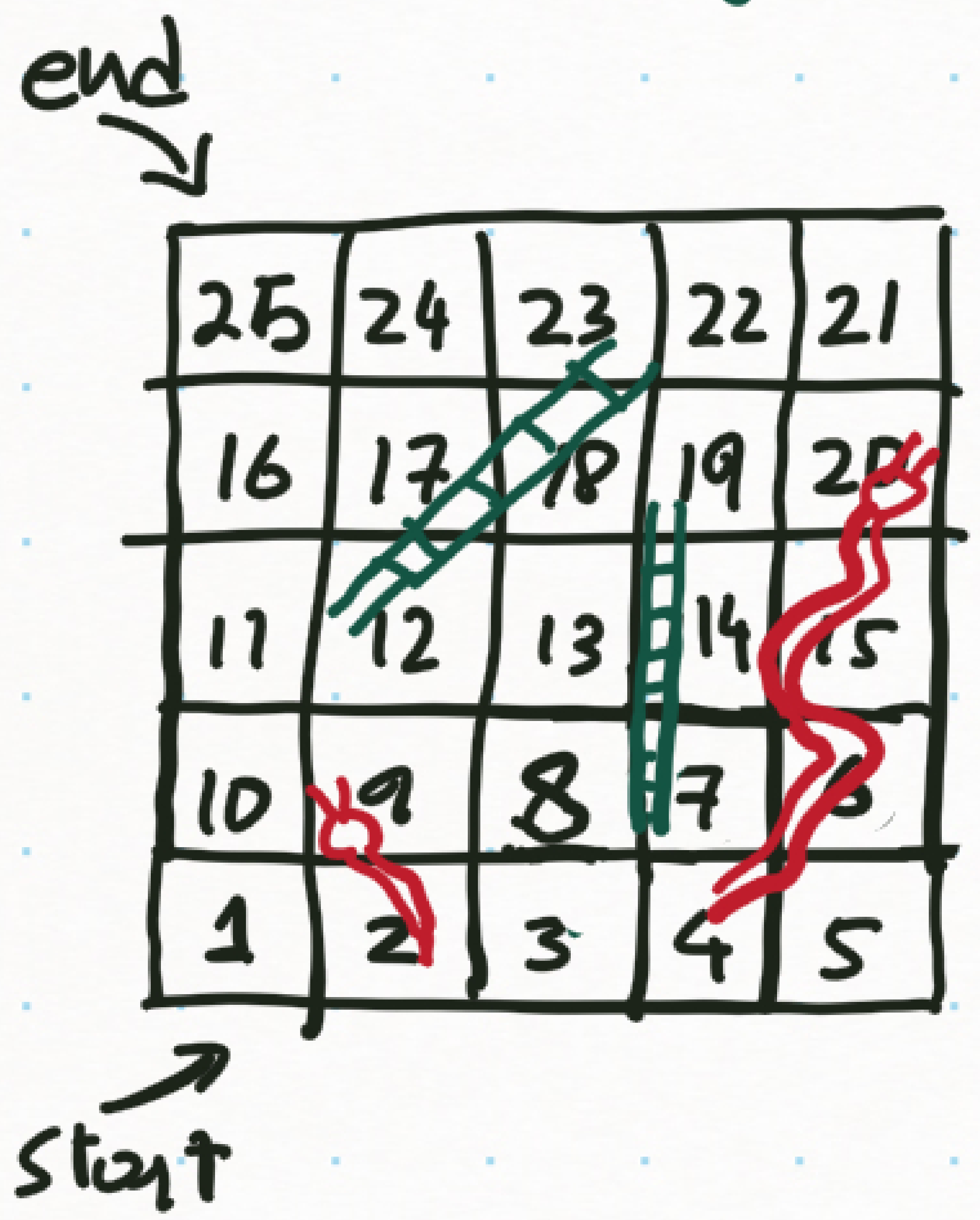
# MATH 380

Hemanshu Kaul

kaul@iit.edu

# Modeling with Markov Chains

We want to study a sequence of random events, e.g. Daily weather, or even more simply a game based on roll of dice such as snakes-and-ladders.



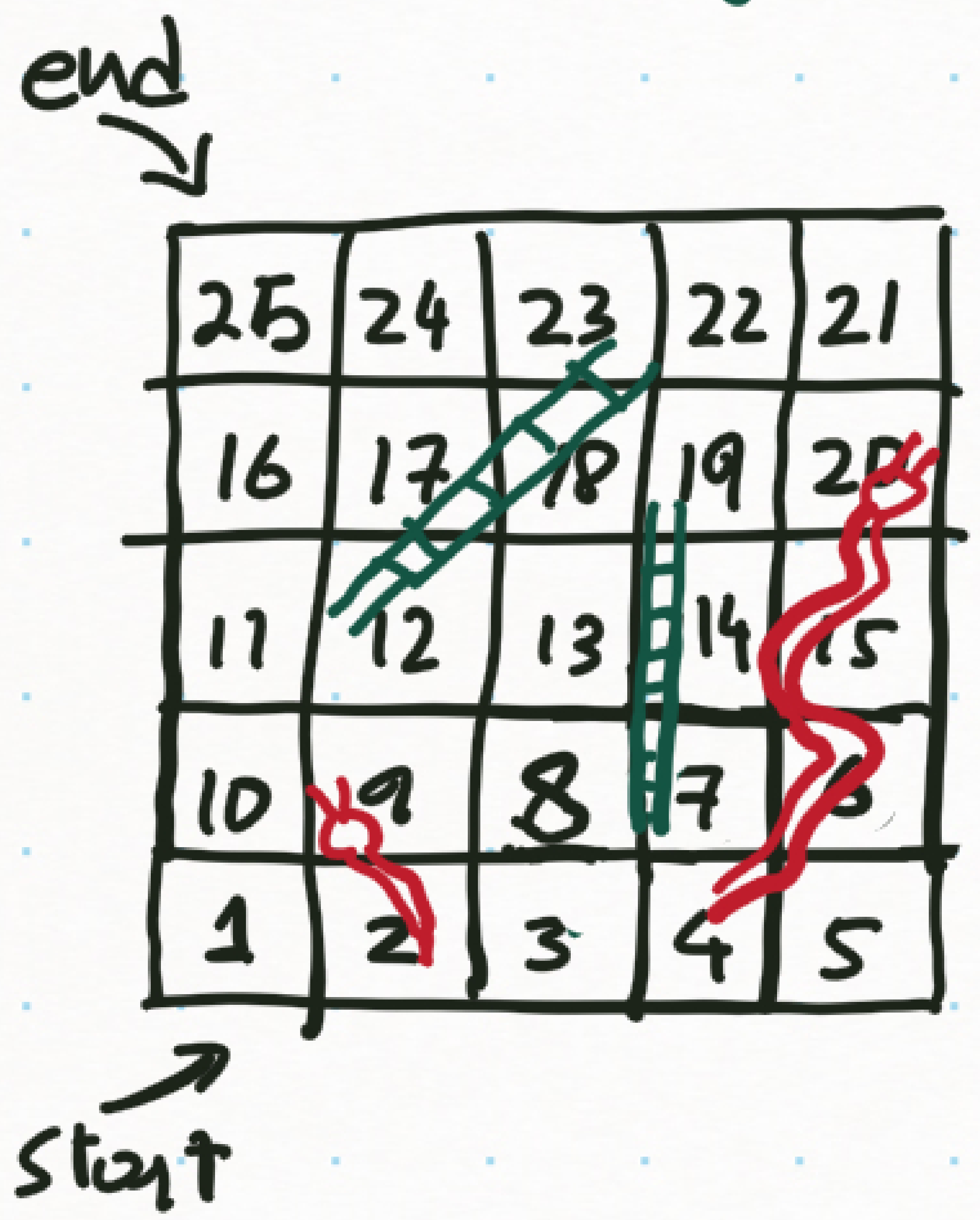

Players next position depends <u>only</u> on

→ current position

→ roll of die

# Modeling with Markov Chains

We want to study a sequence of random events, e.g. Daily weather, or even more simply a game based on roll of dice such as snakes-and-ladders.

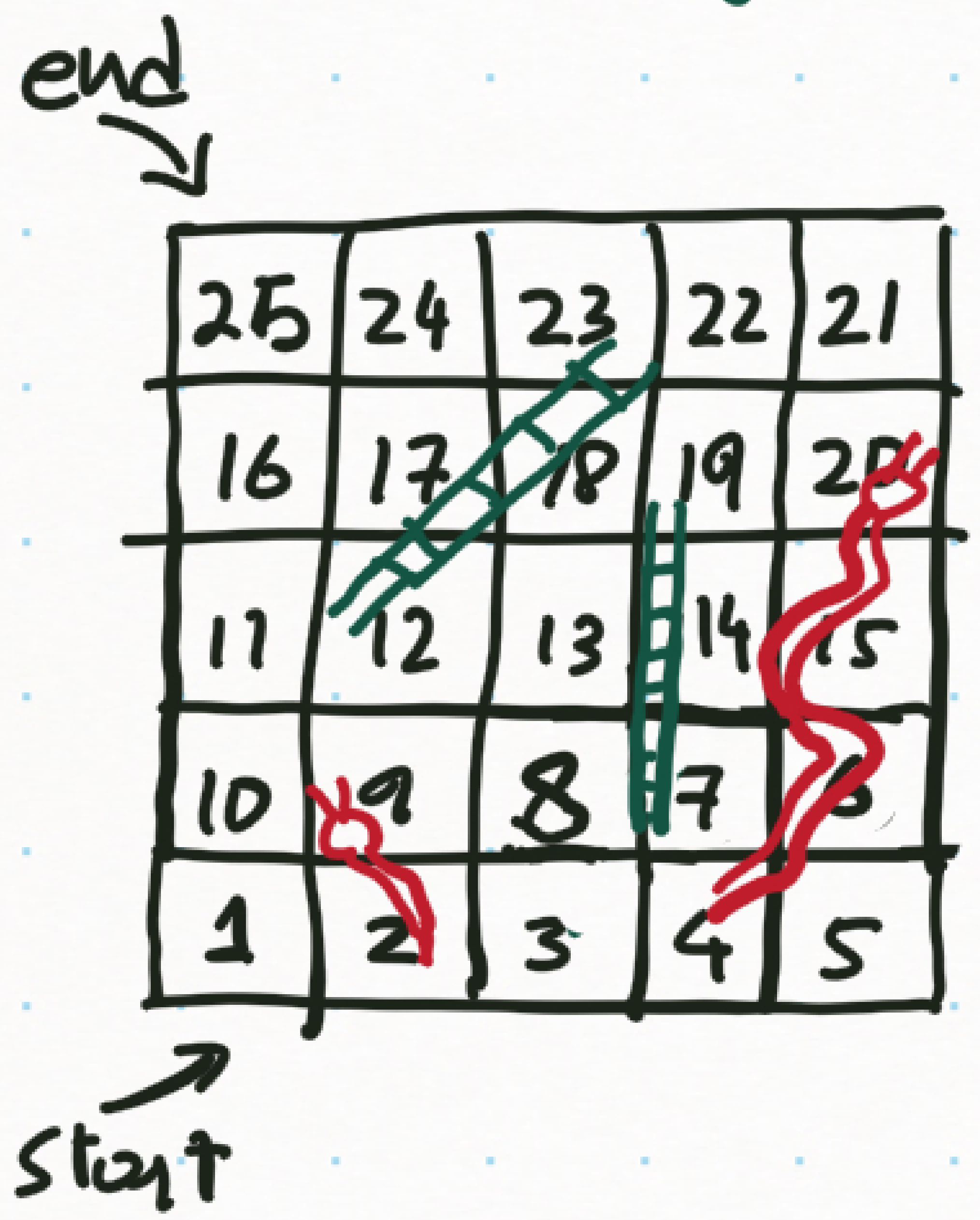



Players next position depends <u>only</u> on → current position
→ roll of die

e.g. If current position is 3
then next position can be ——— ?

# Modeling with Markov Chains

We want to study a sequence of random events, e.g. Daily weather, or even more simply a game based on roll of dice such as snakes-and-ladders.




Players next position depends ***only*** on
→ current position
→ roll of die

e.g. If current position is 3

then next position can be

| | | |
|---|---|---|
| 4 | with probab | 1/6 |
| 5 | — | 1/6 |
| 6 | — | 1/6 |
| ~~7~~ → 19 | — | 1/6 |
| 8 | — | 1/6 |
| ~~9~~ → 2 | — | 1/6 |

All other positions are possible with probability zero.

