

CSC 665: Artificial Intelligence

Probability

Instructor: Pooyan Fazli
San Francisco State University

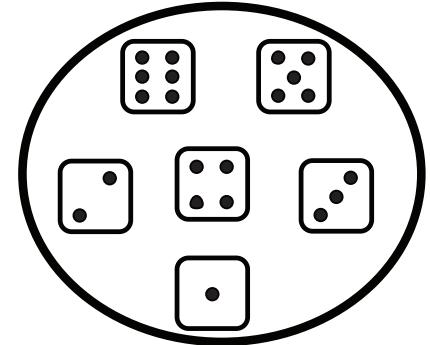
Today

- Probability
 - Random Variables
 - Joint and Marginal Distributions
 - Conditional Distribution
 - Product Rule, Chain Rule, Bayes' Rule
 - Independence
- You'll need all this stuff A LOT for the next couple of weeks, so make sure you go over it now!

Basic Laws of Probability

- Begin with a set Ω of possible worlds (outcomes)

- E.g., 6 possible rolls of a die: $\{1, 2, 3, 4, 5, 6\}$



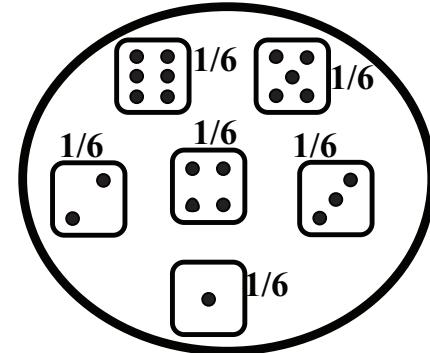
- A Probability Model assigns a number $P(\omega)$ to each world ω

- E.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

- These numbers must satisfy

$$0 \leq P(\omega) \leq 1$$

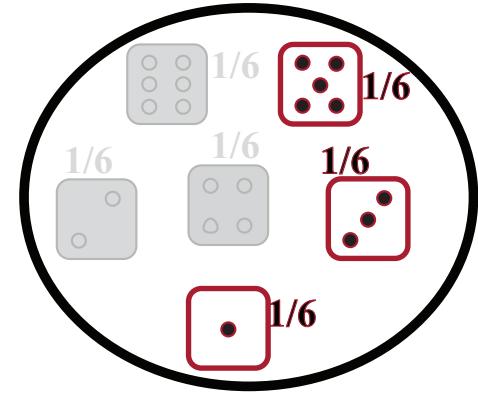
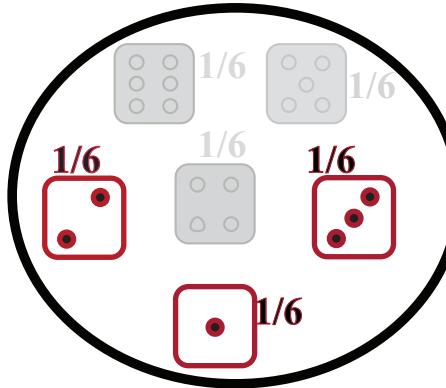
$$\sum_{\omega \in \Omega} P(\omega) = 1$$



Basic Laws of Probability

- An **Event** is any subset of Ω

- E.g., “roll < 4” is the set {1,2,3}
- E.g., “roll is odd” is the set {1,3,5}



- The probability of an event is the **sum** of probabilities over its worlds:

$$P(E) = \sum_{\omega \in E} P(\omega)$$

- E.g., $P(\text{roll} < 4) = P(1) + P(2) + P(3) = \frac{1}{2}$

Random Variables

- A random variable X is some aspect of the world about which we (may) have uncertainty
 - Odd = Is the die roll an odd number?
 - 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
- Random variables have domains (or values)
 - Odd in {true, false}
 - Odd(1) = true, Odd(2)=false
 - 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: **sums to 1**

$$\sum_x P(X = x) = 1$$

Temperature:

$$P(T)$$

T	P
hot	0.5
cold	0.5

Weather:

$$P(W)$$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

- A **probability** is a single number
- A **distribution** is a TABLE of **probabilities** of values

Shorthand notation:

$$P(\text{hot}) = P(T = \text{hot}),$$

$$P(\text{cold}) = P(T = \text{cold}),$$

$$P(\text{rain}) = P(W = \text{rain}),$$

...

