of visits to states

- Define N_i as the number of visits to state i given that $X_0 = i$

$$N_i := \sum_{n=1}^{\infty} \mathbb{I}\{X_n = i\}$$

- If $X_n = i$, this is the last visit to i with probability $1 - f_i$
- Prob. revisiting state i exactly n times is (n visits $\times$ no more visits)

$$\mathrm{P}[N_i = n] = f_i^n(1 - f_i)$$

- Number of visits N_i has a geometric distribution with parameter f_i
- Expected number of visits is

$$\mathbb{E}[N_i] = \sum_{n=1}^{\infty} n f_i^n (1 - f_i) = \frac{1}{1 - f_i}$$

- For recurrent states $N_i = \infty$ almost surely and $\mathbb{E}[N_i] = \infty$ ($f_i = 1$)

- Another way of writing $\mathbb{E}[N_i]$

$$\mathbb{E}[N_i] = \sum_{n=1}^{\infty} \mathbb{E}\Big[\mathbb{I}\{X_n = i\}\Big] = \sum_{n=1}^{\infty} P_{ii}^n$$

- Recall that: for transient states $\mathbb{E}[N_i] = 1/(1-f_1)$
  for recurrent states $\mathbb{E}[N_i] = \infty$
- Therefore proving

Theorem

- *State i is transient if and only if $\sum_{n=1}^{\infty} P_{ii}^n < \infty$*
- *State i is recurrent if and only if $\sum_{n=1}^{\infty} P_{ii}^n = \infty$*

- Number of future visits to transient states is finite
- If number of states is finite some states have to be recurrent

Theorem

If state i is recurrent and $i \leftrightarrow j$, then j is recurrent

Proof.

- If $i \leftrightarrow j$ then there are l, m such that $P_{ji}^{l} > 0$ and $P_{ij}^{m} > 0$
- Then, for any n we have

$$P_{jj}^{l+n+m} \geq P_{ji}^{l} P_{ii}^{n} P_{ij}^{m}$$

- Sum for all n. Note that since i is recurrent $\sum_{n=1}^{\infty} P_{ii}^{n} = \infty$

$$\sum_{n=1}^{\infty} P_{jj}^{l+n+m} \geq \sum_{n=1}^{\infty} P_{ji}^{l} P_{ii}^{n} P_{ij}^{m} = P_{ji}^{l} \left(\sum_{n=1}^{\infty} P_{ii}^{n} \right) P_{ij}^{m} = \infty$$

- Which implies j is recurrent

$\square$

Corollary

If state i is transient and $i \leftrightarrow j$ then j is transient

Proof.

- If j were recurrent, then i would be recurrent from previous theorem

$\square$

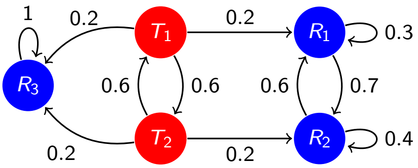
- Since communication defines classes and recurrence is shared by elements of this class, we say that recurrence is a class property
- Likewise, transience is also a class property

- States of a MC are separated in classes of transient and recurrent states

Irreducible Markov chains

- A MC is called irreducible if it has only one class
  - All states communicate with each other
  - If MC also has finite number of states the single class is recurrent
  - If MC infinite, class might be transient
- When it has multiple classes (not irreducible)
  - Classes of transient states $\mathcal{T}_1, \mathcal{T}_2, \ldots$
  - Classes of recurrent states $\mathcal{R}_1, \mathcal{R}_2, \ldots$
  - If MC initialized in a recurrent class $\mathcal{R}_k$, stays within the class
  - If starts in transient class $\mathcal{T}_k$, might stay on $\mathcal{T}_k$ or end up in a recurrent class $\mathcal{R}_l$
- For large time index n, MC restricted to one class
- Can be separated into irreducible components

Example



- Three classes
  - $\Rightarrow \mathcal{T} := \{T_1, T_2\}$, class with transient states
  - $\Rightarrow \mathcal{R}_1 := \{R_1, R_2\}$, class with recurrent states
  - $\Rightarrow \mathcal{R}_2 := \{R_3\}$, class with recurrent states
- Asymptotically in n suffices to study behavior for the irreducible components $\mathcal{R}_1$ and $\mathcal{R}_2$

Messages

- States of a MC can be transient of recurrent
- A MC can be partitioned in classes of states reachable from each other
- Elements of the class are either all recurrent or all transient
- A MC with only one class is irreducible
- If not irreducible can be separated in irreducible components

Limiting distributions

- MCs have one-step memory. Eventually they forget initial state
- What can we say about probabilities for large n?

$$\pi_j := \lim_{n\to\infty} \text{P}\left[X_n = j \,\middle|\, X_0 = i\right] = \lim_{n\to\infty} P_{ij}^n$$

- Implicitly assumed that limit is independent of initial state $X_0 = i$
- We've seen that this problem is related to the matrix power $\mathbf{P}^n$

$$\mathbf{P} := \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix} \qquad \mathbf{P}^7 := \begin{pmatrix} 0.6031 & 0.3969 \\ 0.5953 & 0.4047 \end{pmatrix}$$

$$\mathbf{P}^2 := \begin{pmatrix} 0.7 & 0.3 \\ 0.45 & 0.55 \end{pmatrix} \qquad \mathbf{P}^{30} := \begin{pmatrix} 0.6000 & 0.4000 \\ 0.6000 & 0.4000 \end{pmatrix}$$

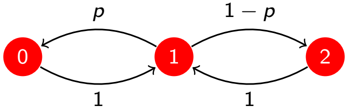
- Matrix product converges $\Rightarrow$ probs. independent of time (large n)
- All columns are equal $\Rightarrow$ probs. independent of initial condition