# Modeling with Markov Chains

We want to study a sequence of random events, e.g. Daily weather, or even more simply a game based on roll of dice such as snakes-and-ladders.

e.g. We describe daily weather as one of

Sunny
Cloudy
Rainy

States

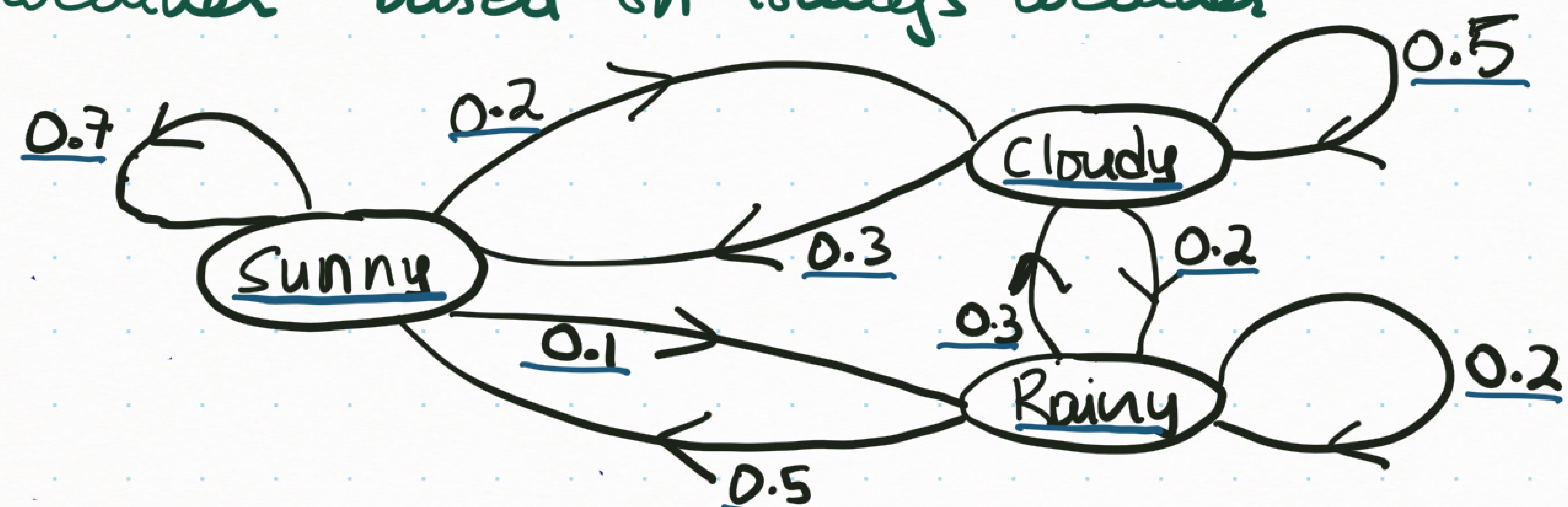
# Modeling with Markov Chains

We want to study a sequence of random events, e.g. Daily weather, or even more simply a game based on roll of dice such as snakes-and-ladders.

e.g. we describe daily weather as one of

$$\left.\begin{array}{l}\text{Sunny}\\ \text{Cloudy}\\ \text{Rainy}\end{array}\right\}\ \text{States}$$

Based on historical weather data, we can predict the next day's weather based on today's weather



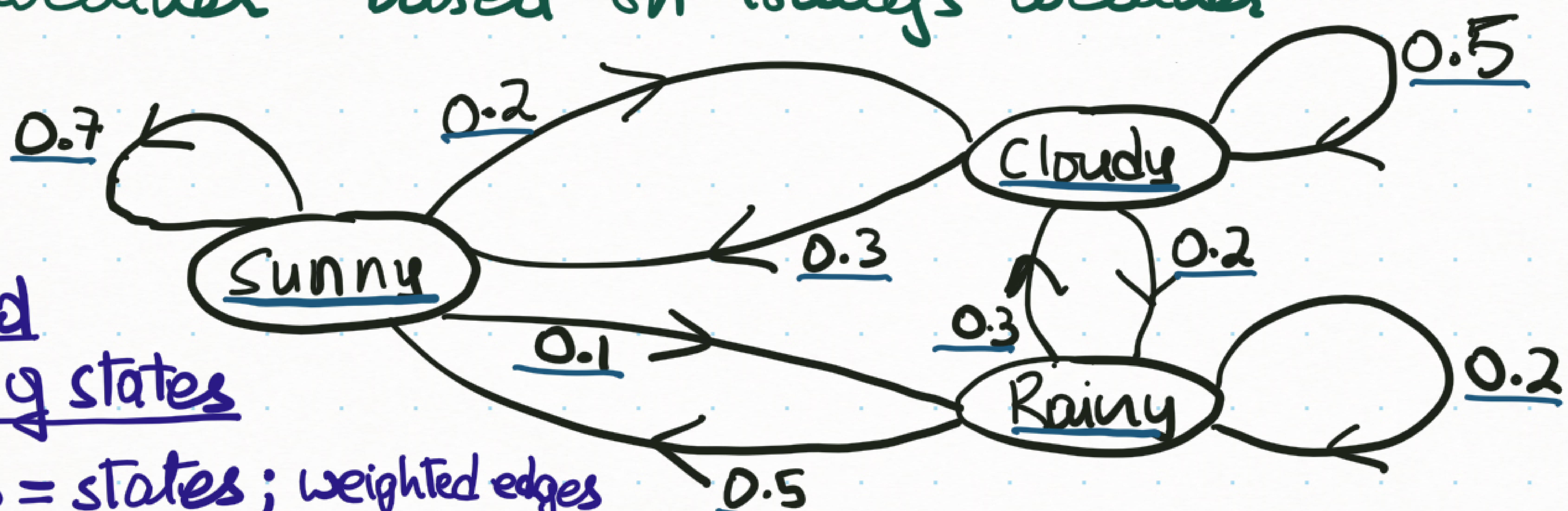
# Modeling with Markov Chains

We want to study a sequence of random events, e.g. Daily weather, or even more simply a game based on roll of dice such as snakes-and-ladders.

e.g. we describe daily weather as one of

$$\left.\begin{array}{l}\text{Sunny}\\ \text{Cloudy}\\ \text{Rainy}\end{array}\right\}\text{States}$$

Based on historical weather data, we can predict the next day's weather based on today's weather




Note the sum of probabilities out of each state equals 1.

Directed graph of states

vertices = states; weighted edges

Markov Chain is a sequence of random outcomes where the next outcome depends only on the current outcome and not on any of the previous outcomes.

(This property is called "memorylessness")

Markov Chain is a sequence of random outcomes where the next outcome depends only on the current outcome and not on any of the previous outcomes.

(This property is called "memorylessness")

To describe a Markov Chain $(X_n)_{n=0}^{\infty}$, $X_0, X_1, X_2, \dots$, we need

→ State space (All possible outcomes / states $X_n$ can take)

$$S = \{s_1, s_2, \dots, s_k\}$$

→ Probability of moving from current state to any possible next state

$$\mathbb{P}[X_{n+1} = s_i \mid X_n = s_j] \quad \forall i, j$$

Markov Chain is a sequence of random outcomes where the next outcome depends only on the current outcome and not on any of the previous outcomes.
(This property is called "memorylessness")

To describe a Markov Chain $(X_n)_{n=0}^{\infty}$, $X_0, X_1, X_2, \dots$, we need

State space $S = \{s_1, \dots, s_k\}$ and Transition probabilities $\mathbb{P}[X_{n+1} = s_j \mid X_n = s_i] = P_{ij}(n)$

(If $P_{ij}(n)$ doesn't depend on $n$ then we simply use $P_{ij}$)

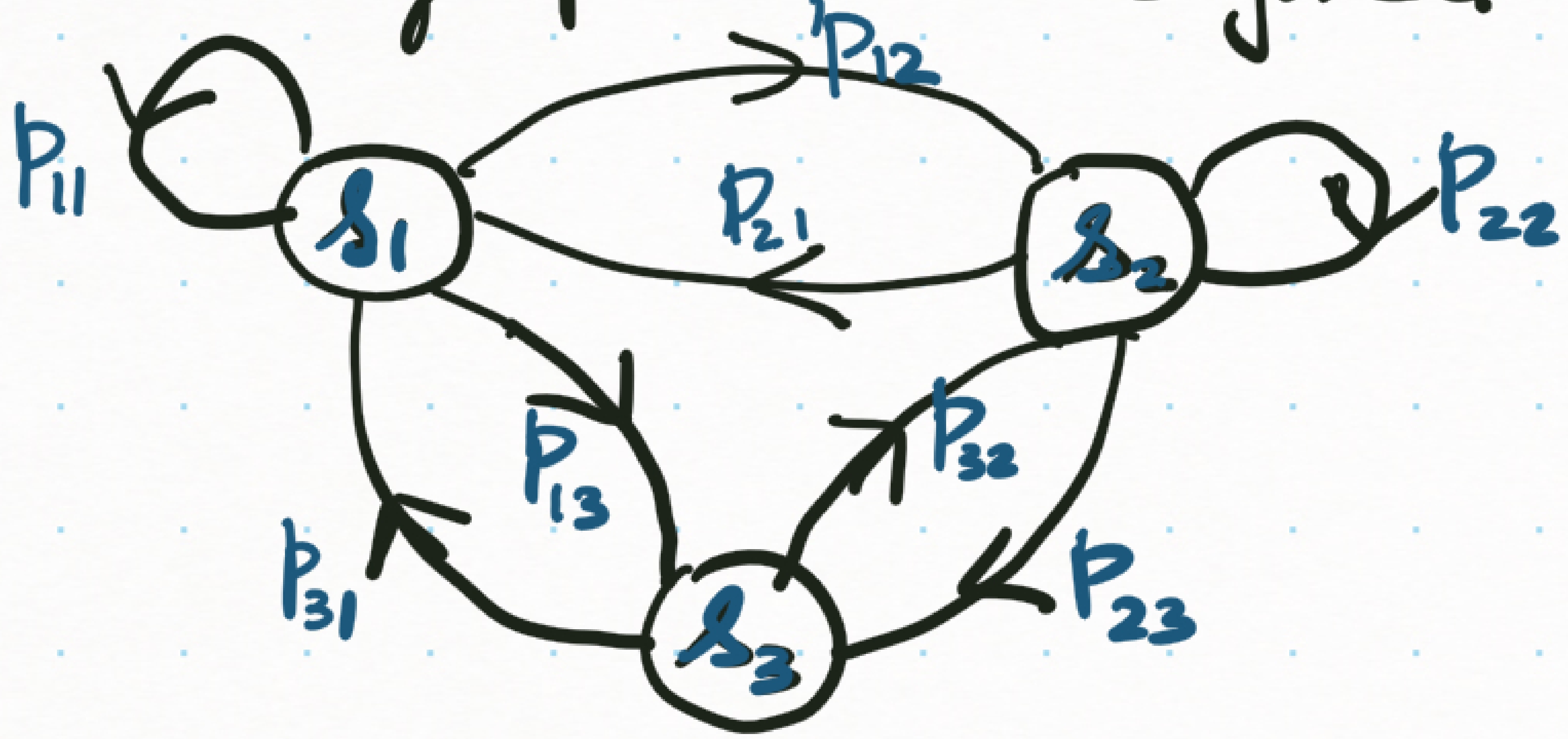
Markov Chain is a sequence of random outcomes where the next outcome depends only on the current outcome and not on any of the previous outcomes.
(This property is called "memorylessness")

To describe a Markov Chain $(X_n)_{n=0}^{\infty}$, $X_0, X_1, X_2, \dots$, we need

State space $S = \{s_1, \dots, s_k\}$ and Transition probabilities $\mathbb{P}[X_{n+1} = s_j \mid X_n = s_i] = p_{ij}(n)$

(If $p_{ij}(n)$ doesn't depend on $n$ then we simply use $p_{ij}$)

Which can be visualized using a directed graph with weighted edges:




Vertex set = $S$

All possible edges except those with transition probability equal to 0.

Markov Chain is a sequence of random outcomes where the next outcome depends only on the current outcome and not on any of the previous outcomes.
(This property is called "memorylessness")

To describe a Markov Chain $(X_n)_{n=0}^{\infty}$, $X_0, X_1, X_2, \cdots$, we need

State space $S = \{s_1, .., s_k\}$ and Transition probabilities $\mathbb{P}[X_{n+1} = s_j \mid X_n = s_i] = P_{ij}(n)$

(If $P_{ij}(n)$ doesn't depend on $n$ then we simply use $P_{ij}$)

which can be captured using the transition matrix

$k \times k$ matrix with $(i,j)$ entry $P_{ij}$

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1k} \\ P_{21} & P_{22} & \cdots & P_{2k} \\ \vdots & & & \\ P_{k1} & P_{k2} & \cdots & P_{kk} \end{bmatrix}_{k \times k}$$

sum of entries in each row equals 1.

Markov Chain is a sequence of random outcomes where the next outcome depends only on the current outcome and not on any of the previous outcomes.
(This property is called "memorylessness")

To describe a Markov Chain $(X_n)_{n=0}^{\infty}$, $X_0, X_1, X_2, \dots$, we need

State space $S = \{s_1, \dots, s_k\}$ and Transition probabilities $\mathbb{P}[X_{n+1} = s_j \mid X_n = s_i] = P_{ij}(n)$

(If $P_{ij}(n)$ doesn't depend on $n$ then we simply use $P_{ij}$)

which can be captured using the transition matrix

$k \times k$ matrix with $(i,j)$ entry $P_{ij}$

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1k} \\ P_{21} & P_{22} & \cdots & P_{2k} \\ \vdots & & & \\ P_{k1} & P_{k2} & \cdots & P_{kk} \end{bmatrix}_{k \times k}$$

sum of entries in each row equals 1.

e.g. $P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}$ $s_1$ = sunny, $s_2$ = cloudy, $s_3$ = rainy

in the weather example.

## Car Rental Company

Two locations in Orlando & Tampa

Cars can be returned to either location, regardless where it was rented.

What will be the long-term impact of this return policy?

What proportion of cars will remain in each location?

# Car Rental Company

Two locations in Orlando & Tampa

Cars can be returned to either location, regardless where it was rented.

What will be the long-term impact of this return policy?

What proportion of cars will remain in each location?

Historical data shows

60% of cars rented in O are returned to O

40% —— " —— O —— " —— T

70% —— " —— T —— " —— T

30% —— " —— T —— " —— O

Will all the cars be in Tampa in the long run?

