Unbiased estimators
Markov chains and PageRank

CSE 312 Spring 26

Lecture 25

General Recipe (single parameter)

1. **Input** Given n i.i.d. samples $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ from parametric model with parameter θ.
2. **Likelihood** Define your likelihood $\mathcal{L}\,(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n;\ \theta)$.
   - For discrete $\mathcal{L}(x_1, \ldots, x_n\ ;\ \theta) = \prod_{i=1}^{n} P(x_i\ ; \theta)$
   - For continuous $\mathcal{L}(x_1, \ldots, x_n\ ;\ \theta) = \prod_{i=1}^{n} f(x_i\ ;\ \theta)$
3. **Log** Compute $\ln \mathcal{L}(x_1, \ldots, x_n\ ;\ \theta)$
4. **Differentiate** Compute $\frac{\mathrm{d}}{\mathrm{d}\theta} \ln \mathcal{L}(x_1, \ldots, x_n\ ;\ \theta\ )$
5. **Solve for** $\hat{\theta}$ by setting derivative to 0 and solving for max.

Do a second derivative test to verify it is a maximum, but we won't ask you to do that in CSE 312.

Definition of Estimator

e.g. $\theta =$ parameter of an Exponential distin

θ: quantity we're trying to estimate

think of these as i.i.d. instances of $X \sim \text{Exp}(\theta)$

This is a constant

$X_1, X_2, \ldots, X_n$: i.i.d. data

$\hat{\theta}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$: estimation of θ based on specific instantiation of the data

This is a r.v. because it's a function of r.v.s

$\hat{\theta}(X_1, X_2, \ldots, X_n)$: estimator of the unknown θ

Sometimes just write $\hat{\theta}$

MLE for pink jelly beans

1000 jelly beans total

X = # pink jelly beans in sample

Q: What is the likelihood function $\boldsymbol{P}\{X = x \mid \theta\}$?

$$\boldsymbol{P}(\mathrm{X} = \mathrm{x}; \theta) = \binom{n}{x}\left(\frac{\theta}{1000}\right)^{x} \cdot \left(1 - \frac{\theta}{1000}\right)^{n-x}$$

Q: What is $\hat{\theta}_{ML}(X = x) = \arg\max_{\theta} \boldsymbol{P}(\mathrm{X} = \mathrm{x}; \theta)$?

$$\hat{\theta}_{ML}(X = x) = \arg\max_{\theta} \boldsymbol{P}(\mathrm{X} = \mathrm{x}; \theta) = \frac{1000x}{n}$$

This holds $\forall x$

$$\rightarrow \boxed{\hat{\theta}_{ML}(X) = \frac{1000X}{n}}$$

MLE estimation for normal distribution

$$\hat{\theta}_{\mu}\,(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

These hold for all $x_1, \dots, x_n$

$$\hat{\theta}_{\sigma^2}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{x_1 + x_2 + \cdots + x_n}{n}\right)^2$$

➔ $$\hat{\theta}_{\mu}\,(X_1, X_2, \dots, X_n) = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

➔ $$\hat{\theta}_{\sigma^2}(X_1, X_2, \dots, X_n) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{X_1 + X_2 + \cdots + X_n}{n}\right)^2$$

When is an estimator good?

Distribution $P(x;\theta)$ → Independent samples $X_1, \ldots, X_n$ from $P(x;\theta)$ → **Estimation Algorithm** → $\hat{\theta}_n$ (Parameter estimate)

θ = unknown parameter

Definition. An estimator of parameter θ is an **unbiased estimator** if

$$\mathbb{E}\left[\hat{\theta}_n(X_1, \ldots, X_n)\right] = \theta.$$

Note: This expectation is over the samples $X_1, \ldots, X_n$

MLE for pink jelly beans

$X \sim \mathrm{Bin}\left(n, \frac{\theta}{1000}\right)$

Q: What is the likelihood function $\boldsymbol{P}\{X = x \mid \theta\}$?

$$\boldsymbol{P}(\mathrm{X} = \mathrm{x}; \theta) = \binom{n}{x}\left(\frac{\theta}{1000}\right)^{x} \cdot \left(1 - \frac{\theta}{1000}\right)^{n-x}$$

1000 jelly beans total

X = # pink jelly beans in sample

Q: What is $\hat{\theta}_{ML}(X = x) = \arg\max_{\theta} \boldsymbol{P}(\mathrm{X} = \mathrm{x}; \theta)$?

$$\hat{\theta}_{ML}(X = x) = \arg\max_{\theta} \boldsymbol{P}(\mathrm{X} = \mathrm{x}; \theta) = \frac{1000x}{n}$$

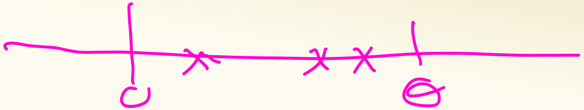
This holds $\forall x$

$\rightarrow$ $$\hat{\theta}_{ML}(X) = \frac{1000X}{n}$$

$$E(\hat{\theta}(X)) = E\left(\frac{1000X}{n}\right) = \frac{1000}{n}E(X) = \frac{1000}{n} \cdot n \cdot \frac{\theta}{1000} = \theta$$