Periodicity

- The period of a state i is defined as ($\dot{d}$ is set of multiples of d)

$$d = \max \left\{ d : P_{ii}^{n} = 0 \text{ for all } n \notin \dot{d} \right\}$$

- State i is periodic with period d if and only if
  - $\Rightarrow$ $P_{ii}^{n} \neq 0$ only if n is a multiple of d ($n \in \dot{d}$)
  - $\Rightarrow$ d is the largest number with this property
- Positive probability of returning to i only every d time steps
- If period $d = 1$ state is aperiodic (most often the case)
- Periodicity is a class property



- State 1 has period 2. So do 0 and 2 (class property)
- One dimensional random walk also has period 2

Positive recurrence and ergodicity

- Recall: state i is recurrent if chain returns to i with probability 1
- Proved it was equivalent to $\sum_{n=1}^{\infty} P_{ii}^{n} = \infty$

- Positive recurrent when expected value of return time is finite

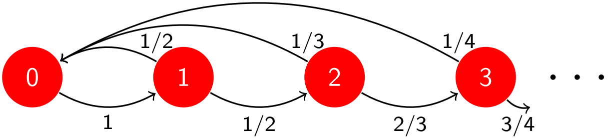
$$\mathbb{E}\left[\text{return time}\right] = \sum_{n=1}^{\infty} nP_{ii}^{n} \prod_{m=0}^{n-1} (1 - P_{ii}^{m}) < \infty$$

- Null recurrent if recurrent but $\mathbb{E}\left[\text{return time}\right] = \infty$

- Positive and null recurrence are class properties
- Recurrent states in a finite-state MC are positive recurrent

- Ergodic states are those that are positive recurrent and aperiodic
- An irreducible MC with ergodic states is said to be an ergodic MC

Example of a null recurrent MC



$$\mathrm{P}\left[\text{return time} = 2\right] = \frac{1}{2} \qquad \mathrm{P}\left[\text{return time} = 3\right] = \frac{1}{2} \times \frac{1}{3}$$

$$\mathrm{P}\left[\text{return time} = 4\right] = \frac{1}{2} \times \frac{2}{3} \times \frac{1}{4} = \frac{1}{3 \times 4} \quad \ldots \quad \mathrm{P}\left[\text{return time} = n\right] = \frac{1}{(n-1) \times n}$$

- It is recurrent because probability of returning is 1 (use induction)

$$\sum_{m=2}^{n} \mathrm{P}\left[\text{return time} = m\right] = \sum_{m=2}^{n} \frac{1}{(m-1) \times m} = \frac{n-1}{n} \to 1$$

- Null recurrent because expected return time is infinite

$$\sum_{n=2}^{\infty} n\mathrm{P}\left[\text{return time} = n\right] = \sum_{n=2}^{\infty} \frac{n}{(n-1) \times n} = \sum_{n=2}^{\infty} \frac{1}{(n-1)} = \infty$$

Limit distribution of ergodic Markov chains

Theorem

For an irreducible ergodic MC, $\lim_{n\to\infty} P_{ij}$ *exists and is independent of the initial state* i. *That is*

$$\pi_j = \lim_{n\to\infty} P_{ij}^n \quad \text{exists}$$

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

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \qquad \sum_{j=0}^{\infty} \pi_j = 1$$

- As observed, limit probs. independent of initial condition exist
- Simple algebraic equations can be solved to find π_j
- No periodic states, transient states, multiple classes or null recurrent

Algebraic relation to determine limit probabilities

- Difficult part of theorem is to prove that $\pi_j = \lim_{n\to\infty} P_{ij}^n$ exists
- To see that algebraic relation is true use theorem of total probability (omit conditioning on X_0 to simplify notation)

$$\begin{aligned} \mathrm{P}\left[X_{n+1}=j\right] &= \sum_{i=1}^{\infty} \mathrm{P}\left[X_{n+1}=j \,\middle|\, X_n=i\right] \mathrm{P}\left[X_n=i\right] \\ &= \sum_{i=1}^{\infty} P_{ij} \mathrm{P}\left[X_n=i\right] \end{aligned}$$

- If limits exists, $\mathrm{P}\left[X_{n+1}=j\right] \approx \mathrm{P}\left[X_n=j\right] \approx \pi_j$ (sufficiently large n)

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}$$

- The other equation is true because the π_j are probabilities

Vector/matrix notation: Matrix limit

- More compact and illuminating on vector/matrix notation
- Finite MC with J states

- First part of theorem says that $\lim_{n\to\infty} \mathbf{P}^n$ exists and

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{pmatrix} \pi_1 & \pi_2 & \ldots & \pi_J \\ \pi_1 & \pi_2 & \ldots & \pi_J \\ \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \ldots & \pi_J \end{pmatrix}$$

- Same probs. for all rows $\Rightarrow$ independent of initial state
- Probability distribution for large n.

$$\lim_{n\to\infty} \mathbf{p}(n) = \lim_{n\to\infty} \mathbf{P}^{T^n} \mathbf{p}(0) = [\pi_0, \pi_1, \ldots, \pi_J]^T$$

- Independent of initial condition

Vector/matrix notation: Eigenvector

- Define vector stationary distribution $\boldsymbol{\pi} := [\pi_0, \pi_1, \ldots, \pi_J]^T$
- Limit distribution is unique solution of ($\mathbf{1} = [1, 1, \ldots]^T$)

$$\boldsymbol{\pi} = \mathbf{P}^T\boldsymbol{\pi}, \qquad \boldsymbol{\pi}^T\mathbf{1} = 1$$