# Car Rental Company

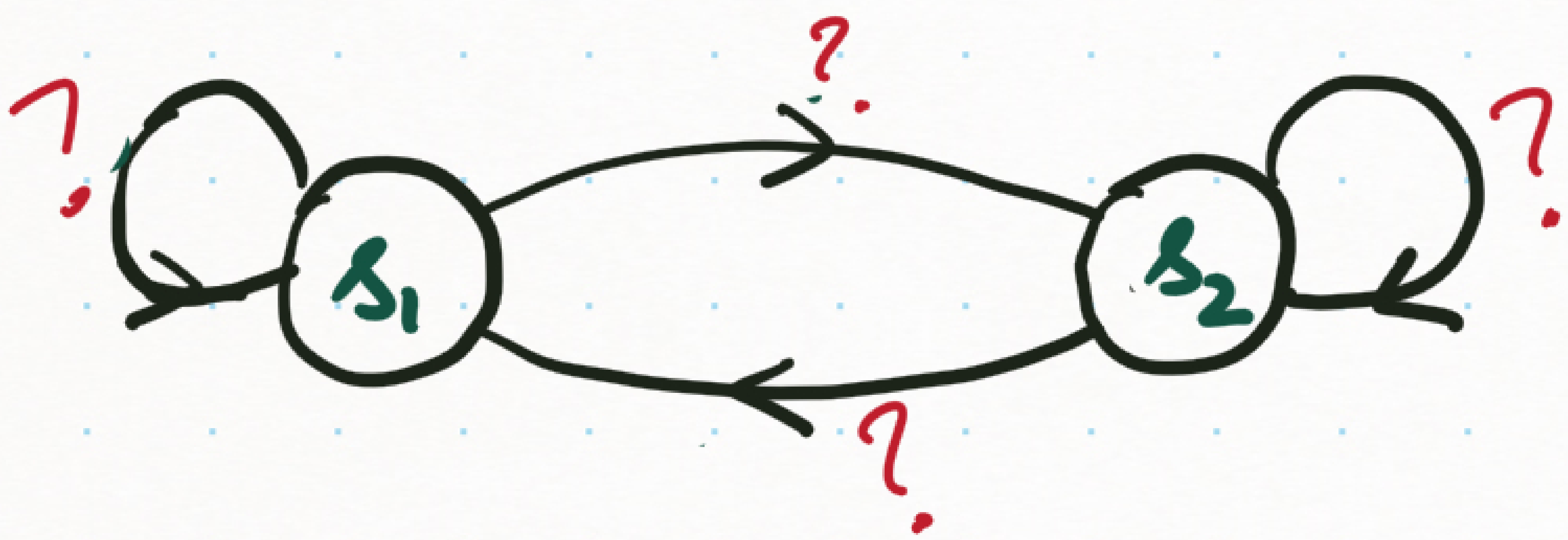
Since the next location of the car only depends on the current locations (& not the past ones) and the current renters, we may model the behavior by a Markov Chain

State Space, $S = ?$

# Car Rental Company

Since the next location of the car only depends on the current locations (& not the past ones) and the current renters, we may model the behavior by a Markov Chain

State Space, $S = \{s_1 = \text{Orlando}, s_2 = \text{Tampa}\}$



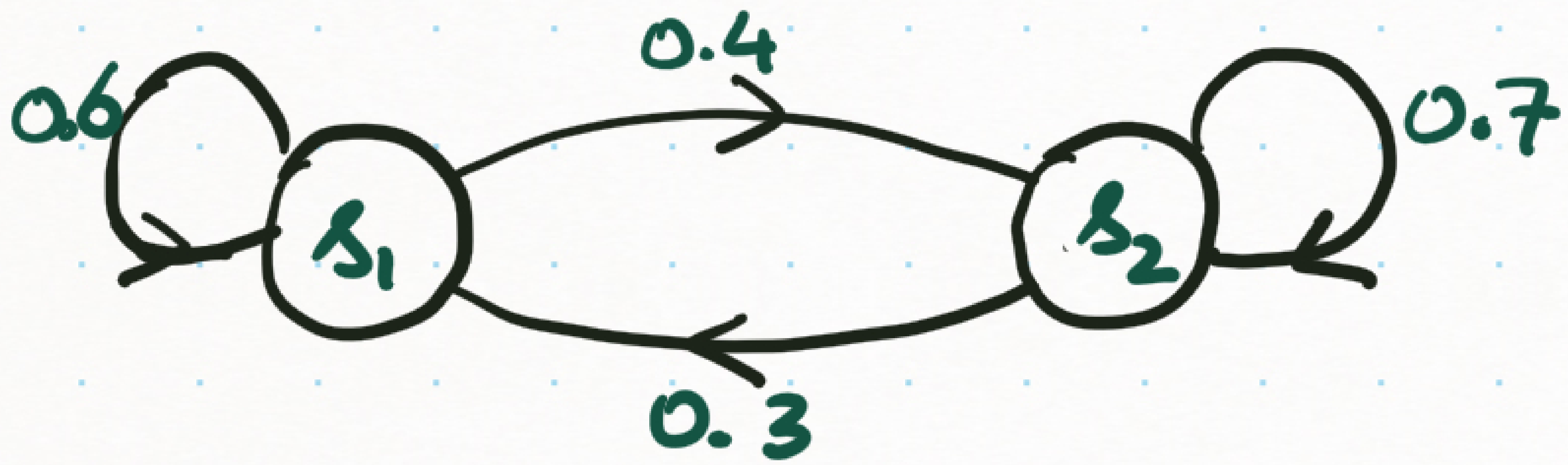

Transition matrix

$$P = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$$

# Car Rental Company

Since the next location of the car only depends on the current locations (& not the past ones) and the current renters, we may model the behavior by a Markov Chain

State Space, $S = \{s_1 = \text{Orlando}, s_2 = \text{Tampa}\}$



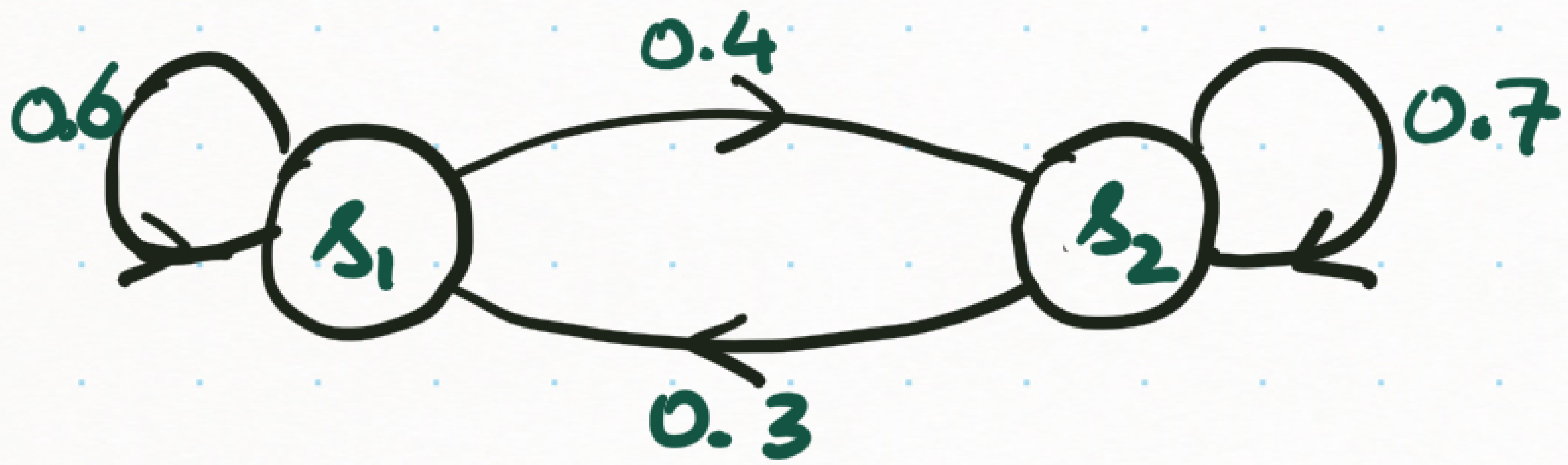

Transition matrix

$$P = \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}$$

## Car Rental Company

Since the next location of the car only depends on the current locations (& not the past ones) and the current renters, we may model the behavior by a Markov Chain.

State Space, $S = \{s_1 = \text{Orlando}, s_2 = \text{Tampa}\}$




Transition matrix

$$P = \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}$$

Let $p_n$ = percentage of cars in Orlando at the end of time period $n$

$q_n$ = percentage of cars in Tampa at the end of time period $n$

$$\left.\begin{aligned} p_{n+1} &= 0.6\,p_n + 0.3\,q_n \\ q_{n+1} &= 0.4\,p_n + 0.7\,q_n \end{aligned}\right] \longleftrightarrow \begin{bmatrix} p_{n+1} \\ q_{n+1} \end{bmatrix} = \begin{bmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{bmatrix} \begin{bmatrix} p_n \\ q_n \end{bmatrix}$$

# Car Rental Company

Based on $p_{n+1} = 0.6\,p_n + 0.3\,q_n$ for $n = 0, 1, 2, \dots$

$q_{n+1} = 0.4\,p_n + 0.7\,q_n$

Numerically compute the values of $p_{n+1}$ & $q_{n+1}$ using the values of $p_n$ & $q_n$ for $n = 0, 1, 2, \dots$ ← Loop.

Use different $(p_0, q_0)$ (say $(1, 0)$ or $(0, 1)$ or $(0.5, 0.5)$)

| $n$ | $p_n$ | $q_n$ |
|---|---|---|
| 0 | 1 | 0 |
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| ⋮ | | |

# Car Rental Company

Based on $p_{n+1} = 0.6\,p_n + 0.3\,q_n$ for $n = 0, 1, 2, \dots$

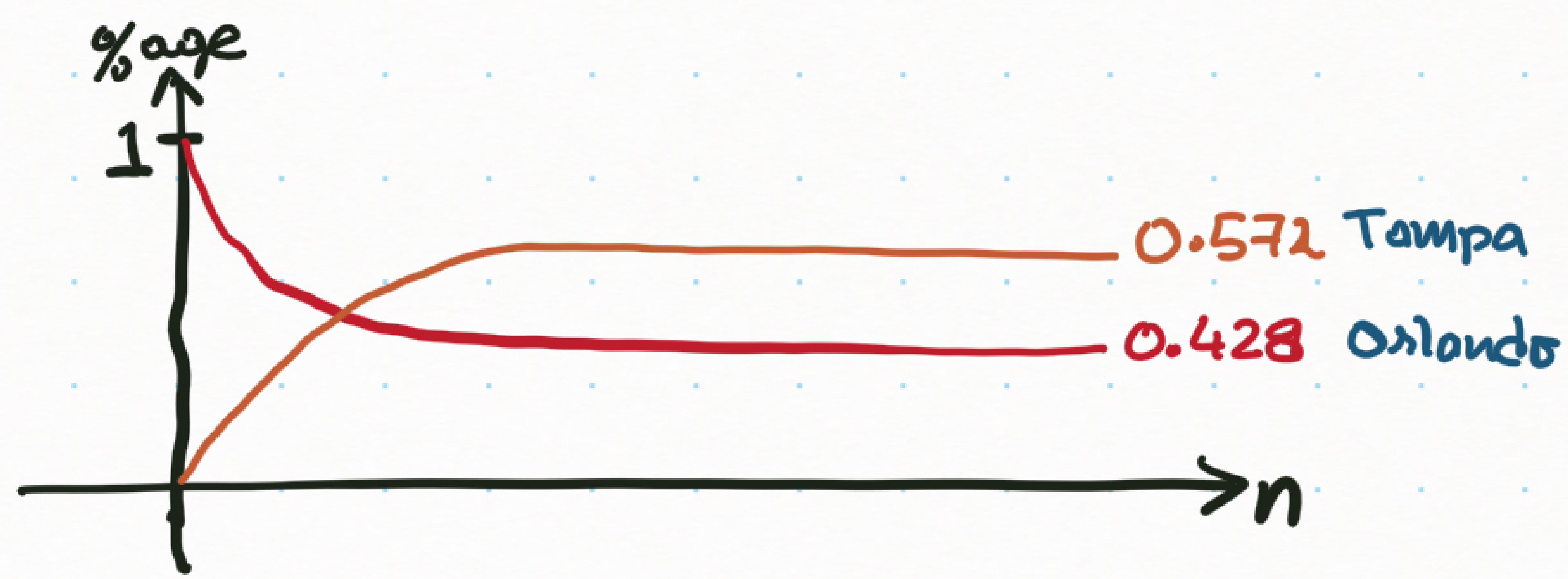
$q_{n+1} = 0.4\,p_n + 0.7\,q_n$

Numerically compute the values of $p_{n+1}$ & $q_{n+1}$ using the values of $p_n$ & $q_n$

for $n = 0, 1, 2, \dots$ ← Loop.

Use different $(p_0, q_0)$ (say $(1,0)$ or $(0,1)$ or $(0.5, 0.5)$)

| n | $p_n$ | $q_n$ |
|---|---|---|
| 0 | 1 | 0 |
| 1 | 0.6 | 0.4 |
| 2 | 0.48 | 0.52 |
| 3 | 0.444 | ⋮ |
| 4 | 0.4332 | ⋮ |
| 5 | 0.4299 | ⋮ |
| 6 | 0.42898 | ⋮ |
| 7 | 0.42869 | ⋮ |
| ⋮ | ⋮ | ⋮ |




In the long-term "steady state",
57.2% of cars will be in Tampa
& 42.8% of cars will be in Orlando.