Fact. $\hat{\theta}_{ML}(X)$ is unbiased

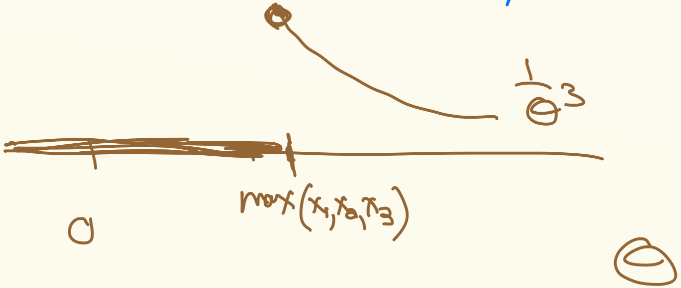
Three samples from $U(0, \theta)$



$$f(x;\theta) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{o.w.} \end{cases}$$

$$\mathcal{L}(X_1 = x_1, X_2 = x_2, X_3 = x_3; \theta) = \prod_{i=1}^{3} f(x_i;\theta) = \begin{cases} \frac{1}{\theta^3} & 0 \le x_1, x_2, x_3 \le \theta \\ 0 & \text{otherwise} \end{cases}$$

($x_1 = 2$, $x_2 = 5$, $x_3 = 1.3$)



$$\hat{\theta}(x_1, x_2, x_3) = \max(x_1, x_2, x_3)$$

$$\hat{\theta}(X_1, X_2, X_3) = \max(X_1, X_2, X_3)$$

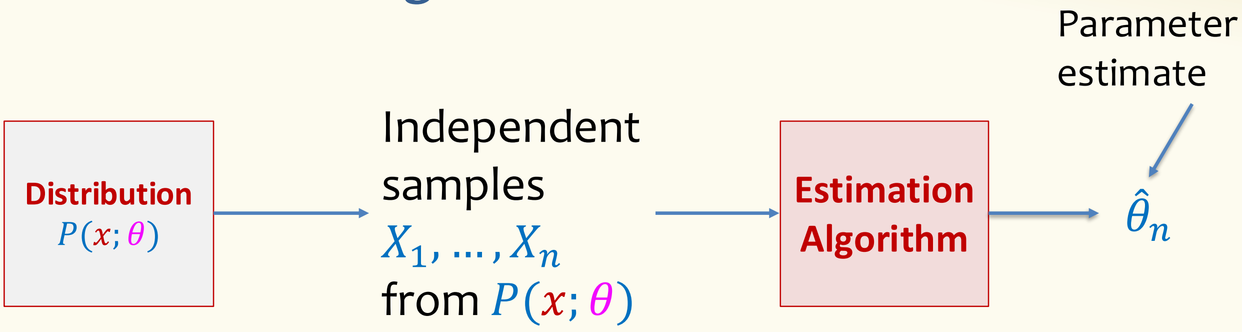
$$E(\hat{\theta}) = E(\max(X_1, X_2, X_3)) = \frac{3}{4}\theta$$

$U(0,\theta)$

$$\hat{\hat{\theta}}(X_1, X_2, X_3) = \frac{4}{3}\max(X_1, X_2, X_3)$$

unbiased estimator

When is an estimator good?



θ = unknown parameter

Definition. An estimator is **unbiased** if $\mathbb{E}[\hat{\theta}_n] = \theta$ for all $n \geq 1$.

Definition. An estimator is **consistent** if $\lim_{n\to\infty} \mathbb{E}[\hat{\theta}_n] = \theta$.

Theorem. MLE estimators are consistent.

(But not necessarily unbiased)

Example – Consistency

Normal outcomes $X_1, \dots, X_n$ i.i.d. according to $\mathcal{N}(\mu, \sigma^2)$ Assume: $\sigma^2 > 0$

$\frac{1}{n}\sum X_i$

$$\widehat{\Theta}_{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\Theta}_\mu)^2$$

$$E(\widehat{\Theta}_{\sigma^2}) = \frac{n-1}{n}\sigma^2$$

Population variance – Biased!

$\widehat{\Theta}_{\sigma^2}$ is "consistent"

Example – Consistency

Normal outcomes $X_1, \ldots, X_n$ i.i.d. according to $\mathcal{N}(\mu, \sigma^2)$ **Assume:** $\sigma^2 > 0$

$\frac{1}{n}\sum X_i$

$$\widehat{\Theta}_{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \widehat{\Theta}_\mu\right)^2 \cdot \frac{n}{n-1}$$

Population variance – Biased!

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \widehat{\Theta}_\mu\right)^2$$

Sample variance – Unbiased!

$\widehat{\Theta}_{\sigma^2}$ converges to same value as S_n^2, i.e., σ^2, as $n \to \infty$.

$\widehat{\Theta}_{\sigma^2}$ is "consistent"

Why does it matter?

- When statisticians are estimating a variance from a sample, they usually divide by $n-1$ instead of n.

