



# 4. Markov Models: Probabilistic Reasoning (over time or space)

COMP 3270  
Artificial Intelligence

Dirk Schnieders

# Outline

- Motivation
- Probabilistic Reasoning
  - Running Example: Ghostbusters
  - Probability Refresher
    - Random Variables, Joint and Marginal Distributions, Conditional Distribution, Product Rule, Chain Rule, Bayes' Rule
  - Inference, Independence
- Probabilistic Reasoning over Time or Space
  - Markov Model
  - Mini-Forward Algorithm
  - Stationary Distributions
- Next chapter: Hidden Markov Models

# Motivation

- AI History
- Applications
  - Diagnosis
    - Physician measures symptoms and infer disease
    - IT experts diagnose printer problem
  - Speech recognition: Noisy signal, people have different pronunciations of words
  - Tracking objects: Combine noisy measurements from multiple sensors to estimate position
  - Genetics: Reproduction
  - Error correcting codes: Electrical signals are noisy

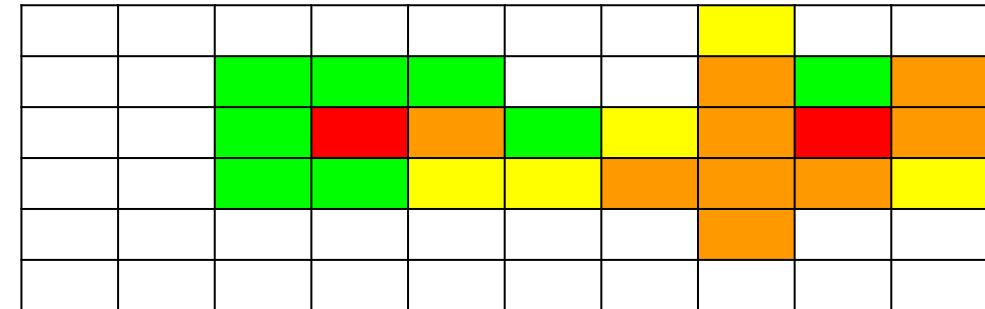


# Running Example: Ghostbusters



# Running Example: Ghostbusters

- A ghost is hiding in the grid somewhere
- Noisy sensor readings tell how close a certain square is to the ghost
  - On the ghost: red
  - 1 or 2 away: orange
  - 3 or 4 away: yellow
  - 5+ away: green
- We know  $P(\text{color} \mid \text{distance})$ 
  - Example for distance = 3



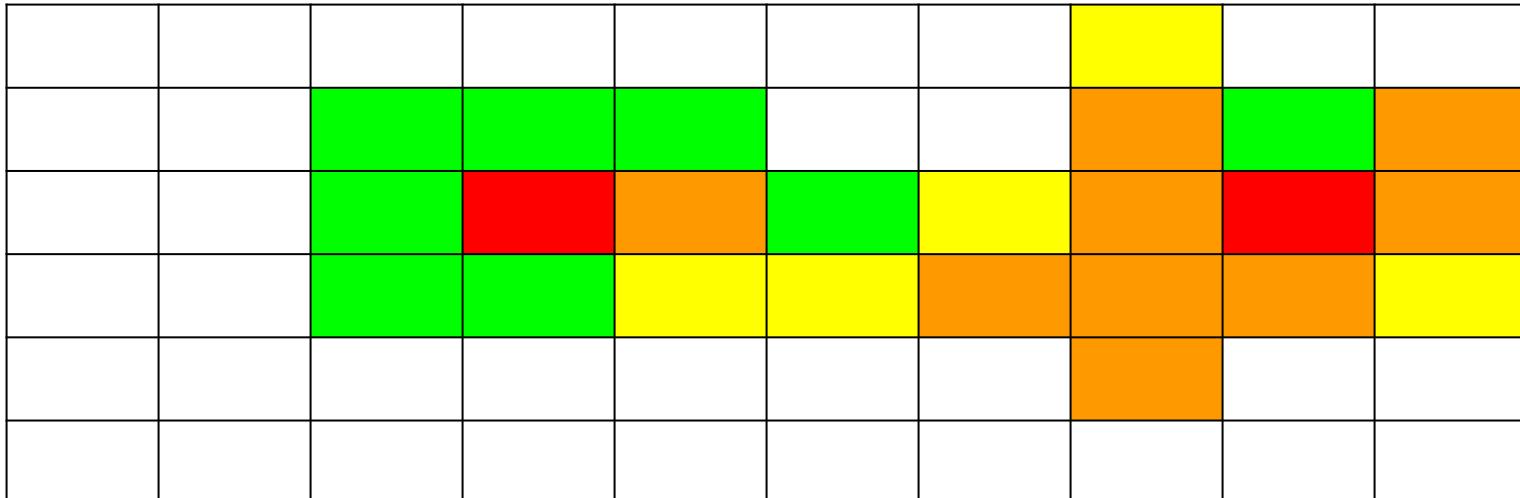
$P(\text{red} \mid 3)$	$P(\text{orange} \mid 3)$	$P(\text{yellow} \mid 3)$	$P(\text{green} \mid 3)$
0.05	0.15	0.5	0.3

Each of the distances has a distribution similar to this one

# Running Example: Ghostbusters

- Sensor readings tell how close a square is to the ghost

- On the ghost: red
- 1 or 2 away: orange
- 3 or 4 away: yellow
- 5+ away: green

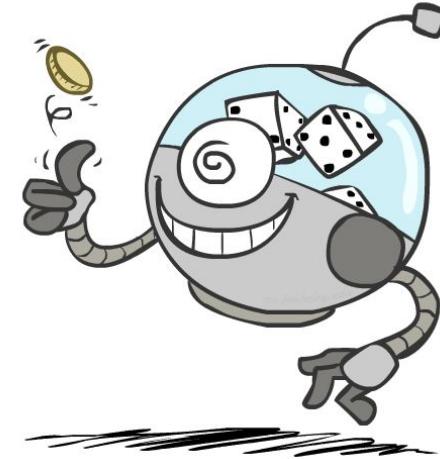


# Uncertainty

- General situation:
  - Observed variables (evidence)  
Agent knows certain things about the state of the world  
(e.g., sensor readings or symptoms)
  - Unobserved variables  
Agent needs to reason about other aspects  
(e.g. where an object is or what disease is present)
  - Model  
Agent knows something about how the known variables relate to the unknown variables
- Probabilistic reasoning gives us a framework for managing our beliefs and knowledge
- Assume model is given (in this chapter)

# Probability Refresher

- Random Variables
- Joint and Marginal Distributions
- Conditional Distribution
- Product Rule, Chain Rule, Bayes' Rule
- Inference
- Independence



The content in this refresher is important. In this and some of the following chapters we assume that you have the above background.

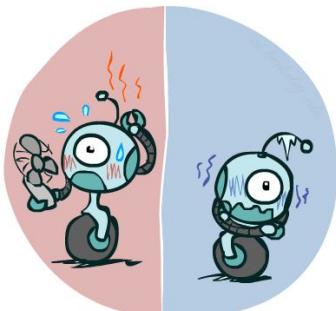
# Random Variables

- A random variable is some aspect of the world about which we (may) have uncertainty
  - $R$  = Is it raining?
  - $T$  = Is it hot or cold?
  - $D$  = How long will it take to drive to work?
  - $L$  = Where is the ghost?
- We denote random variables with capital letters
- Like variables in a CSP, random variables have domains
  - $R$  in {true, false} (often write as  $\{+r, -r\}$ )
  - $T$  in {hot, cold}
  - $D$  in  $[0, \infty)$
  - $L$  in possible locations, maybe  $\{(0,0), (0,1), \dots\}$

# Probability Distributions

- Associate a probability with each value

Temperature



T	P(T)
hot	0.5
cold	0.5

Weather



W	P(W)
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

# Probability Distributions

- Unobserved random variables have distributions
  - A distribution is a table of probabilities of values
- A probability (lower case value) is a single number
  - $P(W = \text{rain}) = 0.1$
- The following must be true

$$\forall x \ P(X = x) \geq 0 \quad \sum_x P(X = x) = 1$$

- Shorthand notation (ok if all domain entries are unique)
  - $P(\text{hot}) = P(T = \text{hot})$
  - $P(\text{cold}) = P(T = \text{cold})$
  - $P(\text{rain}) = P(W = \text{rain})$

W	P(W)
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

T	P(T)
hot	0.5
cold	0.5

# Joint Distributions

- A joint distribution over a set of random variables:  $X_1, X_2, \dots, X_n$  specifies a real number for each assignment (or outcome)

- $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$
- $P(x_1, x_2, \dots, x_n)$

