# 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)

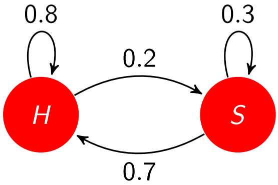
(IV) Stationary random processes (8 lectures)

- 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 $X_{n-1}$, $X_{n-2}$, ...

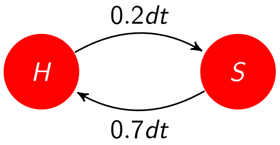
- Can be happy ($X_n = 0$) or sad ($X_n = 1$)
- Tomorrow's mood only affected by today's mood
- Whether happy or sad today, likely to be happy tomorrow
- But when sad, a little less likely so




- 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 $W(t) > 0$

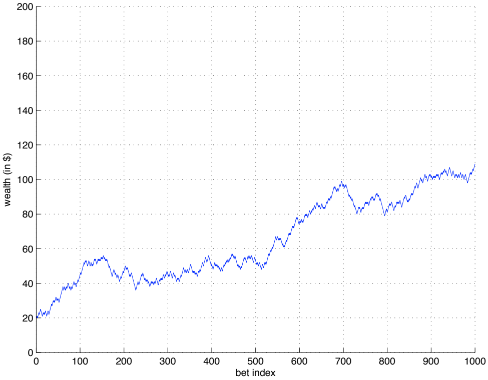
# Coding

```
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?

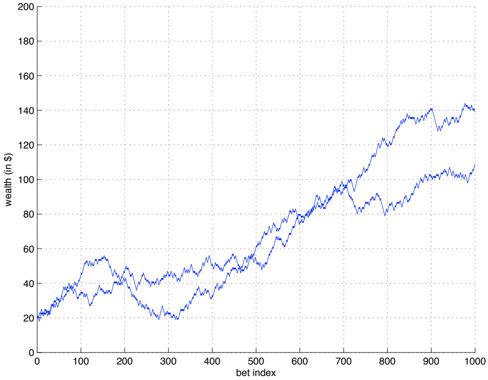
# One lucky player

- 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



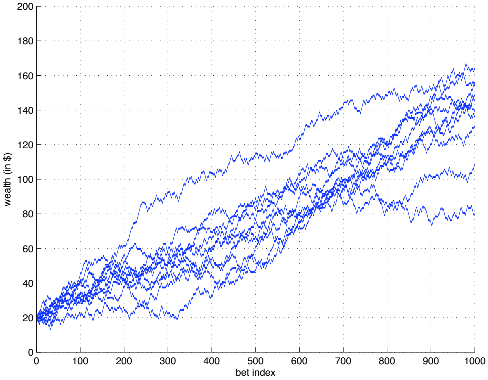
# Two lucky players

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



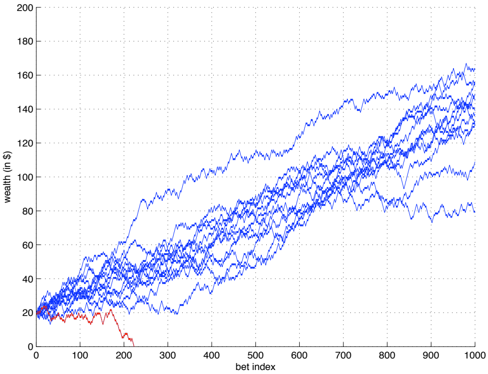
# Ten lucky players

- Wealths $W_j(t)$ after $t = 1000$ bets between 78 and 139
  - $\Rightarrow$ Increasing wealth is definitely a pattern



# One unlucky player

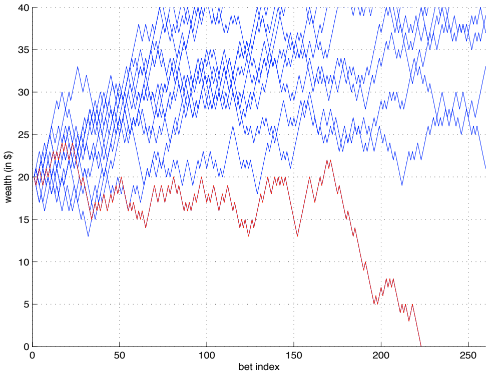
- But this does not mean that all players will turn out as winners
  - $\Rightarrow$ The twelfth player $j = 12$ goes broke



# One unlucky player

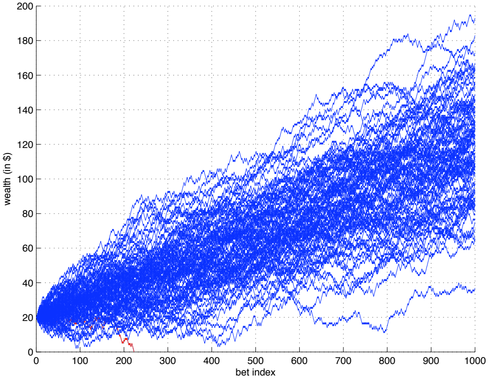
- But this does not mean that all players will turn out as winners

  $\Rightarrow$ The twelfth player $j = 12$ goes broke



# One hundred players

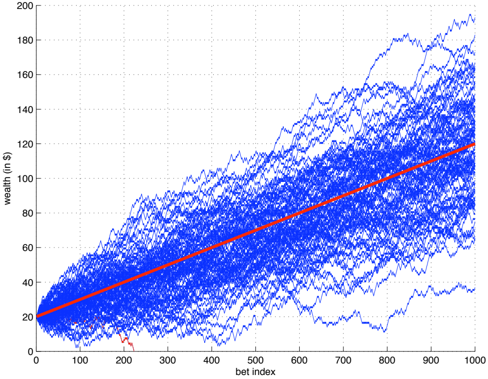
- All players (except for $j = 12$) end up with substantially more money



# Average tendency

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

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



# 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

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