- $\boldsymbol{\pi}$ eigenvector associated with eigenvalue 1 of $\mathbf{P}^T$
  - Eigenvectors are defined up to a constant
  - Normalize to sum 1
- All other eigenvectors of $\mathbf{P}^T$ have modulus smaller than 1
  - If not, $\mathbf{P}^n$ diverges, but we know $\mathbf{P}^n$ contains n-step transition probs.
  - $\boldsymbol{\pi}$ eigenvector associated with largest eigenvalue of $\mathbf{P}^T$
- Computing $\boldsymbol{\pi}$ as eigenvector is computationally efficient and robust in some problems

Vector/matrix notation: Rank

- Can also write as ($\mathbf{I}$ is identity matrix, $\mathbf{0} = [0, 0, \ldots]^T$)

$$\left(\mathbf{I} - \mathbf{P}^T\right) \boldsymbol{\pi} = \mathbf{0} \qquad \boldsymbol{\pi}^T \mathbf{1} = 1$$

- $\boldsymbol{\pi}$ has J elements, but there are $J+1$ equations $\Rightarrow$ overdetermined

- If 1 is eigenvalue of $\mathbf{P}^T$, then 0 is eigenvalue of $\mathbf{I} - \mathbf{P}^T$
  - $\mathbf{I} - \mathbf{P}^T$ is rank deficient, in fact $\text{rank}(\mathbf{I} - \mathbf{P}^T) = J - 1$
  - Then, there are in fact only J equations

- $\boldsymbol{\pi}$ is eigenvector associated with eigenvalue 0 of $\mathbf{I} - \mathbf{P}^T$
  - $\boldsymbol{\pi}$ spans null space of $\mathbf{I} - \mathbf{P}^T$ (not much significance)

Example: Aperiodic, irreducible Markov chain

- MC with transition probability matrix

$$\mathbf{P} := \begin{pmatrix} 0 & 0.3 & 0.7 \\ 0.1 & 0.5 & 0.4 \\ 0.1 & 0.2 & 0.7 \end{pmatrix}$$

- Does P correspond to an ergodic MC?
  - All states communicate with state 2 (full row and column $P_{2j} \neq 0$ and $P_{j2} \neq 0$ for all j)
  - No transient states (irreducible with one recurrent state and finite)
  - Aperiodic (period of state 2 is 1)
- Then, there exist π_1, π_2 and π_3 such that $\pi_j = \lim_{n\to\infty} P_{ij}^n$
- Limit is independent of i

Example: Aperiodic, irreducible MC (continued)

- How do we determine limit probabilities π_j?
- Solve system of linear equations $\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}$ and $\sum_{j=0}^{\infty} \pi_j = 1$

$$\begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 0.1 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.7 & 0.4 & 0.7 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix}$$

- The upper part of matrix above is $\mathbf{P}^T$
- There are three variables and four equations
  - Some equations might be linearly dependent
  - Indeed, summing first three equations: $\pi_1 + \pi_2 + \pi_3 = \pi_1 + \pi_2 + \pi_3$
  - Always true, because probabilities in rows of $\mathbf{P}$ sum up to 1
  - This is because of rank deficiency of $\mathbf{I} - \mathbf{P}^T$
- Solution yields $\pi_1 = 0.0909$, $\pi_2 = 0.2987$ and $\pi_3 = 0.6104$

Stationary distribution

- Limit distributions are sometimes called stationary distributions
- Select initial distribution such that $\mathrm{P}[X_0 = i] = \pi_i$ for all i
- Probabilities at time $n = 1$ follow from theorem of total probability

$$\mathrm{P}[X_1 = i] = \sum_{i=1}^{\infty} \mathrm{P}\left[X_1 = j \,\middle|\, X_0 = i\right] \mathrm{P}[X_0 = i]$$

- Definitions of P_{ij}, and $\mathrm{P}[X_0 = i] = \pi_i$. Algebraic property of π_j

$$\mathrm{P}[X_1 = i] = \sum_{i=1}^{\infty} P_{ij}\pi_i = \pi_j$$

- Probability distribution is unchanged
- Proceeding recursively, system initialized with $\mathrm{P}[X_0 = i] = \pi_i$,
  $\Rightarrow$ Probability distribution constant, $\mathrm{P}[X_n = i] = \pi_i$ for all n
- MC stationary in a probabilistic sense (states change, probs. do not)

Ergodicity

Ergodicity

- Define $T_i^{(n)}$ as fraction of time spent in i-th state up to time n

$$T_i^{(n)} := \frac{1}{n} \sum_{m=1}^{n} \mathbb{I}\{X_m = i\}$$

- Compute expected value of $T_i^{(n)}$

$$\mathbb{E}\left[T_i^{(n)}\right] = \frac{1}{n} \sum_{m=1}^{n} \mathbb{E}\left[\mathbb{I}\{X_m = i\}\right] = \frac{1}{n} \sum_{m=1}^{n} \mathrm{P}\left[X_m = i\right] \to \pi_i$$

- As time $n \to \infty$, probabilities $\mathrm{P}\left[X_m = i\right]$ approach π_i. Then

$$\lim_{t\to\infty} \mathbb{E}\left[T_i^{(n)}\right] = \lim_{t\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathrm{P}\left[X_m = i\right] = \pi_i$$

- For ergodic MCs same is true without expected value $\Rightarrow$ ergodicity

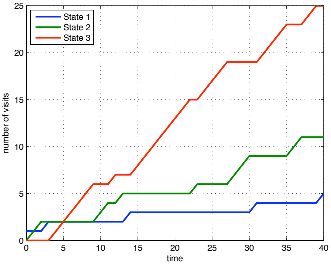
$$\lim_{n\to\infty} T_i^{(n)} = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{I}\{X_m = i\} = \pi_i, \quad \text{a.s.}$$