## Car Rental Company

Based on $p_{n+1} = 0.6\,p_n + 0.3\,q_n$ for $n = 0, 1, 2, \dots$

$q_{n+1} = 0.4\,p_n + 0.7\,q_n$

Numerically compute the values of $p_{n+1}$ & $q_{n+1}$ using the values of $p_n$ & $q_n$ for $n = 0, 1, 2, \dots$, using the Transition matrix

$$\begin{bmatrix} p_{n+1} \\ q_{n+1} \end{bmatrix} = P^T \begin{bmatrix} p_n \\ q_n \end{bmatrix}$$

$$= P^T \left( P^T \begin{bmatrix} p_{n-1} \\ q_{n-1} \end{bmatrix} \right) = (P^T)^2 \begin{bmatrix} p_{n-1} \\ q_{n-1} \end{bmatrix} = (P^T)^2 \left( P^T \begin{bmatrix} p_{n-2} \\ q_{n-2} \end{bmatrix} \right) = \cdots$$

$$\cdots = (P^T)^{n+1} \begin{bmatrix} p_0 \\ q_0 \end{bmatrix}$$

$$\begin{bmatrix} p_n \\ q_n \end{bmatrix} = (P^T)^n \begin{bmatrix} p_0 \\ q_0 \end{bmatrix} \quad \text{for } n \geq 1$$

initial distribution

# Car Rental Company

Based on $p_{n+1} = 0.6\,p_n + 0.3\,q_n$ for $n = 0, 1, 2, \ldots$

$q_{n+1} = 0.4\,p_n + 0.7\,q_n$

Numerically compute the values of $p_{n+1}$ & $q_{n+1}$ using the values of $p_n$ & $q_n$ for $n = 0, 1, 2, \ldots$, using the Transition matrix

$$\begin{bmatrix} p_{n+1} \\ q_{n+1} \end{bmatrix} = P^T \begin{bmatrix} p_n \\ q_n \end{bmatrix}$$

$$= P^T \left( P^T \begin{bmatrix} p_{n-1} \\ q_{n-1} \end{bmatrix} \right) = (P^T)^2 \begin{bmatrix} p_{n-1} \\ q_{n-1} \end{bmatrix} = (P^T)^2 \left( P^T \begin{bmatrix} p_{n-2} \\ q_{n-2} \end{bmatrix} \right) = \cdots$$

$$\cdots = (P^T)^{n+1} \begin{bmatrix} p_0 \\ q_0 \end{bmatrix}$$

$$\begin{bmatrix} p_n \\ q_n \end{bmatrix} = (P^T)^n \begin{bmatrix} p_0 \\ q_0 \end{bmatrix} \quad \text{for } n \geq 1$$

initial distribution

$\boxed{\text{Compute } (P^T)^n \text{ for } n \geq 1 \ldots\ \text{\& multiply with } \begin{bmatrix} p_0 \\ q_0 \end{bmatrix}}$

We can estimate the "steady state" distribution

long-term (as $n \to \infty$)

aka "stationary" distribution

by :

→ Numerically computing $p_{n+1} = f_1(p_n, q_n)$, $q_{n+1} = f_2(p_n, q_n)$ for $n = 0, 1, 2, \ldots$

→ Computing powers of the transition matrix, $(P^T)^n$ for $n = 1, 2, \ldots$

→

We can estimate the "steady state" distribution

long-term (as $n \to \infty$)

aka "stationary" distribution

by :

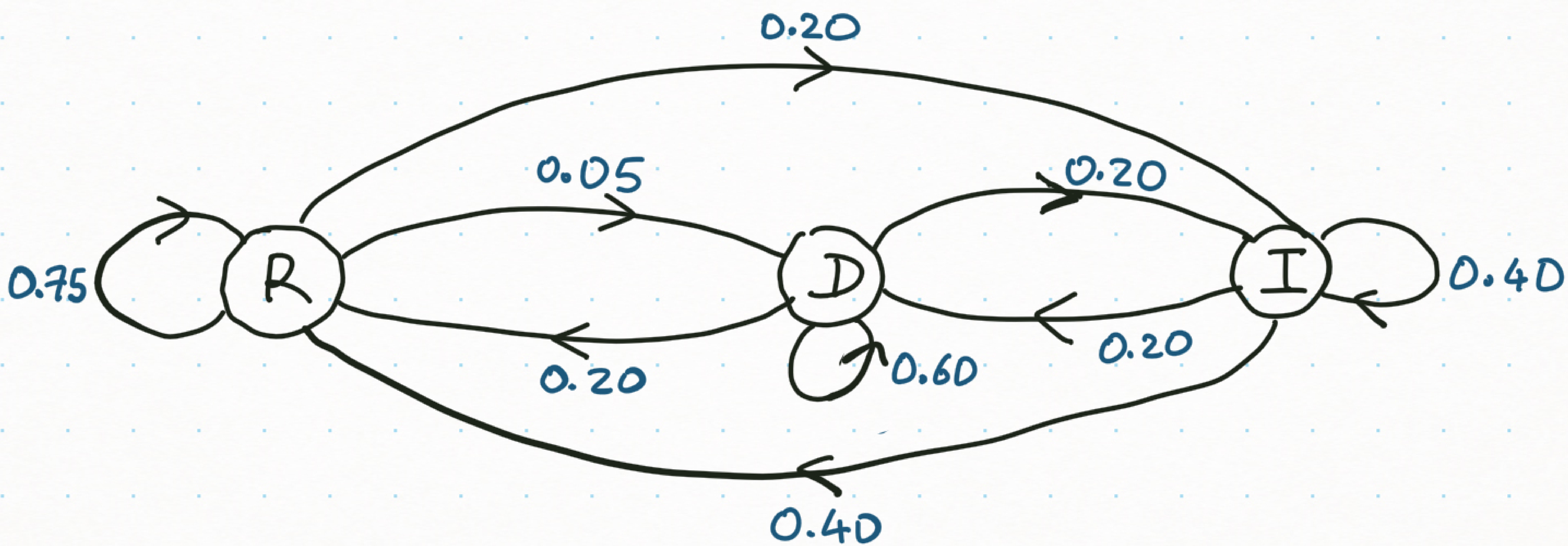
→ Numerically computing $p_{n+1} = f_1(p_n, q_n)$, $q_{n+1} = f_2(p_n, q_n)$ for $n = 0, 1, 2, \ldots$

→ computing powers of the transition matrix, $(P^T)^n$ for $n = 1, 2, \ldots$

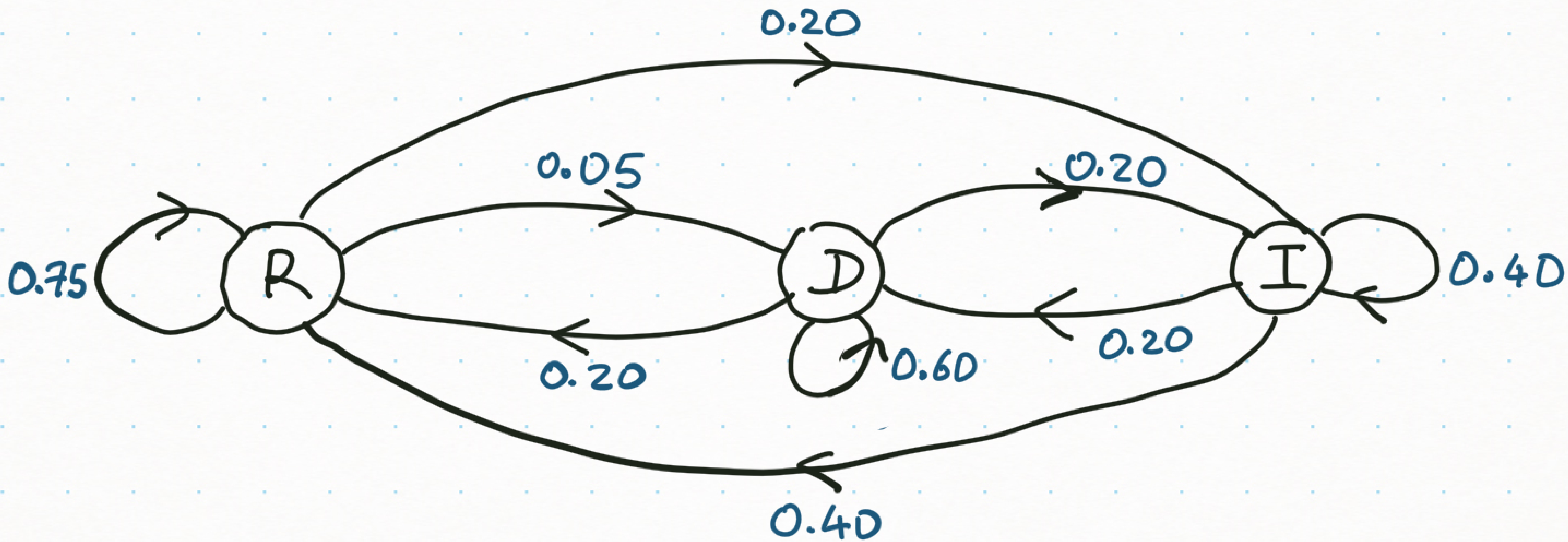
→ You will learn the stationary distribution can also be computed by $\vec{x} P = \vec{x}$ ($\vec{x}$ = eigenvector of $P$ corresponding to eigenvalue 1) whose solution vector $\vec{x}$ gives the stationary distribution for "nice" Markov Chains.

e.g. applying this to the rental car example will give $\vec{x} = (3/7, 4/7)$

# Voting Tendencies with a 3 party system



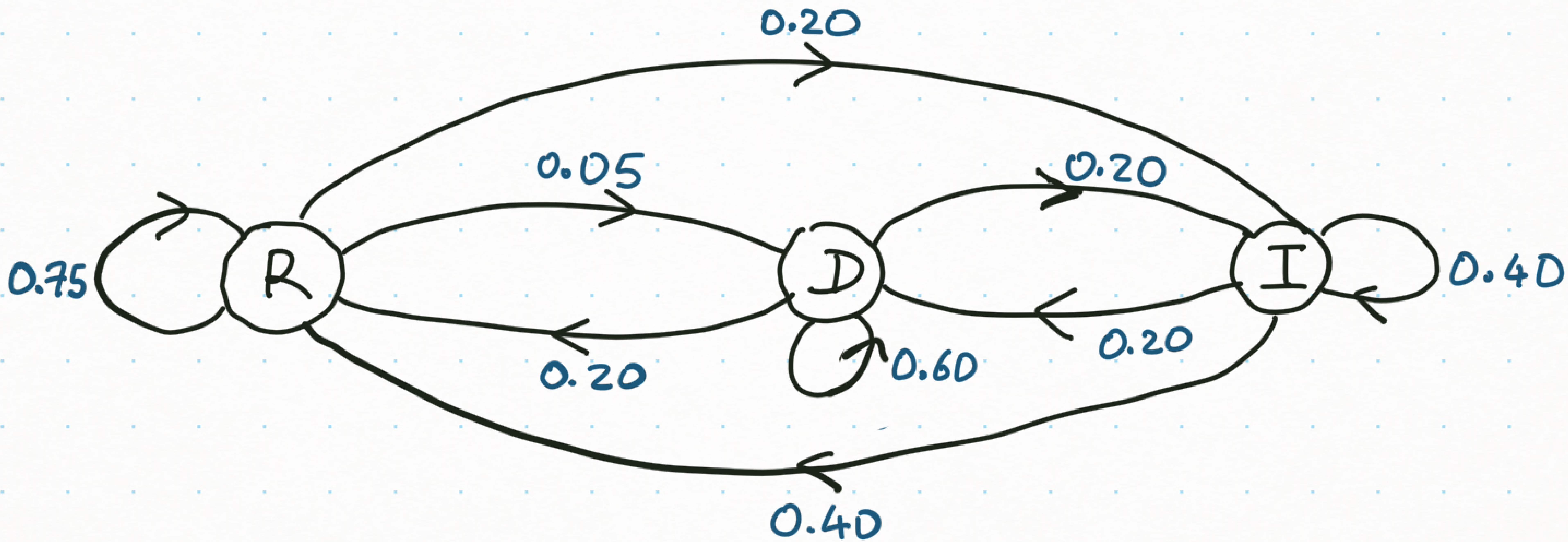
# Voting Tendencies with a 3 party system




## Transition Matrix

$$P = \begin{bmatrix} 0.75 & 0.05 & 0.20 \\ 0.20 & 0.60 & 0.20 \\ 0.40 & 0.20 & 0.40 \end{bmatrix}$$

# Voting Tendencies with a 3 party system




## Transition Matrix

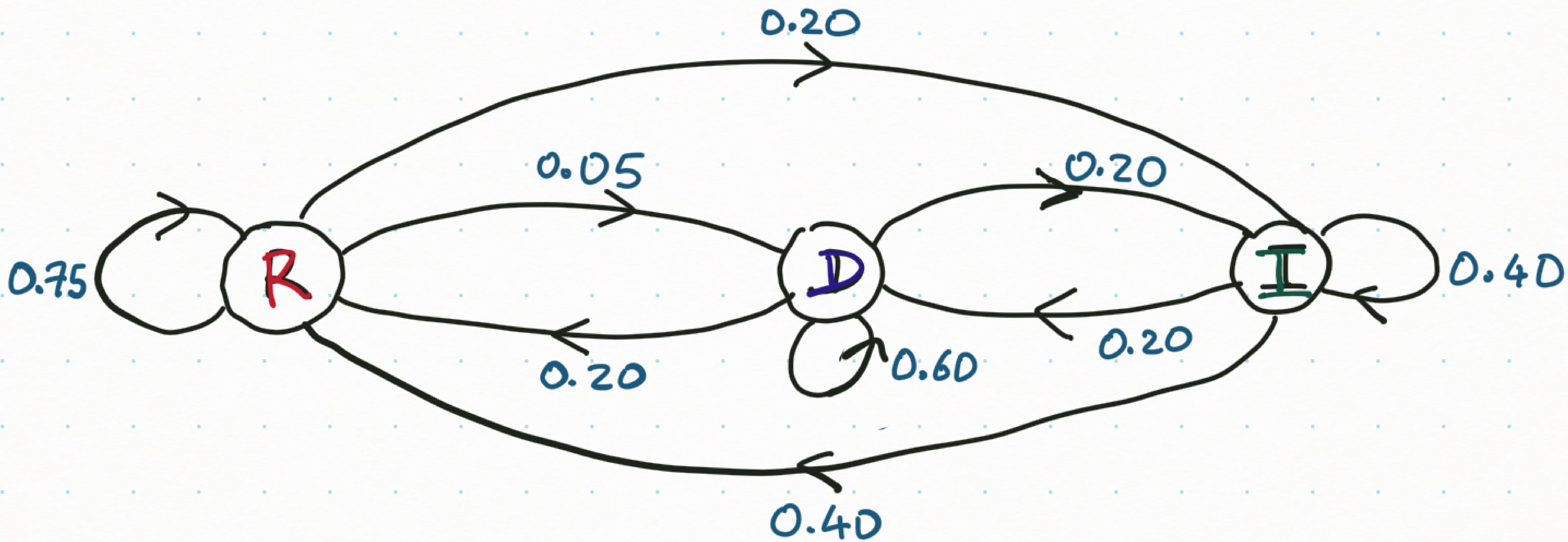
$$P = \begin{bmatrix} 0.75 & 0.05 & 0.20 \\ 0.20 & 0.60 & 0.20 \\ 0.40 & 0.20 & 0.40 \end{bmatrix}$$