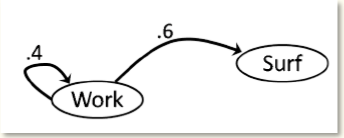
- They and we not only want good estimators (unbiased, consistent)
  - They/we also want **confidence bounds**
    - Upper bounds on the probability that these estimators are far the truth about the underlying distributions
  - Confidence bounds are just like what we wanted for our polling problems, but CLT is usually not the best thing to use to get them (unless the variance is known)

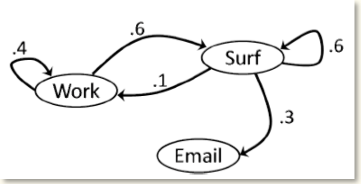
Agenda

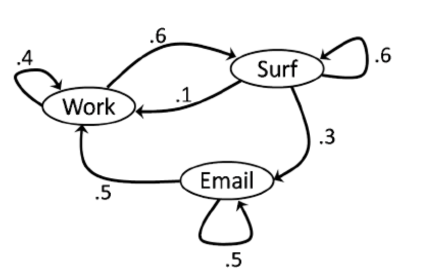
- Unbiased Estimation
- Markov Chains
- Application: PageRank

A typical day in my life… .

time $t = 0$

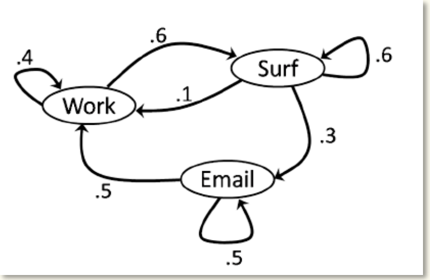






A typical day in my life

How do we interpret this diagram?



This kind of random process is called a **Markov Chain**

At *each* time step t

– I can be in one of 3 **states**

- Work, Surf, Email

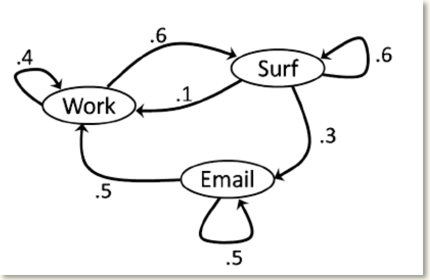
– If I am in some state s at time t

- the **labels of out-edges** of s **give the probabilities** of my moving to each of the states at time $t + 1$ (as well as staying the same)
  - so **labels on out-edges sum to** 1

e.g. If I am in Email, there is a 50-50 chance I will be in each of Work or Email at the next time step, but I will never be in state Surf in the next step.

This diagram looks vaguely familiar if you took CSE 311 …

Markov chains are a special kind of *probabilistic (finite) automaton*

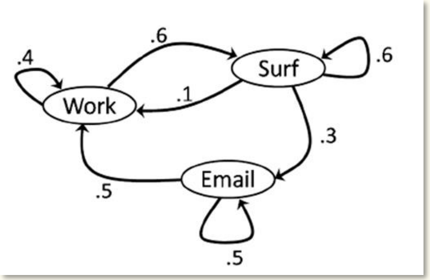


The diagrams look a bit like those of Deterministic Finite Automata (DFAs) you saw in 311 except that…

- There are no input symbols on the edges
  - Think of there being only one kind of input symbol "another tick of the clock" so no need to mark it on the edge
- They have multiple out-edges like an NFA, except that they come with probabilities

But just like DFAs, the only thing they remember about the past is the state they are currently in.

Many interesting questions about Markov Chains

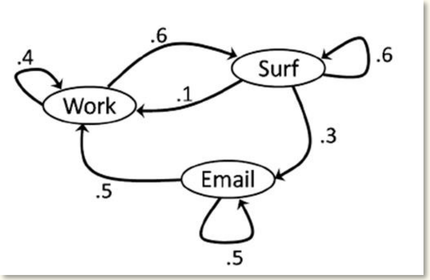


Given: In state Work at time $t = 0$

1. **What is the probability that I am in state s at time 1?**
2. **What is the probability that I am in state s at time 2?**
3. **What is the probability that I am in state s at some time t far in the future?**

To answer these questions, we need to understand how the probability distribution over states I could be in evolves over time.

Many interesting questions about Markov Chains



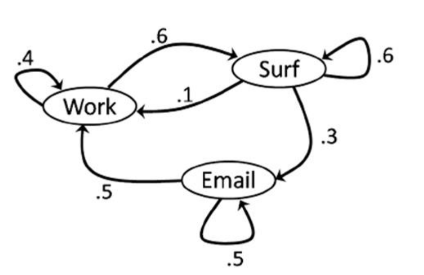
Given: In state Work at time $t = 0$

1. **What is the probability that I am in state s at time 1?**
2. **What is the probability that I am in state s at time 2?**
3. **What is the probability that I am in state s at some time t far in the future?**

To answer these questions, we need to understand how the probability distribution over states I could be in evolves over time.

$$M = \begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} \end{array}$$

An organized way to understand how the distribution evolves



Suppose we know

$$q_W^{(t)} = P(\text{Working at time t})$$

$$q_S^{(t)} = P(\text{Surfing at time t})$$

$$q_E^{(t)} = P(\text{Emailing at time t})$$

By the law of total probability

Pr(Working at t+1) = P(Working at t) ·0.4 + Pr(Surfing at t) · 0.1 + Pr (Emailing at t) · 0.5

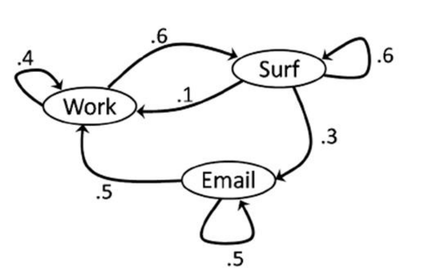
Pr (working at t+1 | working at t)

Pr (working at t+1 | surfing at t)