Example: Ergodic Markov chain

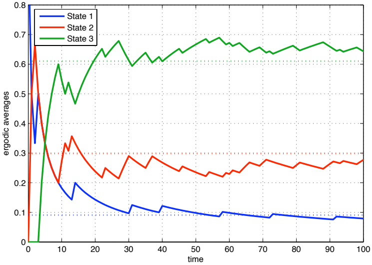
- Recall transition probability matrix

$$\mathbf{P} := \begin{pmatrix} 0 & 0.3 & 0.7 \\ 0.1 & 0.5 & 0.4 \\ 0.1 & 0.2 & 0.7 \end{pmatrix}$$

Visits to states, $nT_i^{(n)}$



Ergodic averages, $T_i^{(n)}$



- Ergodic averages slowly converge to limit probabilities

Function's Ergodic average

- Use of ergodic averages is more general than $T_i^{(n)}$

Theorem

Consider an irreducible Markov chain with states $X_n = 0, 1, 2, \ldots$ and stationary probabilities π_j. Let $f(X_n)$ be a bounded function of the state $X(n)$. Then, with probability 1

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} f(X_m) = \sum_{i=1}^{\infty} f(i) \pi_i$$

- $T_i^{(n)}$ is a particular case with $f(X_m) = \mathbb{I}\{X_m = i\}$
- Think of $f(X_m)$ as a reward associated with state $X(m)$
- $(1/n) \sum_{m=1}^{n} f(X_m)$ is the time average of rewards

Function's Ergodic average (proof)

Proof.

- Because $\mathbb{I}\{X_m = i\} = 1$ if and only if $X_m = i$ we can write

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m) = \frac{1}{n}\sum_{m=1}^{n}\left(\sum_{i=1}^{\infty} f(i)\mathbb{I}\{X_m = i\}\right)$$

- Change order of summations. Use definition of $T_i^{(n)}$

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m) = \sum_{i=1}^{\infty} f(i)\left(\frac{1}{n}\sum_{m=1}^{n}\mathbb{I}\{X_m = i\}\right) = \sum_{i=1}^{\infty} f(i)T_i^{(n)}$$

- Let $n \to \infty$ in both sides
- Use ergodic average result for $\lim_{n\to\infty} T_i^{(n)} = \pi_i$ [cf. page 67]

$\square$

Ensemble and ergodic averages

- There's more depth to ergodic results than meets the eye
- Ensemble average: across different realizations of the MC

$$\mathbb{E}\left[f(X_n)\right] = \sum_{i=1}^{\infty} f(i)\mathrm{P}\left(X_n = i\right) \to \sum_{i=1}^{\infty} f(i)\pi_i$$

- Ergodic average: across time for a single realization of the MC

$$\bar{f}(n) = \frac{1}{n}\sum_{m=1}^{n} f(X_n)$$

- These quantities are fundamentally different but their values coincide asymptotically in n
- Observing one realization of the MC provides as much information as observing all realizations
- Practical consequence: Observe/simulate only one path of the MC

Ergodicity in periodic MCs

- In some sense, still true if MC is periodic
- For irreducible positive recurrent MC (periodic or aperiodic) define

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \qquad \sum_{j=0}^{\infty} \pi_j = 1$$

- A unique solution exists (we say π_j are well defined)
- The fraction of time spent in state i converges to π_i

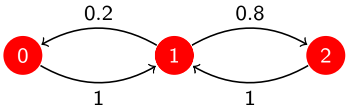
$$\lim_{n\to\infty} T_i^{(n)} = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{I}\left\{X_m = i\right\} \to \pi_i$$

- If MC is periodic, probabilities oscillate, but fraction of time spent in state i converges to π_i

Example: Periodic irreducible Markov chain

- Matrix $\mathbf{P}$ and state diagram of a periodic MC

$$\mathbf{P} := \begin{pmatrix} 0 & 1 & 0 \\ 0.3 & 0 & 0.7 \\ 0 & 1 & 0 \end{pmatrix}$$



- MC has period 2. If initialized with $X_0 = 1$, then

$$P_{11}^{2n+1} = \mathrm{P}\left[X_{2n+1} = 1 \,\middle|\, X_0 = 1\right] = 0,$$
$$P_{11}^{2n} \quad = \mathrm{P}\left[X_{2n} \quad = 1 \,\middle|\, X_0 = 1\right] = 1 \neq 0$$

- Define $\boldsymbol{\pi} := [\pi_1, \pi_2, \pi_3]^T$ as solution of

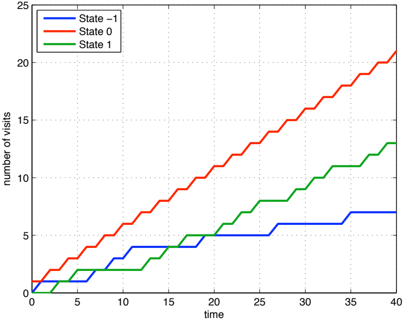
$$\begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 0.3 & 0 \\ 1 & 0 & 1 \\ 0 & 0.7 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix}$$

- Normalized eigenvector for eigenvalue 1 ($\boldsymbol{\pi} = \mathbf{P}^T\boldsymbol{\pi}$, $\boldsymbol{\pi}^T\mathbf{1} = 1$)