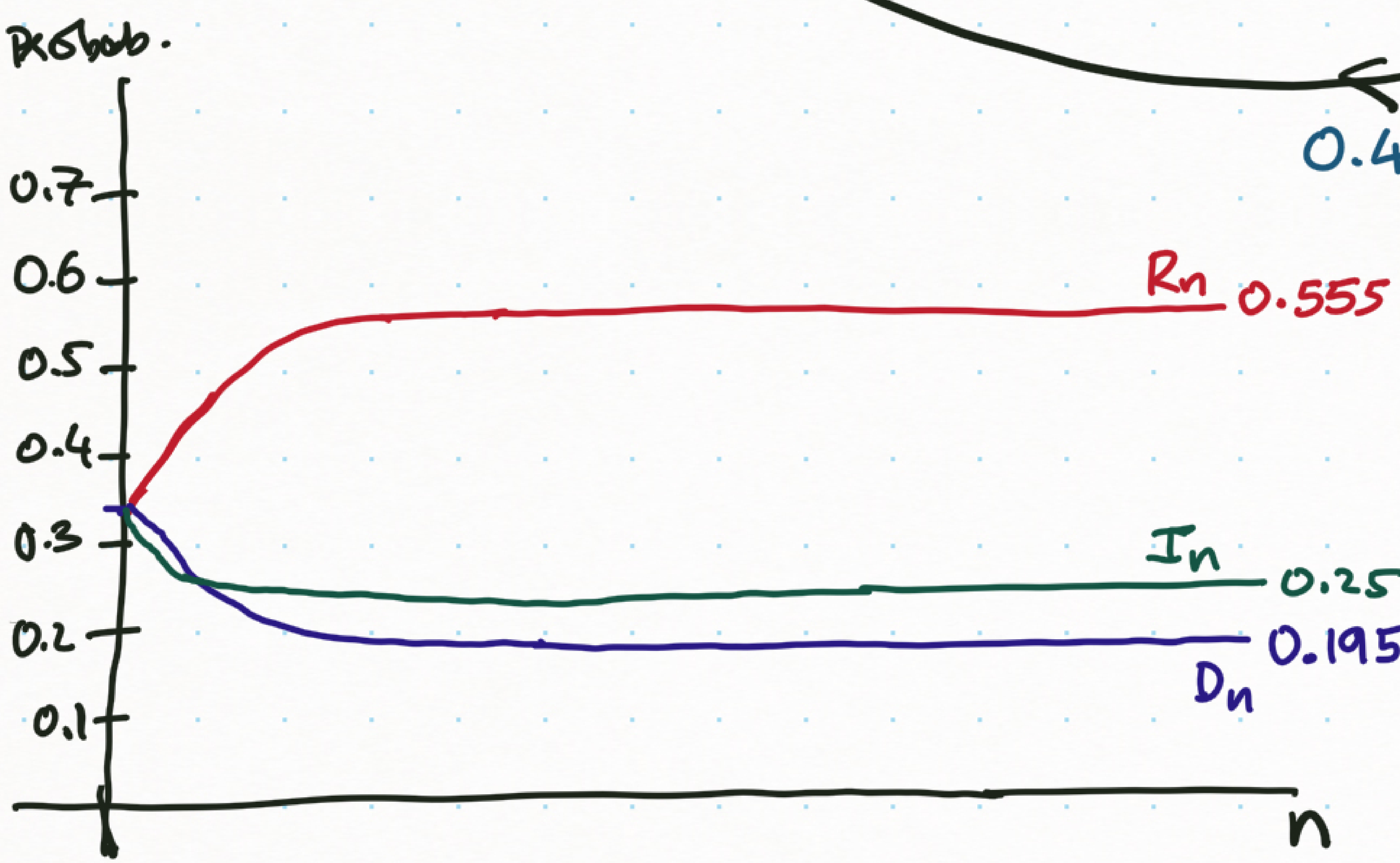
$$R_{n+1} = 0.75R_n + 0.20D_n + 0.40I_n$$

$$D_{n+1} = 0.05R_n + 0.60D_n + 0.20I_n$$

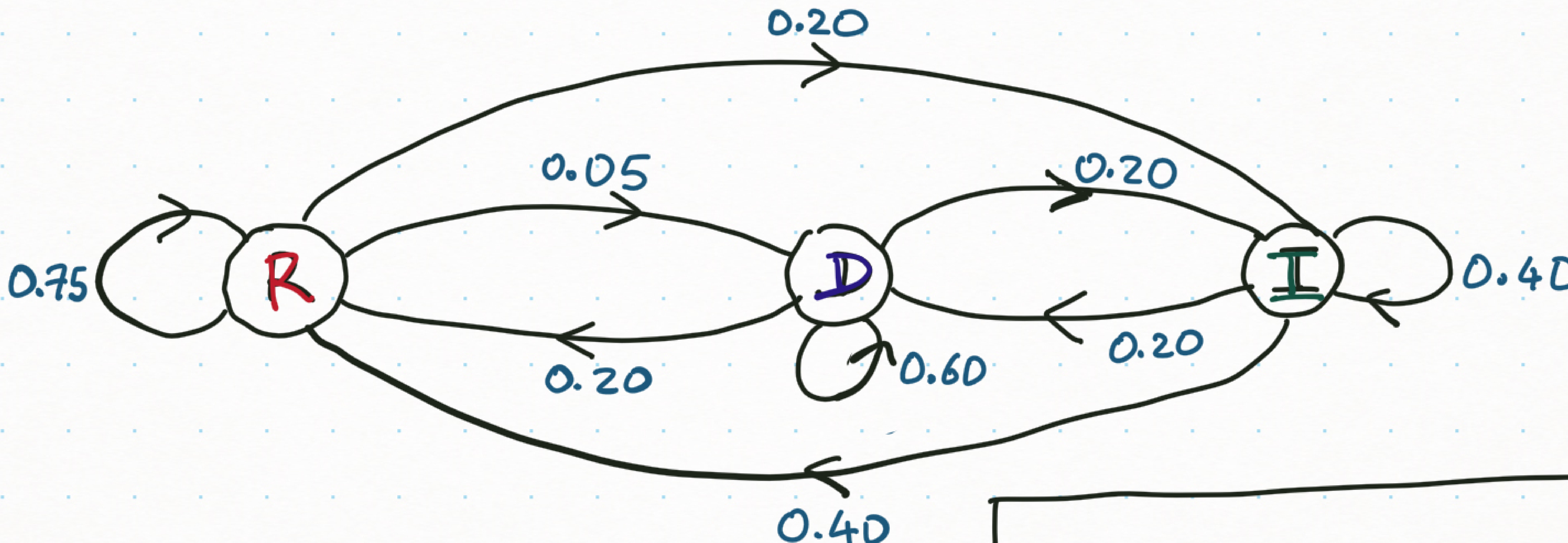
$$I_{n+1} = 0.20R_n + 0.20D_n + 0.40I_n$$

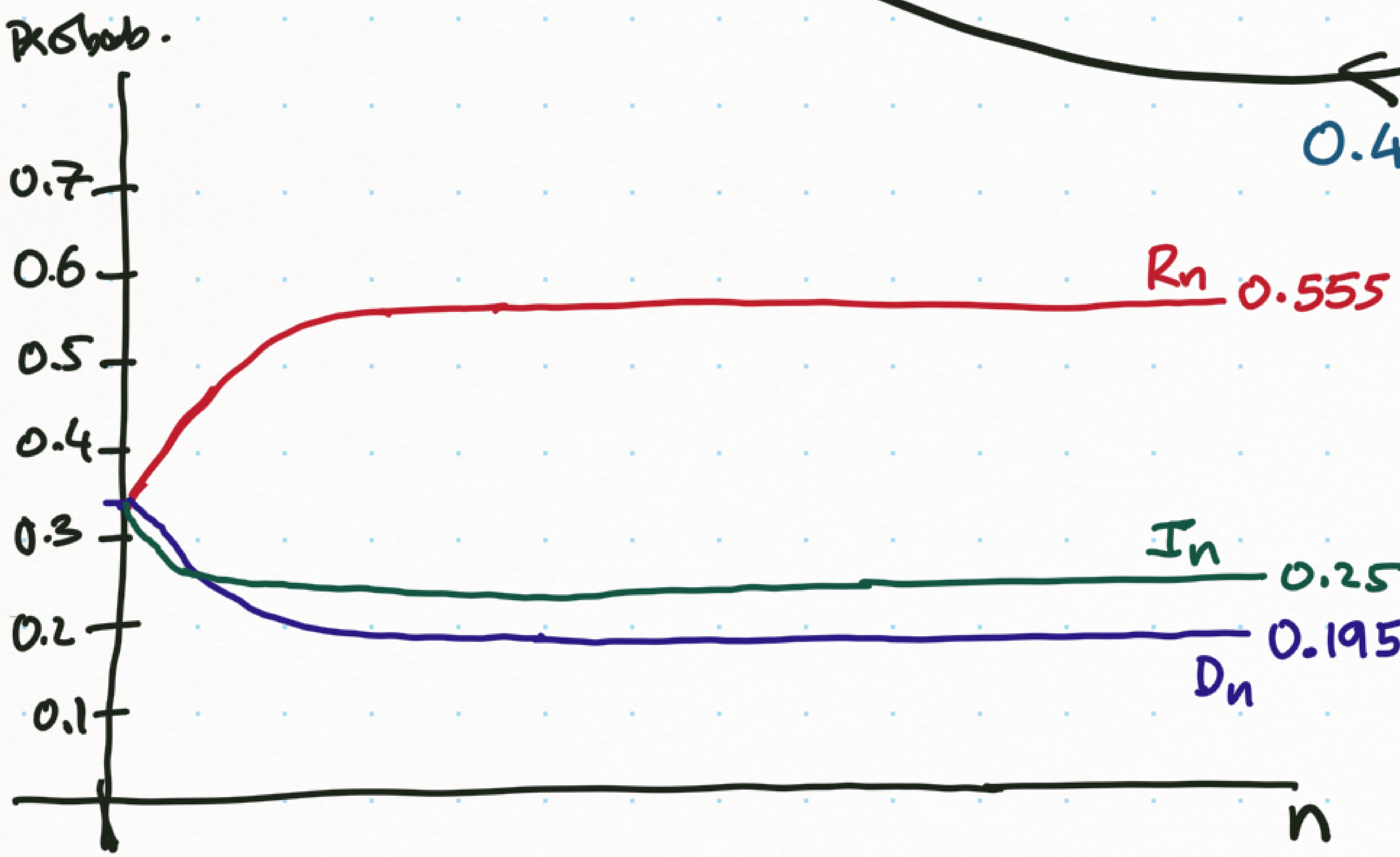
# Voting Tendencies with a 3 party system



# Voting Tendencies with a 3 party system




Underlying assumptions?

Criticisms?

Improvements?

# Sergey Brin & Larry Page's model for web browsing

(1998 : introducing page rank algorithm used by Google)

WWW is a <u>directed graph</u> of webpages with weblinks connecting them.

# Sergey Brin & Larry Page's model for web browsing

(1998 : introducing page rank algorithm used by Google)

WWW is a directed graph of webpages with weblinks connecting them.

A "random web surfer" starts from an arbitrary web page and then keeps on clicking links without using the "back" button until bored or found what they were looking for .......

# Sergey Brin & Larry Page's model for web browsing

(1998 : introducing page rank algorithm used by Google)

WWW is a directed graph of webpages with weblinks connecting them.

A "random web surfer" starts from an arbitrary webpage and then keeps on clicking links without using the "back" button until bored (or found what they were looking for........

Next "click" only depends on current webpage and can only go to a webpage linked from it.

But how to model the probability of clicking a particular webpage?

Our websurfing model can be described as a Markov chain:

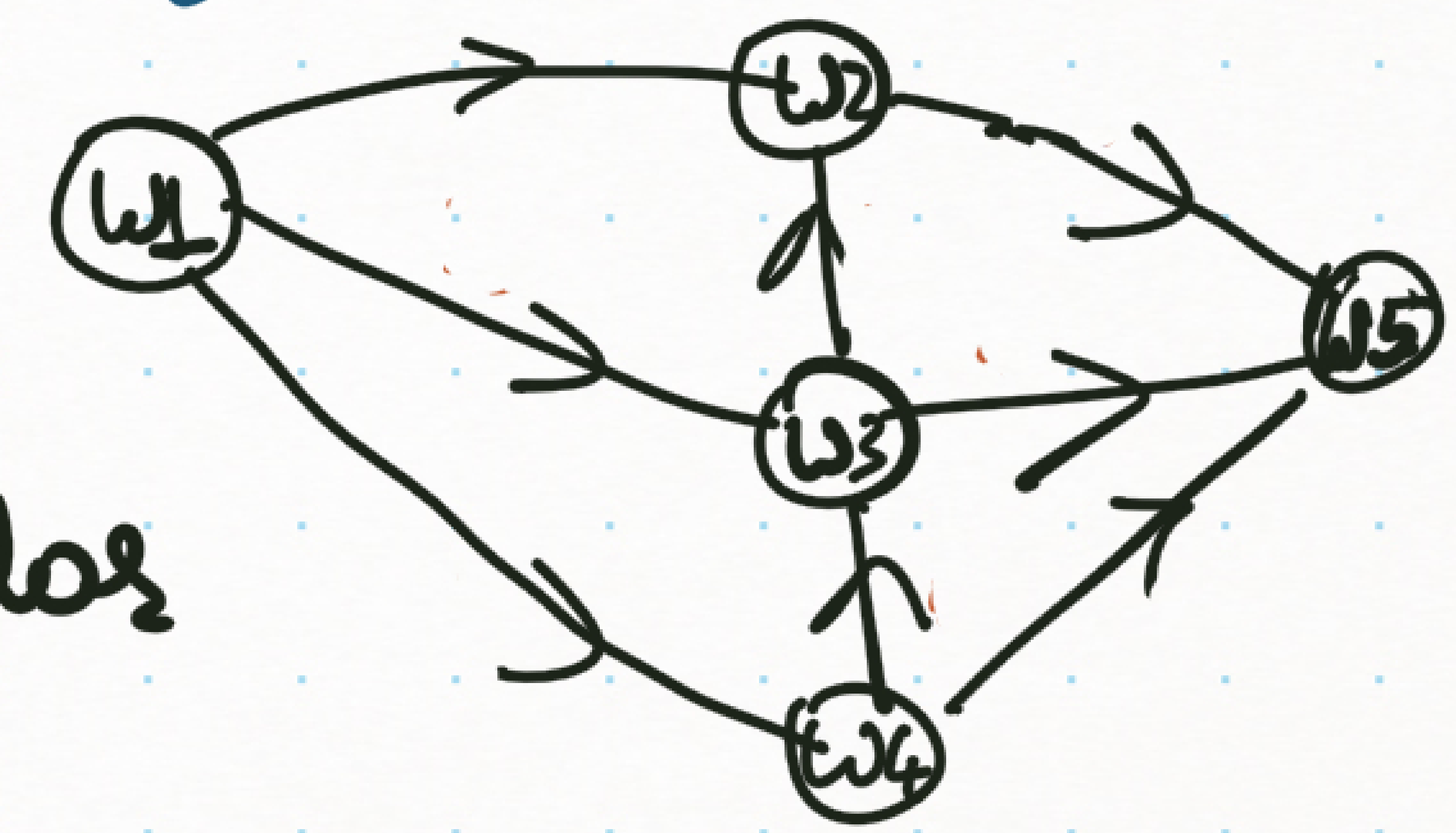
States are the webpages.