- Must obey:

$$P(x_1, x_2, \dots, x_n) \geq 0$$

$$\sum_{(x_1, x_2, \dots, x_n)} P(x_1, x_2, \dots, x_n) = 1$$

T	W	$P(T,W)$
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

- What is the size of the distribution?
  - Let there be  $n$  variables with domain size  $d$

# Probabilistic Models

- A probabilistic model is a joint distribution over a set of random variables
- Probabilistic models:
  - (Random) variables with domains
  - Assignments are called outcomes
  - Joint distributions: say whether assignments (outcomes) are likely
  - Normalized: sum to 1.0
  - Ideally: only certain variables directly interact
- Constraint satisfaction problems:
  - Variables with domains
  - Constraints: state whether assignments are possible
  - Ideally: only certain variables directly interact

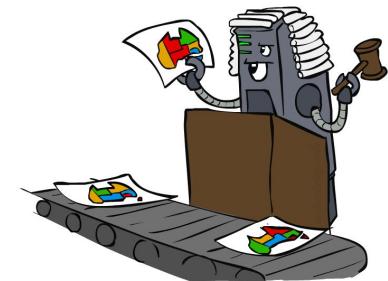
Distribution over T,W

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3



Constraint over T,W

T	W	
hot	sun	T
hot	rain	F
cold	sun	F
cold	rain	T



# Events

- An event is a set E of outcomes

$$P(E) = \sum_{(x_1 \dots x_n) \in E} P(x_1 \dots x_n)$$

- From a joint distribution, we can calculate the probability of any event
  - Probability that it's hot and sunny
  - Probability that it's hot
  - Probability that it's hot or sunny
- Typically, the events we care about are partial assignments, like  $P(T = \text{hot})$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

# Quiz: Events

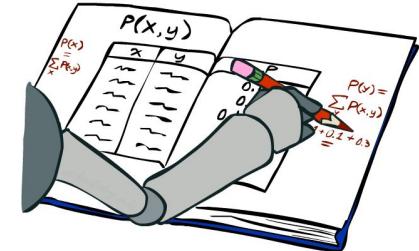
- $P(+x, +y) =$
- $P(+x) =$
- $P(-y \text{ OR } +x) =$

X	Y	$P(X, Y)$
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

# Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding

$$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$$



T	W	P(T, W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

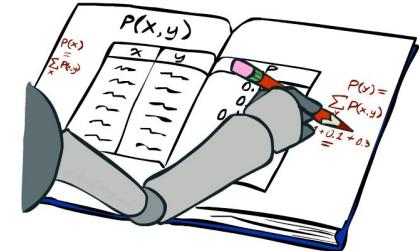
$$\rightarrow P(t) = \sum_s P(t, s) \rightarrow$$

T	P(T)
hot	0.5
cold	0.5

$$\rightarrow P(s) = \sum_t P(t, s) \rightarrow$$

W	P(W)
sun	0.6
rain	0.4

# Quiz: Marginal Distributions



X	Y	P(X, Y)
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

→  $P(x) = \sum_y P(x,y)$  →

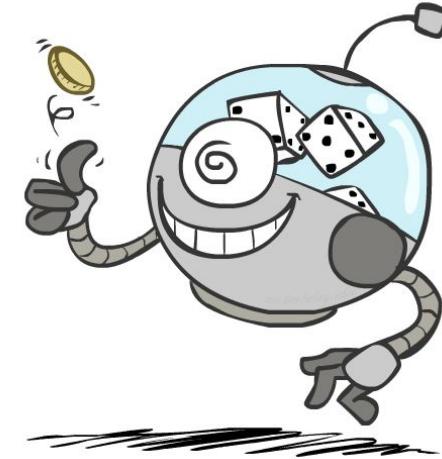
X	P(X)
+x	
-x	

→  $P(y) = \sum_x P(x,y)$  →

Y	P(Y)
+y	
-y	

# Probability Refresher

- Random Variables
- Joint and Marginal Distributions
- Conditional Distribution
- Product Rule, Chain Rule, Bayes' Rule
- Inference
- Independence



The content in this refresher is important. In this and some of the following chapters we assume that you have the above background.

# Conditional Probabilities

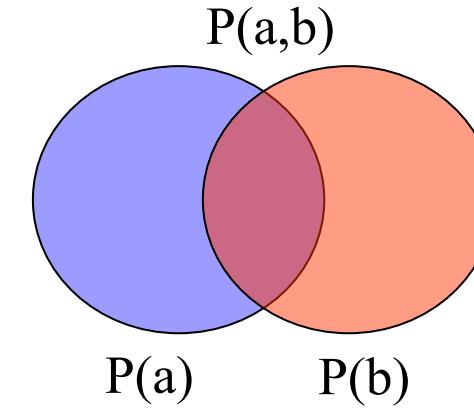
- A simple relation between joint and conditional probabilities
  - Important!
    - If we measure something, what does that tell us about the things we don't measure?

The probability  
of having  $A=a$   
given  $B=b$

The probability  
of having  $A=a$   
and  $B=b$

$$P(a|b) = \frac{P(a, b)}{P(b)}$$

The probability  
of having  $B=b$



# Conditional Probabilities - Example

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(W = s|T = c) = ???$$

$$P(W = s|T = c) = \frac{P(W = s, T = c)}{P(T = c)} = \frac{0.2}{0.5} = 0.4$$

$$\begin{aligned} &= P(W = s, T = c) + P(W = r, T = c) \\ &= 0.2 + 0.3 = 0.5 \end{aligned}$$

# Quiz: Conditional Probabilities

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

$$P(+x \mid +y) =$$

X	Y	P(X,Y)
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

$$P(-x \mid +y) =$$

$$P(-y \mid +x) =$$

# Conditional Distributions

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

- Conditional distributions are probability distributions over some variables given fixed values of others

Conditional Distributions

Set of tables!

		$P(W T)$
		$P(W T = \text{hot})$
$T$	W	
	sun	0.8
		$P(W T = \text{cold})$
$T$	W	
	sun	0.4
		$P(W T = \text{hot})$
$T$	W	
	rain	0.6

Joint Distribution



$T$	$W$	$P(T,W)$
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

# Normalization

- Here is what we have done just now:

$$P(W = s|T = c) = \frac{P(W = s, T = c)}{P(T = c)}$$

Joint Distribution

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$= \frac{P(W = s, T = c)}{P(W = s, T = c) + P(W = r, T = c)}$$

$$= \frac{0.2}{0.2 + 0.3} = 0.4$$

Conditional Distribution

W	P(W T = c)
sun	0.4
rain	0.6

$$P(W = r|T = c) = \frac{P(W = r, T = c)}{P(T = c)}$$

$$= \frac{P(W = r, T = c)}{P(W = s, T = c) + P(W = r, T = c)}$$

$$= \frac{0.3}{0.2 + 0.3} = 0.6$$

# Normalization

- Equivalent approach:

Joint Distribution

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

**SELECT** the joint probabilities matching the evidence

$$P(W = s|T = c) =$$

$$P(W = r|T = c) =$$

T	W	P(c, W)
cold	sun	0.2
cold	rain	0.3

**NORMALIZE** the selection (make it sum to one)

W	P(W T = c)
sun	0.4
rain	0.6

Conditional Distribution

# Quiz: Normalization

X	Y	P(X, Y)
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

$P(X| -y) =$

**SELECT** the joint  
probabilities  
matching the  
evidence



**NORMALIZE** the selection  
(make it sum to one)



# To Normalize

Sum to one

- (Dictionary) To bring or restore to a normal condition
- Procedure:
  - Step 1: Compute  $Z = \text{sum over all entries}$
  - Step 2: Divide every entry by  $Z$

W	P
sun	0.2
rain	0.3

Normalize

$Z = 0.5$

T	W	P
hot	sun	20
hot	rain	5
cold	sun	10
cold	rain	15

Normalize

$Z = 50$

# Probabilistic Inference

- Compute a desired probability from other known probabilities (e.g. conditional from joint)
- We generally compute conditional probabilities
  - $P(\text{on time} \mid \text{no reported accidents}) = 0.90$
  - These represent the agent's beliefs given the evidence
- Probabilities change with new evidence:
  - $P(\text{on time} \mid \text{no accidents, 5 a.m.}) = 0.95$
  - $P(\text{on time} \mid \text{no accidents, 5 a.m., raining}) = 0.80$
  - Observing new evidence causes beliefs to be updated



# Inference by Enumeration

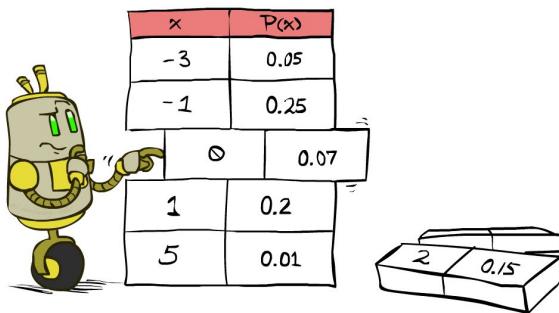
- General case

- Evidence variables:  $E_1 \dots E_k = e_1 \dots e_k$
- Query variable:  $Q$
- Hidden variables:  $H_1 \dots H_r$

