Introduction to Random Processes

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

August 28, 2023

Stochastic systems

- Stochastic system: Anything random that evolves in time
  - $\Rightarrow$ Time can be discrete $n = 0, 1, 2 \ldots$, or continuous $t \in [0, \infty)$
- More formally, random processes assign a function to a random event
- Compare with "random variable assigns a value to a random event"
- Can interpret a random process as a collection of random variables
  - $\Rightarrow$ Generalizes concept of random vector to functions
  - $\Rightarrow$ Or generalizes the concept of function to random settings

Four thematic blocks

(I) Probability theory review (6 lectures)

- Probability spaces, random variables, independence, expectation
- Conditional probability: time $n + 1$ given time n, future given past ...
- Limits in probability, almost sure limits: behavior as $n \to \infty$ ...
- Common probability distributions (binomial, exponential, Poisson, Gaussian)

- Random processes are complicated entities

  $\Rightarrow$ Restrict attention to particular classes that are somewhat tractable

(II) Markov chains (6 lectures)

(III) Continuous-time Markov chains (7 lectures)

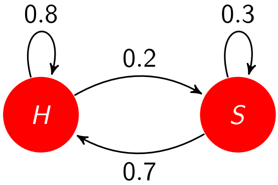
(IV) Stationary random processes (8 lectures)

- Midterm covers up to Markov chains

Markov chains

- Countable set of states $1, 2, \ldots$. At discrete time n, state is X_n
- Memoryless (Markov) property
  - $\Rightarrow$ Probability of next state X_{n+1} depends on current state X_n
  - $\Rightarrow$ But not on past states X_{n-1}, X_{n-2}, ...

- Can be happy ($X_n = 0$) or sad ($X_n = 1$)
- Tomorrow's mood only affected by today's mood
- Whether happy or sad today, likely to be happy tomorrow
- But when sad, a little less likely so

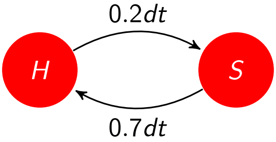


- Of interest: classification of states, ergodicity, limiting distributions
- Applications: Google's PageRank, communication networks, queues, reinforcement learning, ...

Continuous-time Markov chains

- Countable set of states $1, 2, \ldots$. Continuous-time index t, state $X(t)$
  - $\Rightarrow$ Transition between states can happen at any time
  - $\Rightarrow$ Markov: Future independent of the past given the present

- Probability of changing state in an infinitesimal time dt



- Of interest: Poisson processes, exponential distributions, transition probabilities, Kolmogorov equations, limit distributions
- Applications: Chemical reactions, queues, epidemic modeling, traffic engineering, weather forecasting, ...

Stationary random processes

- Continuous time t, continuous state $X(t)$, not necessarily Markov
- Prob. distribution of $X(t)$ constant or becomes constant as t grows
  - $\Rightarrow$ System has a steady state in a random sense
- Of interest: Brownian motion, white noise, Gaussian processes, autocorrelation, power spectral density
- Applications: Black Scholes model for option pricing, radar, face recognition, noise in electric circuits, filtering and equalization, ...

An interesting betting game

- There is a certain game in a certain casino in which ...

  $\Rightarrow$ Your chances of winning are $p > 1/2$
- You place \$1 bets
  (a) With probability p you gain \$1; and
  (b) With probability $1 - p$ you lose your \$1 bet
- The catch is that you either
  (a) Play until you go broke (lose all your money)
  (b) Keep playing forever
- You start with an initial wealth of $\$w_0$
- Q: Shall you play this game?

Modeling

- Let t be a time index (number of bets placed)
- Denote as $X(t)$ the outcome of the bet at time t
  - $\Rightarrow X(t) = 1$ if bet is won (w.p. p)
  - $\Rightarrow X(t) = 0$ if bet is lost (w.p. $1 - p$)
- $X(t)$ is called a Bernoulli random variable with parameter p
- Denote as $W(t)$ the player's wealth at time t. Initialize $W(0) = w_0$
- At times $t > 0$ wealth $W(t)$ depends on past wins and losses
  - $\Rightarrow$ When bet is won $W(t+1) = W(t) + 1$
  - $\Rightarrow$ When bet is lost $W(t+1) = W(t) - 1$
- More compactly can write $W(t+1) = W(t) + (2X(t) - 1)$
  - $\Rightarrow$ Only holds so long as $W(t) > 0$

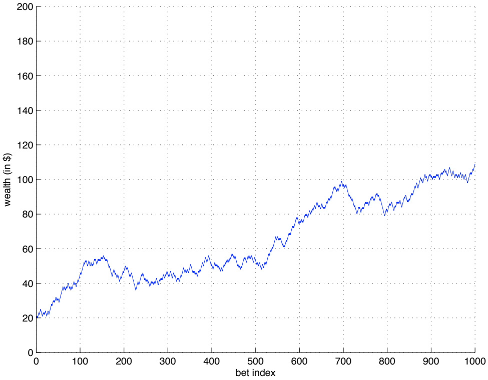
Coding

$t = 0$; $w(t) = w_0$; $max_t = 10^3$; `// Initialize variables`
% repeat while not broke up to time max_t
while $(w(t) > 0)$ & $(t < max_t)$ **do**
    $x(t) =$ random('bino',1,p); % Draw Bernoulli random variable
    if $x(t) == 1$ **then**
        $w(t+1) = w(t) + b$; % If $x = 1$ wealth increases by b
    else
        $w(t+1) = w(t) - b$; % If $x = 0$ wealth decreases by b
    end
    $t = t + 1$;
end

- Initial wealth $w_0 = 20$, bet $b = 1$, win probability $p = 0.55$
- Q: Shall we play?

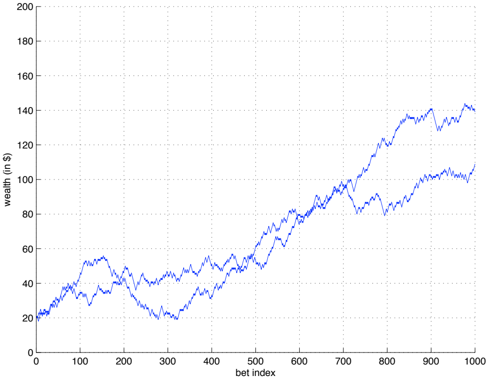
One lucky player

- She didn't go broke. After $t = 1000$ bets, her wealth is $W(t) = 109$
  $\Rightarrow$ Less likely to go broke now because wealth increased



Two lucky players

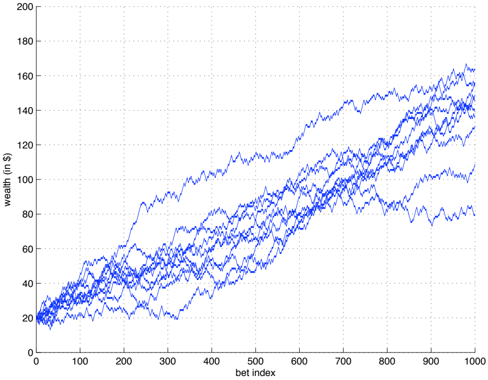
- After $t = 1000$ bets, wealths are $W_1(t) = 109$ and $W_2(t) = 139$
  - $\Rightarrow$ Increasing wealth seems to be a pattern



Ten lucky players

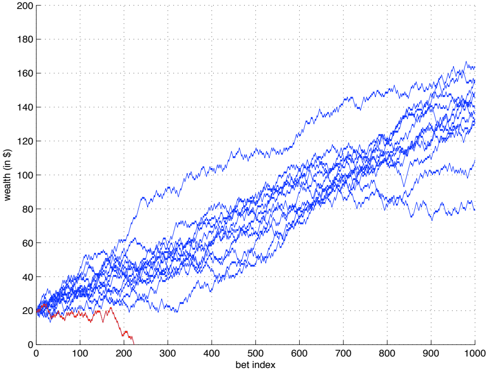
- Wealths $W_j(t)$ after $t = 1000$ bets between 78 and 139

  $\Rightarrow$ Increasing wealth is definitely a pattern



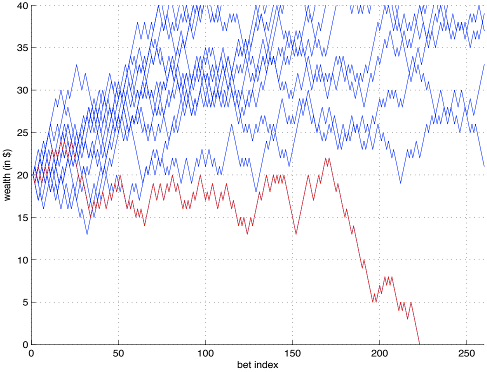
One unlucky player

- But this does not mean that all players will turn out as winners
  - $\Rightarrow$ The twelfth player $j = 12$ goes broke



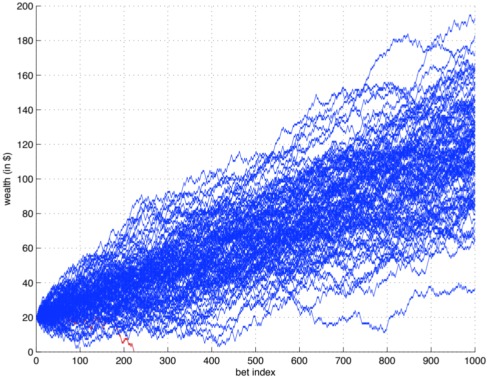
One unlucky player

- But this does not mean that all players will turn out as winners
  - $\Rightarrow$ The twelfth player $j = 12$ goes broke



One hundred players

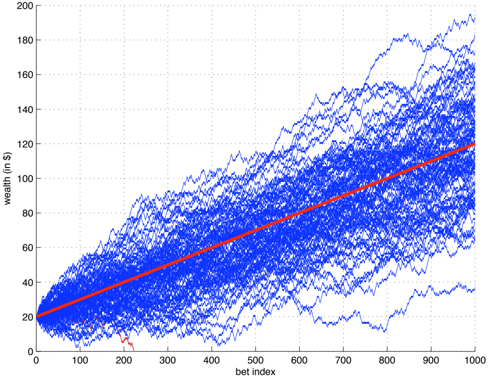
- All players (except for $j = 12$) end up with substantially more money



Average tendency

- It is not difficult to find a line estimating the average of $W(t)$

$$\Rightarrow \bar{w}(t) \approx w_0 + (2p-1)t \approx w_0 + 0.1t \text{ (recall } p = 0.55\text{)}$$



Where does the average tendency come from?