From each state (webpage) we are equally likely to take any weblink out of it.

Our websurfing model can be described as a Markov Chain:

States are the webpages.

From each state (webpage) we are equally likely to take any weblink out of it. e.g.

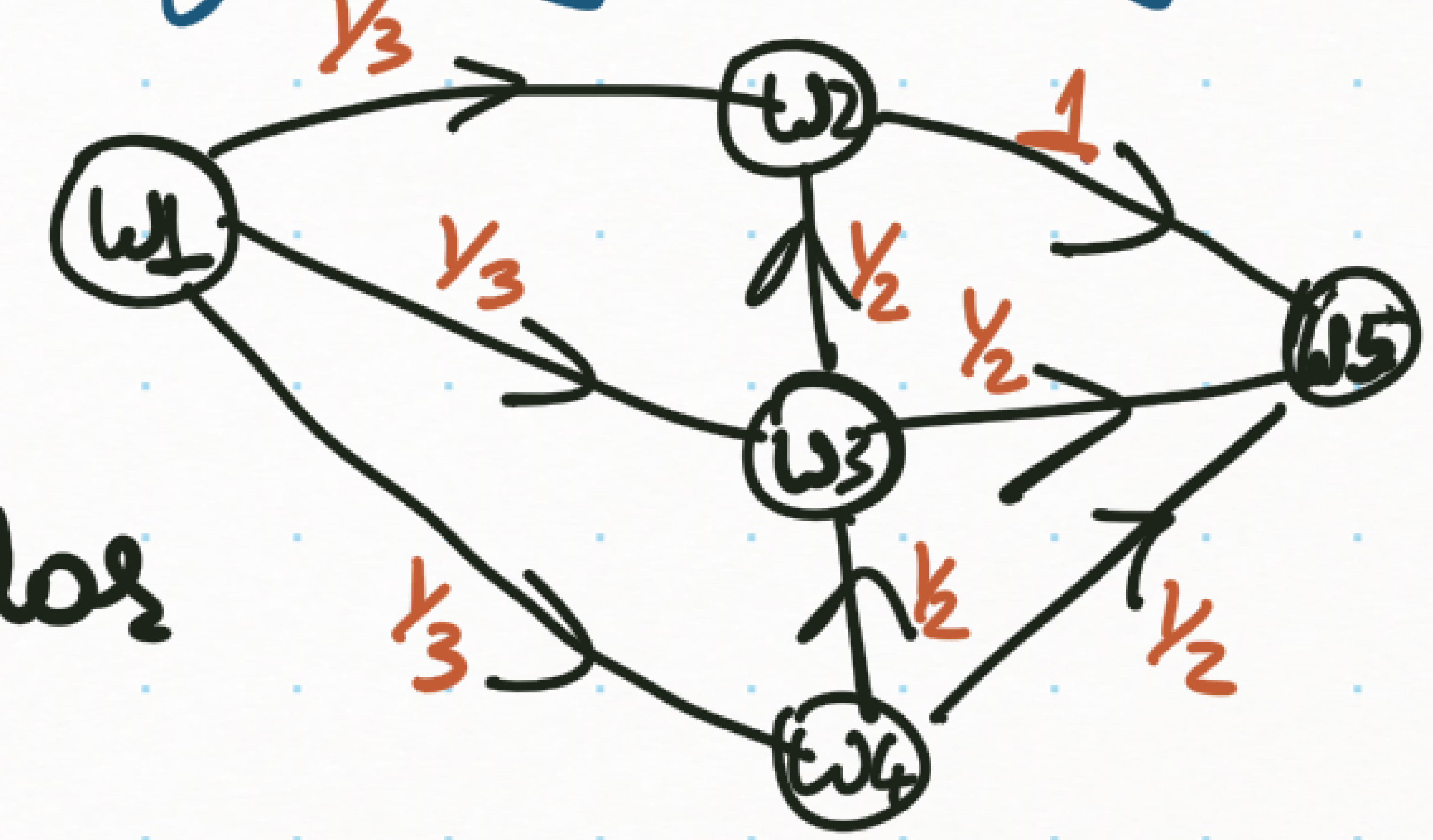



If a webpage W has k outlinks then probability of clicking a particular link is $1/k$.

Our websurfing model can be described as a Markov Chain:

States are the webpages.

From each state (webpage) we are equally likely to take any weblink out of it. e.g.

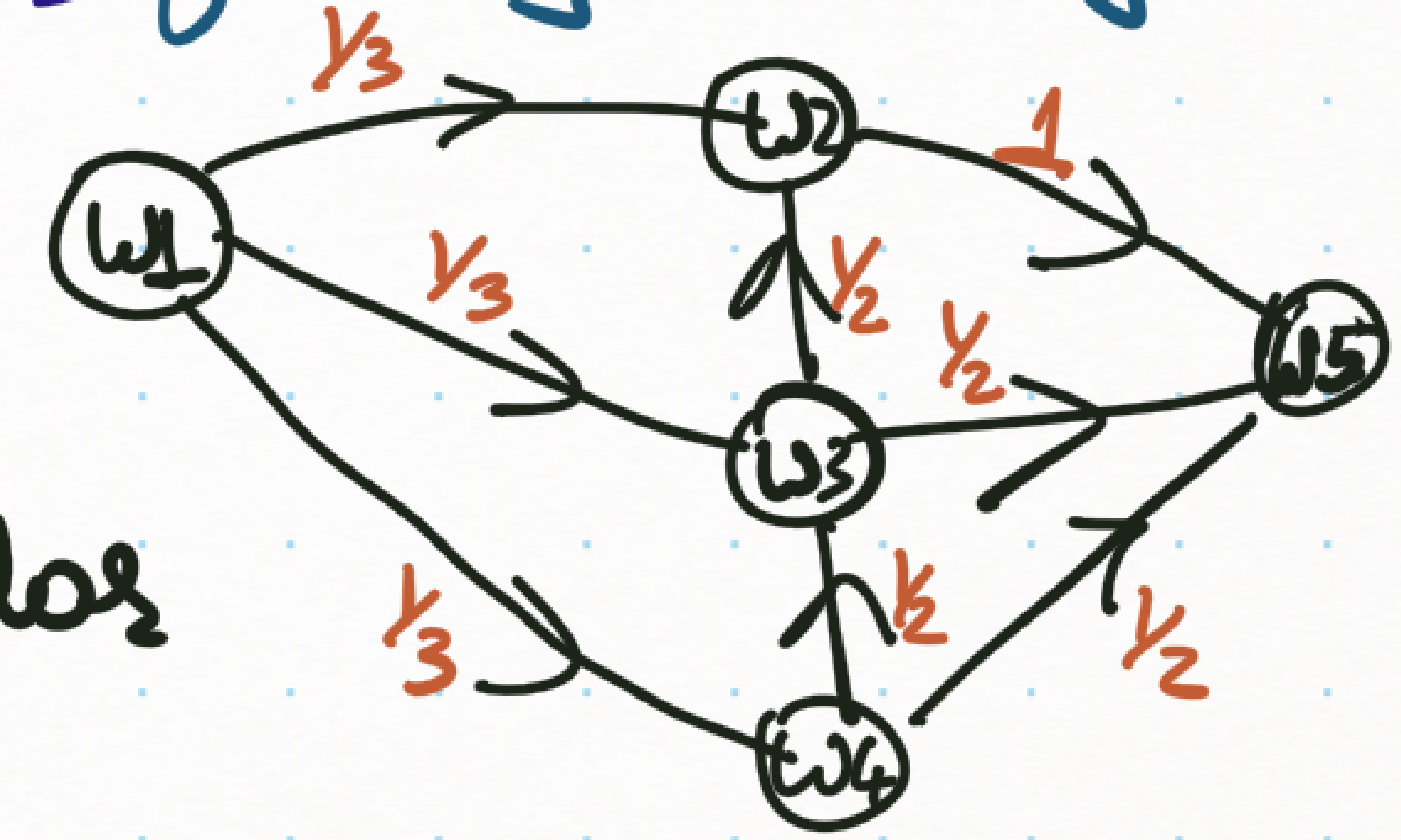



If a webpage W has k outlinks then probability of clicking a particular link is $1/k$.

Our websurfing model can be described as a Markov Chain:

States are the webpages.

From each state (webpage) we are equally likely to take any weblink out of it. e.g.




If a webpage $W$ has $k$ outlinks then probability of clicking a particular link is $1/k$.

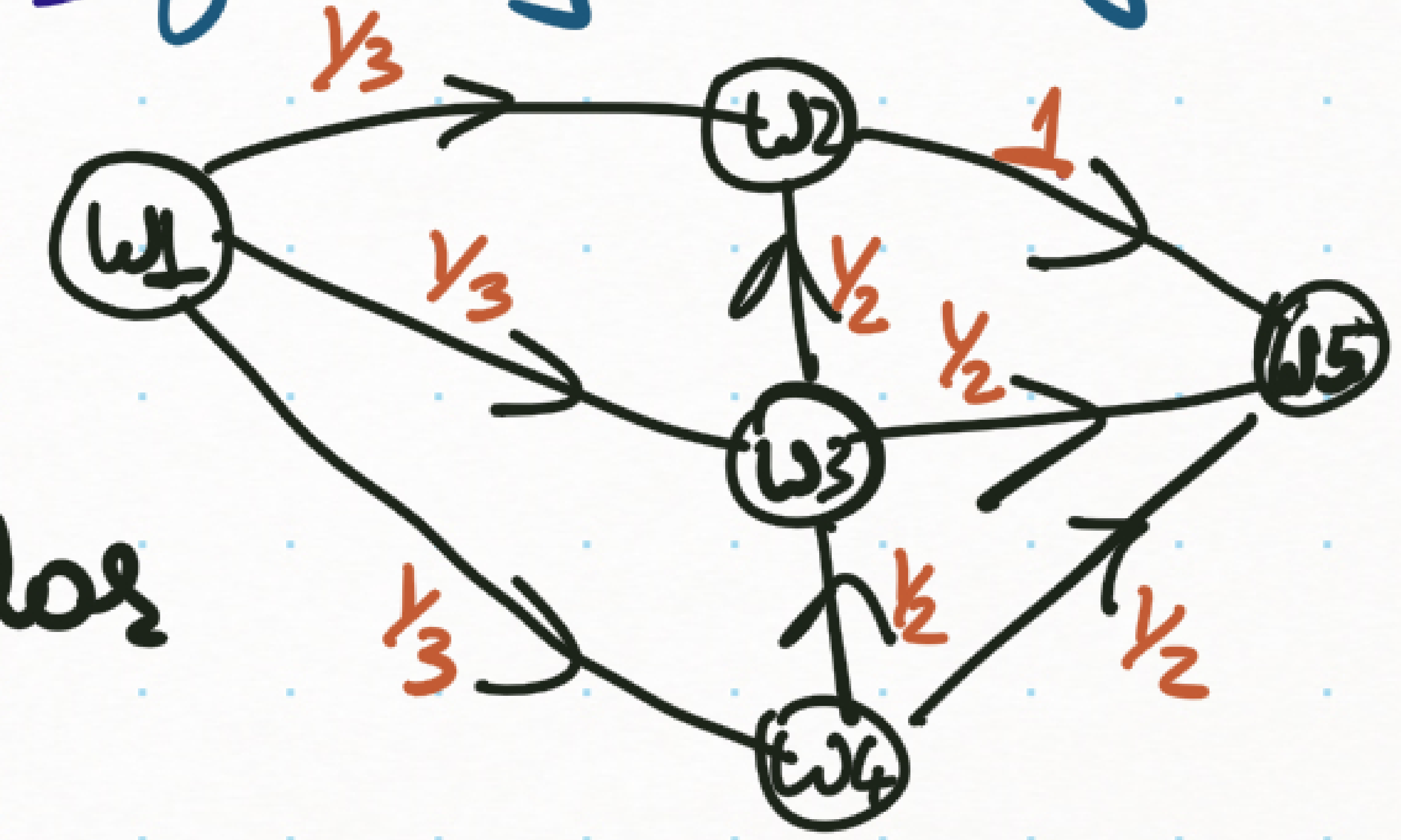
Run this Markov Chain for a long time until it reaches its "stationary distribution" $\boxed{P\vec{x} = \vec{x}}$

e.g. $\vec{x} = (P_1, P_2, P_3, P_4, P_5)$ where $P_i$ = probability of being on webpage $W_i$ in the long-term

Our websurfing model can be described as a Markov Chain:

States are the webpages.

From each state (webpage) we are equally likely to take any weblink out of it. e.g.




If a webpage W has k outlinks then probability of clicking a particular link is $1/k$.

Run this Markov Chain for a long time until it reaches its "stationary distribution" $\boxed{P\vec{x} = \vec{x}}$

e.g. $\vec{x} = (P_1, P_2, P_3, P_4, P_5)$ where $P_i$ = probability of being on webpage $W_i$ in the long-term

(Simplified) Pagerank of webpage $\equiv$ Probability of being on the webpage in the stationary distribution

<u>(Actual) pagerank</u> is based on adjusted probabilities of outclick:

$n = \#$ webpages , $d_i^+ =$ out-degree of webpage $i$

Original Transition Matrix $A = [a_{ij}]_{n \times n}$

where $a_{ij} = \begin{cases} 1/d_i^+ & \text{if } i \to j \\ 0 & \text{otherwise} \end{cases}$

is modifed to $P = \alpha A + (1-\alpha) B$ where $B = [1/n]_{n \times n}$

for some fixed $\alpha \in (0,1)$

e.g. $\alpha = 0.85$

(Actual) pagerank is based on adjusted probabilities of outclick:

$n = \#$ webpages, $d_i^+ =$ out-degree of webpage $i$

Original Transition Matrix $A = [a_{ij}]_{n \times n}$

where $a_{ij} = \begin{cases} 1/d_i^+ & \text{if } i \to j \\ 0 & \text{otherwise} \end{cases}$

is modified to $\boxed{P = \alpha A + (1-\alpha) B}$ where $B = [1/n]_{n \times n}$

for some fixed $\alpha \in (0,1)$
e.g. $\alpha = 0.85$

Now, compute the eigenvector of $P$ corresponding to eigenvalue 1 $\quad P\vec{x} = \vec{x} \quad$ efficiently.