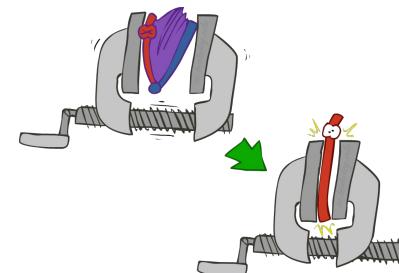
$X_1, X_2, \dots X_n$   
*All variables*

We want:  $P(Q|e_1 \dots e_k)$

Step 1: Select the entries  
consistent with the evidence



Step 2: Sum out H to get joint of  
Query and evidence



$$P(Q, e_1 \dots e_k) = \sum_{h_1 \dots h_r} P(Q, h_1 \dots h_r, e_1 \dots e_k)$$

$X_1, X_2, \dots X_n$

$$\times \frac{1}{Z}$$

$$Z = \sum_q P(Q, e_1 \dots e_k)$$

$$P(Q|e_1 \dots e_k) = \frac{1}{Z} P(Q, e_1 \dots e_k)$$

# Inference by Enumeration - Example

- $P(W) = ?$ 
  - $Q = W, E = \emptyset, H = \{S, T\}$

2	W	P(W)
sun	0.65	
rain	0.35	

3	W	P(W)
sun	0.65	
rain	0.35	

1	S	T	W	P(S, T, W)
summer	hot	sun	0.30	
summer	hot	rain	0.05	
summer	cold	sun	0.10	
summer	cold	rain	0.05	
winter	hot	sun	0.10	
winter	hot	rain	0.05	
winter	cold	sun	0.15	
winter	cold	rain	0.20	

# Inference by Enumeration - Example

- $P(W | \text{winter}) = ?$
- $Q = W, E = S, H = T$

2	W	P(W)
	sun	0.25
	rain	0.25

3	W	P(W)
	sun	0.5
	rain	0.5

1	S	T	W	P(S, T, W)
	summer	hot	sun	0.30
	summer	hot	rain	0.05
	summer	cold	sun	0.10
	summer	cold	rain	0.05
	winter	hot	sun	0.10
	winter	hot	rain	0.05
	winter	cold	sun	0.15
	winter	cold	rain	0.20

# Inference by Enumeration - Example

- $P(W | \text{winter, hot}) = ?$
- $Q = W, E = S, T, H = \emptyset$

2	W	P(W)	
	sun	0.10	
	rain	0.05	

3	W	P(W)	
	sun	0.67	
	rain	0.33	

	S	T	W	P(S, T, W)
	summer	hot	sun	0.30
	summer	hot	rain	0.05
	summer	cold	sun	0.10
	summer	cold	rain	0.05
1	winter	hot	sun	0.10
	winter	hot	rain	0.05
	winter	cold	sun	0.15
	winter	cold	rain	0.20

# Inference by Enumeration

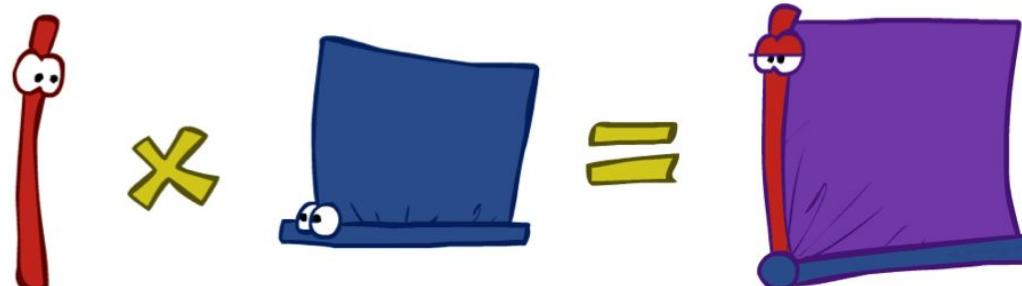
- Time:  $O(d^n)$
- Space:  $O(d^n)$

# The Product Rule

- Sometimes have conditional distributions but want the joint

$$P(x|y) = \frac{P(x,y)}{P(y)} \quad \longleftrightarrow \quad P(y)P(x|y) = P(x,y)$$

Definition of  
Conditional  
Probability



# The Product Rule

$$P(y)P(x|y) = P(x, y)$$

- Example

		T	P(T hot)
		sun	0.8
		rain	0.2
W	P(W)	T	P(T cold)
		sun	0.4
hot	0.5	rain	0.6
cold	0.5		

W	T	P(W,T)
hot	sun	.5 * .8 = .4
	rain	.5 * .2 = .1
cold	sun	.5 * .4 = .2
	rain	.5 * .6 = .3

# Quiz: Product Rule

$$P(y)P(x|y) = P(x,y)$$

	W	P(W)
sun	wet	0.8
rain	dry	0.2

D	P(D sun)
wet	0.1
dry	0.9

D	P(D rain)
wet	0.7
dry	0.3

D	W	P(D,W)
wet	sun	
dry	sun	
wet	rain	
dry	rain	

# The Chain Rule

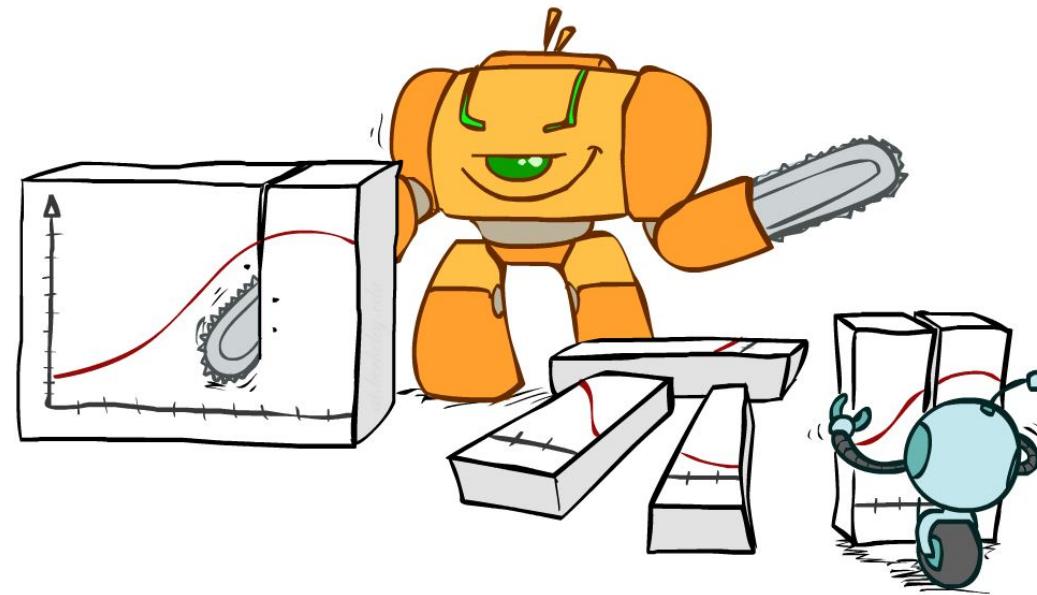
- More generally, can always write any joint distribution as an incremental product of conditional distributions

$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$

$$P(x_1, x_2, \dots, x_n) = \prod_i P(x_i|x_1 \dots x_{i-1})$$

- Why is this always true?