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 world or *outcome*):

Long form: $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$

$P(T, W)$

Short form: $P(x_1, x_2, \dots, x_n)$

- 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
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Probabilities of Events in JPD

- From a joint distribution, we can calculate the probability of any event

$$P(T, W)$$

- Probability that it's hot AND sunny? 0.4
- Probability that it's hot? 0.5
- Probability that it's hot OR sunny? 0.7

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

- Typically, the events we care about are **partial assignments**, like $P(T=\text{hot})$

Quiz: Probabilities of Events in JPD

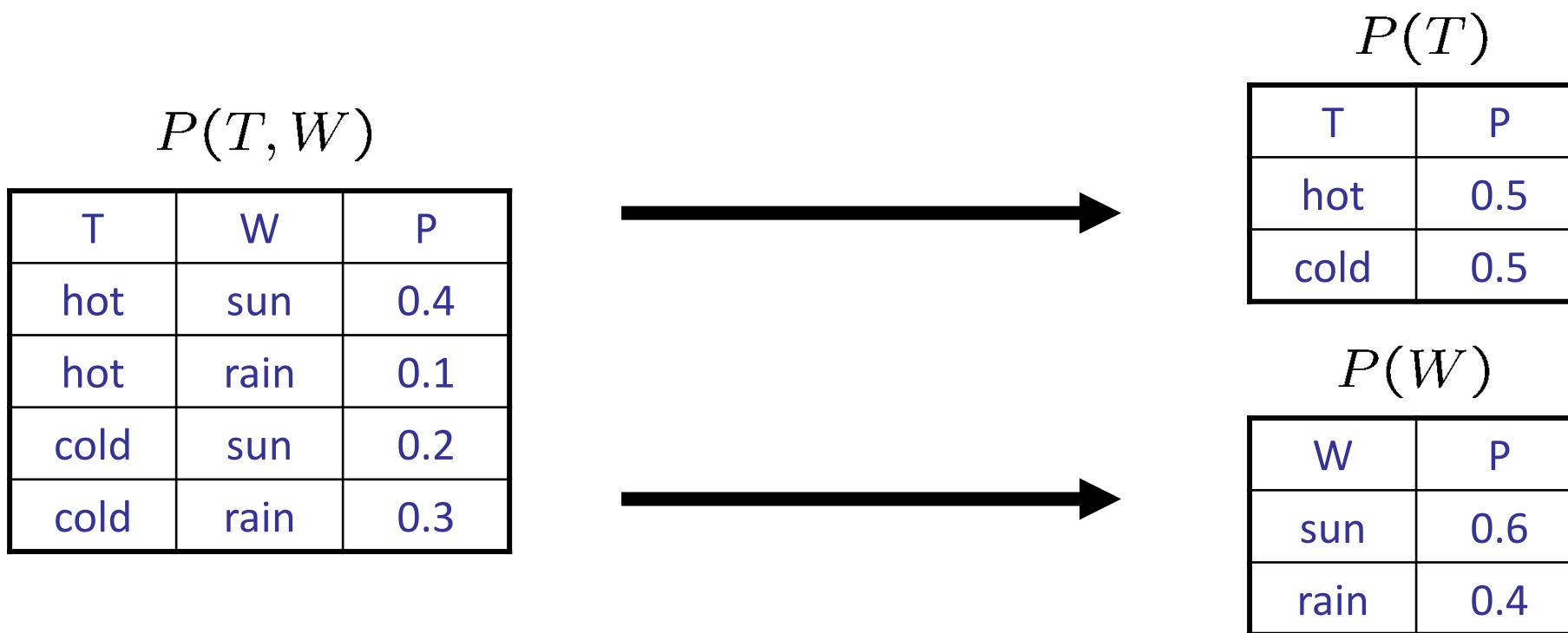
- $P(X=\text{true}, Y=\text{true}) ? \quad 0.2$
- $P(X=\text{true}) ? \quad 0.5$
- $P(Y=\text{false} \text{ OR } X=\text{true}) ? \quad 0.6$

$P(X, Y)$

X	Y	P
true	true	0.2
true	false	0.3
false	true	0.4
false	false	0.1

Marginalization

- Marginalization (summing out): Given the joint distribution, we can compute distributions over smaller sets of variables.