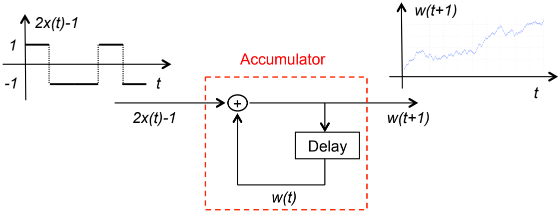
- The expected value of Bernoulli $X(t)$ is

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

- 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

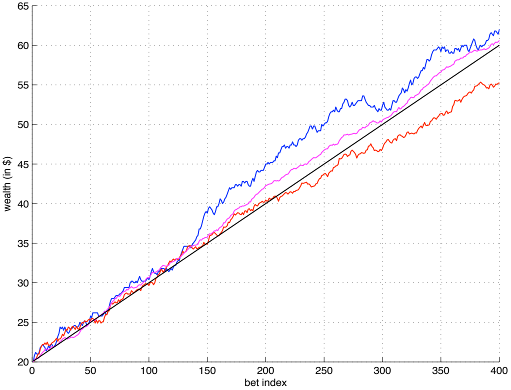
- For a more accurate approximation analyze simulation outcomes
- Consider $J$ experiments. Each yields a wealth history $W_j(t)$
- Can estimate the average outcome via the sample average $\bar{W}_J(t)$

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

- 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 $w^{(0)}, \ldots, w^{(M)}$
- Indicator function of the event $w^{(m)} \leq W_j(t) < w^{(m+1)}$

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

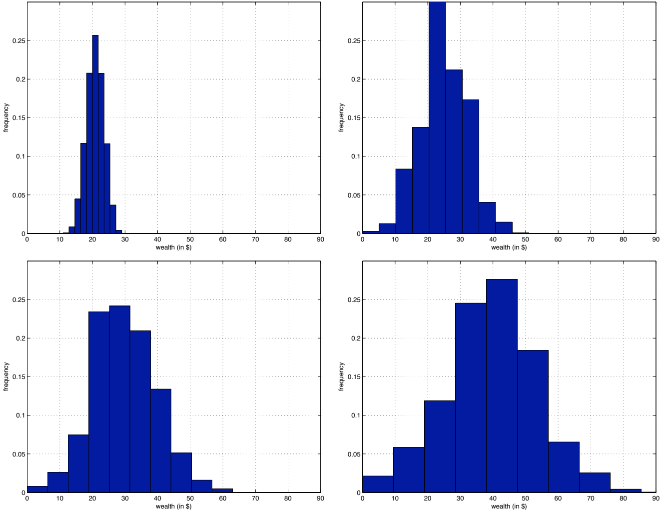
- Histogram is then defined as

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

- 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

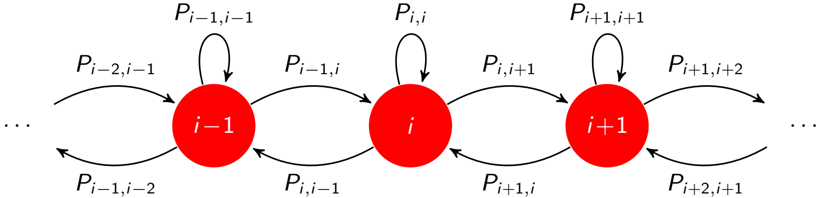
- Group the $P_{ij}$ in a transition probability "matrix" $\mathbf{P}$

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

$\Rightarrow$ 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 $i$

- 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

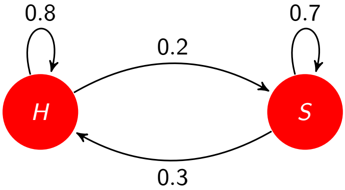
# Example: Happy - Sad

- 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

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

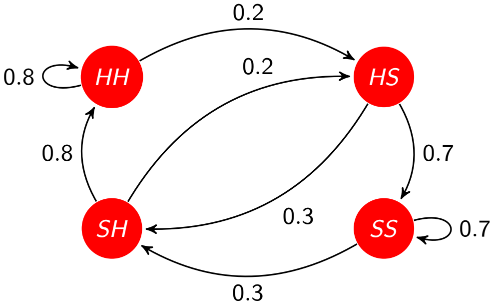



- Inertia $\Rightarrow$ happy or sad today, likely to stay happy or sad tomorrow
- But when sad, a little less likely so ($P_{00} > P_{11}$)

# 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

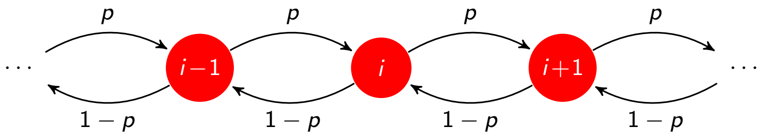
$$\mathbf{P} = \begin{pmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \\ 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \end{pmatrix}$$