# Bayes Rule



# Bayes' Rule

$$P(y)P(x|y) = P(x,y)$$

- Two ways to factor a joint distribution over two variables:

$$P(x,y) = P(x|y)P(y) = P(y|x)P(x)$$

- Dividing, we get:

$$P(x|y) = \frac{P(y|x)}{P(y)}P(x)$$

- Why is this helpful?
  - Lets us build one conditional from its reverse
  - Often one conditional is tricky but the other one is simple

# Example: Inference with Bayes' Rule

$$P(x|y) = \frac{P(y|x)}{P(y)} P(x)$$

- Diagnostic probability from causal probability

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}$$

- M: meningitis, S: stiff neck
- Given:
  - $P(+m) = 0.0001$
  - $P(+s | +m) = 0.8$
  - $P(+s) = 0.01$
- $P(+m | +s) = ?$

$$P(x|y) = \frac{P(y|x)}{P(y)} P(x)$$

# Quiz: Bayes' Rule

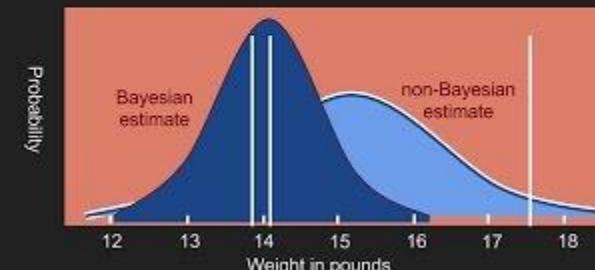
W	P(W)	D	W	P(D W)
sun	0.8	wet	sun	0.1
rain	0.2	dry	sun	0.9
		wet	rain	0.7
		dry	rain	0.3

Distribution!

$$P(W | \text{dry}) = ?$$

# More on Bayes' Rule

## How Bayesian Inference Works



by Brandon Rohrer

$$P(A | B) = \frac{P(B | A) * P(A)}{P(B)}$$

# Ghostbusters, Revisited

- Let's say we have two distributions:
  - Prior distribution over ghost location:  $P(G)$ 
    - Let this be uniform
  - Sensor reading model:  $P(R|G)$ 
    - Given: we know what our sensors do
      - $R$  = color reading
      - $G$  = ghost location
    - E.g.  $P(R = \text{yellow} | G = (1,1)) = 0.1$
- We can calculate the posterior distribution  $P(G|r)$  over ghost locations given a reading  $r$  using Bayes' rule:

$$P(g|r) \propto P(r|g)P(g)$$

$$P(x|y) = \frac{P(y|x)}{P(y)}P(x)$$

# Ghostbusters with Probability

- Sensor readings tell how close a square is to the ghost
    - On the ghost: red
    - 1 or 2 away: orange
    - 3 or 4 away: yellow
    - 5+ away: green

# Ghostbusters with Probability

- Sensor readings tell how close a square is to the ghost
  - On the ghost: red
  - 1 or 2 away: orange
  - 3 or 4 away: yellow
  - 5+ away: green

0.04	0.04	0.04	<0.01	<0.01	<0.01	0.04	0.04	0.04	0.04
0.04	0.04	<0.01	<0.01	<0.01	<0.01	<0.01	0.04	0.04	0.04
0.04	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.04	0.04
<0.01	<0.01	<0.01	<0.01	<0.01	0.04	<0.01	<0.01	<0.01	0.04
0.04	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.04	0.04
0.04	0.04	<0.01	<0.01	<0.01	<0.01	<0.01	0.04	0.04	0.04

# Ghostbusters with Probability

- Sensor readings tell how close a square is to the ghost
  - On the ghost: red
  - 1 or 2 away: orange
  - 3 or 4 away: yellow
  - 5+ away: green

0.02	0.02	0.02	<0.01	<0.01	0.01	0.13	0.01	<0.01	<0.01
0.02	0.02	<0.01	<0.01	<0.01	<0.01	0.01	0.13	0.01	<0.01
0.02	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.13	0.01
<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.13
0.02	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.02	0.13
0.02	0.02	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.02	0.02

# Ghostbusters with Probability

- Sensor readings tell how close a square is to the ghost
    - On the ghost: red
    - 1 or 2 away: orange
    - 3 or 4 away: yellow
    - 5+ away: green

# Ghostbusters with Probability

- Sensor readings tell how close a square is to the ghost
    - On the ghost: red
    - 1 or 2 away: orange
    - 3 or 4 away: yellow
    - 5+ away: green

# Independence

- Two variables are independent in a joint distribution if

$$P(X, Y) = P(X)P(Y)$$

$$X \perp\!\!\!\perp Y$$

$$\forall x, y P(x, y) = P(x)P(y)$$

- I.e., the joint distribution factors into a product of two simple ones
  - Note: Usually, in practice, variables are not independent
- Can use independence as a modeling assumption
  - Simplifying assumption
  - Empirical joint distributions: at best “close” to independent
  - What could we assume for {Weather, Traffic, Cavity}?
- Independence is like something from CSPs: what?



# Example: Independence check

- Given: Distribution
- Task: Check for independence

T	W	$P(T, W)$
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

T	$P(T)$
hot	0.5
cold	0.5

W	$P(W)$
sun	0.6
rain	0.4

T	W	$P = P(T) P(W)$
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

# Example: Independence

- N fair, independent coin flips

$$P(X_1)$$

H	0.5
T	0.5

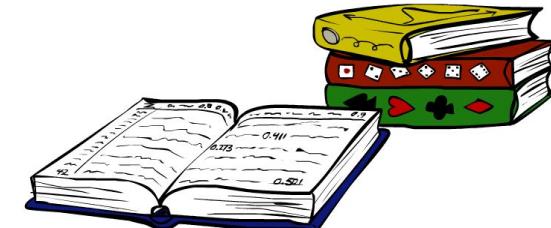
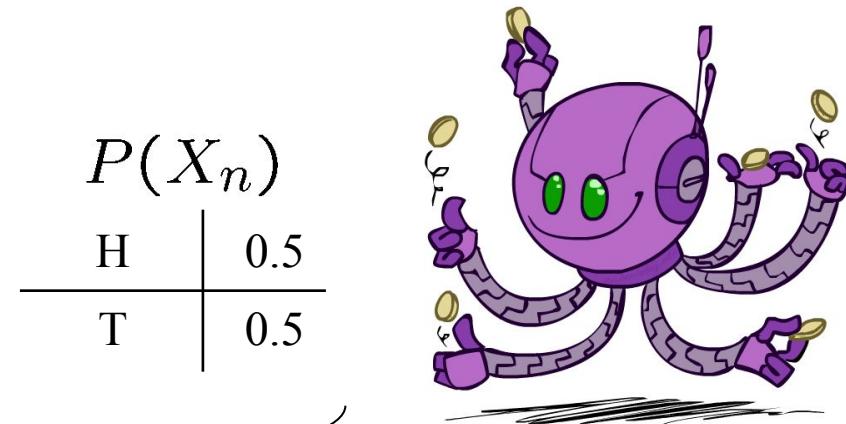
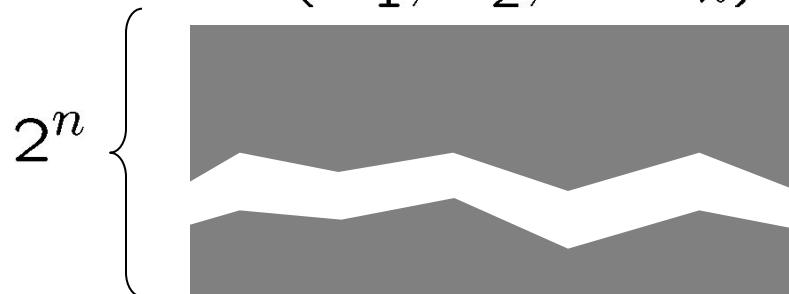
$$P(X_2)$$

H	0.5
T	0.5

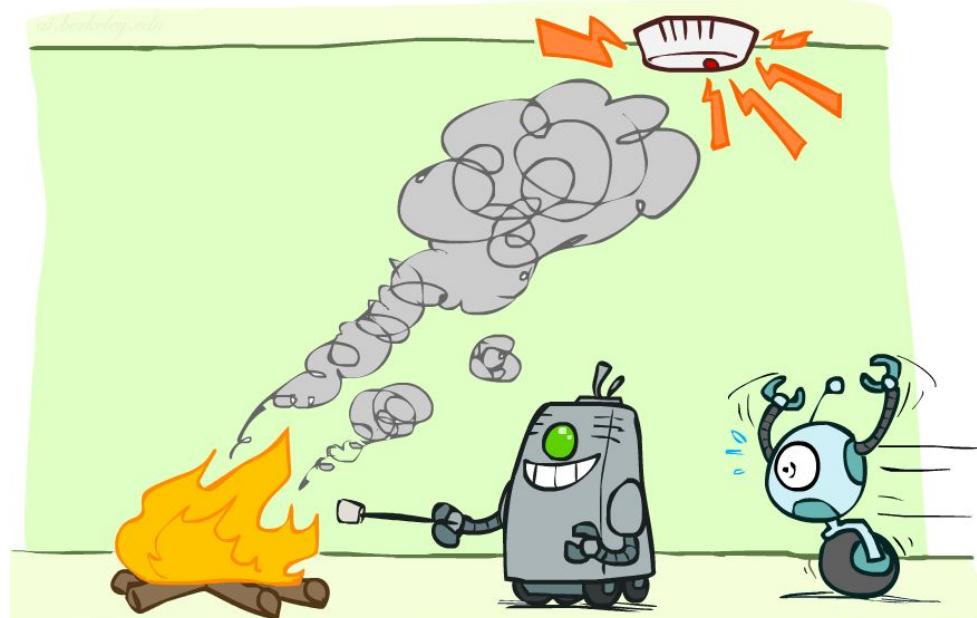
...