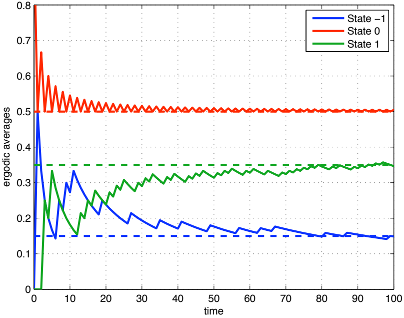
Example: Periodic irreducible MC (continued)

- Solution yields $\pi_1 = 0.15$, $\pi_2 = 0.50$ and $\pi_3 = 0.35$

Visits to states, $nT_i^{(n)}$



Ergodic averages, $T_i^{(n)}$



- Ergodic averages $T_i^{(n)}$ converge to the ergodic limits π_i

Example: Periodic irreducible MC (continued)

- Powers of the transition probability matrix do not converge

$$\mathbf{P}^2 = \begin{pmatrix} 0.3 & 0 & 0.7 \\ 0 & 1 & 0 \\ 0.3 & 0 & 0.7 \end{pmatrix} \qquad \mathbf{P}^3 = \begin{pmatrix} 0 & 1 & 0 \\ 0.3 & 0 & 0.7 \\ 0 & 1 & 0 \end{pmatrix} = \mathbf{P}$$

- In general we have $\mathbf{P}^{2n} = \mathbf{P}^2$ and $\mathbf{P}^{2n+1} = \mathbf{P}$

- At least one other eigenvalue of the transition probability matrix has modulus 1

$$\left|\text{eig}_2\left(\mathbf{P}^T\right)\right| = 1$$

- In this example, eigenvalues of $\mathbf{P}^T$ are 1, -1 and 0

Reducible Markov chains

- If MC is not irreducible it can be decomposed in transient ($\mathcal{T}_k$), ergodic ($\mathcal{E}_k$), periodic ($\mathcal{P}_k$) and null recurrent ($\mathcal{N}_k$) components
  - All of these are class properties
- Limit probabilities for transient states are null $\mathrm{P}\left[X_n = i\right] \to 0$, for all $X_n \in \mathcal{T}_k$
- For arbitrary ergodic component $\mathcal{E}_k$, define conditional limits

$$\pi_i = \lim_{n\to\infty} \mathrm{P}\left[X_n = i \,\middle|\, X_0 \in \mathcal{E}_k\right], \quad \text{for all } i \in \mathcal{E}_k$$

- Result in page 58 is true with this (re)defined π_i

Reducible Markov chains (continued)

- Likewise, for arbitrary periodic component $\mathcal{P}_k$ (re)define π_j as

$$\pi_j = \sum_{i \in \mathcal{P}_k} \pi_i P_{ij}, \quad \sum_{j \in \mathcal{P}_k} \pi_j = 1, \quad \text{for all } j \in \mathcal{P}_k$$

- A conditional version of the result in page 72 is true

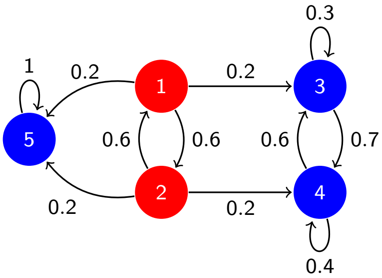
$$\lim_{n\to\infty} T_i^{(n)} := \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{I}\left\{X_m = i \,\middle|\, X_0 \in \mathcal{P}_k\right\} \to \pi_i$$

- For null recurrent components limit probabilities are null $\mathrm{P}\left[X_n = i\right] \to 0$, for all $X_n \in \mathcal{N}_k$

Example: Reducible Markov chain

- Transition matrix and state diagram of a reducible MC

$$\mathbf{P} := \begin{pmatrix} 0 & 0.6 & 0.2 & 0 & 0.2 \\ 0.6 & 0 & 0 & 0.2 & 0.2 \\ 0 & 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$



- States 1 and 2 are transient $\mathcal{T} = \{1, 2\}$
- States 3 and 4 form an ergodic class $\mathcal{E}_1 = \{3, 4\}$
- State 5 is a separate ergodic class $\mathcal{E}_2 = \{5\}$

Example: Reducible MC - matrix powers

- 10-step and 20 step transition probabilities

$$\mathbf{P}^5 = \begin{pmatrix} 0 & 0.08 & 0.24 & 0.22 & 0.46 \\ 0.08 & 0 & 0.19 & 0.27 & 0.46 \\ 0 & 0 & 0.46 & 0.54 & 0 \\ 0 & 0 & 0.46 & 0.54 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{P}^{10} = \begin{pmatrix} 0.00 & 0 & 0.23 & 0.27 & 0.50 \\ 0 & 0.00 & 0.23 & 0.27 & 0.50 \\ 0 & 0 & 0.46 & 0.54 & 0 \\ 0 & 0 & 0.46 & 0.54 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- Transition into transient states is vanishing (columns 1 and 2)
- Transition from 3 and 4 into 3 and 4 only
  - If initialized in ergodic class $\mathcal{E}_1 = \{3, 4\}$ stays in $\mathcal{E}_1$
- Transition from 5 only into 5
- From transient states $\mathcal{T} = \{1, 2\}$ can go into either ergodic component $\mathcal{E}_1 = \{3, 4\}$ or $\mathcal{E}_2 = \{5\}$

Example: Reducible MC - matrix decomposition

- Matrix $\mathbf{P}$ can be separated in blocks

$$\mathbf{P} = \begin{pmatrix} 0 & 0.6 & 0.2 & 0 & 0.2 \\ 0.6 & 0 & 0 & 0.2 & 0.2 \\ 0 & 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{P}_{\mathcal{T}} & \mathbf{P}_{\mathcal{T}\mathcal{E}_1} & \mathbf{P}_{\mathcal{T}\mathcal{E}_2} \\ \mathbf{0} & \mathbf{P}_{\mathcal{E}_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_{\mathcal{E}_2} \end{pmatrix}$$