- Assuming we do not go broke, we can write

$$W(t+1) = W(t) + \Big(2X(t) - 1\Big), \quad t = 0, 1, 2, \ldots$$

  - The assumption is incorrect as we saw, but suffices for simplicity

- Taking expectations on both sides and using linearity of expectation

$$\mathbb{E}\left[W(t+1)\right] = \mathbb{E}\left[W(t)\right] + \Big(2\mathbb{E}\left[X(t)\right] - 1\Big)$$

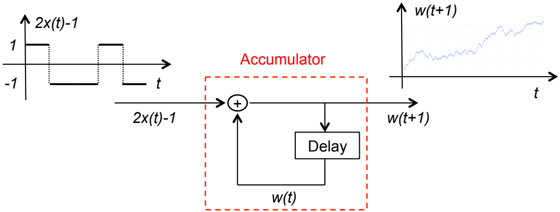
- The expected value of Bernoulli $X(t)$ is

$$\mathbb{E}\left[X(t)\right] \; = \; 1 \times \mathrm{P}\left(X(t) = 1\right) + 0 \times \mathrm{P}\left(X(t) = 0\right) \; = \; p$$

- Which yields $\Rightarrow \mathbb{E}\left[W(t+1)\right] = \mathbb{E}\left[W(t)\right] + (2p - 1)$
- Applying recursively $\Rightarrow \mathbb{E}\left[W(t+1)\right] = w_0 + (2p - 1)(t + 1)$

Gambling as LTI system with stochastic input

- Recall the evolution of wealth $W(t+1) = W(t) + \Big(2X(t) - 1\Big)$



- View $W(t+1)$ as output of LTI system with random input $2X(t) - 1$
- Recognize accumulator $\Rightarrow W(t+1) = w_0 + \sum_{\tau=0}^{t} \Big(2X(\tau) - 1\Big)$
  - Useful, a lot we can say about sums of random variables
- Filtering random processes in signal processing, communications, ...

Numerical analysis of simulation outcomes

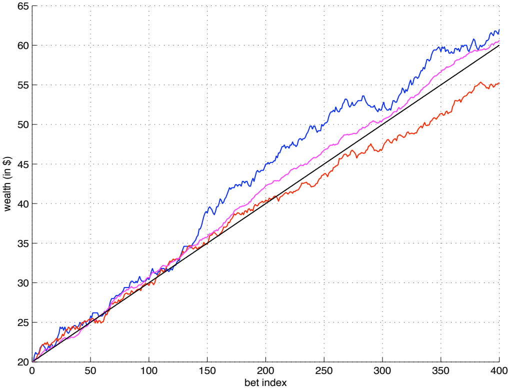
- For a more accurate approximation analyze simulation outcomes
- Consider J experiments. Each yields a wealth history $W_j(t)$
- Can estimate the average outcome via the sample average $\bar{W}_J(t)$

$$\bar{W}_J(t) := \frac{1}{J}\sum_{j=1}^{J} W_j(t)$$

- Do not confuse $\bar{W}_J(t)$ with $\mathbb{E}[W(t)]$
  - $\bar{W}_J(t)$ is computed from experiments, it is a random quantity in itself
  - $\mathbb{E}[W(t)]$ is a property of the random variable $W(t)$
  - We will see later that for large J, $\bar{W}_J(t) \to \mathbb{E}[W(t)]$

Analysis of simulation outcomes: mean

- Expected value $\mathbb{E}[W(t)]$ in black
- Sample average for $J = 10$ (blue), $J = 20$ (red), and $J = 100$ (magenta)



Analysis of simulation outcomes: distribution

- There is more information in the simulation's output
- Estimate the distribution function of $W(t)$ $\Rightarrow$ Histogram
- Consider a grid of points $w^{(0)}, \dots, w^{(M)}$
- Indicator function of the event $w^{(m)} \leq W_j(t) < w^{(m+1)}$

$$\mathbb{I}\left\{w^{(m)} \leq W_j(t) < w^{(m+1)}\right\} = \begin{cases} 1, & \text{if } w^{(m)} \leq W_j(t) < w^{(m+1)} \\ 0, & \text{otherwise} \end{cases}$$

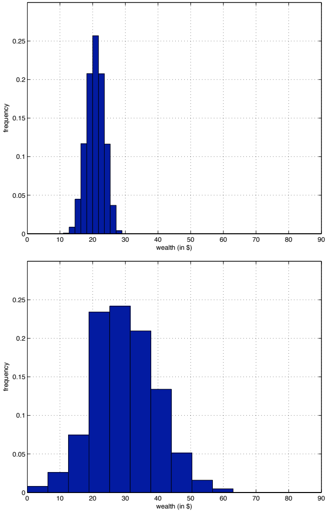
- Histogram is then defined as

$$H\left[t; w^{(m)}, w^{(m+1)}\right] = \frac{1}{J} \sum_{j=1}^{J} \mathbb{I}\left\{w^{(m)} \leq W_j(t) < w^{(m+1)}\right\}$$

- Fraction of experiments with wealth $W_j(t)$ between $w^{(m)}$ and $w^{(m+1)}$

Histogram

- Distribution broadens and shifts to the right ($t = 10, 50, 100, 200$)



What is this class about?

- Analysis and simulation of stochastic systems

  $\Rightarrow$ A system that evolves in time with some randomness

- They are usually quite complex $\Rightarrow$ Simulations
- We will learn how to model stochastic systems, e.g.,
  - $X(t)$ Bernoulli with parameter p
  - $W(t+1) = W(t) + 1$, when $X(t) = 1$
  - $W(t+1) = W(t) - 1$, when $X(t) = 0$
- ... how to analyze their properties, e.g., $\mathbb{E}\left[W(t)\right] = w_0 + (2p-1)t$
- ... and how to interpret simulations and experiments, e.g.,
  - Average tendency through sample average
  - Estimate probability distributions via histograms

Markov chains in discrete time

- Consider discrete-time index $n = 0, 1, 2, \ldots$
- Time-dependent random state X_n takes values on a countable set
  - In general, states are $i = 0, \pm 1, \pm 2, \ldots$, i.e., here the state space is $\mathbb{Z}$
  - If $X_n = i$ we say "the process is in state i at time n"
- Random process is $X_{\mathbb{N}}$, its history up to n is $\mathbf{X}_n = [X_n, X_{n-1}, \ldots, X_0]^T$
- **Def:** process $X_{\mathbb{N}}$ is a Markov chain (MC) if for all $n \geq 1$, i, j, $\mathbf{x} \in \mathbb{Z}^n$

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

- Future depends only on current state X_n (memoryless, Markov property)

  $\Rightarrow$ Future conditionally independent of the past, given the present

Observations

- Given X_n, history $\mathbf{X}_{n-1}$ irrelevant for future evolution of the process
- From the Markov property, can show that for arbitrary $m > 0$

$$\mathrm{P}\left(X_{n+m} = j \,\middle|\, X_n = i, \mathbf{X}_{n-1} = \mathbf{x}\right) = \mathrm{P}\left(X_{n+m} = j \,\middle|\, X_n = i\right)$$

- Transition probabilities P_{ij} are constant (MC is time invariant)

$$\mathrm{P}\left(X_{n+1} = j \,\middle|\, X_n = i\right) = \mathrm{P}\left(X_1 = j \,\middle|\, X_0 = i\right) = P_{ij}$$

- Since P_{ij}'s are probabilities they are non-negative and sum up to 1

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

$\Rightarrow$ Conditional probabilities satisfy the axioms

Matrix representation

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

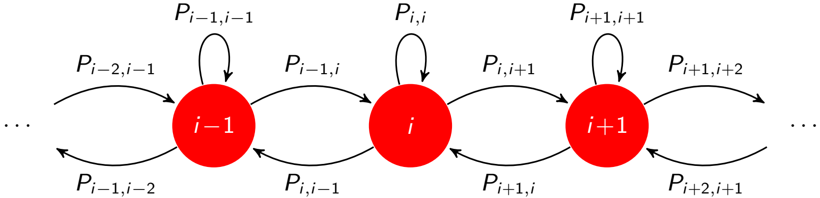
$$\mathbf{P} = \begin{pmatrix} P_{00} & P_{01} & P_{02} & \dots & P_{0j} & \dots \\ P_{10} & P_{11} & P_{12} & \dots & P_{1j} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{i0} & P_{i1} & P_{i2} & \dots & P_{ij} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$\Rightarrow$ Not really a matrix if number of states is infinite