$$P(X_n)$$

H	0.5
T	0.5

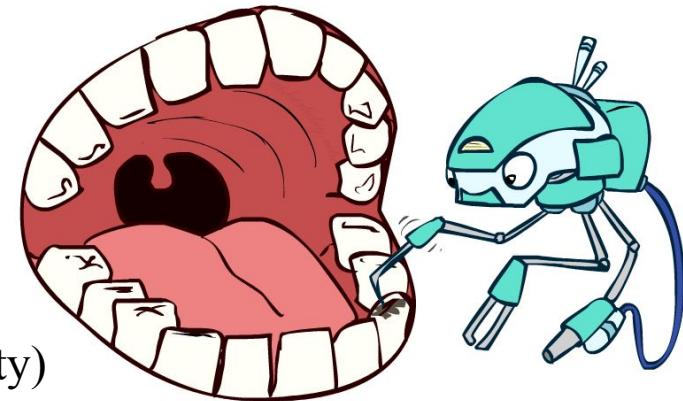


# Conditional Independence



# Conditional Independence

- $P(\text{toothache}, \text{cavity}, \text{catch})$
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - $P(+\text{catch} | +\text{toothache}, +\text{cavity}) = P(+\text{catch} | +\text{cavity})$
- The same independence holds if I don't have a cavity:
  - $P(+\text{catch} | +\text{toothache}, -\text{cavity}) = P(+\text{catch} | -\text{cavity})$
- Catch is conditionally independent of toothache given cavity:
  - $P(\text{catch} | \text{toothache}, \text{cavity}) = P(\text{catch} | \text{cavity})$
  - We also write:  $\text{catch} \perp\!\!\!\perp \text{toothache} | \text{cavity}$
- Equivalent statement:
  - $P(\text{toothache} | \text{catch}, \text{cavity}) = P(\text{toothache} | \text{cavity})$
  - We also write:  $\text{toothache} \perp\!\!\!\perp \text{catch} | \text{cavity}$



# Conditional Independence

- Unconditional (absolute) independence very rare
- Conditional independence is our most basic and robust form of knowledge about uncertain environments

$$X \perp\!\!\!\perp Y | Z$$

- $X$  is conditionally independent of  $Y$  given  $Z$  iff

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

or, equivalently, iff

$$\forall x, y, z : P(x|z, y) = P(x|z)$$

# Quiz - Conditional Independence Proof

- Show that the statement  $\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$  is equivalent to  $\forall x, y, z : P(x|z, y) = P(x|z)$

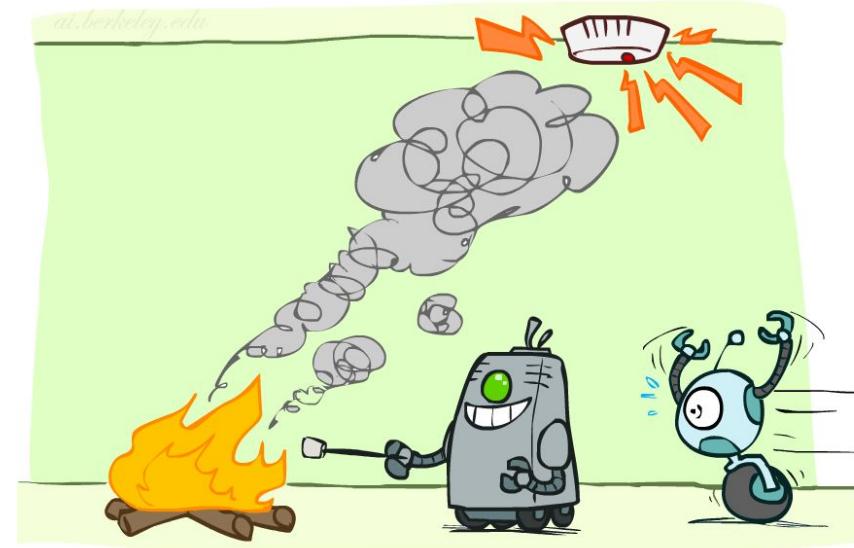
# Quiz: Conditional Independence

- What about this domain
  - Traffic (T)
  - Umbrella (U)
  - Raining (R)



# Quiz: Conditional Independence

- What about this domain
  - Fire (F)
  - Smoke (S)
  - Alarm (A)



# Probability Recap

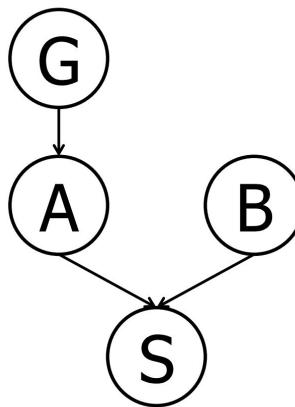
- Conditional probability  $P(x|y) = \frac{P(x,y)}{P(y)}$
- Product rule  $P(x,y) = P(x|y)P(y)$
- Chain rule 
$$\begin{aligned} P(X_1, X_2, \dots, X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\dots \\ &= \prod_{i=1}^n P(X_i|X_1, \dots, X_{i-1}) \end{aligned}$$
- X, Y independent if and only if:  $\forall x, y : P(x,y) = P(x)P(y)$
- X and Y are conditionally independent given Z if and only if:  
$$\begin{aligned} \forall x, y, z : P(x, y|z) &= P(x|z)P(y|z) \\ \forall x, y, z : P(x|z, y) &= P(x|z) \end{aligned} \qquad \qquad \qquad X \perp\!\!\!\perp Y | Z$$

# Quiz - G, S, A, B

- Suppose that a patient can have a symptom (S) that can be caused by two different diseases (A and B)
- It is known that the variation of gene G plays a big role in the manifestation of disease A
- The conditional probability tables for this situation are shown below
- Calculate the following
  - 1.  $P(+g, +a, +b, +s)$
  - 2.  $P(+a)$
  - 3.  $P(+a|+b)$
  - 4.  $P(+a|+s, +b)$
  - 5.  $P(+g|+a)$
  - 6.  $P(+g|+b)$

$\mathbb{P}(G)$	
$+g$	0.1
$-g$	0.9

$\mathbb{P}(A G)$		
$+g$	$+a$	1.0
$+g$	$-a$	0.0
$-g$	$+a$	0.1
$-g$	$-a$	0.9



$\mathbb{P}(B)$	
$+b$	0.4
$-b$	0.6

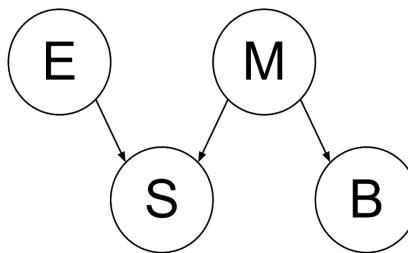
$\mathbb{P}(S A, B)$			
$+a$	$+b$	$+s$	1.0
$+a$	$+b$	$-s$	0.0
$+a$	$-b$	$+s$	0.9
$+a$	$-b$	$-s$	0.1
$-a$	$+b$	$+s$	0.8
$-a$	$+b$	$-s$	0.2
$-a$	$-b$	$+s$	0.1
$-a$	$-b$	$-s$	0.9

# Quiz - E, S, M, B

- A smell of sulphur (S) can be caused either by rotten eggs (E) or as a sign of the doom brought by the Mayan Apocalypse (M)
- The Mayan Apocalypse also causes the oceans to boil (B)
- The conditional probability tables for this situation are shown below
- Calculate the following
  - $P(\neg e, \neg s, \neg m, \neg b)$
  - $P(+b)$
  - $P(+m| + b)$
  - $P(+m| + s, +b, +e)$
  - $P(+e| + m)$

$P(E)$	
$+e$	0.4
$\neg e$	0.6

$P(S E, M)$			
$+e$	$+m$	$+s$	1.0
$+e$	$+m$	$\neg s$	0.0
$+e$	$\neg m$	$+s$	0.8
$+e$	$\neg m$	$\neg s$	0.2
$\neg e$	$+m$	$+s$	0.3
$\neg e$	$+m$	$\neg s$	0.7
$\neg e$	$\neg m$	$+s$	0.1
$\neg e$	$\neg m$	$\neg s$	0.9

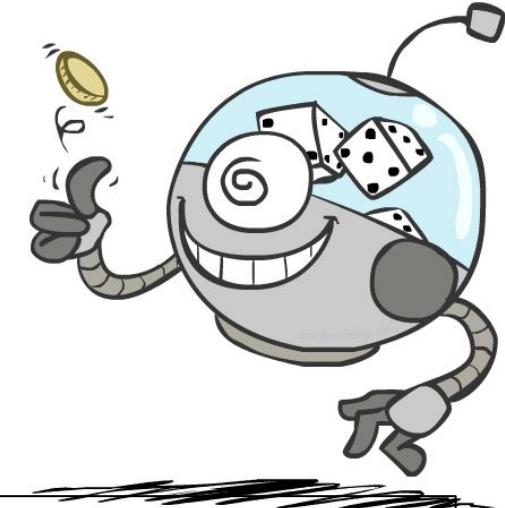


$P(M)$	
$+m$	0.1
$\neg m$	0.9

$P(B M)$		
$+m$	$+b$	1.0
$+m$	$\neg b$	0.0
$\neg m$	$+b$	0.1
$\neg m$	$\neg b$	0.9

