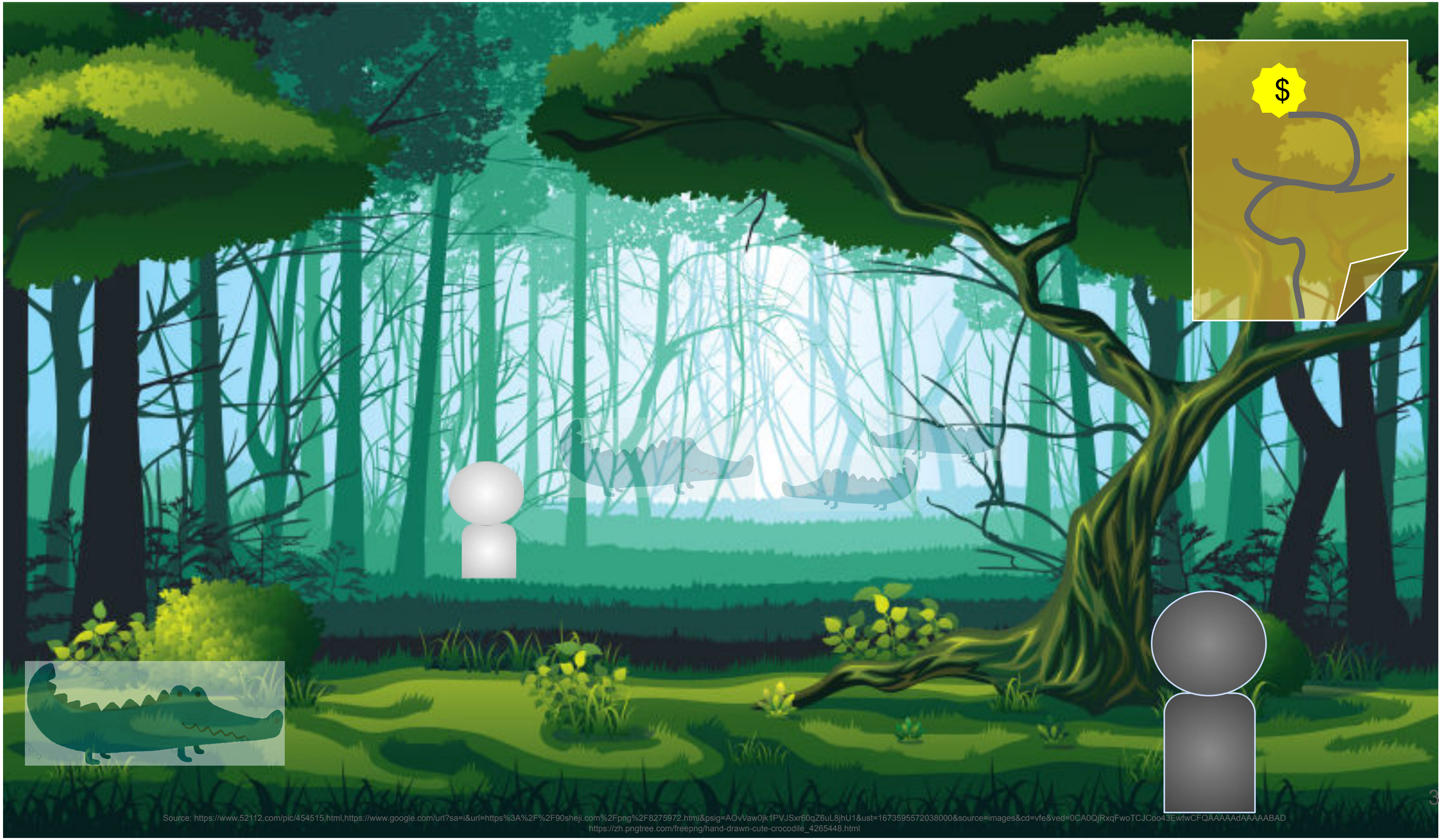
Grading Policy

- Final report [10%]
  - # of team members
    - If there is a graduate student in the team, you could have 2-3 members
    - Otherwise, you could have 3-4 members
  - Submit your team member list **by April 7** (or assigned randomly)
  - Announce the paper list on **April 16**
    - International conference papers published in 2023-2024
  - Submit your top 3 papers **by April 23** (or assigned randomly)
  - Announce the assigned paper before **April 30**
  - Presentation on **June 4/June 11**


$
Source: https://www.52112.com/pic/454515.html,https://www.google.com/url?sa=i&url=https%3A%2F%2F90sheji.com%2Fpng%2F8275972.html&psig=AOvVaw0jk1PVJSxr60qZ6uL8jhU1&ust=1673595572038000&source=images&cd=vfe&ved=0CA0QjRxqFwoTCJCoo43EwfwCFQAAAAAdAAAAABAD
https://zh.pngtree.com/freepng/hand-drawn-cute-crocodile_4265448.html

