CSE 312

Foundations of Computing II

Lecture 26: Markov Chain

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Rachel Lin, Hunter Schafer

Slide Credit: Based on Stefano Tessaro's slides for 312 19au
incorporating ideas from Alex Tsun's and Anna Karlin's slides for 312 20su and 20au

So far, a single-shot random process

Random Process → **Outcome Distribution** D

Today:
See a very special type of DTSP called **Markov Chains**

Many-step random process

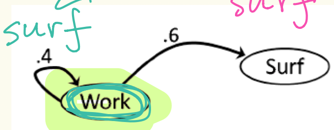
Random Process 1 → **Outcome Distribution** D_1 → Random Process 2 → **Outcome Distribution** D_2 → Random Process 3 → **Outcome Distribution** D_3 …

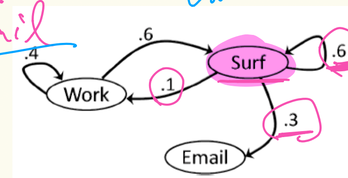
Definition: A **discrete-time stochastic process** (DTSP) is a sequence of random variables $X^{(0)}, X^{(1)}, X^{(2)}, \ldots$ where $X^{(t)}$ is the value at time t.

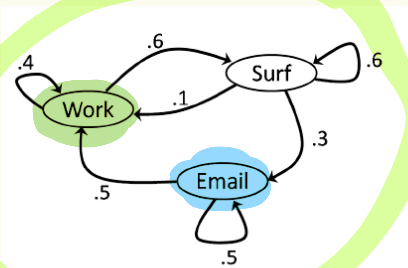
A boring day in pandemic

t = 0 → t=1 → t=2 → t=3 → t=4 → t=5 → t=6

work, surf, surf, Email, Email, work







This graph defines a **Markov Chain**

Special Property: The random process at each step is the same!

t=0 / t=6

For ANY $t \geq 0$,

if I was working at time t, then at t+1

with probability 0.4 I continue working

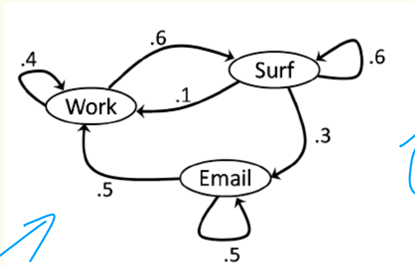
with probability 0.6, I move to surfing, and

with probability 0, I move to emailing

This is called History Independent (similar to memoryless)

A boring day in pandemic

t=0 work.



Many interesting questions?

1. What is the probability that I work at time 1?
2. What is the probability that I work at time 2?

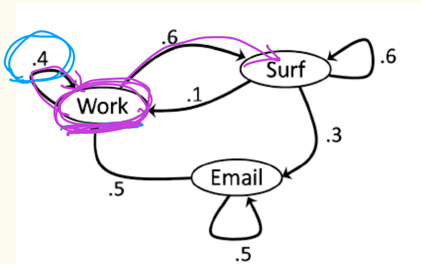
   0.22

3. What is the probability that I work at time t=100?
4. What is the probability that I work at time $t \to \infty$? Does it always converge?

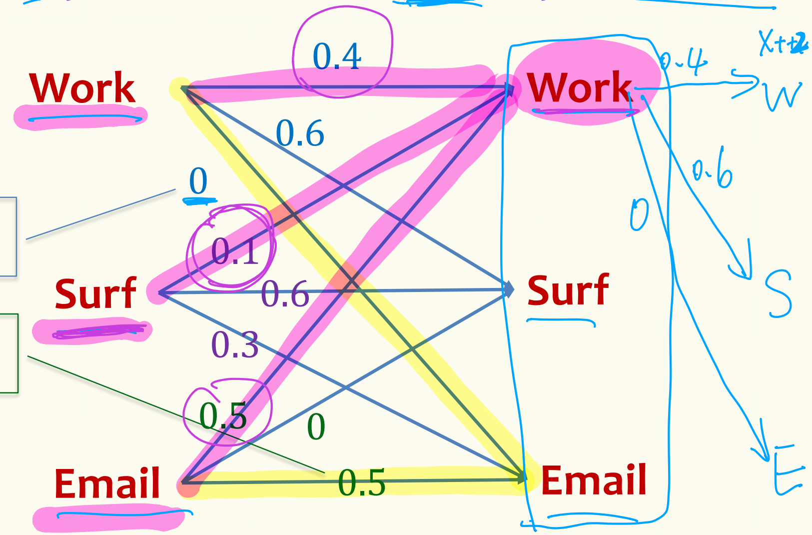
t=0 work,

t=1 w.p 0.4. work

Formalizing Markov Chain



$\boldsymbol{X^{(t)}}$ my state at t $\qquad$ $\boldsymbol{X^{(t+1)}}$ my state at t+1