# Probability Refresher

- Random Variables
- Joint and Marginal Distributions
- Conditional Distribution
- Product Rule, Chain Rule, Bayes' Rule
- Inference
- Independence

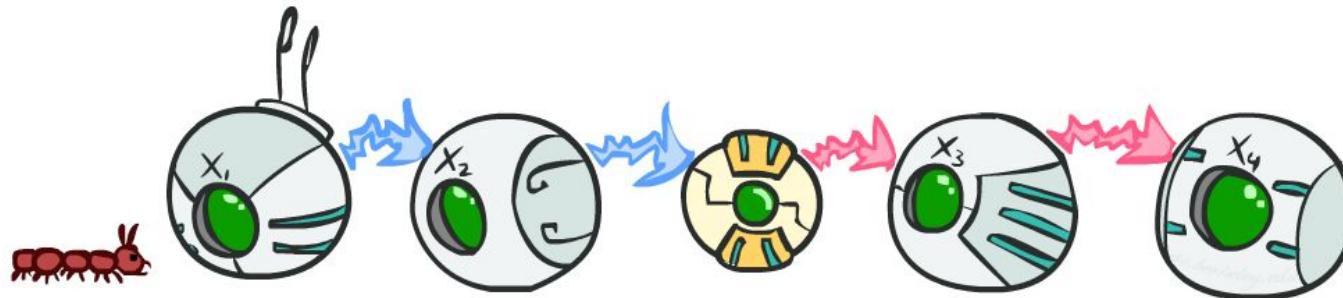


DONE

# Outline

- Motivation
- Probabilistic Reasoning
  - Running Example: Ghostbusters
  - Probability Refresher
    - Random Variables, Joint and Marginal Distributions, Conditional Distribution, Product Rule, Chain Rule, Bayes' Rule
  - Inference, Independence
- Probabilistic Reasoning over Time or Space
  - Markov Model
  - Mini-Forward Algorithm
  - Stationary Distributions
- Next chapter: Hidden Markov Models

# Markov Model



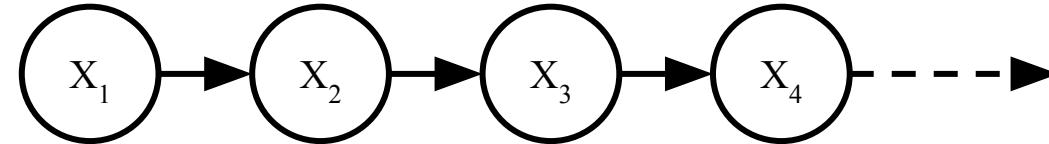
Introducing time (or space) into our models

# Reasoning over Time or Space

- Often, we want to reason about a sequence of observations
  - Robot localization
    - From noisy readings of sensors
  - User attention
    - When is a good time to interrupt (popup-ad) a user?
  - Speech recognition
    - Inference on the acoustic sequence of words
  - Medical monitoring
    - Is there a medical emergency occurring ?
- Need to introduce time (or space) into our models

# Markov Model

- Chain structure

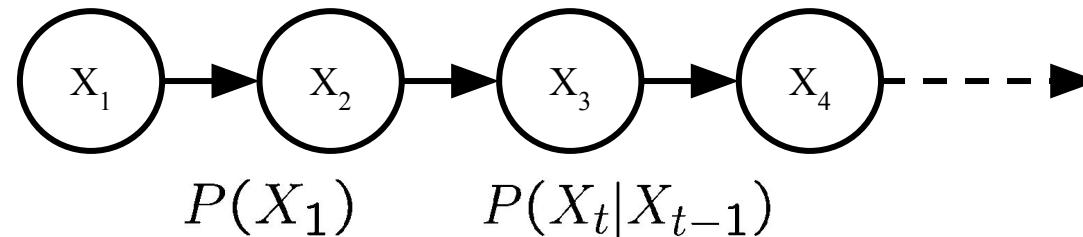


- Value of  $X$  at a given time is called the state
- We assume that future states depend only on the current state
  - not on previous states

$$P(X_1) \quad P(X_t | X_{t-1})$$

- Parameters: called transition probabilities or dynamics, specify how the state evolves over time (also, initial state probabilities)
  - Stationarity assumption
    - transition probabilities **the same at all times**
  - Same as MDP transition model, but no choice of action

# Joint Distribution of a Markov Model



- Joint distribution:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

- More generally:

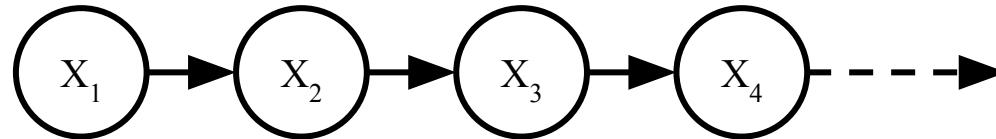
$$P(X_1, X_2, \dots, X_T) = P(X_1)P(X_2|X_1)P(X_3|X_2)\dots P(X_T|X_{T-1})$$

$$= P(X_1) \prod_{t=2}^T P(X_t|X_{t-1})$$

- Questions to be resolved:

- Does this indeed define a joint distribution?
- Can every joint distribution be factored this way, or are we making some assumptions about the joint distribution by using this factorization?

# Chain Rule and Markov Models



- From the chain rule, every joint distribution over  $X_1, X_2, X_3, X_4$  can be written as:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)P(X_4|X_1, X_2, X_3)$$

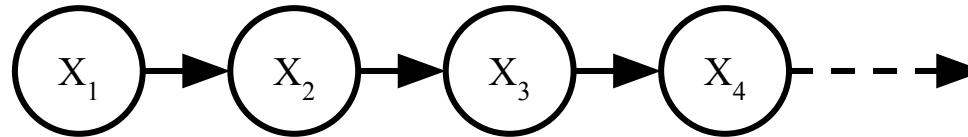
- Assuming that

$$X_3 \perp\!\!\!\perp X_1 \mid X_2 \quad \text{and} \quad X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$$

results in the expression

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

# Chain Rule and Markov Models



- From the chain rule, every joint distribution over  $X_1, X_2, \dots, X_T$  can be written as:

$$P(X_1, X_2, \dots, X_T) = P(X_1) \prod_{t=2}^T P(X_t | X_1, X_2, \dots, X_{t-1})$$

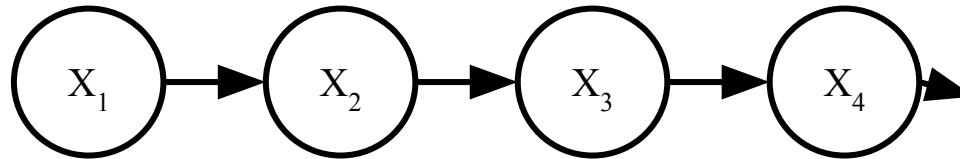
- Assuming that

$$X_t \perp\!\!\!\perp X_1, \dots, X_{t-2} \mid X_{t-1}$$

results in the expression

$$P(X_1, X_2, \dots, X_T) = P(X_1) \prod_{t=2}^T P(X_t | X_{t-1})$$

# Quiz - Chain Rule and Markov Models



- We assumed  $X_3 \perp\!\!\!\perp X_1 | X_2$  and  $X_4 \perp\!\!\!\perp X_1, X_2 | X_3$
- Do we also have  $X_1 \perp\!\!\!\perp X_3, X_4 | X_2$  ?