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

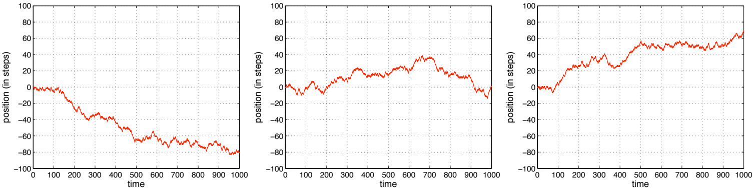
$$P_{i,i+1} = p, \qquad P_{i,i-1} = 1 - p$$

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

# Random (drunkard's) walk (continued)

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

$p = 0.45$ $\quad$ $p = 0.50$ $\quad$ $p = 0.55$




$\Rightarrow$ 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

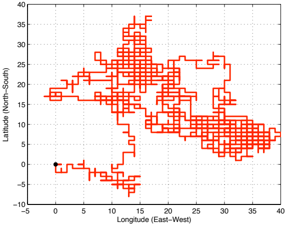
- Take a step in random direction E, W, S or N
  - $\Rightarrow$ E, W, S, N chosen with equal probability
- States are pairs of coordinates $(X_n, Y_n)$
  - $X_n = 0, \pm 1, \pm 2, \ldots$ and $Y_n = 0, \pm 1, \pm 2, \ldots$
- Transiton probs. $\neq 0$ only for adjacent points

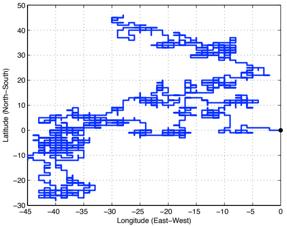
East: $\mathrm{P}\left(X_{n+1} = i+1, Y_{n+1} = j \,\middle|\, X_n = i, Y_n = j\right) = \frac{1}{4}$

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

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

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



# 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
  - $\Rightarrow$ 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 $\pm 1$, $\mathrm{P}\left(Y_n = 1\right) = p$, $\mathrm{P}\left(Y_n = -1\right) = 1 - p$

- 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*

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

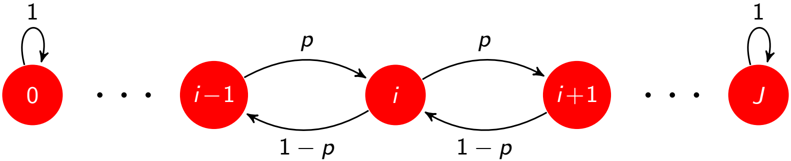
*Then $X_{\mathbb{N}}$ is a Markov chain with transition probabilities*

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

- 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

# 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:
  - ⇒ $\mathbf{P}^{(m)}$ with elements $P_{ij}^{m}$
  - ⇒ $\mathbf{P}^{(n)}$ with elements $P_{ij}^{n}$
  - ⇒ $\mathbf{P}^{(m+n)}$ 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.,*