$$p_{WE} = \mathrm{P}(X^{(t+1)} = E \mid X^{t} = W\ ) = 0$$

$$p_{EE} = \mathrm{P}(X^{(t+1)} = E \mid X^{t} = E\ ) = 0.5$$

1. What is the probability that I work at time 1? $p_W^{(1)} = P(X^{(1)} = W)$
2. What is the probability that I work at time 2? $p_W^{(2)} = P(X^{(2)} = W)$

LTP $\quad p_W^{(2)} = P(X^{(2)} = W \mid X^{(1)} = W)\, P(X^{(1)} = W)$ 0.4 $\quad = 0.16 + 0.06$

0.4

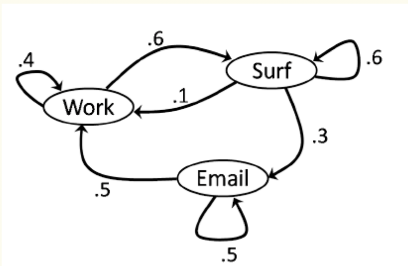
$+ P(X^{(2)} = W \mid X^{(1)} = S)\, P(X^{(1)} = S)$ 0.6 $\quad = 0.22$

0.1

$$+ P(X^{(2)} = W \mid X^{(1)} = E) P(X^{(1)} = E)$$

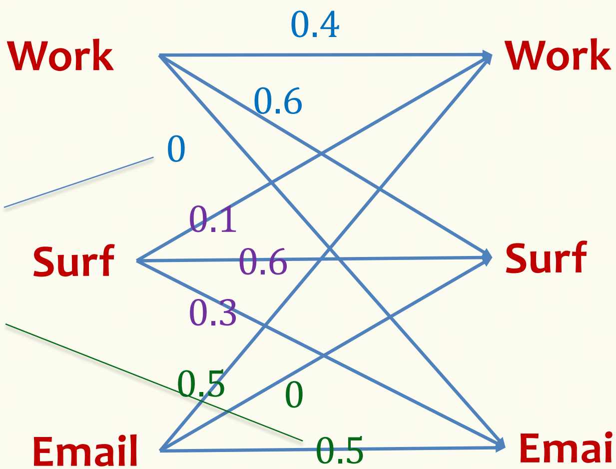
0.5 0

Formalizing Markov Chain



$X^{(t)}$ my state at t

X^{t+1} my state at t+1



$$p_{WE} = P(X^{(t+1)} = E \mid X^t = W) = 0$$

$$p_{EE} = P(X^{(t+1)} = E \mid X^t = E) = 0.5$$

3. What is the prob that I work at t+1 = 100?

By LTP: $p_W^{(t+1)} = P(X^{(t+1)} = W) = \Sigma_{U\in\{W,S,E\}} P(X^{(t+1)} = W \mid X^{(t)} = U) P(X^{(t)} = U)$

Vectors and Matrixes

$v = (v_1 \cdots \underline{v_i} \cdots v_n)$

$v[i]$

$$\begin{array}{cc} & \text{column } j \\ A = \text{row } i & \begin{pmatrix} & \vdots & \\ \cdots & a_{ij} & \\ & & \end{pmatrix} \end{array}$$

$A[i,j]$

$a_1b_1 + a_2b_4 + a_3b_7$

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \begin{bmatrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_7 & b_8 & b_9 \end{bmatrix} = \begin{bmatrix} c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 \\ c_7 & c_8 & c_9 \end{bmatrix}$$

1

1

Vectors and Matrixes

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{pmatrix}$$

2nd column ↓

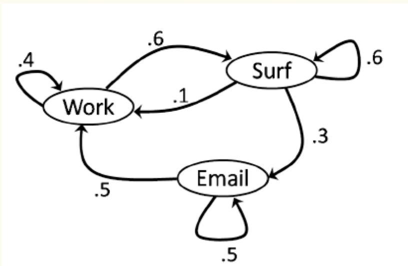
$$\mathbf{C} = \begin{pmatrix} a_{11}b_{11} + \cdots + a_{1n}b_{n1} & a_{11}b_{12} + \cdots + a_{1n}b_{n2} & \cdots & a_{11}b_{1p} + \cdots + a_{1n}b_{np} \\ a_{21}b_{11} + \cdots + a_{2n}b_{n1} & a_{21}b_{12} + \cdots + a_{2n}b_{n2} & \cdots & a_{21}b_{1p} + \cdots + a_{2n}b_{np} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{11} + \cdots + a_{mn}b_{n1} & a_{m1}b_{12} + \cdots + a_{mn}b_{n2} & \cdots & a_{m1}b_{1p} + \cdots + a_{mn}b_{np} \end{pmatrix}$$