# Markov Models Recap

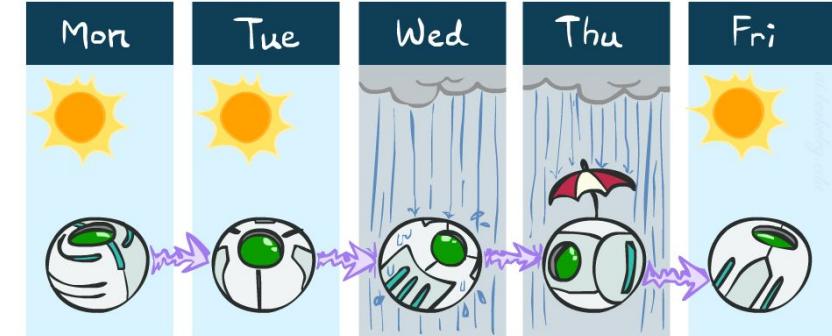
- Explicit assumption for all t:  $X_t \perp\!\!\!\perp X_1, \dots, X_{t-2} \mid X_{t-1}$
- Consequence, joint distribution can be written as:

$$\begin{aligned} P(X_1, X_2, \dots, X_T) &= P(X_1)P(X_2|X_1)P(X_3|X_2)\dots P(X_T|X_{T-1}) \\ &= P(X_1) \prod_{t=2}^T P(X_t|X_{t-1}) \end{aligned}$$

- Implied conditional independencies:
  - Past variables independent of future variables given the present
  - i.e., if  $t_1 < t_2 < t_3$  or  $t_1 > t_2 > t_3$  then:  $X_{t_1} \perp\!\!\!\perp X_{t_3} \mid X_{t_2}$
- Additional explicit assumption:  $P(X_t \mid X_{t-1})$  is the same for all t

# Example Markov Chain: Weather

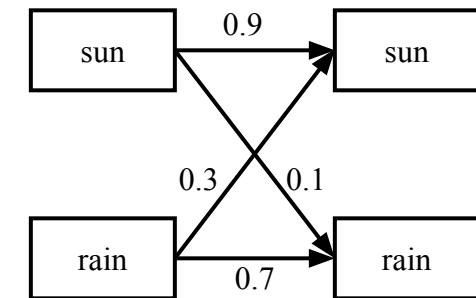
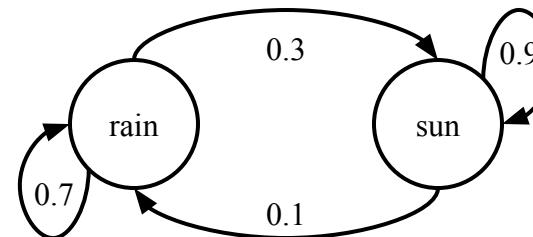
- States:  $X = \{\text{rain, sun}\}$
- Initial distribution: 1.0 sun



Conditional Probability Table (CPT):

$X_{t-1}$	$X_t$	$P(X_t X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Two new ways of representing the same CPT

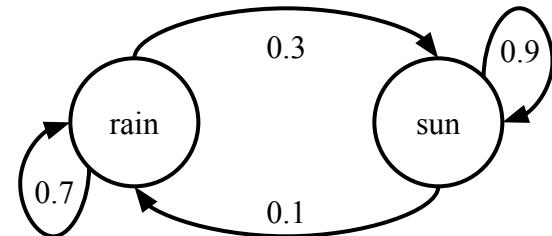


# Example Markov Chain: Weather

- Initial distribution: 1.0 sun
- What is the probability distribution after one step?

$$P(X_2 = \text{sun}) = ?$$

$$P(X_2 = \text{rain}) = ?$$



$$P(X_2 = \text{sun}) = P(X_2 = \text{sun} | X_1 = \text{sun})P(X_1 = \text{sun}) + P(X_2 = \text{sun} | X_1 = \text{rain})P(X_1 = \text{rain})$$

$$P(X_2 = \text{sun}) = .9 * 1.0 + 0.3 * 0.0 = 0.9$$

$$P(X_2 = \text{rain}) = P(X_2 = \text{rain} | X_1 = \text{sun})P(X_1 = \text{sun}) + P(X_2 = \text{rain} | X_1 = \text{rain})P(X_1 = \text{rain})$$

$$P(X_2 = \text{rain}) = .1 * 1.0 + 0.7 * 0.0 = 0.1$$

# Quiz - Markov Chain: Weather

- Initial distribution: 1.0 sun
- What is the probability distribution after two steps?

Distribution after one step:

0.9 sun

0.1 rain

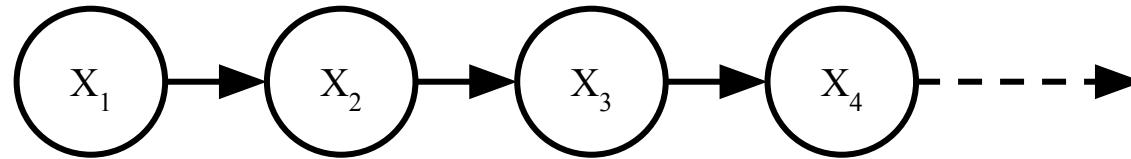
$$P(X_3 = \text{sun}) = ?$$

$$P(X_3 = \text{rain}) = ?$$

$X_{t-1}$	$X_t$	$P(X_t   X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

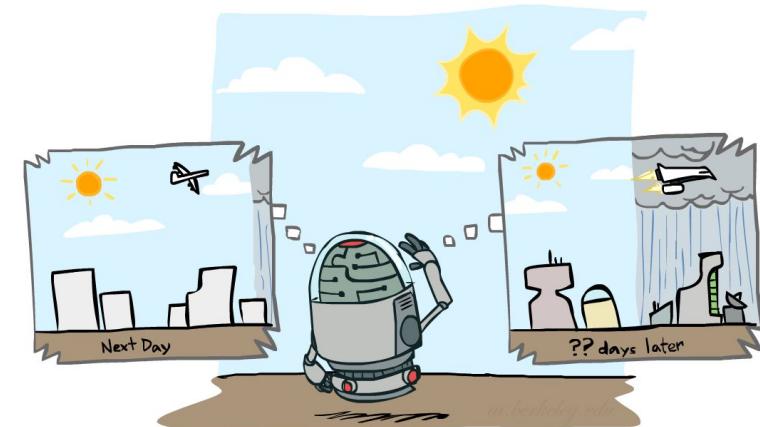
# Mini-Forward Algorithm

- Question: What's  $P(X)$  on some day  $t$ ?



$P(x_1) = \text{known}$

$$\begin{aligned} P(x_t) &= \sum_{x_{t-1}} P(x_{t-1}, x_t) \\ &= \sum_{x_{t-1}} P(x_t \mid x_{t-1}) P(x_{t-1}) \end{aligned}$$



# Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\begin{array}{c} \left\langle \begin{array}{c} 1.0 \\ 0.0 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.9 \\ 0.1 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.84 \\ 0.16 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.804 \\ 0.196 \end{array} \right\rangle \longrightarrow \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle \\ P(X_1) \qquad P(X_2) \qquad P(X_3) \qquad P(X_4) \qquad \qquad \qquad P(X_\infty) \end{array}$$

- From initial observation of rain

$$\begin{array}{c} \left\langle \begin{array}{c} 0.0 \\ 1.0 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.3 \\ 0.7 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.48 \\ 0.52 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.588 \\ 0.412 \end{array} \right\rangle \longrightarrow \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle \\ P(X_1) \qquad P(X_2) \qquad P(X_3) \qquad P(X_4) \qquad \qquad \qquad P(X_\infty) \end{array}$$

- From initial observation of rain

$$\begin{array}{c} \left\langle \begin{array}{c} p \\ 1-p \end{array} \right\rangle \qquad \dots \qquad \longrightarrow \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle \\ P(X_1) \qquad \qquad \qquad \qquad \qquad \qquad P(X_\infty) \end{array}$$

# Stationary Distributions

- For most chains:
  - Influence of the initial distribution gets less and less over time
  - The distribution we end up in is independent of the initial distribution
- Stationary distribution:
  - The distribution we end up with is called the stationary distribution of the chain
  - It satisfies
$$P_\infty(X) = P_{\infty+1}(X) = \sum_x P(X|x)P_\infty(x)$$



# Example: Stationary Distributions

- We know:  $P_\infty(X) = P_{\infty+1}(X) = \sum_x P(X|x)P_\infty(x)$
- Question: What's  $P(X)$  at time  $t = \text{infinity}$ ?