Pr (working at t+1 | emailing at t)

$$q_W^{(t+1)} = q_W^{(t)} \cdot 0.4 + q_S^{(t)} \cdot 0.1 + q_E^{(t)} \cdot 0.5$$

An organized way to understand how the distribution evolves



$q_W^{(t)} = P(\text{in state Work at time t})$

$q_S^{(t)} = P(\text{in state Surf at time t})$

$q_E^{(t)} = P(\text{in state Email at time t})$

Pr(Surfing at t+1) = P(Working at t) ·0.6 + Pr(Surfing at t) · 0.6 + Pr (email at t) · 0

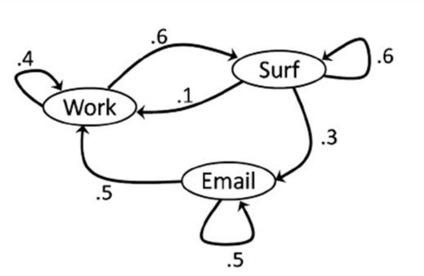
Pr (surfing at t+1 | working at t)

Pr (surfing at t+1 | surfing at t)

Pr (surfing at t+1 | emailing at t)

$$q_S^{(t+1)} = q_W^{(t)} \cdot 0.6 + q_S^{(t)} \cdot 0.6 + q_E^{(t)} \cdot 0$$

An organized way to understand the distribution at time t



$$q_W^{(t+1)} = q_W^{(t)} \cdot 0.4 + q_S^{(t)} \cdot 0.1 + q_E^{(t)} \cdot 0.5$$

$$q_S^{(t+1)} = q_W^{(t)} \cdot 0.6 + q_S^{(t)} \cdot 0.6 + q_E^{(t)} \cdot 0$$

$$q_E^{(t+1)} = q_W^{(t)} \cdot 0 + q_S^{(t)} \cdot 0.3 + q_E^{(t)} \cdot 0.5$$

$q_W^{(t)} = P(\text{in state Work at time t})$

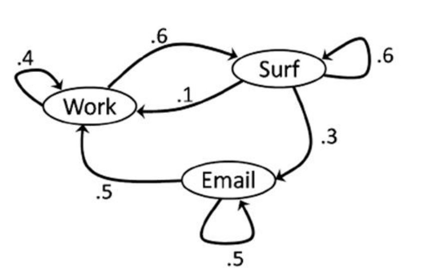
$q_S^{(t)} = P(\text{in state Surf at time t})$

$q_E^{(t)} = P(\text{in state Email at time t})$

$$[q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}] = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \overbrace{\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}}^{\boldsymbol{M}}$$

$$M = \begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} \end{array}$$

An organized way to understand the distribution at time t



$$q_W^{(t+1)} = q_W^{(t)} \cdot 0.4 + q_S^{(t)} \cdot 0.1 + q_E^{(t)} \cdot 0.5$$

$$q_S^{(t+1)} = q_W^{(t)} \cdot 0.6 + q_S^{(t)} \cdot 0.6 + q_E^{(t)} \cdot 0$$

$$q_E^{(t+1)} = q_W^{(t)} \cdot 0 + q_S^{(t)} \cdot 0.3 + q_E^{(t)} \cdot 0.5$$

$$q_W^{(t)} = P(\text{in state Work at time t})$$

$$q_S^{(t)} = P(\text{in state Surf at time t})$$

$$q_E^{(t)} = P(\text{in state Email at time t})$$

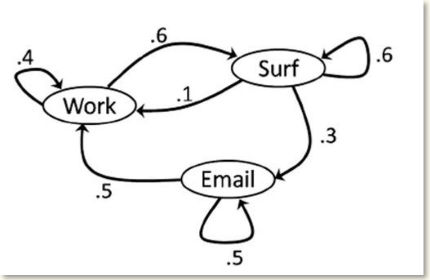
$$[q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}] = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \overset{\boldsymbol{M}}{\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}}$$

$$M = \begin{matrix} & W & S & E \\ W & 0.4 & 0.6 & 0 \\ S & 0.1 & 0.6 & 0.3 \\ E & 0.5 & 0 & 0.5 \end{matrix}$$

Write $\boldsymbol{q}^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$

$$\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)} \boldsymbol{M}$$

Many interesting questions about Markov Chains



Given: In state Work at time $t = 0$

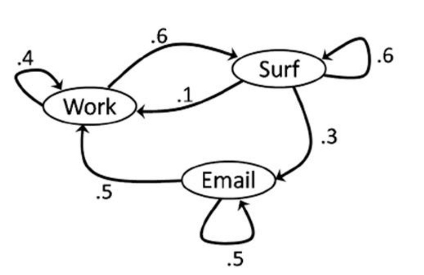
1. **What is the probability that I am in state s at time 1?**

In other words, what is $\boldsymbol{q}^{(1)} = [q_W^{(1)}, q_S^{(1)}, q_E^{(1)}]$?

2. **What is the probability that I am in state s at time 2?**

In other words, what is $\boldsymbol{q}^{(2)} = [q_W^{(2)}, q_S^{(2)}, q_E^{(2)}]$?