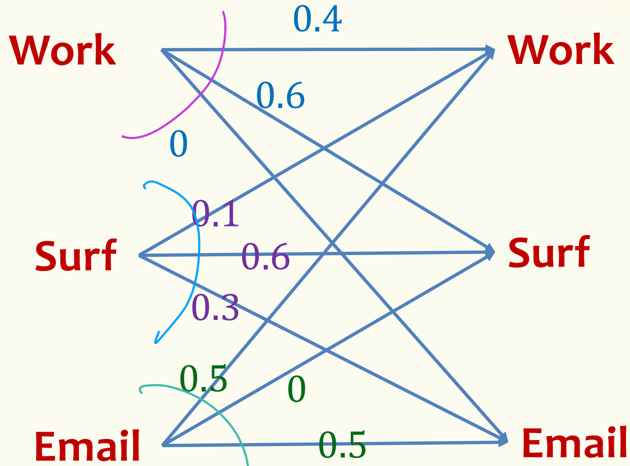
m th row

$$c[m,2] = \sum_{k=1}^{n} a_{\cancel{p}}[m,k]\, b[k,2]$$

Transition Matrix

$X^{(t)}$ my state at t $\quad$ X^{t+1} my state at t+1





$$P = \begin{array}{cc} & \begin{array}{ccc} W & S & E \end{array} \\ \begin{array}{c} W \\ S \\ E \end{array} & \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix} \end{array}$$

$$X^{(t)} = \left(p_W^{(t)} \quad p_S^{(t)} \quad p_E^{(t)}\right) = X^{(t)}[1] + X^{(t)}[2] + X^{(t)}[3] = 1$$

By LTP: $p_W^{(t+1)} = P(X^{(t+1)} = W) = \Sigma_{U \in \{W,S,E\}} \mathrm{P}\big(X^{(t+1)} = W \big| \mathrm{X}^{(t)} = U\big)\, \mathrm{P}(\mathrm{X}^{(t)} = \mathrm{U})$

➔ $X^{(t+1)} = X^{(t)} P$

Transition Matrix

$$P = \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$

$$X^{(t)} = \left(p_w^{(t)} \quad p_S^{(t)} \quad p_E^{(t)}\right)$$

$U = W$

$P_W^{(t)}$

$\|$

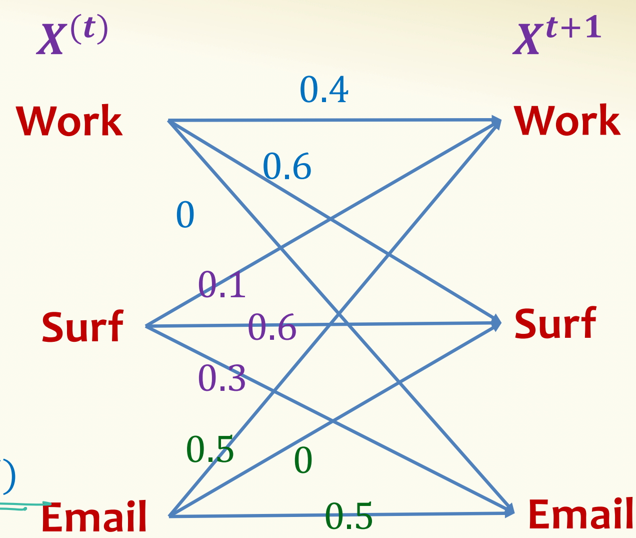
☆ LTP: $p_W^{(t+1)} = P\left(X^{(t+1)} = W\right)$

$= \Sigma_{U\in\{W,S,E\}} \mathrm{P}\left(X^{(t+1)} = W \,\middle|\, \mathrm{X}^{(t)} = U\right) \mathrm{P}(\mathrm{X}^{(t)} = \mathrm{U})$

➔ $X^{(t+1)} = X^{(t)} P$

$$\left(p_w^{(t+1)} \quad p_S^{(t+1)} \quad p_E^{(t+1)}\right) = \left(p_w^{(t)} \quad p_S^{(t)} \quad p_E^{(t)}\right) \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$