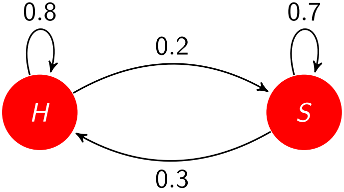
$$\mathbf{P}^{(n)} = \mathbf{P}^n$$

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

# Example: Happy-Sad

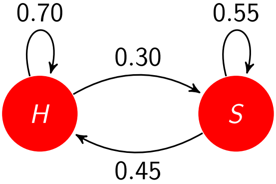
- Mood transitions in one day

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




- 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 $p_j(n)$

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

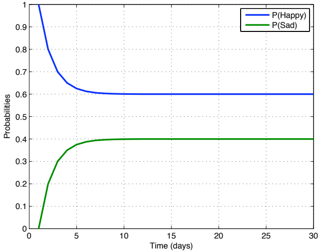
- In matrix form (define vector $\mathbf{p}(n) = [p_1(n), p_2(n), \ldots]^T$)

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

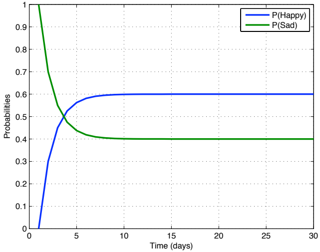
# Example: Happy-Sad

- 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 $\mathbf{p}(0)$

# 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 $\Delta t$
  $\Rightarrow$ $n$-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 $\Delta t$ 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 $\Delta t$ is $\mu = \bar{\mu}\Delta t$
- Assume no simultaneous arrival and departure (no concurrence)
  $\Rightarrow$ Reasonable for small $\Delta t$ ($\mu$ and $\lambda$ 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

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

# 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

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

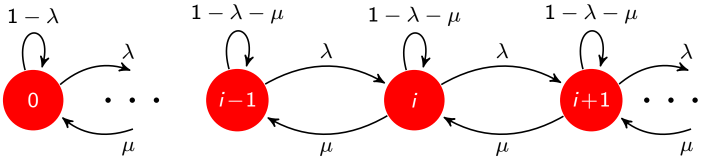
$\Rightarrow$ No departures when $Q_n = 0$ explain second equation

# Queue as a Markov chain

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

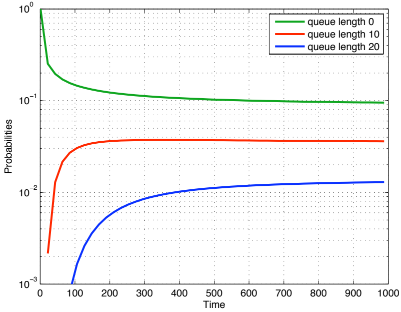
$$P_{i,i-1} = \mu, \qquad P_{i,i} = 1 - \lambda - \mu, \qquad P_{i,i+1} = \lambda$$

- 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

$$\mathrm{P}\left(Q_n = i \,\middle|\, Q_0 = 0\right) = p_i(n)$$

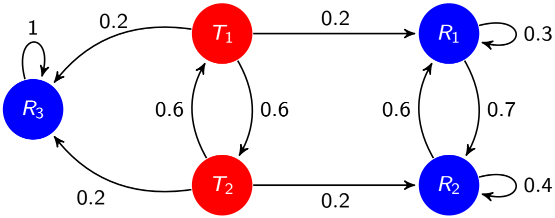