$X_{t-1}$	$X_t$	$P(X_t X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

$$P_\infty(\text{sun}) = P(\text{sun}|\text{sun})P_\infty(\text{sun}) + P(\text{sun}|\text{rain})P_\infty(\text{rain})$$

$$P_\infty(\text{rain}) = P(\text{rain}|\text{sun})P_\infty(\text{sun}) + P(\text{rain}|\text{rain})P_\infty(\text{rain})$$

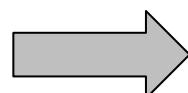
$$P_\infty(\text{sun}) = 0.9P_\infty(\text{sun}) + 0.3P_\infty(\text{rain})$$

$$P_\infty(\text{rain}) = 0.1P_\infty(\text{sun}) + 0.7P_\infty(\text{rain})$$

$$P_\infty(\text{sun}) = 3P_\infty(\text{rain})$$

$$P_\infty(\text{rain}) = 1/3P_\infty(\text{sun})$$

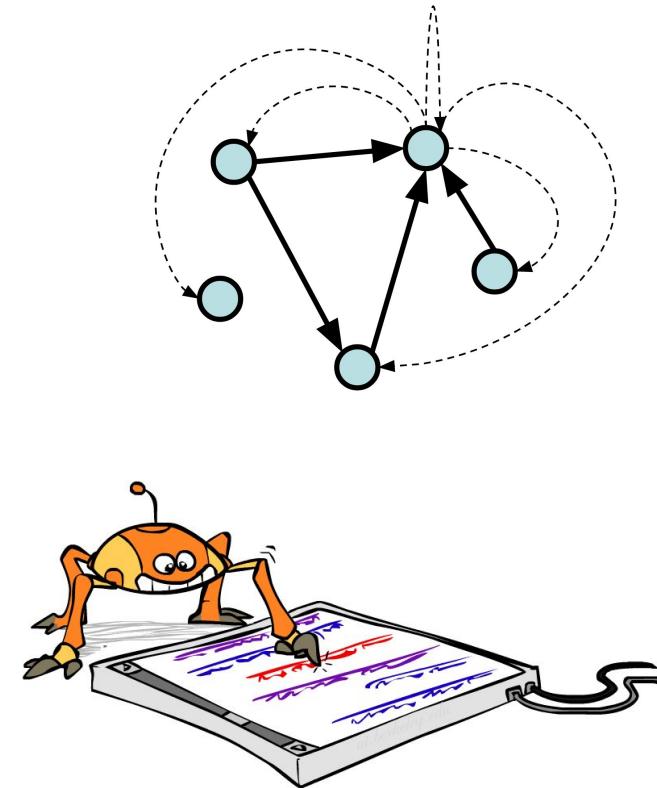
$$P_\infty(\text{sun}) + P_\infty(\text{rain}) = 1$$



$$\begin{aligned}P_\infty(\text{sun}) &= 3/4 \\P_\infty(\text{rain}) &= 1/4\end{aligned}$$

# Application of Stationary Distribution: Web Link Analysis

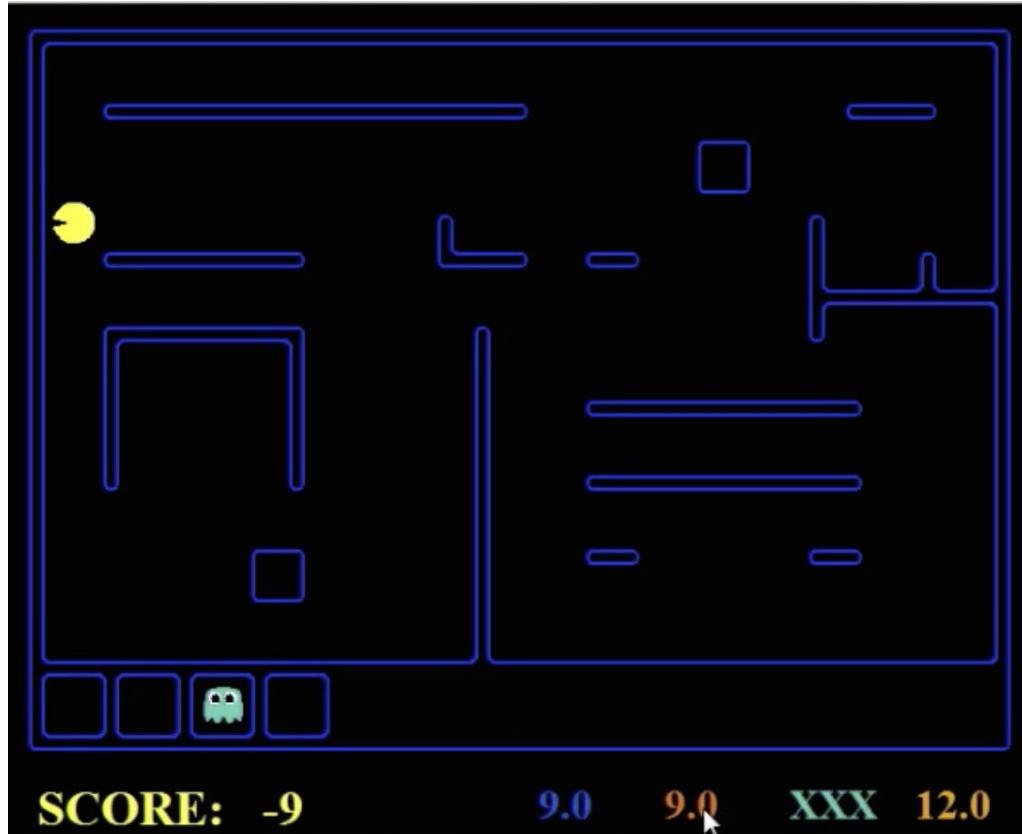
- PageRank over a web graph
  - Each web page is a state
  - Initial distribution: uniform over pages
- Transitions:
  - With prob.  $c$ , uniform jump to a random page (dotted lines, not all shown)
  - With prob.  $1-c$ , follow a random outlink (solid lines)
- Stationary distribution
  - Will spend more time on highly reachable pages
  - E.g. many ways to get to the hku.hk page
  - Somewhat robust to link spam
  - Google 1.0 returned the set of pages containing all your keywords in decreasing rank, now all search engines use link analysis along with many other factors (rank actually getting less important over time)



# Next chapter: Hidden Markov Models







**SCORE:** -9

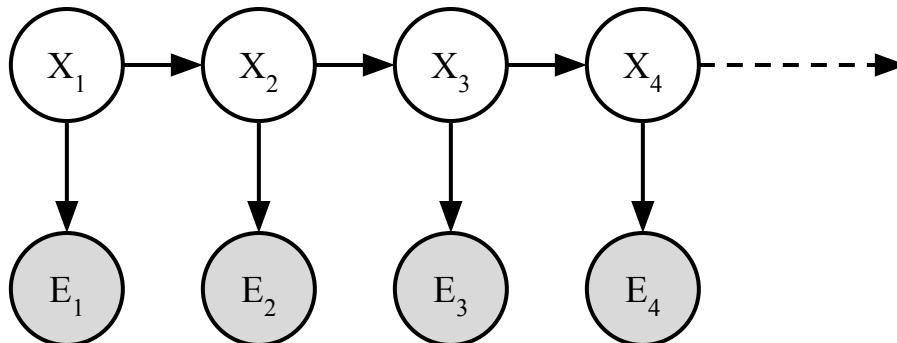
9.0

9.0

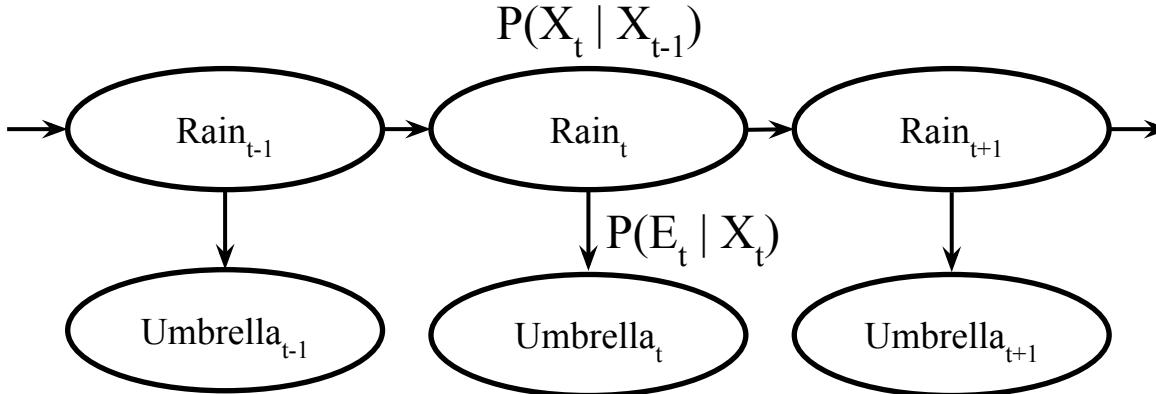
XXX 12.0

# Hidden Markov Models

- Markov chains not so useful for most agents
  - Need observations to update your beliefs
- Hidden Markov models (HMMs)
  - Underlying Markov chain over states  $X$
  - You observe outputs (effects) at each time step



# Example: Weather HMM



R <sub>t</sub>	R <sub>t+1</sub>	P(R <sub>t+1</sub>   R <sub>t</sub> )
+r	+r	0.7
+r	-r	0.3
-r	+r	0.3
-r	-r	0.7

R <sub>t</sub>	U <sub>t</sub>	P(U <sub>t</sub>   R <sub>t</sub> )
+r	+u	0.9
+r	-u	0.1
-r	+u	0.2
-r	-u	0.8

- An HMM is defined by
  - Initial distribution P(X<sub>1</sub>)
  - Transitions P(X<sub>t</sub> | X<sub>t-1</sub>)
  - Emissions P(E<sub>t</sub> | X<sub>t</sub>)