- Row-wise sums should be equal to one, i.e., $\sum_{j=0}^{\infty} P_{ij} = 1$ for all i

Graph representation

- A graph representation or state transition diagram is also used



- Useful when number of states is infinite, skip arrows if $P_{ij} = 0$
- Again, sum of per-state outgoing arrow weights should be one

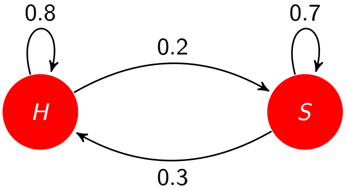
Example: Happy - Sad

- I can be happy ($X_n = 0$) or sad ($X_n = 1$)

  $\Rightarrow$ My mood tomorrow is only affected by my mood today

- 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
- But when sad, a little less likely so ($P_{00} > P_{11}$)

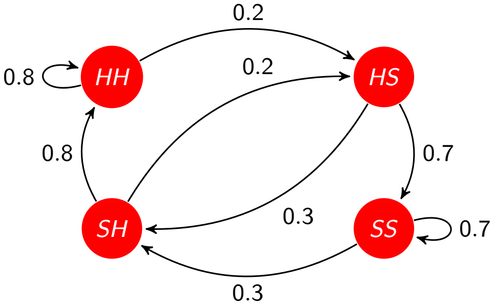
Example: Happy - Sad with memory

- Happiness tomorrow affected by today's and yesterday's mood

  $\Rightarrow$ Not a Markov chain with the previous state space

- 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

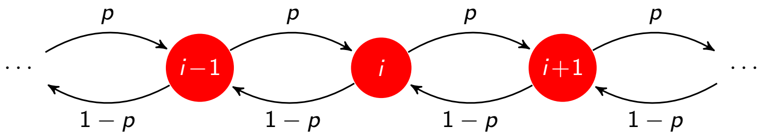
$$\mathbf{P} = \begin{pmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \\ 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \end{pmatrix}$$



- **Key:** can capture longer time memory via state augmentation

Random (drunkard's) walk

- Step to the right w.p. p, to the left w.p. $1-p$
  $\Rightarrow$ Not that drunk to stay on the same place



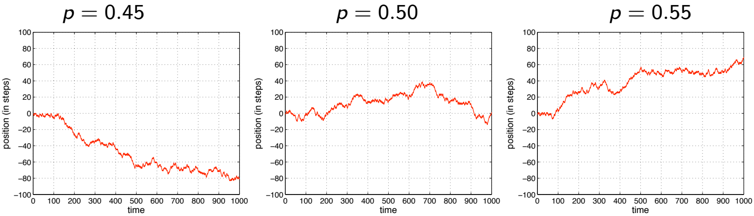
- States are $0, \pm 1, \pm 2, \dots$ (state space is $\mathbb{Z}$), infinite number of states
- Transition probabilities are

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

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

Random (drunkard's) walk (continued)

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



$\Rightarrow$ With $p > 1/2$ diverges to the right ($\nearrow$ almost surely)

$\Rightarrow$ With $p < 1/2$ diverges to the left ($\searrow$ almost surely)

$\Rightarrow$ With $p = 1/2$ always come back to visit origin (almost surely)

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

Two dimensional random walk

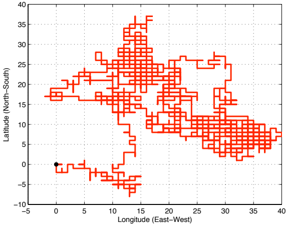
- Take a step in random direction E, W, S or N
  - $\Rightarrow$ E, W, S, N chosen with equal probability
- States are pairs of coordinates (X_n, Y_n)
  - $X_n = 0, \pm 1, \pm 2, \ldots$ and $Y_n = 0, \pm 1, \pm 2, \ldots$
- Transiton probs. $\neq 0$ only for adjacent points

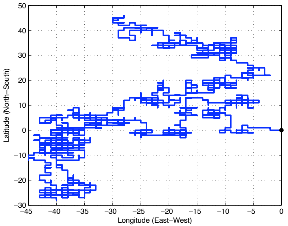
East: $\mathrm{P}\left(X_{n+1} = i+1, Y_{n+1} = j \,\middle|\, X_n = i, Y_n = j\right) = \dfrac{1}{4}$

West: $\mathrm{P}\left(X_{n+1} = i-1, Y_{n+1} = j \,\middle|\, X_n = i, Y_n = j\right) = \dfrac{1}{4}$

North: $\mathrm{P}\left(X_{n+1} = i, Y_{n+1} = j+1 \,\middle|\, X_n = i, Y_n = j\right) = \dfrac{1}{4}$

South: $\mathrm{P}\left(X_{n+1} = i, Y_{n+1} = j-1 \,\middle|\, X_n = i, Y_n = j\right) = \dfrac{1}{4}$





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

  $\Rightarrow$ Will almost surely return home
- In more than two dimensions, probability of returning to origin is < 1

  $\Rightarrow$ In three dimensions probability of returning to origin is 0.34

  $\Rightarrow$ Then 0.19, 0.14, 0.10, 0.08, ...

Another representation of a random walk

- Consider an i.i.d. sequence of RVs $Y_{\mathbb{N}} = Y_1, Y_2, \ldots, Y_n, \ldots$
- Y_n takes the value ± 1, $\mathrm{P}\left(Y_n = 1\right) = p$, $\mathrm{P}\left(Y_n = -1\right) = 1 - p$

- Define $X_0 = 0$ and the cumulative sum

$$X_n = \sum_{k=1}^{n} Y_k$$

$\Rightarrow$ The process $X_{\mathbb{N}}$ is a random walk (same we saw earlier)

$\Rightarrow$ $Y_{\mathbb{N}}$ are i.i.d. steps (increments) because $X_n = X_{n-1} + Y_n$

- Q: Can we formally establish the random walk is a Markov chain?
- A: Since $X_n = X_{n-1} + Y_n$, $n \geq 1$, and Y_n independent of $\mathbf{X}_{n-1}$

$$\begin{aligned} \mathrm{P}\left(X_n = j \,\middle|\, X_{n-1} = i, \mathbf{X}_{n-2} = \mathbf{x}\right) &= \mathrm{P}\left(X_{n-1} + Y_n = j \,\middle|\, X_{n-1} = i, \mathbf{X}_{n-2} = \mathbf{x}\right) \\ &= \mathrm{P}\left(Y_1 = j - i\right) := P_{ij} \end{aligned}$$

General result to identify Markov chains

Theorem

Suppose $Y_{\mathbb{N}} = Y_1, Y_2, \ldots, Y_n, \ldots$ are i.i.d. and independent of X_0. Consider the random process $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$ of the form

$$X_n = f(X_{n-1}, Y_n), \quad n \geq 1$$

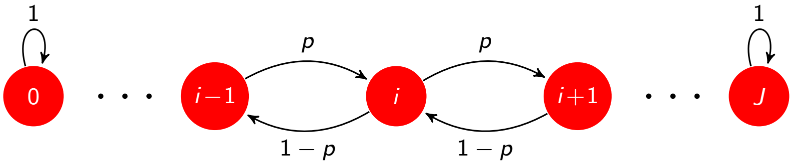
Then $X_{\mathbb{N}}$ is a Markov chain with transition probabilities

$$P_{ij} = \mathrm{P}\left(f(i, Y_1) = j\right)$$

- Useful result to identify Markov chains
  - $\Rightarrow$ Often simpler than checking the Markov property
- Proof similar to the random walk special case, i.e., $f(x, y) = x + y$

Random walk with boundaries (gambling)

- As a random walk, but stop moving when $X_n = 0$ or $X_n = J$
  - 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

- Transition probabilities are

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

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

- States 0 and J are called absorbing. Once there stay there forever

  $\Rightarrow$ The rest are transient states. Visits stop almost surely

Chapman-Kolmogorov equations

Definition and examples

Chapman-Kolmogorov equations

Gambler's ruin problem

Queues in communication networks: Transition probabilities

Classes of states

Multiple-step transition probabilities

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

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

$\Rightarrow$ Caution: P_{ij}^2 is just notation, $P_{ij}^2 \neq P_{ij} \times P_{ij}$

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

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

- Relation between n-, m-, and $(m+n)$-step transition probabilities

  $\Rightarrow$ 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 probabilities between two time slots

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

- Using the law of total probability

$$P_{ij}^2 = \sum_{k=0}^{\infty} \mathrm{P}\left(X_{n+2} = j \,\middle|\, X_{n+1} = k, X_n = i\right) \mathrm{P}\left(X_{n+1} = k \,\middle|\, X_n = i\right)$$

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

$$P_{ij}^2 = \sum_{k=0}^{\infty} \mathrm{P}\left(X_{n+2} = j \,\middle|\, X_{n+1} = k\right) \mathrm{P}\left(X_{n+1} = k \,\middle|\, X_n = i\right)$$