An organized way to understand the distribution at time t



$$q^{(0)} = [1, \ 0, \ 0]$$

Start out working

$$q^{(1)} = q^{(0)}M$$

$$q^{(1)} = [0.4, \ 0.6, \ 0] = [1, \ 0, \ 0]\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$q_W^{(t)} = P$(in state Work at time t)

$q_S^{(t)} = P$(in state Surf at time t)

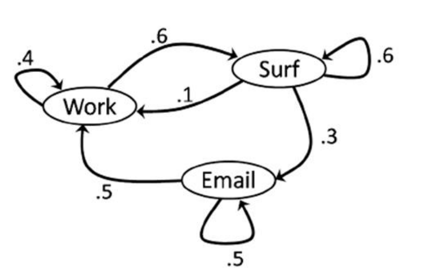
$q_E^{(t)} = P$(in state Email at time t)

$$M = \begin{array}{c|ccc} & W & S & E \\ W & 0.4 & 0.6 & 0 \\ S & 0.1 & 0.6 & 0.3 \\ E & 0.5 & 0 & 0.5 \end{array}$$

$$[q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}] = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \overbrace{\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}}^{M}$$

$$q^{(t+1)} = q^{(t)}M$$

An organized way to understand the distribution at time t



$$q^{(0)} = [1, \;\; 0, \;\; 0]$$

Start out working

$$q^{(1)} = q^{(0)} M$$

$$q^{(1)} = [0.4, \;\; 0.6, \;\; 0] = [1, \;\; 0, \;\; 0] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$$q^{(2)} = q^{(1)} M$$

$$q^{(2)} = [0.22, \;\; 0.6, \;\; 0.18] = [0.4, \;\; 0.6, \;\; 0] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$q_W^{(t)} = P(\text{in state Work at time t})$

$q_S^{(t)} = P(\text{in state Surf at time t})$

$q_E^{(t)} = P(\text{in state Email at time t})$

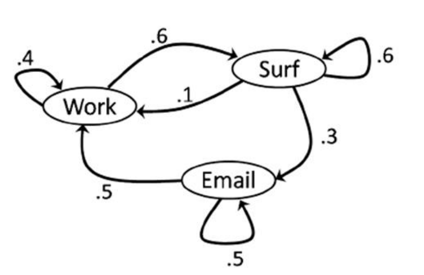
	W	S	E
W	0.4	0.6	0
S	0.1	0.6	0.3
E	0.5	0	0.5

$$M = \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$$[q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}] = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \overset{\boldsymbol{M}}{\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}}$$

$$q^{(t+1)} = q^{(t)} M$$

An organized way to understand the distribution at time t



$$q^{(0)} = [1, \ 0, \ 0]$$ Start out working

$$q^{(1)} = q^{(0)} M$$

$$q^{(1)} = [0.4, \ 0.6, \ 0] = [1, \ 0, \ 0] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$$q^{(2)} = q^{(1)} M$$

$$q^{(2)} = [0.22, \ 0.6, \ 0.18] = [0.4, \ 0.6, \ 0] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$q_W^{(t)} = P(\text{in state Work at time t})$

$q_S^{(t)} = P(\text{in state Surf at time t})$

$q_E^{(t)} = P(\text{in state Email at time t})$

$$q^{(2)} = [1, \ 0, \ 0] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

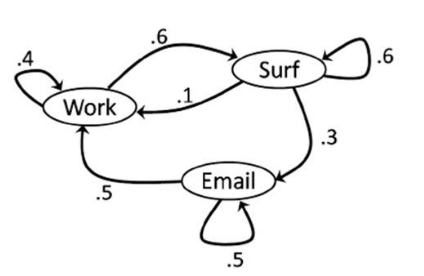
$$= q^{(0)} M^2$$

	W	S	E
W	0.4	0.6	0
S	0.1	0.6	0.3
E	0.5	0	0.5

$$M = \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$$q^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$$

An organized way to understand the distribution at time t



M^2
captures 2-step transition probabilities

Start out working

$q^{(1)} = [0.4, \ 0.6, \ \ldots$ $= [1, \ 0, \ 0] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$

$q^{(2)} = \ldots M$

$q^{(2)} = [0.22, \ 0.6, \ 0.18] = [0.\ldots, \ 0.6, \ 0] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$

$q^{(t)}_W = P(\text{in state Work at time t})$

$q^{(t)}_S = P(\text{in state Surf at time t})$

$q^{(t)}_E = P(\text{in state Email at time t})$

$$q^{(2)} = [1, \ 0, \ 0] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

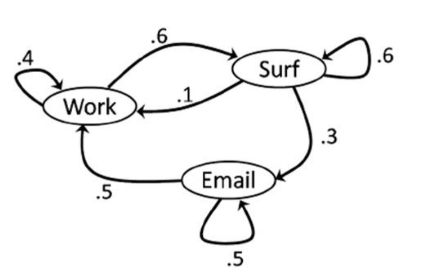
$$= q^{(0)} M^2$$

	W	S	E
W	0.4	0.6	0
S	0.1	0.6	0.3
E	0.5	0	0.5

$$M = \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$$q^{(t)} = [q^{(t)}_W, q^{(t)}_S, q^{(t)}_E]$$

An organized way to understand the distribution at time t



$$\mathbf{M}$$

$$[q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}] = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

Write $\boldsymbol{q}^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$

Then for all $t \geq 0$, $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}\boldsymbol{M}$

$q_W^{(t)} =$ P(in state Work at time t)

$q_S^{(t)} =$ P(in state Surf at time t)