- Block $\mathbf{P}_{\mathcal{T}}$ describes transition between transient states
- Blocks $\mathbf{P}_{\mathcal{E}_1}$ and $\mathbf{P}_{\mathcal{E}_2}$ describe transitions in ergodic components
- Blocks $\mathbf{P}_{\mathcal{T}\mathcal{E}_1}$ and $\mathbf{P}_{\mathcal{T}\mathcal{E}_2}$ describe transitions from $\mathcal{T}$ to $\mathcal{E}_1$ and $\mathcal{E}_2$

- Powers of n can be written as

$$\mathbf{P}^n = \begin{pmatrix} \mathbf{P}^n_{\mathcal{T}} & \mathbf{Q}_{\mathcal{T}\mathcal{E}_1} & \mathbf{Q}_{\mathcal{T}\mathcal{E}_2} \\ \mathbf{0} & \mathbf{P}^n_{\mathcal{E}_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}^n_{\mathcal{E}_2} \end{pmatrix}$$

- The transient transition block converges to 0, $\lim_{n\to\infty} \mathbf{P}^n_{\mathcal{T}} = \mathbf{0}$

Example: Reducible MC - limiting behavior

- As n grows the MC hits an ergodic state with probability 1
- Henceforth, MC stays within ergodic component

$$\mathrm{P}\left[X_{n+m} \in \mathcal{E}_i \,\middle|\, X_n \in \mathcal{E}_i\right] = 1, \quad \text{for all } m$$

- For large n suffices to study ergodic components
  - $\Rightarrow$ MC behaves like a MC with transition probabilities $\mathbf{P}_{\mathcal{E}_1}$
  - $\Rightarrow$ Or like one with transition probabilities $\mathbf{P}_{\mathcal{E}_2}$

- We can think that all MCs as ergodic
- Ergodic behavior cannot be inferred a priori (before observing)
- Becomes known a posteriori (after observing sufficiently large time)

Culture micro: Something is known a priori if its knowledge is independent of experience (MCs exhibit ergodic behavior). A posteriori knowledge depends on experience (values of the ergodic limits). They are inherently different forms of knowledge (search for Immanuel Kant)

Queues in communication systems

Non concurrent communication queue

- Communication system: Move packets from source to destination
- Between arrival and transmission hold packets in a memory buffer
- Example problem, buffer design: Packets arrive at a rate of 0.45 packets per unit of time and depart at a rate of 0.55. How many packets the buffer needs to hold to have a drop rate smaller than 10^{-6} (one packet dropped for every million packets handled)

- Time slotted in intervals of duration Δt
- During each time slot n

  $\Rightarrow$ A packet arrives with prob. λ, arrival rate is $\lambda/\Delta t$

  $\Rightarrow$ A packet is transmitted with prob. μ, departure rate is $\mu/\Delta t$
- No concurrence: No simultaneous arrival and departure (small Δt)

Queue evolution probabilities (reminder)

- Future queue lengths depend on current length only
- Probability of queue length increasing

$$\mathrm{P}\left[q_{n+1} = i+1 \,\middle|\, q_n = i\right] = \lambda, \qquad \text{for all } i$$

- Queue length might decrease only if $q_n > 0$. Probability is

$$\mathrm{P}\left[q_{n+1} = i-1 \,\middle|\, q_n = i\right] = \mu, \qquad \text{for all } i > 0$$

- Queue length stays the same if it neither increases nor decreases

$$\mathrm{P}\left[q_{n+1} = i \,\middle|\, q_n = i\right] = 1 - \lambda - \mu, \qquad \text{for all } i > 0$$
$$\mathrm{P}\left[q_{n+1} = 0 \,\middle|\, q_n = 0\right] = 1 - \lambda$$

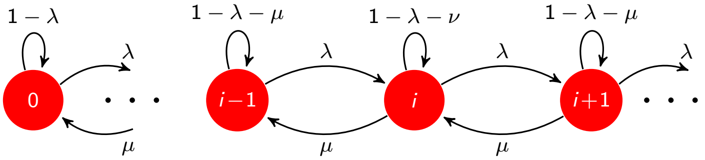
- No departures when $q_n = 0$ explain second equation

Queue as a Markov chain (reminder)

- MC with states $0, 1, 2, \ldots$. Identify states with queue lengths
- Transition probabilities for $i \neq 0$ are

$$P_{i,i-1} = \lambda, \qquad P_{i,i} = 1 - \lambda - \mu, \qquad P_{i,i+1} = \mu$$

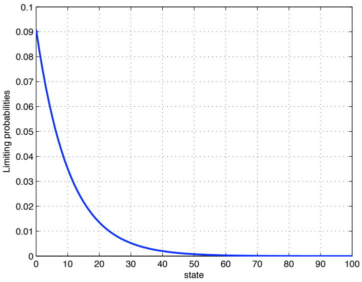
- For $i = 0$ $P_{0,0} = 1 - \lambda$ and $P_{01} = \lambda$