- Which by definition of transition probabilities yields

$$P_{ij}^2 = \sum_{k=0}^{\infty} P_{kj} P_{ik}$$

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

- Same argument works (condition on X_0 w.l.o.g., time invariance)

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

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

$$P_{ij}^{m+n} = \sum_{k=0}^{\infty} \mathrm{P}\left(X_{m+n} = j \,\middle|\, X_m = k, X_0 = i\right) \mathrm{P}\left(X_m = k \,\middle|\, X_0 = i\right)$$

$$P_{ij}^{m+n} = \sum_{k=0}^{\infty} \mathrm{P}\left(X_{m+n} = j \,\middle|\, X_m = k\right) \mathrm{P}\left(X_m = k \,\middle|\, X_0 = i\right)$$

$$P_{ij}^{m+n} = \sum_{k=0}^{\infty} P_{kj}^{n} P_{ik}^{m} \quad \text{for all } i, j \text{ and } n, m \geq 0$$

$\Rightarrow$ These are the Chapman-Kolmogorov equations

Interpretation

- Chapman-Kolmogorov equations are intuitive. Recall

$$P_{ij}^{m+n} = \sum_{k=0}^{\infty} P_{ik}^{m} P_{kj}^{n}$$

- Between times 0 and $m + n$, time m occurred
- At time m, the Markov 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, just sum over all k

Chapman-Kolmogorov equations in matrix form

- Define the following three matrices:
  - $\Rightarrow$ $\mathbf{P}^{(m)}$ with elements P_{ij}^{m}
  - $\Rightarrow$ $\mathbf{P}^{(n)}$ with elements P_{ij}^{n}
  - $\Rightarrow$ $\mathbf{P}^{(m+n)}$ with elements P_{ij}^{m+n}
- Matrix product $\mathbf{P}^{(m)}\mathbf{P}^{(n)}$ has (i,j)-th element $\sum_{k=0}^{\infty} P_{ik}^{m} P_{kj}^{n}$
- Chapman Kolmogorov in matrix form

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

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

Computing n-step transition probabilities

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

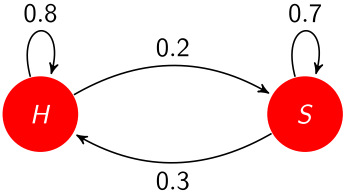
$$\mathbf{P}^{(n)} = \mathbf{P}^n$$

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

Example: Happy-Sad

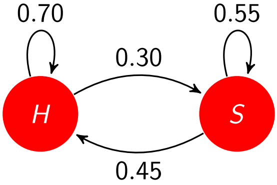
- Mood transitions in one day

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



- 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

  $\Rightarrow$ Note that this is a regular limit

- After a month transition from H to H and from S to H w.p. 0.6

  $\Rightarrow$ State becomes independent of initial condition (H w.p. 0.6)

- Rationale: 1-step memory $\Rightarrow$ Initial condition eventually forgotten
  - More about this soon

Unconditional probabilities

- All probabilities so far are conditional, i.e., $P_{ij}^n = \text{P}\left(X_n = j \,\middle|\, X_0 = i\right)$
  $\Rightarrow$ May want unconditional probabilities $p_j(n) = \text{P}\left(X_n = j\right)$
- Requires specification of initial conditions $p_i(0) = \text{P}\left(X_0 = i\right)$
- Using law of total probability and definitions of P_{ij}^n and $p_j(n)$

$$
\begin{aligned}
p_j(n) = \text{P}\left(X_n = j\right) &= \sum_{i=0}^{\infty} \text{P}\left(X_n = j \,\middle|\, X_0 = i\right) \text{P}\left(X_0 = i\right) \\
&= \sum_{i=0}^{\infty} P_{ij}^n p_i(0)
\end{aligned}
$$

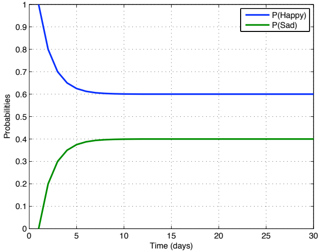
- In matrix form (define vector $\mathbf{p}(n) = [p_1(n), p_2(n), \ldots]^T$)

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

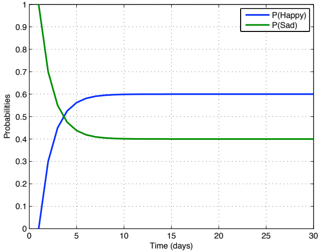
Example: Happy-Sad

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

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



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



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

Gambler's ruin problem

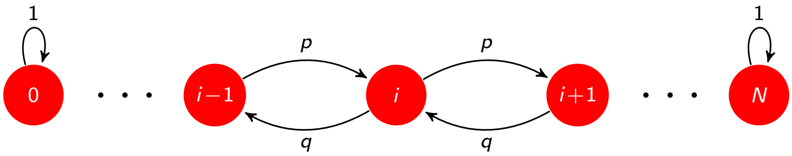
Gambler's ruin problem

- You place $\$1$ bets
  (i) With probability p you gain $\$1$, and
  (ii) With probability $q = 1 - p$ you loose your $\$1$ bet
- Start with an initial wealth of $\$i$
- 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)
  - $\alpha = 1$ game is fair
- You keep playing until
  (a) You go broke (loose all your money)
  (b) You reach a wealth of $\$N$ (same as first lecture, HW1 for $N \to \infty$)
- Prob. S_i of reaching $\$N$ before going broke for initial wealth $\$i$?
  - S stands for success, or successful betting run (SBR)

Gambler's Markov chain

- Model wealth 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
  - $\Rightarrow$ Sates 0 and N are absorbing. Eventually end up in one of them
  - $\Rightarrow$ Remaining states are 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

- Total probability to relate S_i with S_{i+1}, S_{i-1} from adjacent states
  $\Rightarrow$ Condition on first bet X_1, Markov chain homogeneous

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

- Recall $p + q = 1$ and 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

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

Probability of successful betting run

- Geometric series can be summed in closed form, assuming $\alpha \neq 1$

$$S_i = \left(\sum_{k=0}^{i-1} \alpha^k\right) S_1 = \frac{1-\alpha^i}{1-\alpha} S_1$$

- When in state N, $S_N = 1$ and so

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

- Substitute S_1 above into expression for probability of SBR S_i

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

- For $\alpha = 1$ $\Rightarrow S_i = iS_1$, $1 = S_N = NS_1$, $\Rightarrow S_i = \frac{i}{N}$

Analysis for large N

- Recall

$$S_i = \begin{cases} (1-\alpha^i)/(1-\alpha^N), & \alpha \neq 1, \\ i/N, & \alpha = 1 \end{cases}$$

- Consider exit bound N arbitrarily large

(i) For $\alpha > 1$, $S_i \approx (\alpha^i - 1)/\alpha^N \to 0$

(ii) Likewise for $\alpha = 1$, $S_i = i/N \to 0$

- If win probability p does not exceed loose probability q

  $\Rightarrow$ Will almost surely loose all money

(iii) For $\alpha < 1$, $S_i \to 1 - \alpha^i$

- If win probability p exceeds loose probability q

  $\Rightarrow$ For sufficiently high initial wealth i, will most likely win

- This explains what we saw on first lecture and HW1

Queues in communication systems

Definition and examples

Chapman-Kolmogorov equations

Gambler's ruin problem

Queues in communication networks: Transition probabilities

Classes of states

Queues in communication systems

- General communication systems goal

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

- Between arrival and departure we hold packets in a memory buffer

  $\Rightarrow$ Want to design buffers appropriately

Non-concurrent queue

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

Queue evolution equations

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

- If the queue is empty $Q_n = 0$ then there are no departures
  $\Rightarrow$ 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

$$Q_{n+1} = Q_n + \mathbb{A}_n - \mathbb{D}_n, \quad \text{if } Q_n > 0$$

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

$$\mathrm{P}(\mathbb{A}_n = 1) = \lambda, \qquad \mathrm{P}(\mathbb{D}_n = 1) = \mu$$

Queue evolution probabilities

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

$$P\left(Q_{n+1} = i+1 \,\middle|\, Q_n = i\right) = P\left(\mathbb{A}_n = 1\right) = \lambda, \qquad \text{for all } i$$

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

$$P\left(Q_{n+1} = i-1 \,\middle|\, Q_n = i\right) = P\left(\mathbb{D}_n = 1\right) = \mu, \qquad \text{for all } i > 0$$

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

$$P\left(Q_{n+1} = i \,\middle|\, Q_n = i\right) = 1 - \lambda - \mu, \qquad \text{for all } i > 0$$
$$P\left(Q_{n+1} = 0 \,\middle|\, Q_n = 0\right) = 1 - \lambda$$

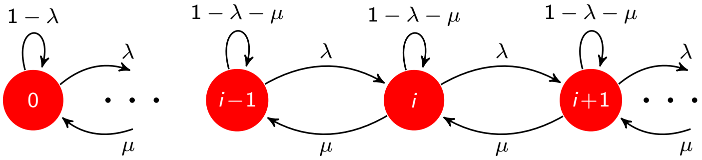
$\Rightarrow$ No departures when $Q_n = 0$ explain second equation