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

Quiz: Marginalization

$P(X, Y)$

X	Y	P
true	true	0.2
true	false	0.3
false	true	0.4
false	false	0.1

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

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

$P(X)$

X	P
true	
false	

$P(Y)$

Y	P
true	
false	

Quiz: Marginalization

$P(X, Y)$

X	Y	P
true	true	0.2
true	false	0.3
false	true	0.4
false	false	0.1

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

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

$P(X)$

X	P
true	0.5
false	0.5

$P(Y)$

Y	P
true	0.6
false	0.4

Conditional Probabilities

- A simple relation between joint and conditional probabilities

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

$P(T, W)$

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

$$\begin{aligned} P(W = s | T = c) &= \frac{P(W = s, T = c)}{P(T = c)} = \frac{0.2}{0.5} = 0.4 \\ &= P(W = s, T = c) + P(W = r, T = c) \\ &= 0.2 + 0.3 = 0.5 \end{aligned}$$

Quiz: Conditional Probabilities

$P(X, Y)$

X	Y	P
true	true	0.2
true	false	0.3
false	true	0.4
false	false	0.1

- $P(X=\text{true} \mid Y=\text{true}) ? \quad 0.333$

- $P(X=\text{false} \mid Y=\text{true}) ? \quad 0.667$

- $P(Y=\text{false} \mid X=\text{true}) ? \quad 0.6$

Conditional Distributions

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

Joint Distribution

$P(T, W)$

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

Conditional Distributions

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

Conditional Distributions: Normalization Trick

$P(T, W)$

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

$$\begin{aligned} P(W = s|T = c) &= \frac{P(W = s, T = c)}{P(T = c)} \\ &= \frac{P(W = s, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\ &= \frac{0.2}{0.2 + 0.3} = 0.4 \end{aligned}$$



$P(W|T = c)$

W	P
sun	0.4
rain	0.6

$$\begin{aligned} P(W = r|T = c) &= \frac{P(W = r, T = c)}{P(T = c)} \\ &= \frac{P(W = r, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\ &= \frac{0.3}{0.2 + 0.3} = 0.6 \end{aligned}$$

Conditional Distributions: Normalization Trick

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

$$\begin{aligned} P(W = s|T = c) &= \frac{P(W = s, T = c)}{P(T = c)} \\ &= \frac{P(W = s, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\ &= \frac{0.2}{0.2 + 0.3} = 0.4 \end{aligned}$$

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

SELECT the joint probabilities matching the evidence
→

T	W	P
cold	sun	0.2
cold	rain	0.3

NORMALIZE the selection
(make it sum to one)
→

W	P
sun	0.4
rain	0.6

$$\begin{aligned} P(W = r|T = c) &= \frac{P(W = r, T = c)}{P(T = c)} \\ &= \frac{P(W = r, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\ &= \frac{0.3}{0.2 + 0.3} = 0.6 \end{aligned}$$

Quiz: Normalization Trick

- $P(X | Y=\text{false})$?

$P(X, Y)$

X	Y	P
true	true	0.2
true	false	0.3
false	true	0.4
false	false	0.1

SELECT the joint probabilities matching the evidence



X	Y	P
true	false	0.3
false	false	0.1

NORMALIZE the selection
(make it sum to one)



X	P
true	0.75
false	0.25

To Normalize

- (Dictionary) To bring or restore to a **normal condition**

All entries sum to ONE

- Procedure:

- Step 1: Compute $Z = \text{sum over all entries}$
- Step 2: Divide every entry by Z

- Example 1

W	P
sun	0.2
rain	0.3

Normalize \rightarrow $Z = 0.5$

W	P
sun	0.4
rain	0.6

- Example 2

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

Normalize \rightarrow $Z = 50$

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

The Product Rule

- Sometimes have conditional distributions but want the joint

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

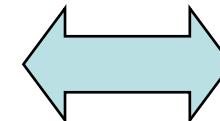
The Product Rule

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

- Example:

W	P
sun	0.8
rain	0.2

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3



$$P(D, W)$$

D	W	P
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

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

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

Bayes' Rule

Bayes' Rule

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

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

That's my rule!

- Dividing, we get:

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

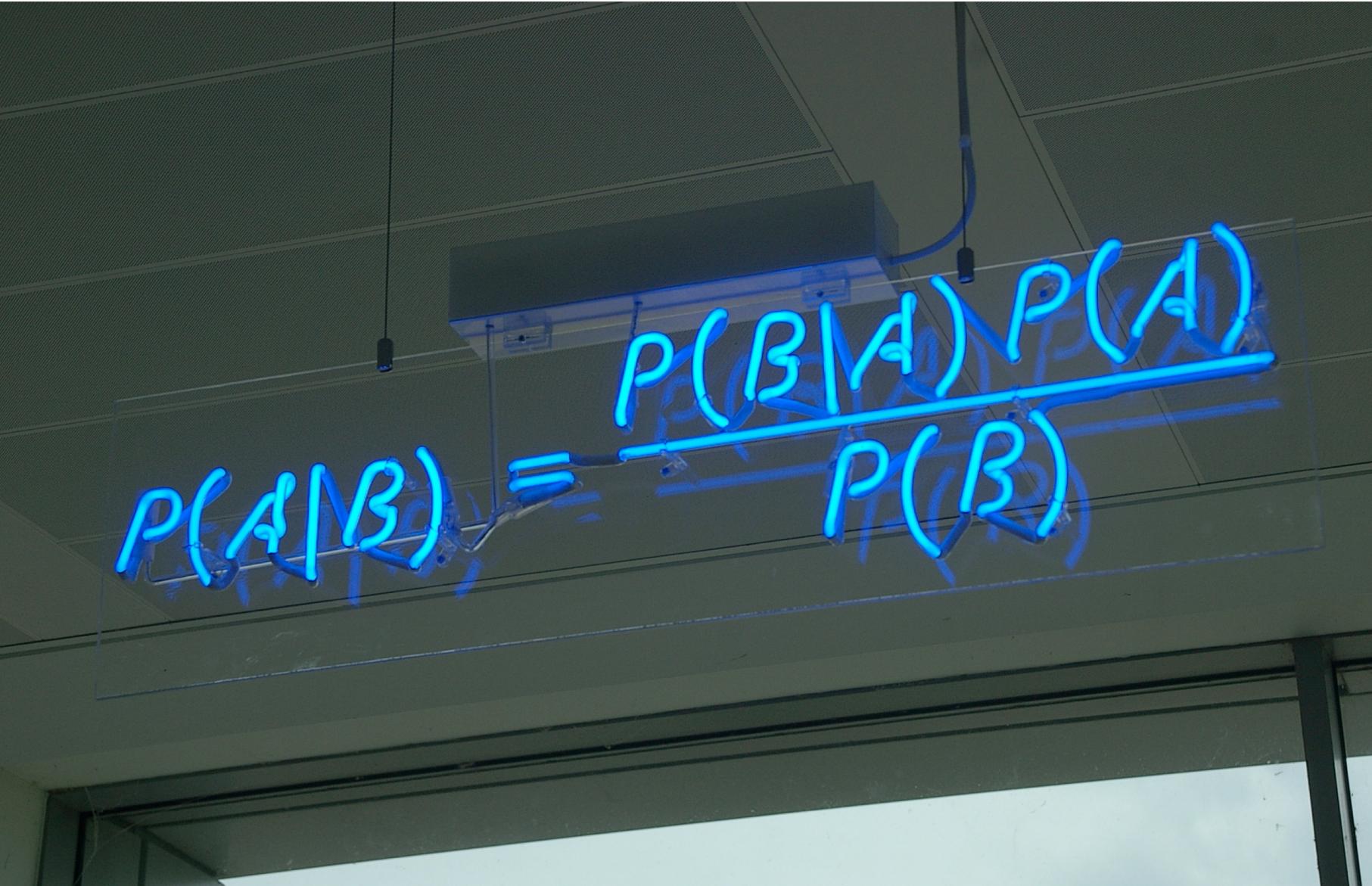
- Why is this at all helpful?

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

- In the running for most important AI equation!



Bayes' Rule



Example: Bayes' Rule

- On Average, the alarm rings once a year
- If there is a fire, the alarm will almost always ring
- On average, we have a fire every 10 years
- The fire alarm rings, what is the probability there is a fire?

Example: Bayes' Rule

- On Average, the alarm rings once a year
 - $P(\text{alarm}) = 1/365$
- If there is a fire, the alarm will almost always ring
 - $P(\text{alarm} \mid \text{fire}) = 0.999$
- On average, we have a fire every 10 years
 - $P(\text{fire}) = 1/3650$
- The fire alarm rings, what is the probability there is a fire?

0.999

0.9

0.0999

0.01

Example: Bayes' Rule

- On Average, the alarm rings once a year
 - $P(\text{alarm}) = 1/365$
- If there is a fire, the alarm will almost always ring
 - $P(\text{alarm} \mid \text{fire}) = 0.999$
- On average, we have a fire every 10 years
 - $P(\text{fire}) = 1/3650$
- The fire alarm rings, what is the probability there is a fire?

$$P(\text{fire} \mid \text{alarm}) = \frac{P(\text{alarm} \mid \text{fire}) \times P(\text{fire})}{P(\text{alarm})} = \frac{0.999 \times 1/3650}{1/365} = 0.0999$$

Even though the alarm rings the chance for a fire is only about 10%

Independence

Independence

- Two variables X and Y are *independent* if:

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

- Knowing the value of one does not tell you anything about the other
- Another form:

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

- This says that their joint distribution *factors* into a product of two simpler distributions
- We write:

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

Example: Independence

- Variables W (weather) and R (result of a die throw)
- Let's compare $P(W)$ vs. $P(W | R = 6)$
- What is $P(W=\text{cloudy})$?