$q_E^{(t)} =$ P(in state Email at time t)

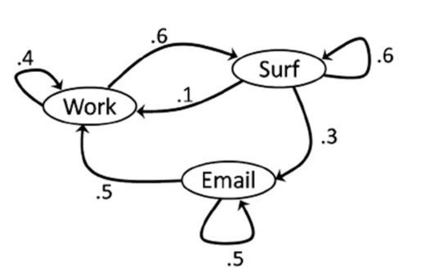
$\boldsymbol{q}^{(1)} = \boldsymbol{q}^{(0)}\boldsymbol{M}$

$\boldsymbol{q}^{(2)} = \boldsymbol{q}^{(1)}\boldsymbol{M} = \boldsymbol{q}^{(0)}\boldsymbol{M}\boldsymbol{M} = \boldsymbol{q}^{(0)}\boldsymbol{M}^2$

$\boldsymbol{q}^{(3)} = \boldsymbol{q}^{(2)}\boldsymbol{M} = \boldsymbol{q}^{(0)}\boldsymbol{M}^2\boldsymbol{M} = \boldsymbol{q}^{(0)}\boldsymbol{M}^3$

$\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)}\boldsymbol{M}^t$ for all $t \geq 0$

An organized way to understand the distribution at time t



$$[q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}] = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \; \boldsymbol{M}$$

$$\boldsymbol{M} = \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ & 0 & 0.5 \end{bmatrix}$$

$\boldsymbol{M}^t$

captures t -step transition probabilities

ite $\boldsymbol{q}^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$

The all $t \geq 0$, $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}\boldsymbol{M}$

$q_W^{(t)} =$ P(in state Work at time t)

$q_S^{(t)} =$ P(in state Surf at time t)

$q_E^{(t)} =$ P(in state Email at time t)

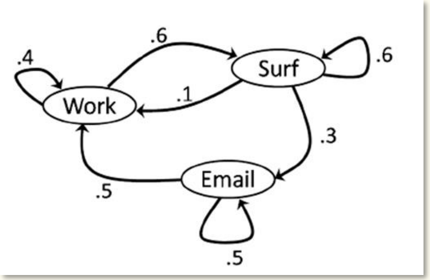
$\boldsymbol{q}^{(1)} = \boldsymbol{q}^{(0)}\boldsymbol{M}$

$\boldsymbol{q}^{(2)} = \boldsymbol{q}^{(1)}\boldsymbol{M}$ $\boldsymbol{q}^{(0)}\boldsymbol{M}\boldsymbol{M} = \boldsymbol{q}^{(0)}\boldsymbol{M}^2$

$\boldsymbol{q}^{(3)} = \boldsymbol{q}^{(2)}\boldsymbol{M} = \boldsymbol{q}^{(0)}\boldsymbol{M}^2\boldsymbol{M} = \boldsymbol{q}^{(0)}\boldsymbol{M}^3$

$$\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)}\boldsymbol{M}^t \text{ for all } t \geq 0$$

Many interesting questions about Markov Chains



1. What is the probability that I am in state s at time 1?
2. What is the probability that I am in state s at time 2?
3. What is the probability that I am in state s at some time t far in the future?

Given: In state Work at time $t = 0$

$$\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)} \boldsymbol{M}^{t} \text{ for all } t \geq 0$$

$$\boldsymbol{q}^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$$

What does $\boldsymbol{q}^{(t)}$ look like for really big t ?

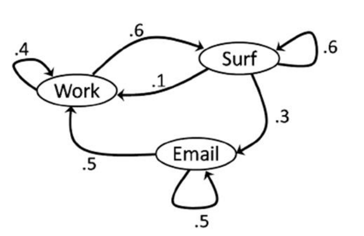
$q_W^{(t)} =$ P(in state Work at time t)

$q_S^{(t)} =$ P(in state Surf at time t)

$q_E^{(t)} =$ P(in state Email at time t)

$q^{(t)} = q^{(0)} M^t$ for all $t \geq 0$

M^t as t grows



M

$$\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

M^2

	W	S	E
W	.22	.6	.18
S	.25	.42	.33
E	.45	.3	.25

M^3

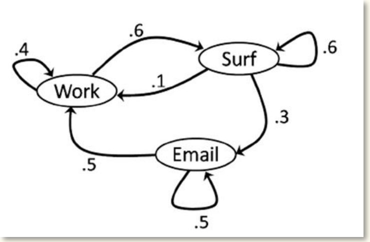
	W	S	E
W	.238	.492	.270
S	.307	.402	.291
E	.335	.450	.215

M^{10}

	W	S	E
W	.2940	.4413	.2648
S	.2942	.4411	.2648
E	.2942	.4413	.2648

$q^{(t)} = q^{(0)} M^t$ for all $t \geq 0$

M^t as t grows



M

$$\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

M^2

	W	S	E
W	.22	.6	.18
S	.25	.42	.33
E	.45	.3	.25

M^3

	W	S	E
W	.238	.492	.270
S	.307	.402	.291
E	.335	.450	.215

M^{10}

	W	S	E
W	.2940	.4413	.2648
S	.2942	.4411	.2648
E	.2942	.4413	.2648

M^{30}

	W	S	E
W	.29411764705	.44117647059	.26470588235
S	.29411764706	.44117647058	.26470588235
E	.29411764706	.44117647059	.26470588235

M^{60}

	W	S	E
W	.294117647058823	.441176470588235	.264705882352941
S	.294117647068823	.441176470588235	.264705882352941
E	.294117647068823	.441176470588235	.264705882352941

What does this say about $q^{(t)}$?