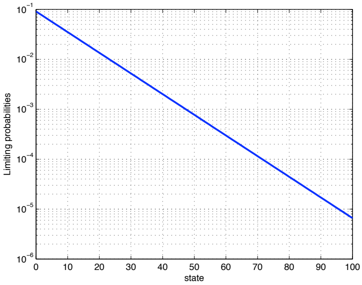
Numerical example: Limit probabilities

- Build matrix $\mathbf{P}$ truncating at maximum queue length $L = 100$
- Arrival rate $\lambda = 0.3$. Departure rate $\mu = 0.33$
- Find eigenvector of $\mathbf{P}^T$ associated with largest eigenvalue (i.e., 1)
- Yields limit probabilities $\boldsymbol{\pi} = \lim_{n\to\infty} \mathbf{p}(n)$

linear scale

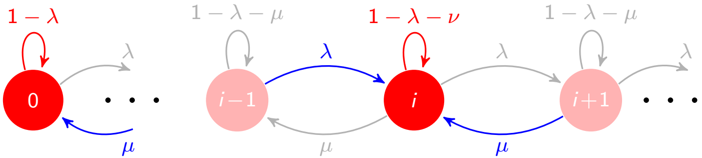


logarithmic scale



- Limit probabilities appear linear in logarithmic scale

  $\Rightarrow$ Seemingly implying an exponential expression $\pi_i \propto \alpha^i$



- Limit distribution equations for state 0 (empty queue)

$$\pi_0 = (1-\lambda)\pi_0 + \mu\pi_1$$

- For the remaining states

$$\pi_i = \lambda\pi_{i-1} + (1-\lambda-\mu)\pi_i + \mu\pi_{i+1}$$

- Propose candidate solution $\pi_i = c\alpha^i$

Verification of candidate solution

- Substitute candidate solution $\pi_i = c\alpha^i$ in equation for π_0

$$c\alpha^0 = (1-\lambda)c\alpha^0 + \mu c\alpha^1 \quad \Rightarrow \quad 1 = (1-\lambda) + \mu\alpha$$

- The above equation is true if we make $\alpha = \lambda/\mu$
- Does $\alpha = \lambda/\mu$ verify the remaining equations ?
- From the equation for generic π_i (divide by $c\alpha^{i-1}$)

$$\begin{aligned} c\alpha^i &= \lambda c\alpha^{i-1} + (1-\lambda-\mu)c\alpha^i + \mu c\alpha^{i+1} \\ \mu\alpha^2 &- (\lambda+\mu)\alpha + \lambda = 0 \end{aligned}$$

- The above quadratic equation is satisfied by $\alpha = \lambda/\mu$
  - And $\alpha = 1$, which is irrelevant

Compute normalization constant

- Determine c so that probabilities sum 1 ($\sum_{i=0}^{\infty} \pi_i = 1$)

$$\sum_{i=0}^{\infty} \pi_i = \sum_{i=0}^{J} c(\lambda/\mu)^i = \frac{c}{1 - \lambda/\mu} = 1$$

- Used geometric sum
- Solving for c and substituting in $\pi_i = c\alpha^i$ yields

$$\pi_i = (1 - \lambda/\mu) \left(\frac{\lambda}{\mu}\right)^i$$

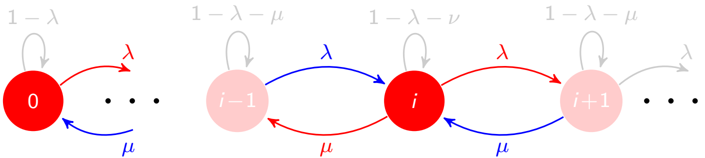
- The ratio μ/λ is the queues' stability margin
- Larger $\mu/\lambda \Rightarrow$ larger probability of having few queued packets

- Rearrange terms in equation for limit probabilities [cf. page 87]

$$\lambda\pi_0 = \mu\pi_1$$
$$(\lambda+\mu)\pi_i = \lambda\pi_{i-1} + \mu\pi_{i+1}$$

- $\lambda\pi_0$ is average rate at which the queue leaves state 0
- Likewise $(\lambda+\mu)\pi_i$ is the rate at which queue leaves state i
- $\mu\pi_0$ is average rate at which the queue enters state 0
- $\lambda\pi_{i-1} + \mu\pi_{i+1}$ is rate at which queue enters state i
- Limit equations prove validity of queue balance equations

Rate at which leaves = Rate at which enters



Concurrent arrival and departures

- Packets may arrive and depart in same time slot (concurrence)
- Queue evolution equations remain the same, [cf. 35]
- But queue probabilities change [cf. 84]

- Probability of queue length increasing (for all i)

$$\mathrm{P}\left[q_{n+1} = i+1 \,\middle|\, q_n = i\right] = \mathrm{P}\left[\mathbb{A}_n = 1\right]\mathrm{P}\left[\mathbb{D}_n = 0\right] = \lambda(1-\mu)$$

- Queue length might decrease only if $q_n > 0$ (for all $i > 0$)

$$\mathrm{P}\left[q_{n+1} = i-1 \,\middle|\, q_n = i\right] = \mathrm{P}\left[\mathbb{D}_n = 1\right]\mathrm{P}\left[\mathbb{D}_n = 0\right] = \mu(1-\lambda)$$

- Queue length stays the same if it neither increases nor decreases

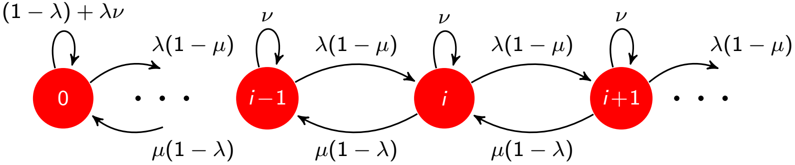
$$\mathrm{P}\left[q_{n+1} = i \,\middle|\, q_n = i\right] = \lambda\mu + (1-\lambda)(1-\mu) = \nu, \qquad \text{for all } i > 0$$
$$\mathrm{P}\left[q_{n+1} = 0 \,\middle|\, q_n = 0\right] = (1-\lambda) + \lambda\mu$$