$= P_W^{(t)} \cdot 0.4 + P_S^{(t)} \cdot 0.1 + P_E^{(t)} \cdot 0.5$



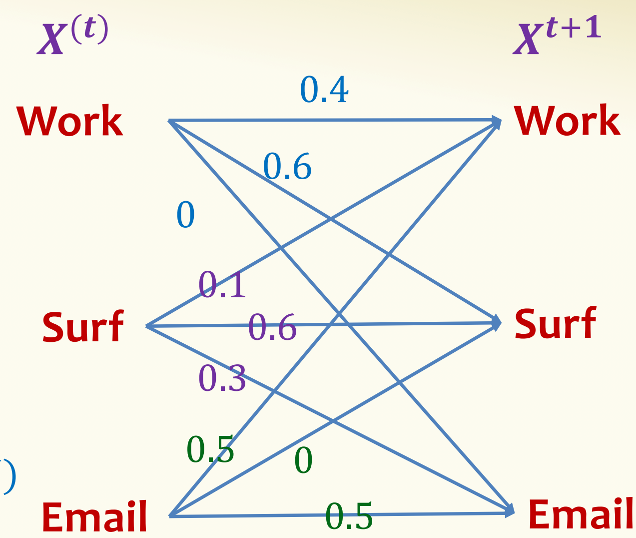
Transition Matrix

$$P = \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix} \quad \begin{pmatrix} 0.4 & 0.6 & 0 \\ 0.1, & 0.6, & 0.3 \\ 0.5, & 0, & 0.5 \end{pmatrix}$$

$$X^{(t)} = (p_w^{(t)} \quad p_S^{(t)} \quad p_E^{(t)})$$

LTP: $p_W^{(t+1)} = P\left(X^{(t+1)} = W\right)$

$= \Sigma_{U \in \{W,S,E\}} \mathrm{P}\left(X^{(t+1)} = W \middle| \mathrm{X}^{(t)} = U\right) \mathrm{P}(\mathrm{X}^{(t)} = \mathrm{U})$



➔ $X^{(t+1)} = X^{(t)} P$ ✱

➔ $\boxed{X^{(t)} = X^{(0)} P^t}$

3. What is the prob that I work at t = 100?

Closed formula: $p_W^{(t)} = X^{(t)}[1] = (X^{(0)} P^t)\,[1]$

$$X^{(1)} = X^{(0)} \cdot P$$

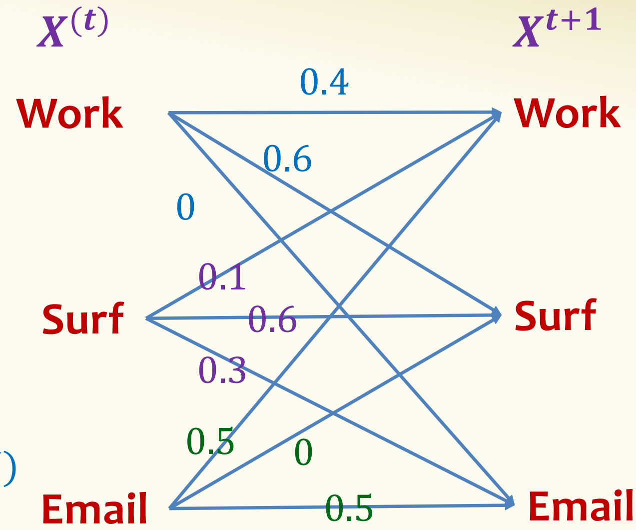
$$X^{(2)} = X^{(1)} \cdot P = \left(X^{(0)} \cdot P\right) \cdot P = \underline{X^{(0)} \cdot P^2}$$

$$X^{(3)} = X^{(2)} \cdot P = (X^{(0)} \cdot P^2) \cdot P = X^{(0)} \cdot P^3$$

Transition Matrix

$$P = \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$

$$X^{(t)} = (p_w^{(t)} \quad p_S^{(t)} \quad p_E^{(t)})$$



LTP: $p_W^{(t+1)} = P\big(X^{(t+1)} = W\big)$

$= \Sigma_{U \in \{W,S,E\}} \mathrm{P}\big(X^{(t+1)} = W \big| \mathrm{X}^{(t)} = U\big) \, \mathrm{P}(\mathrm{X}^{(t)} = \mathrm{U})$