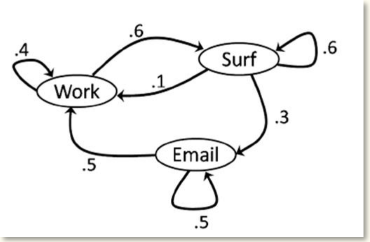
M^t as t grows

$$q^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$$

$q_W^{(t)} =$ P(in state Work at time t)

$q_S^{(t)} =$ P(in state Surf at time t)

$q_E^{(t)} =$ P(in state Email at time t)



$$q^{(60)} = q^{(0)} M^{60}$$

$q^{(0)}$

$$[1, 0, 0] \cdot \begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{pmatrix} .294117647058823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \end{pmatrix} \end{array}$$

$$= (.294117647058823 \quad .441176470588235 \quad .264705882352941)$$

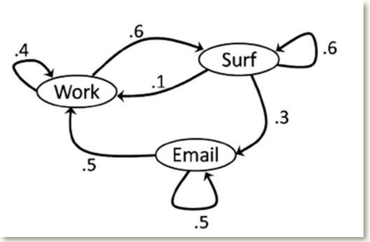
$\boldsymbol{M}^t$ as t grows

$$\boldsymbol{q}^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$$

$q_W^{(t)} =$ P(in state Work at time t)

$q_S^{(t)} =$ P(in state Surf at time t)

$q_E^{(t)} =$ P(in state Email at time t)



$$\boldsymbol{q}^{(60)} = \boldsymbol{q}^{(0)}\,\boldsymbol{M}^{60}$$

$$\underbrace{[0.5, 0.5, 0\,]}_{\boldsymbol{q}^{(0)}} \cdot \begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{ccc} W & S & E \\ \left(\begin{array}{ccc} .294117647058823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \end{array}\right) \end{array}$$

$$= (\ .294117647058823 \quad .441176470588235 \quad .264705882352941\ )$$

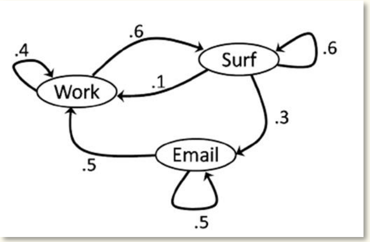
M^t as t grows

$$\boldsymbol{q}^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$$

$q_W^{(t)} = P(\text{in state Work at time t})$

$q_S^{(t)} = P(\text{in state Surf at time t})$

$q_E^{(t)} = P(\text{in state Email at time t})$



$$\boldsymbol{q}^{(60)} = \boldsymbol{q}^{(0)} \boldsymbol{M}^{60}$$

$\boldsymbol{q}^{(0)}$

$$[0.4, 0.5, 0.1] \cdot \begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{ccc} W & S & E \\ \hline \end{array}$$

	W	S	E
W	.294117647058823	.441176470588235	.264705882352941
S	.294117647068823	.441176470588235	.264705882352941
E	.294117647068823	.441176470588235	.264705882352941

$$= (\ .294117647058823 \quad .441176470588235 \quad .264705882352941\ )$$

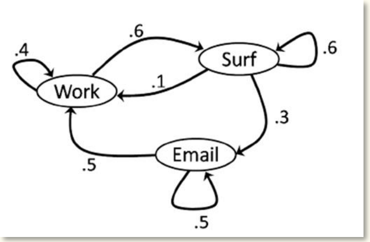
M^t as t grows

$$\boldsymbol{q}^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$$

$q_W^{(t)} = P$(in state Work at time t)

$q_S^{(t)} = P$(in state Surf at time t)

$q_E^{(t)} = P$(in state Email at time t)



$$\boldsymbol{q}^{(60)} = \boldsymbol{q}^{(0)} \boldsymbol{M}^{60}$$

$$[q_W^{(60)}, q_S^{(60)}, q_E^{(60)}] = [q_W^{(0)}, q_S^{(0)}, q_E^{(0)}] \cdot \begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{pmatrix} .294117647058823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \end{pmatrix} \end{array}$$

$$\forall \boldsymbol{q}^{(0)} \quad [q_W^{(60)}, q_S^{(60)}, q_E^{(60)}] = [\ .294117647058823 \quad .441176470588235 \quad .264705882352941\ ]$$

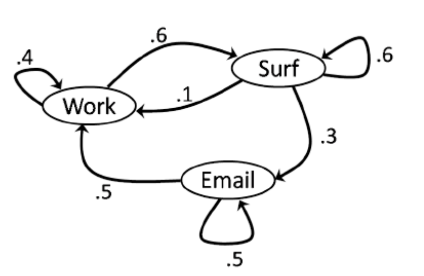
- In the long run, the starting state doesn't really matter!!
- In particular, at any point in the (slightly distant) future, the chance I'm surfing the web is about 44%
- The distribution on states converges to $[\ .294117647058823 \quad .441176470588235 \quad .264705882352941\ ]$

- Suppose that we believe the distribution on states converges to some fixed probability vector $\boldsymbol{\pi} = (\pi_W, \pi_S, \pi_E)$

In other words, $\lim_{t \to \infty} \boldsymbol{q}^{(t)} = \boldsymbol{\pi} = (\pi_W, \pi_S, \pi_E)$

- Can we figure out what $\boldsymbol{\pi}$ vis just by looking at M?