Queue as a Markov chain

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

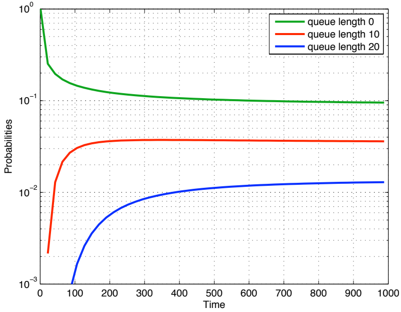
$$P_{i,i-1} = \mu, \qquad P_{i,i} = 1 - \lambda - \mu, \qquad P_{i,i+1} = \lambda$$

- For $i = 0$: $P_{00} = 1 - \lambda$ and $P_{01} = \lambda$



Numerical example: Probability propagation

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



- Propagate probabilities $(\mathbf{P}^n)^T\mathbf{p}(0)$
- 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 decreases with i

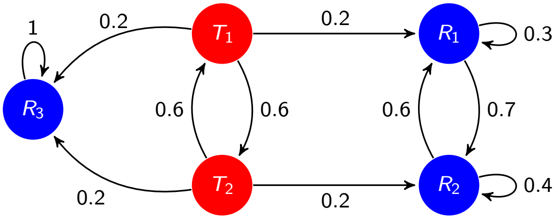
Classes of states

Transient and recurrent states

- States of a MC can be recurrent or transient

- Transient states might be visited early on but visits eventually 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 $f_i = 1$

  $\Rightarrow$ Process reenters i again and again (a.s.). Infinitely often

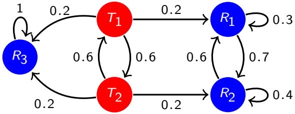
- State i is transient if $f_i < 1$

  $\Rightarrow$ Positive probability $1 - f_i > 0$ of never coming back to i

Recurrent states example

- State R_3 is recurrent because it is absorbing $\mathrm{P}\left(X_1 = R_3 \,\middle|\, X_0 = R_3\right) = 1$

- State R_1 is recurrent because



$$\mathrm{P}\left(X_1 = R_1 \,\middle|\, X_0 = R_1\right) = 0.3$$
$$\mathrm{P}\left(X_2 = R_1, X_1 \neq R_1 \,\middle|\, X_0 = R_1\right) = (0.7)(0.6)$$
$$\mathrm{P}\left(X_3 = R_1, X_2 \neq R_1, X_1 \neq R_1 \,\middle|\, X_0 = R_1\right) = (0.7)(0.4)(0.6)$$
$$\vdots$$
$$\mathrm{P}\left(X_n = R_1, X_{n-1} \neq R_1, \ldots, X_1 \neq R_1 \,\middle|\, X_0 = R_1\right) = (0.7)(0.4)^{n-2}(0.6)$$

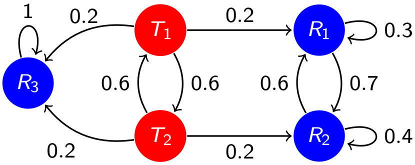
- Sum up: $$f_i = \sum_{n=1}^{\infty} \mathrm{P}\left(X_n = R_1, X_{n-1} \neq R_1, \ldots, X_1 \neq R_1 \,\middle|\, X_0 = R_1\right)$$
$$= 0.3 + 0.7\left(\sum_{n=2}^{\infty} 0.4^{n-2}\right) 0.6 = 0.3 + 0.7\left(\frac{1}{1-0.4}\right) 0.6 = 1$$

Transient state example

- States T_1 and T_2 are transient
- Probability of returning to T_1 is $f_{T_1} = (0.6)^2 = 0.36$

  $\Rightarrow$ Might come back to T_1 only if it goes to T_2 (w.p. 0.6)

  $\Rightarrow$ Will come back only if it moves back from T_2 to T_1 (w.p. 0.6)



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

Expected number of visits to states

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

$$N_i := \sum_{n=1}^{\infty} \mathbb{I}\left\{X_n = i \,\middle|\, X_0 = i\right\}$$

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

$$\mathrm{P}\left(N_i = n\right) = f_i^n(1 - f_i)$$

$\Rightarrow$ Number of visits $N_i + 1$ is geometric with parameter $1 - f_i$

- Expected number of visits is

$$\mathbb{E}\left[N_i\right] + 1 = \frac{1}{1 - f_i} \;\Rightarrow\; \mathbb{E}\left[N_i\right] = \frac{f_i}{1 - f_i}$$

$\Rightarrow$ For recurrent states $N_i = \infty$ a.s. and $\mathbb{E}\left[N_i\right] = \infty$ $(f_i = 1)$

Alternative transience/recurrence characterization

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

$$\mathbb{E}[N_i] = \sum_{n=1}^{\infty} \mathbb{E}\Big[\mathbb{I}\{X_n = i \,\big|\, X_0 = i\}\Big] = \sum_{n=1}^{\infty} P_{ii}^n$$

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

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

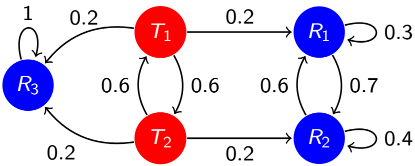
  $\Rightarrow$ If number of states is finite some states have to be recurrent

Accessibility

- **Def:** State j is accessible from state i if $P_{ij}^n > 0$ for some $n \geq 0$

  $\Rightarrow$ It is possible to enter j if MC initialized at $X_0 = i$

- Since $P_{ii}^0 = \mathrm{P}\left(X_0 = i \,\middle|\, X_0 = i\right) = 1$, 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 R_3 accessible from itself

Communication

- **Def:** States i and j are said to communicate ($i \leftrightarrow j$) if
  - $\Rightarrow$ j is accessible from i, i.e., $P_{ij}^n > 0$ for some n; and
  - $\Rightarrow$ i is accessible from j, i.e., $P_{ji}^m > 0$ for some m
- Communication is an equivalence relation
- Reflexivity: $i \leftrightarrow i$
  - Holds 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)
  - $\Rightarrow$ What are these classes?

Recurrence and communication

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 $\sum_{n=1}^{\infty} P_{ii}^{n} = \infty$

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

$\Rightarrow$ Which implies j is recurrent

□

Recurrence and transience are class properties

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

□

- Recurrence is shared by elements of a communication class

  $\Rightarrow$ We say that recurrence is a class property

- Likewise, transience is also a class property

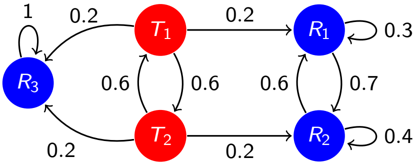
- MC states are separated in classes of transient and recurrent states

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 MC starts in transient class $\mathcal{T}_k$, then it might
  (a) Stay on $\mathcal{T}_k$ (only if $|\mathcal{T}_k| = \infty$)
  (b) End up in another transient class $\mathcal{T}_r$ (only if $|\mathcal{T}_r| = \infty$)
  (c) End up in a recurrent class $\mathcal{R}_l$
- For large time index n, MC restricted to one class

  $\Rightarrow$ Can be separated into irreducible components

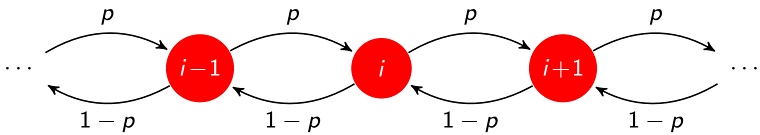
Communication classes example



- 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 state
- For large n suffices to study the irreducible components $\mathcal{R}_1$ and $\mathcal{R}_2$

Example: Random walk

- Step right with probability p, left with probability $q = 1 - p$



- All states communicate $\Rightarrow$ States either all transient or all recurrent
- To see which, consider initially $X_0 = 0$ and note for any $n \geq 1$

$$P_{00}^{2n} = \binom{2n}{n} p^n q^n = \frac{(2n)!}{n!n!} p^n q^n$$

$\Rightarrow$ Back to 0 in $2n$ steps $\Leftrightarrow$ n steps right and n steps left

Example: Random walk (continued)

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

  $\Rightarrow$ Approximate probability P_{00}^{2n} of returning home as

$$P_{00}^{2n} = \frac{(2n)!}{n!n!} p^n q^n \approx \frac{(4pq)^n}{\sqrt{n\pi}}$$

- Symmetric random walk ($p = q = 1/2$)

$$\sum_{n=1}^{\infty} P_{00}^{2n} = \sum_{n=1}^{\infty} \frac{1}{\sqrt{n\pi}} = \infty$$

  $\Rightarrow$ State 0 (hence all states) are recurrent

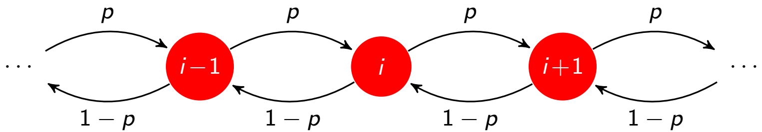
- Biased random walk ($p > 1/2$ or $p < 1/2$), then $pq < 1/4$ and

$$\sum_{n=1}^{\infty} P_{00}^{2n} = \sum_{n=1}^{\infty} \frac{(4pq)^n}{\sqrt{n\pi}} < \infty$$

  $\Rightarrow$ State 0 (hence all states) are transient

Example: Right-biased random walk

- Alternative proof of transience of right-biased random walk ($p > 1/2$)