➔ $X^{(t+1)} = X^{(t)} \, P$

➔ $X^{(t)} = X^{(0)} \, P^t$

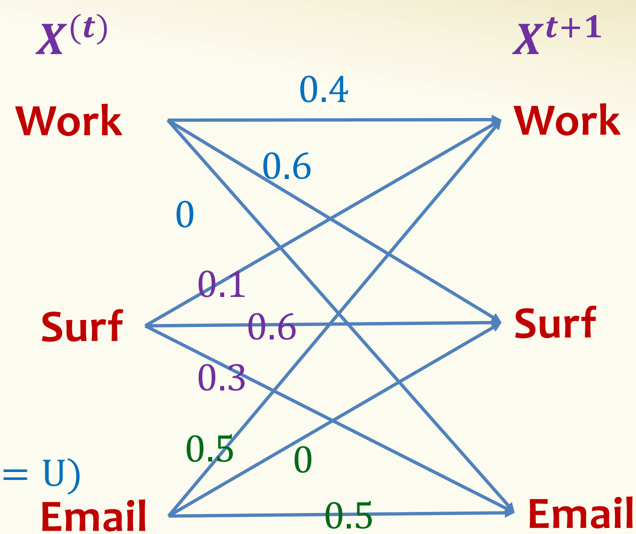
4. What is the probability that I work at time $t \to \infty$?
Does it always converge?

Poll: A. Yes it converges (orderly universe)
B. No, it does not converge (anarchic universe)

Stationary Distribution

$$P = \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$

$$X^{(t)} = \left(p_w^{(t)} \quad p_S^{(t)} \quad p_E^{(t)}\right)$$



LTP: $p_W^{(t+1)} = P\left(X^{(t+1)} = W\right)$

$= \Sigma_{U \in \{W,S,E\}} \mathrm{P}\left(X^{(t+1)} = W \middle| \mathrm{X}^{(t)} = U\right) \mathrm{P}(\mathrm{X}^{(t)} = \mathrm{U})$

➔ $X^{(t+1)} = X^{(t)}\ P$

➔ $X^{(t)} = X^{(0)}\ P^t$

➔ As t $\rightarrow \infty$, $X^{(t)} \rightarrow \pi$ the stationary distribution

4. What is the probability that I work at time $t \rightarrow \infty$?
Does it always converge?

Solving for Stationary Distribution

$X^{(t+1)} = X^{(t)} \cdot P$ $\quad t \to \infty$

$$P = \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$

Constraints on Stationary Distribution

- $\pi = \pi P$
- $\pi[1] + \pi[2] + \pi[3] = 1$

➔ $\pi[1] = \frac{15}{34}$, $\pi[2] = \frac{10}{34}$, $\pi[3] = \frac{9}{34}$

LTP: $p_W^{(t+1)} = P\left(X^{(t+1)} = W\right)$

$= \Sigma_{U \in \{W,S,E\}} \mathrm{P}\left(X^{(t+1)} = W \middle| \mathrm{X}^{(t)} = U\right) \mathrm{P}(\mathrm{X}^{(t)} = \mathrm{U})$

➔ $X^{(t+1)} = X^{(t)}\ P$

$[\pi[1] \cdots \pi[3]] = [\pi[1]\ \pi[2]\ \pi[3]] \left(P \right)$

➔ $X^{(t)} = X^{(0)}\ P^t$

$[\pi[1], \pi[2], \pi[3]]$

➔ As t → ∞, $\boxed{X^{(t)} \to \pi}$ the stationary distribution

$p_W \quad p_S \quad p_E$

General Markov Chain

- A set of n states $\{1, 2, 3, \ldots\ n\}$
- The state at time t $X^{(t)}$
- A transition matrix P, dimension $n \times n$

$$P[i, j] = \Pr\left(X^{(t+1)} = j \middle| X^{(t)} = i\right)$$

- Transition: LTP ➔ $X^{(t+1)} = X^{(t)} P \implies X^{(t)} = X^{(0)} P^t$
- A stationary distribution π is the solution to:

$$\pi = \pi P, \text{ normalized so that } \Sigma_{i \in [n]} \pi[i] = 1$$

{ W, S, E }

$X^{(t)}$

The Fundamental Theorem of Markov Chain

If a Markov Chain is Irreducible and aperiodic, then it has a unique stationary distribution.

Moreover, $t \to \infty$, for all i, j, $P^t[i, j] \to \pi[j]$

$X^{(t)} = (p_1, p_2, \cdots, p_n)$

$\pi =$

$$P^t = \begin{bmatrix} \pi[1] & \pi[i] & \cdots & \pi[n] \\ \pi[1] & \pi[i] & \cdots & \pi[n] \\ \cdots & \cdots & & \\ \pi[1] & \pi[i] & \cdots & \pi[n] \end{bmatrix}$$

$X^{(t)} = X^{(0)} P^t$