Observation



If $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}$ then it will never change again!

Proof: $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)} \boldsymbol{M}$

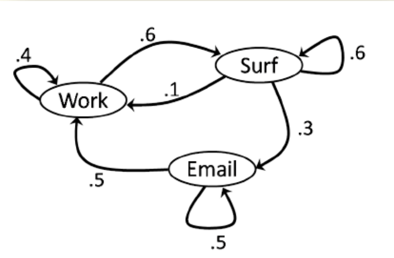
$\boldsymbol{q}^{(t+2)} = \boldsymbol{q}^{(t+1)} \boldsymbol{M} = \boldsymbol{q}^{(t)} \boldsymbol{M} = \boldsymbol{q}^{(t+1)}$

Such a $\boldsymbol{q}^{(t)}$ (that never changes) is called a **stationary distribution** and has a special name $\boldsymbol{\pi} = (\pi_W, \pi_S, \pi_E)$

It is the solution to $\boldsymbol{\pi} = \boldsymbol{\pi} \boldsymbol{M}$

Solving for Stationary Distribution

$$(\pi_W, \pi_S, \pi_E) = (\pi_W, \pi_S, \pi_E) \begin{pmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{pmatrix}$$



$$\pi_W = \pi_W \cdot 0.4 + \pi_S \cdot 0.1 + \pi_E \cdot 0.5$$

$$\pi_S = \pi_W \cdot 0.6 + \pi_S \cdot 0.6 + \pi_E \cdot 0$$

$$\pi_E = \pi_W \cdot 0 + \pi_S \cdot 0.3 + \pi_E \cdot 0.5$$

$$\pi_W + \pi_S + \pi_E = 1$$

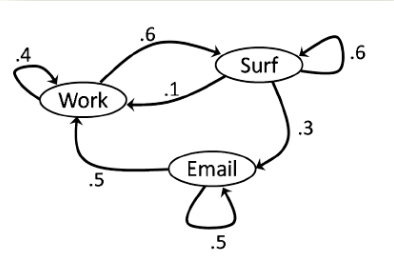
Solving for Stationary Distribution

$$(\pi_W, \pi_S, \pi_E)\begin{pmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{pmatrix} = (\pi_W, \pi_S, \pi_E)$$

$$\pi_W = \pi_W \cdot 0.4 + \pi_S \cdot 0.1 + \pi_E \cdot 0.5$$

$$\pi_S = \pi_W \cdot 0.6 + \pi_S \cdot 0.6 + \pi_E \cdot 0$$

$$\pi_E = \pi_W \cdot 0 + \pi_S \cdot 0.3 + \pi_E \cdot 0.5$$



$$\pi_W + \pi_S + \pi_E = 1$$

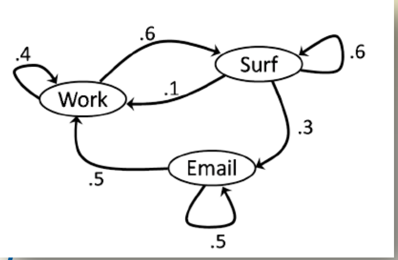
$$\rightarrow \pi_W = \frac{10}{34}, \quad \pi_S = \frac{15}{34}, \quad \pi_E = \frac{9}{34}$$

As $t \to \infty$, $\boldsymbol{q}^{(t)} \to \boldsymbol{\pi}$ no matter what distribution $\boldsymbol{q}^{(0)}$ is !!

Markov Chains summary

	W	S	E
W	0.4	0.6	0
S	0.1	0.6	0.3
E	0.5	0	0.5

$$M = \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$



- A set of n **states** $\{1, 2, 3, \dots\ n\}$
- A square **transition matrix** $\boldsymbol{M}$, dimension $n \times n$

$$\boldsymbol{M}_{ij} = P(\text{transition from } i \text{ to } j)$$

- $\boldsymbol{M}^t{}_{ij}$ = Pr (in state j after t steps| start in state i).
- Nice Markov chains are not sensitive to initial distribution of states. $\boldsymbol{M}^t \to \boldsymbol{W}$, where all rows in $\boldsymbol{W}$ are the same probability vector $\boldsymbol{\pi}$
- A **stationary distribution** $\boldsymbol{\pi}$ is the solution to:

$\boldsymbol{\pi} = \boldsymbol{\pi}\,\boldsymbol{M}$, normalized so that $\Sigma_{i \in [n]} \pi_i = 1$

$\boldsymbol{M}^{60}$

	W	S	E
W	.294117647058823	.441176470588235	.264705882352941
S	.294117647068823	.441176470588235	.264705882352941
E	.294117647068823	.441176470588235	.264705882352941

The Fundamental Theorem of Markov Chains

Theorem. Any nice* Markov chain

has a unique stationary distribution $\boldsymbol{\pi}$.

Moreover, as $t \to \infty$, for all i, j, $\lim_{t\to\infty} \boldsymbol{M}_{ij}^{t} = \pi_j$

**aperiodic and irreducible: these concepts are beyond us but they turn out to cover a very large class of Markov chains of practical importance.*