- A few $i$'s (0, 10, 20) shown
- Probability of empty queue $\approx 0.1$
- Occupancy decreases with $i$

# 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 $i$

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

- State $i$ is recurrent if $f_i = 1$

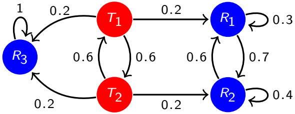
  $\Rightarrow$ Process reenters $i$ again and again (a.s.). Infinitely often

- 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

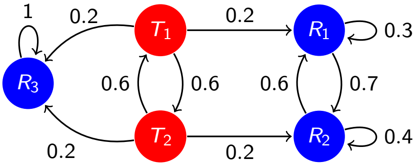
$$\mathrm{P}\left(X_1 = R_1 \,\middle|\, X_0 = R_1\right) = 0.3$$
$$\mathrm{P}\left(X_2 = R_1, X_1 \neq R_1 \,\middle|\, X_0 = R_1\right) = (0.7)(0.6)$$
$$\mathrm{P}\left(X_3 = R_1, X_2 \neq R_1, X_1 \neq R_1 \,\middle|\, X_0 = R_1\right) = (0.7)(0.4)(0.6)$$
$$\vdots$$
$$\mathrm{P}\left(X_n = R_1, X_{n-1}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

- 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 $\mathbb{E}[N_i]$

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

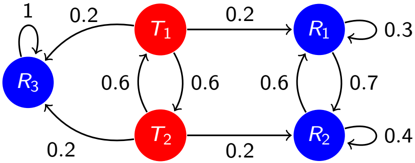
- Recall that: for transient states $\mathbb{E}[N_i] = f_i/(1-f_i) < \infty$
  for recurrent states $\mathbb{E}[N_i] = \infty$

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

- **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 $P_{ii}^0 = \text{P}\left(X_0 = i \,\middle|\, X_0 = i\right) = 1$, 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?

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

$\square$

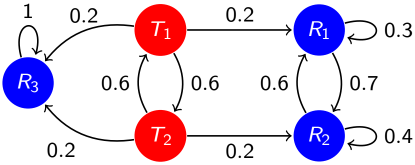
- Recurrence is shared by elements of a communication class
  - $\Rightarrow$ We say that recurrence is a class property
- Likewise, transience is also a class property
- MC states are separated in classes of transient and recurrent states

# 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 $\mathcal{R}_2$