Limit distribution / queue balance equations

- Write limit distribution equations $\Rightarrow$ queue balance equations
- Rate at which leaves = rate at which enters

$$\lambda(1-\mu)\pi_0 = \mu(1-\lambda)\pi_1$$
$$\big(\lambda(1-\mu)+\mu(1-\lambda)\big)\pi_i = \lambda(1-\mu)\pi_{i-1} + \mu(1-\lambda)\pi_{i+1}$$

- Propose exponential solution $\pi = c\alpha^i$



Solving for limit distribution

- Substitute candidate solution in equation for π_0

$$\lambda(1-\mu)c = \mu(1-\lambda)c\alpha \quad \Rightarrow \quad \alpha = \frac{\lambda(1-\mu)}{\mu(1-\lambda)}$$

- Same substitution in equation for generic π_i

$$\mu(1-\lambda)c\alpha^2 + \big(\lambda(1-\mu) + \mu(1-\lambda)\big)c\alpha + \lambda(1-\mu)c = 0$$

- which as before is solved for $\alpha = \lambda(1-\mu)/\mu(1-\lambda)$
- Find constant c to ensure $\sum_{i=0}^{\infty} c\alpha^i = 1$ (geometric series). Yields

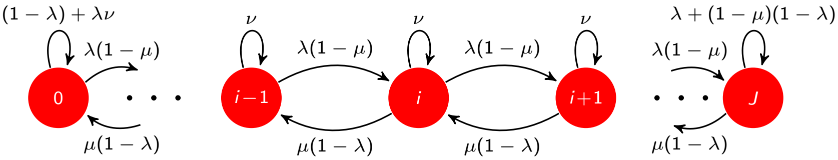
$$\pi_i = (1-\alpha)\alpha^i = \left(1 - \frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)\left(\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)^i$$

Limited queue size

- Packets dropped if there are too many packets in queue
- Too many packets in queue, then delays too large, packets useless when they arrive. Also preserve memory
- Equation for state J requires modification (rate leaves = rate enters)

$$\mu(1-\lambda)\pi_J = \lambda(1-\mu)\pi_{J-1}$$

- $\pi_i = c\alpha^i$ with $\alpha = \lambda(1-\mu)/\mu(1-\lambda)$ also solve this equation (Yes!)



Compute limit distribution

- Limit probabilities are not the same because constant c is different
- To compute c, sum a finite geometric series

$$1 = \sum_{i=0}^{J} c\alpha^i = c\frac{1-\alpha^{J+1}}{1-\alpha} \quad \Rightarrow \quad c = \frac{1-\alpha}{1-\alpha^{J+1}}$$

- Limit distributions for the finite queue are then

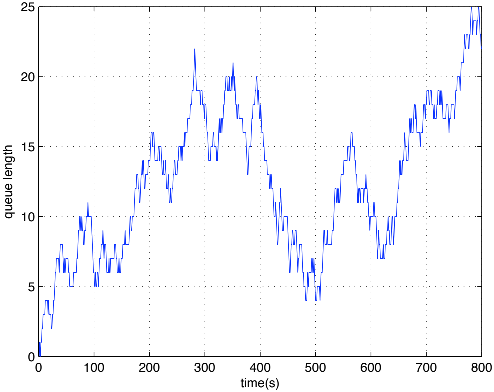
$$\pi_i = \frac{1-\alpha}{1-\alpha^{J+1}}\alpha^i \approx (1-\alpha)\alpha^i$$

- with $\alpha = \lambda(1-\mu)/\mu(1-\lambda)$, and approximation valid for large J
- Approximation for large J yields same result as infinite length queue
  - As it should

Simulations

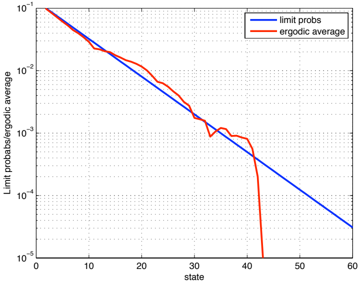
- Arrival rate $\lambda = 0.3$. Departure rate $\mu = 0.33$. Resulting $\alpha \approx 0.87$
- Maximum queue length $J = 100$. Initial state $q_0 = 0$ (queue empty)
  - Not the same as initial probability distribution

Queue lenght as function of time

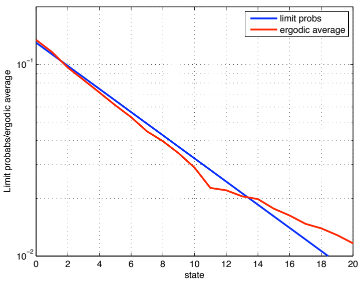


- Average time spent at each queue state is predicted by limit distribution
- For $i = 60$ occupancy probability is $\pi \approx 10^{-5}$.
  - Explains inaccurate prediction for large i

60 states



20 states



Buffer overflow

- If $\lambda = 0.45$ and $\mu = 0.55$ how many packets the buffer needs to hold to have a drop rate smaller than 10^{-6} (one packet dropped for every million packets handled)

- What is the probability of buffer overflow?
- Packet discarded if queue is in state J and a new packet arrives

$$\mathrm{P}\left[\text{overflow}\right] = \lambda \pi_J = \frac{1-\alpha}{1-\alpha^{J+1}} \lambda \alpha^J \approx (1-\alpha)\lambda\alpha^J$$

- With $\lambda = 0.45$ and $\mu = 0.55$, $\alpha \approx 0.82 \Rightarrow J \approx 57$

- A final caveat
  - Still assuming only 1 packet arrives per time slot
  - Lifting this assumption requires introduction of continuous time Markov chains