(Actual) PageRank is based on adjusted probabilities of outclick:

$n = \#$ webpages, $d_i^+ =$ out-degree of webpage $i$

Original Transition Matrix $A = [a_{ij}]_{n \times n}$

where $a_{ij} = \begin{cases} 1/d_i^+ & \text{if } i \to j \\ 0 & \text{otherwise} \end{cases}$

is modified to $\boxed{P = \alpha A + (1-\alpha) B}$ where $B = [1/n]_{n \times n}$

for some fixed $\alpha \in (0,1)$
e.g. $\alpha = 0.85$

PR = PageRank

$$PR(i) = (1-\alpha) + \alpha \sum_{j : j \in N^+(i)} PR(j) \frac{1}{d^+(j)}, \ \ldots\ldots$$

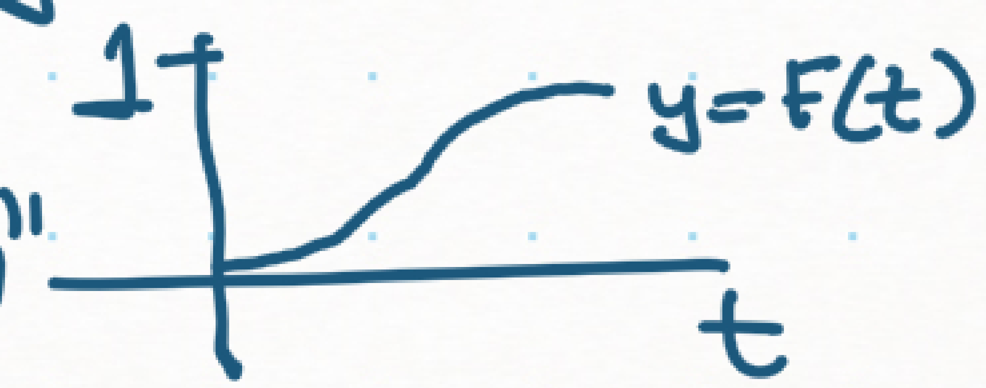
# Modeling Component & System Reliability

**Reliability** of a component is the probability that it will not fail for a certain time period.

$R(t)$ = reliability (upto time $t$) $= 1 - F(t)$

where $F(t)$ = probability of failing before time $t$

"cumulative probability distribution"



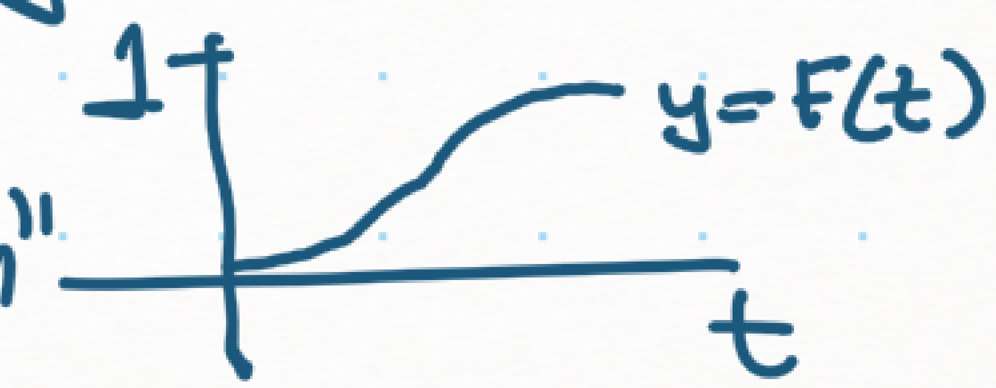
# Modeling Component & System Reliability

**Reliability** of a component is the probability that it will not fail for a certain time period.

$R(t)$ = reliability (upto time $t$) $= 1 - F(t)$

where $F(t)$ = probability of failing before time $t$

"cumulative probability distribution"




We consider systems built from components with independent failure / reliability.

## Series System

requires all of its components to work — if one component fails then the whole system fails.

# Series System

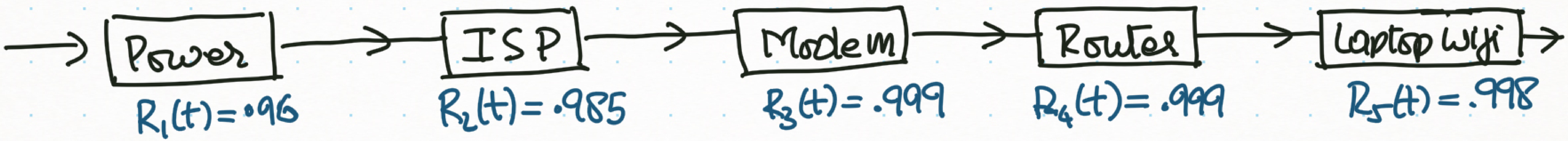
requires all of its components to work — if one component fails then the whole system fails.

$\rightarrow$ Power $\rightarrow$ ISP $\rightarrow$ Modem $\rightarrow$ Router $\rightarrow$ Laptop wifi $\rightarrow$

$R_1(t) = .96$ $\quad R_2(t) = .985$ $\quad R_3(t) = .999$ $\quad R_4(t) = .999$ $\quad R_5(t) = .998$

# Series System

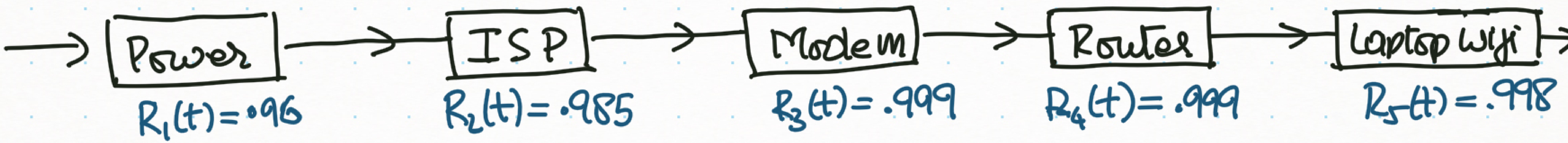
requires all of its components to work — if one component fails then the whole system fails.




System Reliability $= R_1(t)\,R_2(t)\ldots R_5(t)\ldots$

## Series System

requires all of its components to work — if one component fails then the whole system fails.

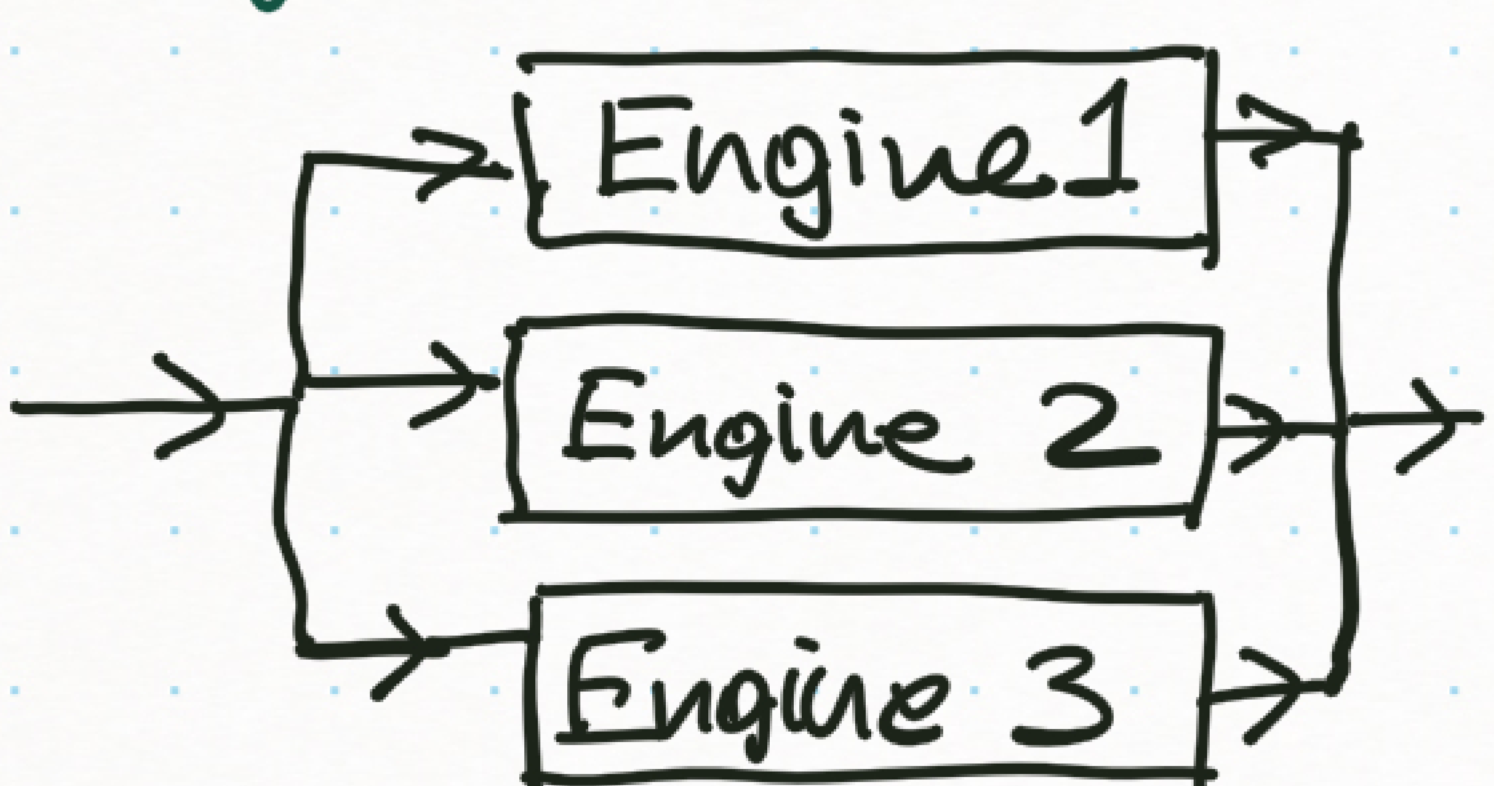



System Reliability $= R_1(t)\,R_2(t) \ldots R_5(t) \ldots$

## Parallel System

requires at least one of its components to work — if all components fail then the whole system fails.

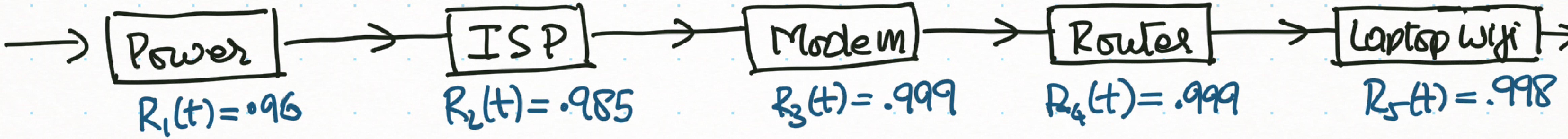
$$F(t) = F_1(t)\,F_2(t)\,F_3(t) \ldots\ldots$$




Airplane

## Series System

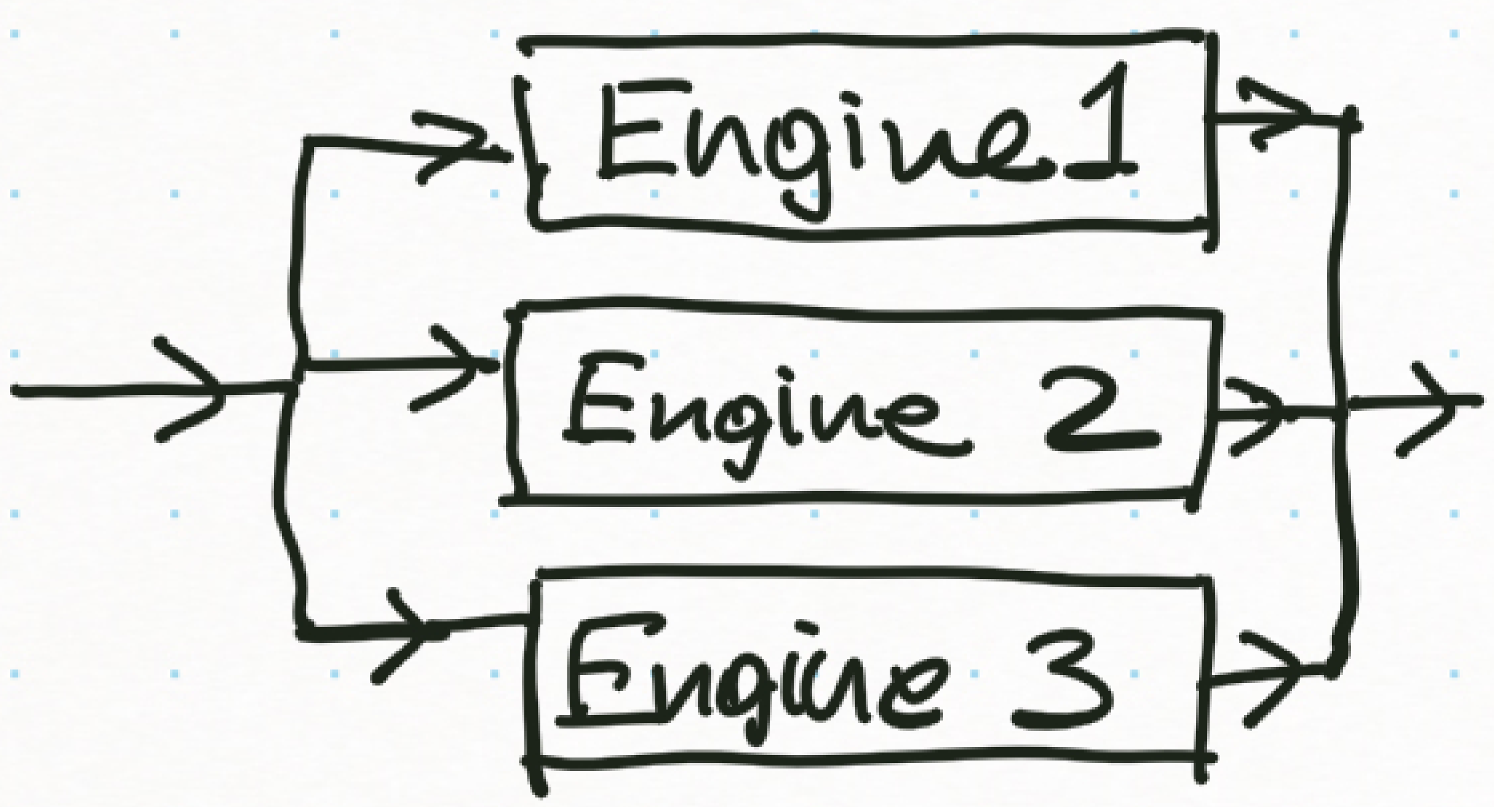
requires all of its components to work — if one component fails then the whole system fails.