0.066 0.1 0.4 0.6

Weather W	Result R	$P(W,R)$
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Example: Independence

- Variables W (weather) and R (result of a die throw)
- Let's compare $P(W)$ vs. $P(W | R = 6)$
- What is $P(W=\text{cloudy})$?
 - $P(W=\text{cloudy}) = \sum_{r \in R} P(W=\text{cloudy}, R = r)$
 $= 0.1+0.1+0.1+0.1+0.1+0.1 = 0.6$
- What is $P(W=\text{cloudy} | R=6)$?

0.066/0.166

0.066+0.1

0.1/0.166

0.1/0.6

Weather W	Result R	$P(W,R)$
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Example: Independence

- Variables W (weather) and R (result of a die throw)

- Let's compare $P(W)$ vs. $P(W | R = 6)$

- What is $P(W=\text{cloudy})$?

- $$P(W=\text{cloudy}) = \sum_{r \in R} P(W=\text{cloudy}, R = r)$$
$$= 0.1+0.1+0.1+0.1+0.1+0.1 = 0.6$$

- What is $P(W=\text{cloudy} | R=6)$?

$$P(W=\text{cloudy}|R=6) = \frac{P(W=\text{cloudy} \wedge R=6)}{P(R=6)}$$

$$P(W=\text{cloudy} \wedge R=6) = 0.1 \text{ (from table)}$$

$$P(R=6) = 0.166 \text{ (marginal, } 0.1+0.066\text{)}$$

$$\text{Thus, } P(W=\text{cloudy}|R=6) = 0.1/0.166 = 0.6$$

Weather W	Result R	$P(W,R)$
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Example: Independence

- Variables W (weather) and R (result of a die throw)
- Let's compare $P(W)$ vs. $P(W | R = 6)$
- What is $P(W=\text{cloudy})$?
 - $P(W=\text{cloudy}) = \sum_{r \in R} P(W=\text{cloudy}, R = r)$
 $= 0.1+0.1+0.1+0.1+0.1+0.1 = 0.6$
- What is $P(W=\text{cloudy} | R=6)$?
 $P(W=\text{cloudy}|R=6) = \frac{P(W=\text{cloudy} \wedge R=6)}{P(R=6)}$
 $P(W=\text{cloudy} \wedge R=6) = 0.1$ (from table)
 $P(R=6) = 0.166$ (marginal, $0.1+0.066$)
Thus, $P(W=\text{cloudy}|R=6) = 0.1/0.166 = 0.6$

Weather W	Result R	$P(W,R)$
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Example: Independence

- Variables W (weather) and R (result of a die throw)
- Let's compare $P(W)$ vs. $P(W | R = 6)$
- The two distributions are identical: Knowing the result of the die does not change our belief in the weather

Weather W	$P(W)$
sunny	0.4
cloudy	0.6

Weather W	$P(W R=6)$
sunny	$0.066/0.166=0.4$
cloudy	$0.1/0.166=0.6$

Weather W	Result R	$P(W,R)$
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Exploiting Independence

- If X_1, X_2, \dots, X_n are independent, then we can represent their JPD as a product of marginal distributions

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i)$$

Conditional Independence

Conditional Independence

- X is conditionally independent of Y given Z $X \perp\!\!\!\perp Y | Z$

if and only if:

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

- learning that $Y=y$ does not change your belief in X when we already know $Z=z$
- and this is true for all values y that Y could take and all values z that Z could take
- or, equivalently, if and only if

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

Conditional Independence

- What about this domain:

- Traffic
- Umbrella
- Raining

Traffic ind Umbrella | Raining

Conditional Independence

- What about this domain:
 - Fire
 - Smoke
 - Alarm
- Fire ind Alarm | Smoke

Conditional Independence and the Chain Rule

- Chain rule: $P(X_1, X_2, \dots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\dots$

- Trivial decomposition:

$$\begin{aligned} P(\text{Traffic, Rain, Umbrella}) &= \\ P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain, Traffic}) \end{aligned}$$

- With assumption of conditional independence:

$$\begin{aligned} P(\text{Traffic, Rain, Umbrella}) &= \\ P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}) \end{aligned}$$

- Conditional Independence allows us to write them **compactly**

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

- X and Y are conditionally independent given Z if and only if:

$$\begin{aligned} \forall x, y, z : P(x|z, y) &= P(x|z) \\ \forall x, y, z : P(x, y|z) &= P(x|z)P(y|z) \end{aligned}$$

Reading

- Read Sections 13.1-5 in the AIMA textbook (Third Edition)