- Write current position of random walker as $X_n = \sum_{k=1}^{n} Y_k$

  $\Rightarrow$ Y_k are the i.i.d. steps: $\mathbb{E}[Y_k] = 2p - 1$, $\text{var}[Y_k] = 4p(1-p)$

- From Central Limit Theorem ($\Phi(x)$ is cdf of standard Normal)

$$\mathrm{P}\left(\frac{\sum_{k=1}^{n} Y_k - n(2p-1)}{\sqrt{n4p(1-p)}} \leq a\right) \to \Phi(a)$$

Example: Right-biased random walk (continued)

- Choose $a = \frac{\sqrt{n}(1-2p)}{\sqrt{4p(1-p)}} < 0$, use Chernoff bound $\Phi(a) \leq \exp(-a^2/2)$

$$\mathrm{P}\left(X_n \leq 0\right) = \mathrm{P}\left(\sum_{k=1}^{n} Y_k \leq 0\right) \to \Phi\left(\frac{\sqrt{n}(1-2p)}{\sqrt{4p(1-p)}}\right) < e^{-\frac{n(1-2p)^2}{8p(1-p)}} \to 0$$

- Since $P_{00}^n \leq \mathrm{P}\left(X_n \leq 0\right)$, sum over n

$$\sum_{n=1}^{\infty} P_{00}^n \leq \sum_{n=1}^{\infty} \mathrm{P}\left(X_n \leq 0\right) < \sum_{n=1}^{\infty} e^{-\frac{n(1-2p)^2}{8p(1-p)}} < \infty$$

- This establishes state 0 is transient

  $\Rightarrow$ Since all states communicate, all states are transient

Take-home messages

- States of a MC can be transient or recurrent
- A MC can be partitioned into classes of communicating states
  - $\Rightarrow$ Class members are either all transient or all recurrent
  - $\Rightarrow$ Recurrence and transience are class properties
  - $\Rightarrow$ A finite MC has at least one recurrent class
- A MC with only one class is irreducible
  - $\Rightarrow$ If reducible it can be separated into irreducible components

Glossary

- Markov chain
- State space
- Markov property
- Transition probability matrix
- State transition diagram
- State augmentation
- Random walk
- n-step transition probabilities
- Chapman-Kolmogorov eqs.
- Initial distribution
- Gambler's ruin problem
- Communication system
- Non-concurrent queue
- Queue evolution model
- Recurrent and transient states
- Accessibility
- Communication
- Equivalence relation
- Communication classes
- Class property
- Irreducible Markov chain
- Irreducible components

Limiting distributions

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

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

$\Rightarrow$ Assumed that limit is independent of initial state $X_0 = i$

- We've seen that this problem is related to the matrix power $\mathbf{P}^n$

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

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

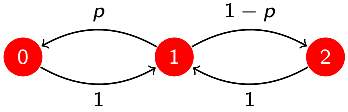
- Matrix product converges $\Rightarrow$ probs. independent of time (large n)
- All rows are equal $\Rightarrow$ probs. independent of initial condition

Periodicity

- **Def:** Period d of a state i is (gcd means greatest common divisor)

$$d = \text{gcd}\left\{n : P_{ii}^n \neq 0\right\}$$

- 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
  - $\Rightarrow d$ is the largest number with this property
- Positive probability of returning to i only every d time steps
  - $\Rightarrow$ If period $d = 1$ state is aperiodic (most often the case)
  - $\Rightarrow$ Periodicity is a class property



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

Periodicity example

Example

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 0.5 & 0.5 \end{pmatrix}, \quad \mathbf{P}^2 = \begin{pmatrix} 0.50 & 0.50 \\ 0.25 & 0.75 \end{pmatrix}, \quad \mathbf{P}^3 = \begin{pmatrix} 0.250 & 0.750 \\ 0.375 & 0.625 \end{pmatrix}$$

- $P_{11} = 0$, but $P_{11}^2, P_{11}^3 \neq 0$ so $\gcd\{2, 3, \ldots\} = 1$. State 1 is aperiodic
- $P_{22} \neq 0$. State 2 is aperiodic (had to be, since $1 \leftrightarrow 2$)

Example

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{P}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{P}^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ldots$$

- $P_{11}^{2n+1} = 0$, but $P_{11}^{2n} \neq 0$ so $\gcd\{2, 4, \ldots\} = 2$. State 1 has period 2
- The same is true for state 2 (since $1 \leftrightarrow 2$)

Positive recurrence and ergodicity

- **Recall:** state i is recurrent if the MC returns to i with probability 1
  - $\Rightarrow$ Define the return time to state i as

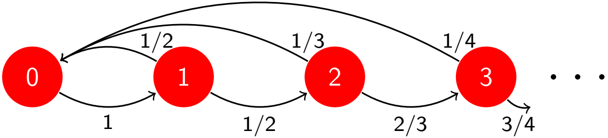
$$T_i = \min\{n > 0 : X_n = i \,\big|\, X_0 = i\}$$

- **Def:** State i is positive recurrent when expected value of T_i is finite

$$\mathbb{E}\left[T_i \,\big|\, X_0 = i\right] = \sum_{n=1}^{\infty} n \mathrm{P}\left(T_i = n \,\big|\, X_0 = i\right) < \infty$$

- **Def:** State i is null recurrent if recurrent but $\mathbb{E}\left[T_i \,\big|\, X_0 = i\right] = \infty$
  - $\Rightarrow$ Positive and null recurrence are class properties
  - $\Rightarrow$ Recurrent states in a finite-state MC are positive recurrent
- **Def:** Jointly positive recurrent and aperiodic states are ergodic
  - $\Rightarrow$ Irreducible MC with ergodic states is said to be an ergodic MC

Null recurrent Markov chain example



$$P\left(T_0 = 2 \,\middle|\, X_0 = 0\right) = \frac{1}{2} \qquad P\left(T_0 = 3 \,\middle|\, X_0 = 0\right) = \frac{1}{2} \times \frac{1}{3}$$

$$P\left(T_0 = 4 \,\middle|\, X_0 = 0\right) = \frac{1}{2} \times \frac{2}{3} \times \frac{1}{4} = \frac{1}{3 \times 4} \quad \ldots \quad P\left(T_0 = n \,\middle|\, X_0 = 0\right) = \frac{1}{(n-1) \times n}$$

- State 0 is recurrent because probability of not returning is 0

$$P\left(T_0 = \infty \,\middle|\, X_0 = 0\right) = \lim_{n \to \infty} \frac{1}{(n-1) \times n} \to 0$$

- Also null recurrent because expected return time is infinite

$$\mathbb{E}\left[T_0 \,\middle|\, X_0 = 0\right] = \sum_{n=2}^{\infty} n P\left(T_0 = n \,\middle|\, X_0 = 0\right) = \sum_{n=2}^{\infty} \frac{1}{(n-1)} = \infty$$

Limit distribution of ergodic Markov chains

Theorem

For an ergodic (i.e., irreducible, aperiodic and positive recurrent) MC, $\lim_{n\to\infty} P_{ij}^n$ exists and is independent of the initial state i, i.e.,

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

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

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

- Limit probs. independent of initial condition exist for ergodic MC
  $\Rightarrow$ Simple algebraic equations can be solved to find π_j
- No periodic, transient, null recurrent states, or multiple classes

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 total probability

$$P_{kj}^{n+1} = \sum_{i=0}^{\infty} \mathrm{P}\left(X_{n+1} = j \,\middle|\, X_n = i, X_0 = k\right) P_{ki}^n$$
$$= \sum_{i=0}^{\infty} P_{ij} P_{ki}^n$$

- If limits exists, $P_{kj}^{n+1} \approx \pi_j$ and $P_{ki}^n \approx \pi_i$ (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 using vector/matrix notation

  $\Rightarrow$ 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 probabilities for all rows $\Rightarrow$ Independent of initial state

- Probability distribution for large n

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

  $\Rightarrow$ Independent of initial condition $\mathbf{p}(0)$

Vector/matrix notation: Eigenvector

- **Def:** Vector limit (steady-state) distribution is $\boldsymbol{\pi} := [\pi_1, \ldots, \pi_J]^T$
- 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 scaling factor
  - Normalize to sum 1
- All other eigenvalues 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 often computationally efficient

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 linearly independent 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)

Ergodic Markov chain example

- MC with transition probability matrix

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

- Q: Does **P** correspond to an ergodic MC?
  - Irreducible: all states communicate with state 2 ✓
  - Positive recurrent: irreducible and finite ✓
  - Aperiodic: period of state 2 is 1 ✓
- Then, there exist π_1, π_2 and π_3 such that $\pi_j = \lim_{n\to\infty} P_{ij}^n$

  $\Rightarrow$ Limit is independent of i

Ergodic Markov chain example (continued)

- Q: How do we determine the limit probabilities π_j?
- Solve system of linear equations $\pi_j = \sum_{i=1}^{3} \pi_i P_{ij}$ and $\sum_{j=1}^{3} \pi_j = 1$

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

$\Rightarrow$ The blue block in the 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
  - A manifestation of the 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
  $\Rightarrow$ Select initial distribution to $\mathrm{P}(X_0 = i) = \pi_i$ for all i
- Probabilities at time $n = 1$ follow from law of total probability