System Reliability $= R_1(t)\,R_2(t) \ldots R_5(t) \ldots$

## Parallel System

requires at least one of its components to work — if all components fail then the whole system fails.




Airplane

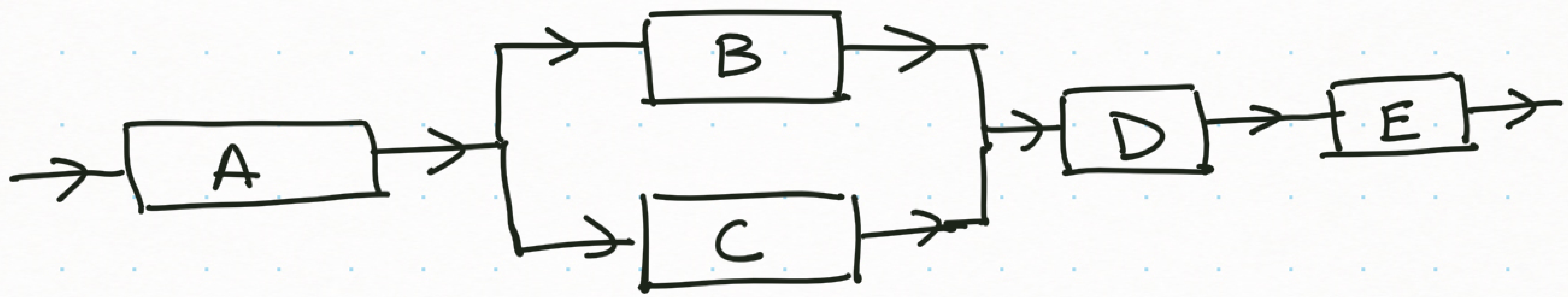
$$F(t) = F_1(t)\,F_2(t)\,F_3(t) \ldots\ldots$$

System Reliability

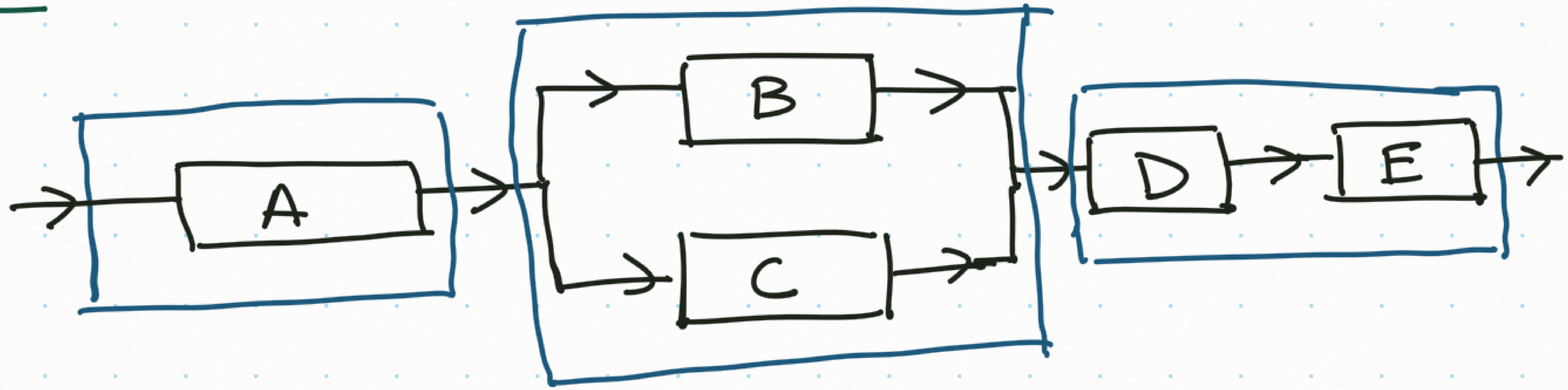
$$= 1 - F(t) = 1 - \left(F_1(t)\,F_2(t) \ldots F_k(t)\right)$$

$$= 1 - \left((1 - R_1(t))(1 - R_2(t)) \ldots (1 - R_k(t))\right)$$

example



example




$$\text{System Reliability} = R_A(t)\left[R_{BC}(t)\right] R_D(t)\, R_E(t)$$

$$= R_A(t)\left[1-\left(\left(1-R_B(t)\right)\left(1-R_C(t)\right)\right)\right] R_D(t)\, R_E(t)$$

# Linear Regression Model

Given data $(x_i, y_i)$, $i = 1, \dots, m$

<u>Model</u> $y = ax + b$ for some $a, b$ that minimizes $\sum_{i=1}^{m} (y_i - ax_i)^2$

Usual <u>least sq. criterion</u>
under the assumption
one observation $y_i$
for each $x_i$ (ind. var.)

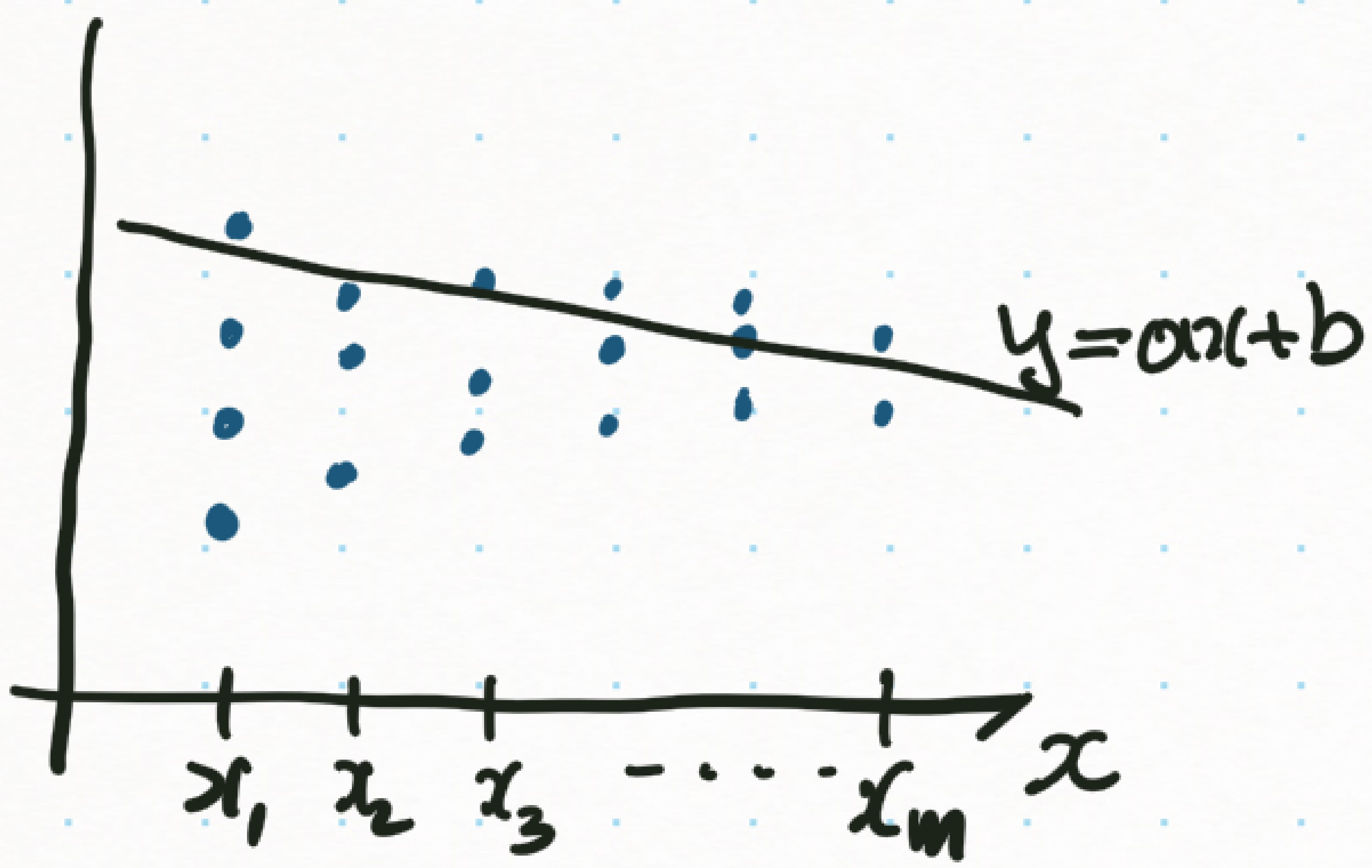
# Linear Regression Model

Given data $(x_i, y_i)$, $i = 1, \cdots, m$

<u>Model</u> $y = ax + b$ for some $a, b$ that minimizes $\sum_{i=1}^{m} (y_i - ax_i)^2$

Usual <u>least sq. criterion</u> under the assumption one observation $y_i$ for each $x_i$ (ind. var.)

Linear regression handles LSC when we allow multiple observations per $x_i$.




Repeated trials with random measurement errors.

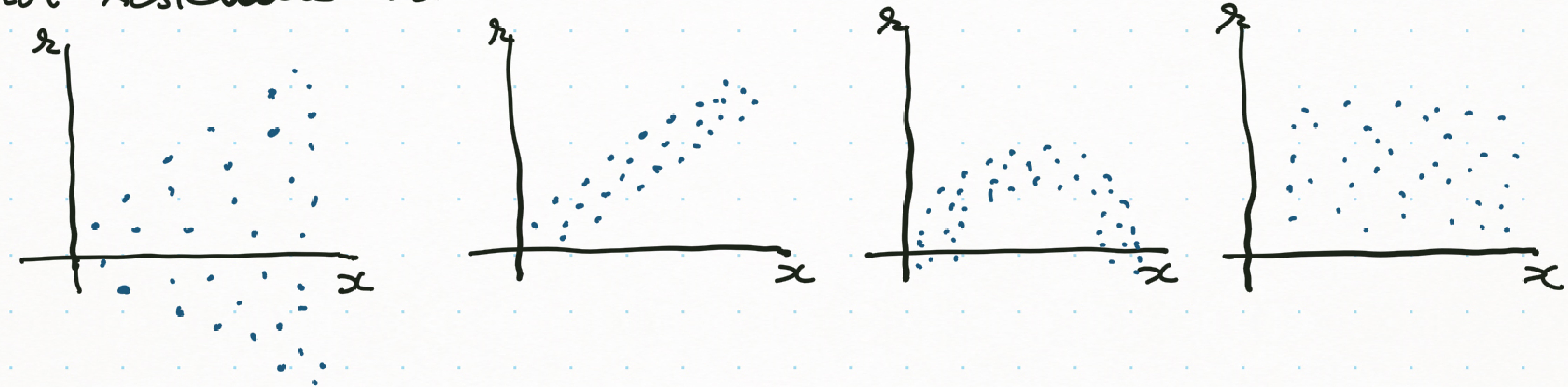
we can do the usual LSC setup $y = ax + b$ & derive normal eqn.s for $a$ & $b$.

But there are <u>more statistical measures of fit.</u>

## The residuals $r_i = y_i - (ax_i + b)$, $i = 1, \ldots, m$

If residuals have a trend, then our model is missing an important factor.

Plot residuals vs. ind. variable.



# Error Sum of Squares

$$SSE = \sum_{i=1}^{m} [y_i - (ax_i + b)]^2$$

"variation around the linear model"

## Error Sum of Squares

$$SSE = \sum_{i=1}^{m} [y_i - (ax_i + b)]^2$$

"variation around the linear model"

## Total Corrected sum of squares

$$SST = \sum_{i=1}^{m} [y_i - \bar{y}]^2$$

"variation around $\underline{\text{the average value}}$"

constant approximation model

where $\bar{y} = \frac{1}{m} \sum_{i=1}^{m} y_i$

"average value of the linear model over the range of data".

Note that $SST \geq SSE$

## Error Sum of Squares

$$SSE = \sum_{i=1}^{m} [y_i - (ax_i + b)]^2$$

"variation around the linear model"

## Total Corrected sum of squares

$$SST = \sum_{i=1}^{m} [y_i - \bar{y}]^2$$

"variation around the average value"

constant approximation model

where $\bar{y} = \frac{1}{m} \sum_{i=1}^{m} y_i$

"average value of the linear model over the range of data".

## Note that $SST \geq SSE$

## Regression of sum squares,

$SSR = SST - SSE$

"how much variation in y-values is captured by our linear model"

## Error Sum of Squares

$$SSE = \sum_{i=1}^{m} [y_i - (ax_i + b)]^2$$

"variation around the linear model"

## Total Corrected sum of squares

$$SST = \sum_{i=1}^{m} [y_i - \bar{y}]^2$$

"variation around the average value"
constant approximation model

where $\bar{y} = \frac{1}{m} \sum_{i=1}^{m} y_i$

"average value of the linear model over the range of data".

Note that $SST \geq SSE$

Regression of sum squares, $SSR = SST - SSE$

"how much variation in y-values is captured by our linear model"

Coefficient of Regression, $R^2 = 1 - \frac{SSE}{SST}$ "what proportion of variation is captured by our linear model"

Note $R^2$ is unit-less & $R^2 \leq 1$. R close to 1 indicates a good fit.