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

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

$$\mathrm{P}(X_1 = j) = \sum_{i=1}^{\infty} P_{ij}\pi_i = \pi_j$$

$\Rightarrow$ Probability distribution is unchanged

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

Ergodicity

Limiting distributions

Ergodicity

Queues in communication networks: Limit probabilities

Ergodicity

- **Def:** Fraction of time $T_i^{(n)}$ spent in i-th state by time n is

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

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

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

- As $n \to \infty$, probabilities $\mathrm{P}\left(X_m = i\right) \to \pi_i$ (ergodic MC). Then

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

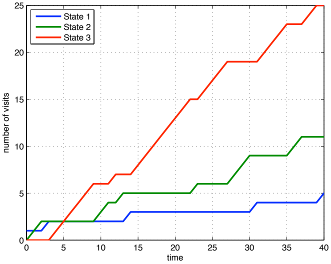
- For ergodic MCs same is true without expected value $\Rightarrow$ Ergodicity

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

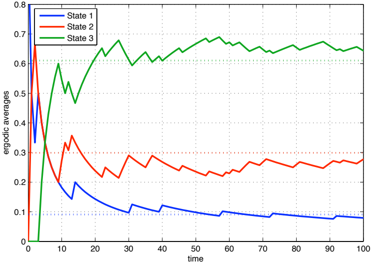
- Recall transition probability matrix

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

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



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



- Ergodic averages slowly converge to $\boldsymbol{\pi} = [0.09, 0.29, 0.61]^T$

Function's ergodic average

Theorem

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

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} f(X_m) = \sum_{j=1}^{\infty} f(j)\pi_j, \quad a.s.$$

- Ergodic average $\to$ Expectation under stationary distribution $\boldsymbol{\pi}$
- Use of ergodic averages is more general than $T_i^{(n)}$
  - $\Rightarrow$ $T_i^{(n)}$ is a particular case with $f(X_m) = \mathbb{I}\{X_m = i\}$
- Think of $f(X_m)$ as a reward (or cost) associated with state X_m
  - $\Rightarrow$ $(1/n)\sum_{m=1}^{n} f(X_m)$ is the time average of rewards (costs)

Function's ergodic average (cheat's 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 $n \to \infty$, use ergodicity result for $\lim_{n\to\infty} T_i^{(n)} = \pi_i$ [cf. page 17]

□

Ensemble and ergodic averages

- 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

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

- These quantities are fundamentally different
  - $\Rightarrow$ But $\mathbb{E}\left[f(X_n)\right] = \bar{f}_n$ almost surely, asymptotically in n
- One realization of the MC as informative as all realizations
  - $\Rightarrow$ Practical value: observe/simulate only one path of the MC

Ergodicity in periodic Markov chains

- Ergodic averages still converge if the MC is periodic
- For irreducible, positive recurrent MC (periodic or aperiodic) define

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

- **Claim 1:** A unique solution exists (we say π_j are well defined)
- **Claim 2:** The fraction of time spent in state i converges to π_i

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

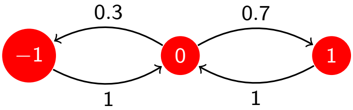
- If MC is periodic the probabilities P_{ij}^n oscillate

  $\Rightarrow$ But fraction of time spent in state i converges to π_i

Periodic irreducible Markov chain example

- Matrix **P** and state transition 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 $X_0 = 0$, then

$$\begin{aligned} P_{00}^{2n+1} &= \mathrm{P}\left(X_{2n+1} = 0 \,\middle|\, X_0 = 0\right) = 0, \\ P_{00}^{2n} &= \mathrm{P}\left(X_{2n} \quad = 0 \,\middle|\, X_0 = 0\right) = 1 \neq 0 \end{aligned}$$

- Define $\boldsymbol{\pi} := [\pi_{-1}, \pi_0, \pi_1]^T$ as solution of

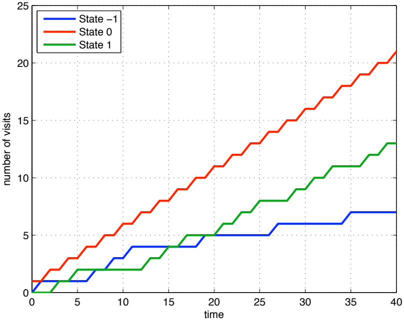
$$\begin{pmatrix} \pi_{-1} \\ \pi_0 \\ \pi_1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 0.3 & 0 \\ 1 & 0 & 1 \\ 0 & 0.7 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \pi_{-1} \\ \pi_0 \\ \pi_1 \end{pmatrix}$$

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

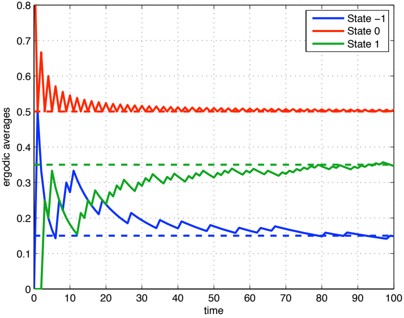
Periodic irreducible MC example (continued)

- Solution yields $\pi_{-1} = 0.15$, $\pi_0 = 0.50$ and $\pi_1 = 0.35$

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



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



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

Periodic irreducible MC example (continued)

- Powers of the transition probability matrix do not converge

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

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

- At least one other eigenvalue of $\mathbf{P}^T$ has modulus 1

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

$\Rightarrow$ 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
  - $\Rightarrow$ All these are (communication) class properties
- Limit probabilities for transient states are null

$$\mathrm{P}\left(X_n = i\right) \to 0, \text{ for all } i \in \mathcal{T}_k$$

- For arbitrary ergodic component $\mathcal{E}_k$, define conditional limits

$$\pi_j = \lim_{n\to\infty} \mathrm{P}\left(X_n = j \,\middle|\, X_0 \in \mathcal{E}_k\right), \quad \text{for all } j \in \mathcal{E}_k$$

- Results in pages 8 and 19 are true with this (re)defined π_j, where

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

Reducible Markov chains (continued)

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

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

- Probabilities $\mathrm{P}\left(X_n = j \,\middle|\, X_0 \in \mathcal{P}_k\right)$ do not converge (they oscillate)
- A conditional version of the result in page 22 is true

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

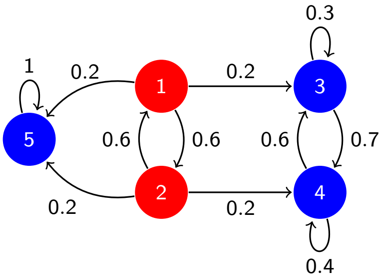
- Limit probabilities for null-recurrent states are null

$$\mathrm{P}\left(X_n = i\right) \to 0, \text{ for all } i \in \mathcal{N}_k$$

Reducible Markov chain example

- 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 (absorbing) is a separate ergodic class $\mathcal{E}_2 = \{5\}$

Reducible MC example - Matrix powers

- 5-step and 10-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)

  $\Rightarrow$ From $\mathcal{T} = \{1, 2\}$ will end up in either $\mathcal{E}_1 = \{3, 4\}$ or $\mathcal{E}_2 = \{5\}$

- Transition from 3 and 4 into 3 and 4 only

  $\Rightarrow$ If initialized in ergodic class $\mathcal{E}_1 = \{3, 4\}$ stays in $\mathcal{E}_1$

- Transition from 5 only into 5 (absorbing state)

Reducible MC example - Matrix decomposition

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

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

  (a) Block $\mathbf{P}_{\mathcal{T}}$ describes transition between transient states
  (b) Blocks $\mathbf{P}_{\mathcal{E}_1}$ and $\mathbf{P}_{\mathcal{E}_2}$ describe transitions within ergodic components
  (c) 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

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

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

Reducible MC example - Limiting behavior

- As n grows the MC hits an ergodic state almost surely
  - $\Rightarrow$ 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
  - $\Rightarrow$ 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 of all MCs as ergodic
- Ergodic behavior cannot be inferred a priori (before observing)
- Becomes known a posteriori (after observing sufficiently large time)

Cultural aside: 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

Limiting distributions

Ergodicity

Queues in communication networks: Limit probabilities

Non-concurrent communication queue

- Communication system: Move packets from source to destination
- Between arrival and transmission hold packets in a memory buffer

- Example engineering problem, buffer design:
  - Packets arrive at a rate of 0.45 packets per unit of time
  - Packets depart at a rate of 0.55 packets per unit of time
  - How big should the buffer be to have a drop rate smaller than 10^{-6}? (i.e., one packet dropped for every million packets handled)

- Model: Time slotted in intervals of duration Δt. 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 equations (reminder)

- Q_n denotes number of packets in queue (backlog) in n-th time slot
- $\mathbb{A}_n =$ nr. of packet arrivals, $\mathbb{D}_n =$ nr. of departures (during n-th slot)
- If the queue is empty $Q_n = 0$ then there are no departures
  $\Rightarrow$ 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

$$Q_{n+1} = Q_n + \mathbb{A}_n - \mathbb{D}_n, \quad \text{if } Q_n > 0$$

- $\mathbb{A}_n \in \{0,1\}$, $\mathbb{D}_n \in \{0,1\}$ and either $\mathbb{A}_n = 1$ or $\mathbb{D}_n = 1$ but not both
  $\Rightarrow$ 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 (reminder)

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

$$\mathrm{P}\left(Q_{n+1} = i+1 \,\middle|\, Q_n = i\right) = \mathrm{P}\left(\mathbb{A}_n = 1\right) = \lambda, \qquad \text{for all } i$$

- 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

$$\mathrm{P}\left(Q_{n+1} = i \,\middle|\, Q_n = i\right) = 1 - \lambda - \mu, \qquad \text{for all } i > 0$$
$$\mathrm{P}\left(Q_{n+1} = 0 \,\middle|\, Q_n = 0\right) = 1 - \lambda$$

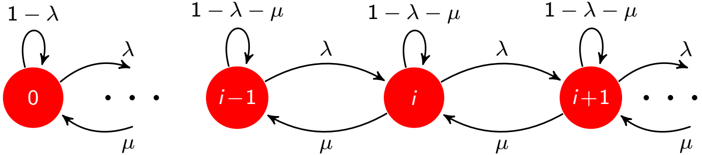
$\Rightarrow$ No departures when $Q_n = 0$ explain second equation

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

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

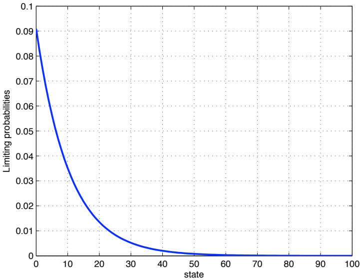
- For $i = 0$: $P_{00} = 1 - \lambda$ and $P_{01} = \lambda$



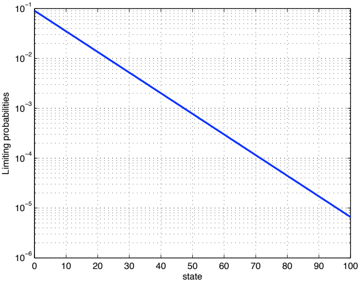
Numerical example: Limit probabilities

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

linear scale

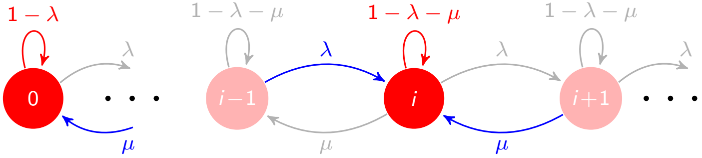


logarithmic scale



- Limit probabilities appear linear in logarithmic scale
  - $\Rightarrow$ Seemingly implying an exponential expression $\pi_i \propto \alpha^i$ $(0 < \alpha < 1)$

Limit distribution equations



- Total probability yields

$$P\left(X_{n+1}=i\right)=\sum_{j=i-1}^{i+1} P\left(X_{n+1}=i \,\middle|\, X_n=j\right) P\left(X_n=j\right)$$

- Limit distribution equations for state 0 (empty queue)

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

- For the remaining states $i \neq 0$

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

Verification of candidate solution

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

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

$\Rightarrow$ The above equation holds for $\alpha = \lambda/\mu$

- Q: Does $\alpha = \lambda/\mu$ verify the remaining equations?
- From the equation for generic π_i (divide by $c\alpha^{i-1}$)

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

$\Rightarrow$ The above quadratic equation is satisfied by $\alpha = \lambda/\mu$

$\Rightarrow$ And $\alpha = 1$, which is irrelevant

Compute normalization constant

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

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

  $\Rightarrow$ Used geometric sum, need $\lambda/\mu < 1$ (queue stability condition)

- 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 queue's stability margin

  $\Rightarrow$ Probability of having fewer queued packets grows with μ/λ

Queue balance equations

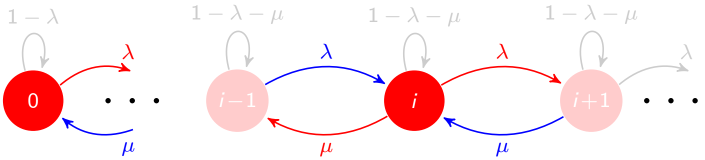
- Rearrange terms in equation for limit probabilities [cf. page 38]

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

  - $\lambda\pi_0$ is average rate at which the queue leaves state 0
  - Likewise $(\lambda+\mu)\pi_i$ is the rate at which the queue leaves state i
  - $\mu\pi_1$ is average rate at which the queue enters state 0
  - $\lambda\pi_{i-1} + \mu\pi_{i+1}$ is rate at which the 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)
  - $\Rightarrow$ Queue evolution equations remain the same [cf. page 34]
  - $\Rightarrow$ But queue probabilities change [cf. page 35]

- 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 $i > 0$)

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

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

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

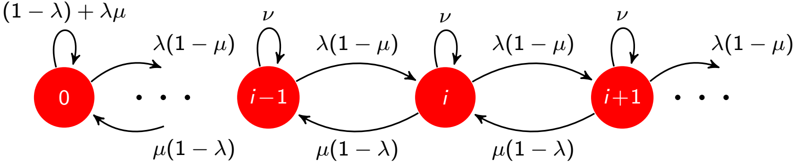
$$\mathrm{P}\left(Q_{n+1} = 0 \,\middle|\, Q_n = 0\right) = (1-\lambda) + \lambda\mu$$

Limit distribution from queue balance equations

- Write limit distribution equations $\Rightarrow$ Queue balance equations

  $\Rightarrow$ Rate at which leaves = Rate at which enters

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



- Again, try an exponential solution $\pi_i = c\alpha^i$

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

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

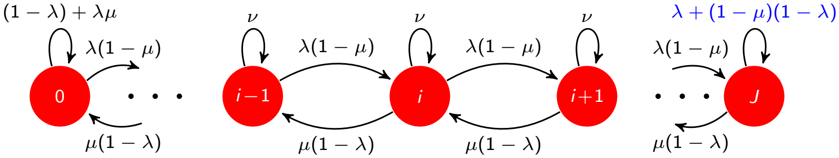
$\Rightarrow$ As before is solved for $\alpha = \lambda(1-\mu)/\mu(1-\lambda)$

- 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 queue backlog exceeds buffer size J
  - $\Rightarrow$ Many packets $\rightarrow$ large delays $\rightarrow$ packets useless upon arrival
  - $\Rightarrow$ Also preserve memory



- Should modify equation for state J (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 solves 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 probabilities for the finite queue thus are

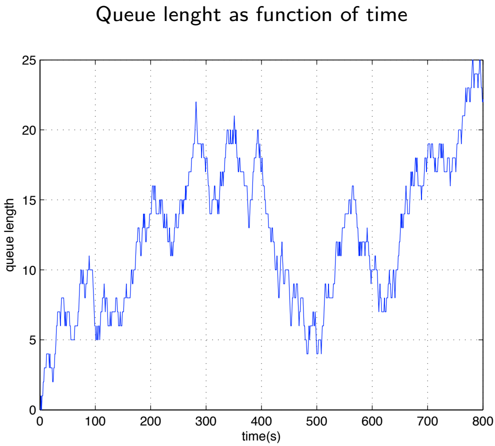
$$\pi_i = \frac{1-\alpha}{1-\alpha^{J+1}}\alpha^i \approx (1-\alpha)\alpha^i$$

$\Rightarrow$ Recall $\alpha = \lambda(1-\mu)/\mu(1-\lambda)$, and $\approx$ valid for large J

- Large J approximation yields same result as infinite length queue

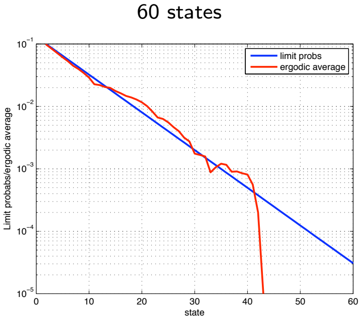
Simulations: Process realization

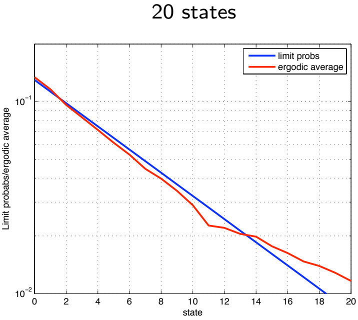
- Arrival rate $\lambda = 0.3$. Departure rate $\mu = 0.33$. Resulting $\alpha \approx 0.87$
- Maximum queue length $J = 100$. Initial state $Q_0 = 0$ (queue empty)



Simulations: Average occupancy and limit distribution

- Can estimate average time spent at each queue state
  - $\Rightarrow$ Should coincide with the limit (stationary) distribution $\boldsymbol{\pi}$





- For $i = 60$ occupancy probability is $\pi_i \approx 10^{-5}$
  - $\Rightarrow$ Explains inaccurate prediction for large i (rarely visit state i)

Buffer overflow

- Closing the loop, recall our buffer design problem
  - Arrival rate $\lambda = 0.45$ and departure rate $\mu = 0.55$
  - How big should the buffer be to have a drop rate smaller than 10^{-6}? (i.e., one packet dropped for every million packets handled)

- Q: What is the probability of buffer overflow (non-concurrent case)?
- A: Packet discarded if queue is in state J and a new packet arrives

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

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

- A final caveat

  $\Rightarrow$ Still assuming only 1 packet arrives per time slot

  $\Rightarrow$ Lifting this assumption requires continuous-time MCs