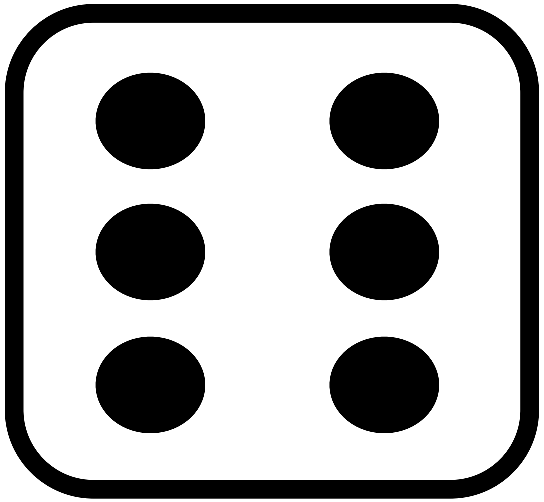
Uncertainty

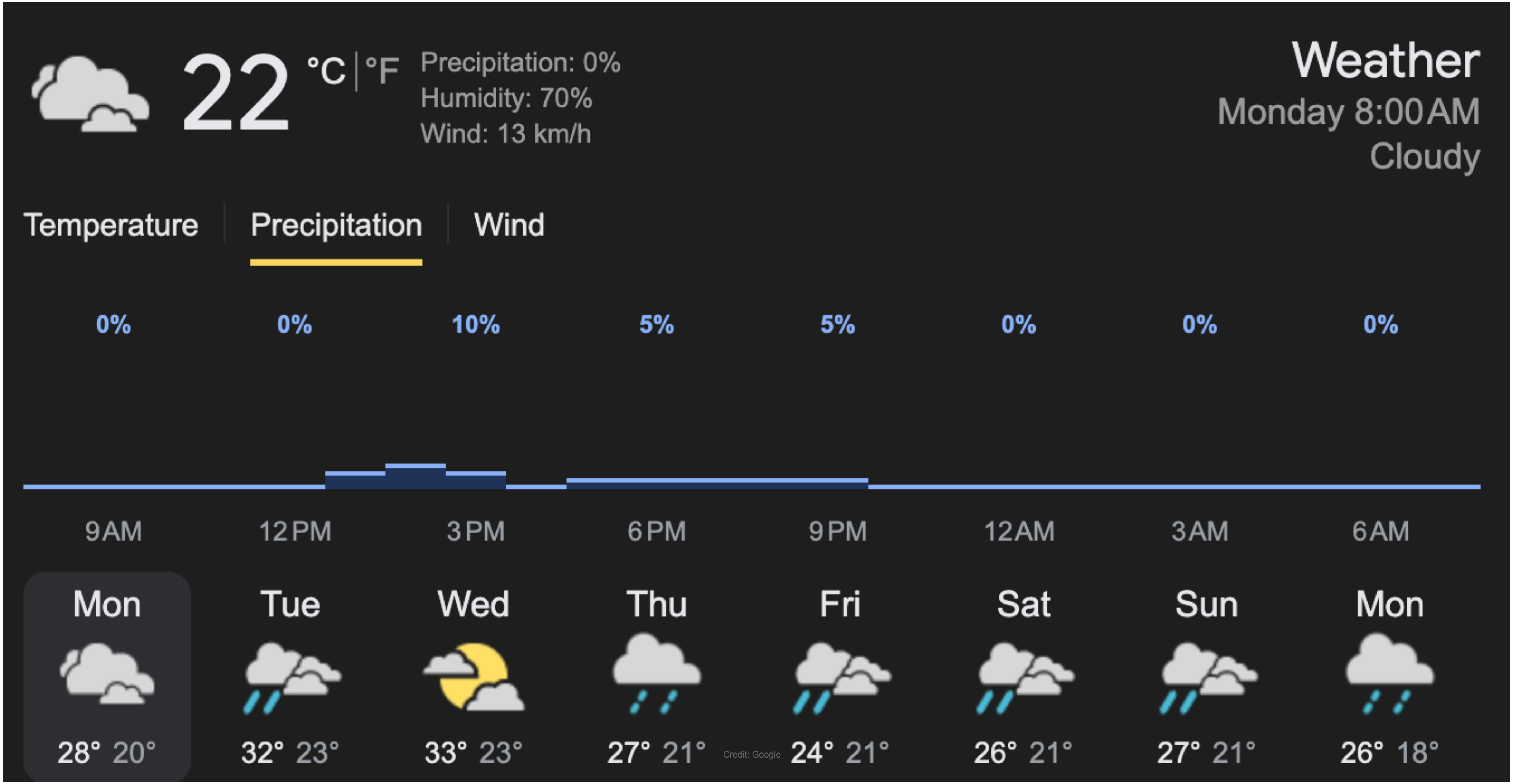
Real World

Partial observability

Nondeterminism

Adversaries




22
°C | °F
Precipitation: 0%
Humidity: 70%
Wind: 13 km/h
Weather
Monday 8:00AM
Cloudy
Temperature
Precipitation
Wind
0%
0%
10%
5%
5%
0%
0%
0%
9AM
12PM
3PM
6PM
9PM
12AM
3AM
6AM
Mon
28° 20°
Tue
32° 23°
Wed
33° 23°
Thu
27° 21°
Fri
24° 21°
Sat
26° 21°
Sun
27° 21°
Mon
26° 18°
Credit: Google

Evidence

Source: https://www.analyticssteps.com/blogs/weather-forecasting-how-do-big-data-analytics-magnify-it

Reasoning

Probabilistic Reasoning

Possible World

ω

e.g. 1

All Possible Worlds

$\{\omega\}$

$\{1, 2, 3, 4, 5, 6\}$

All Possible Worlds

(Sample Space)

$$\Omega = \{\omega\}$$

Probability Model

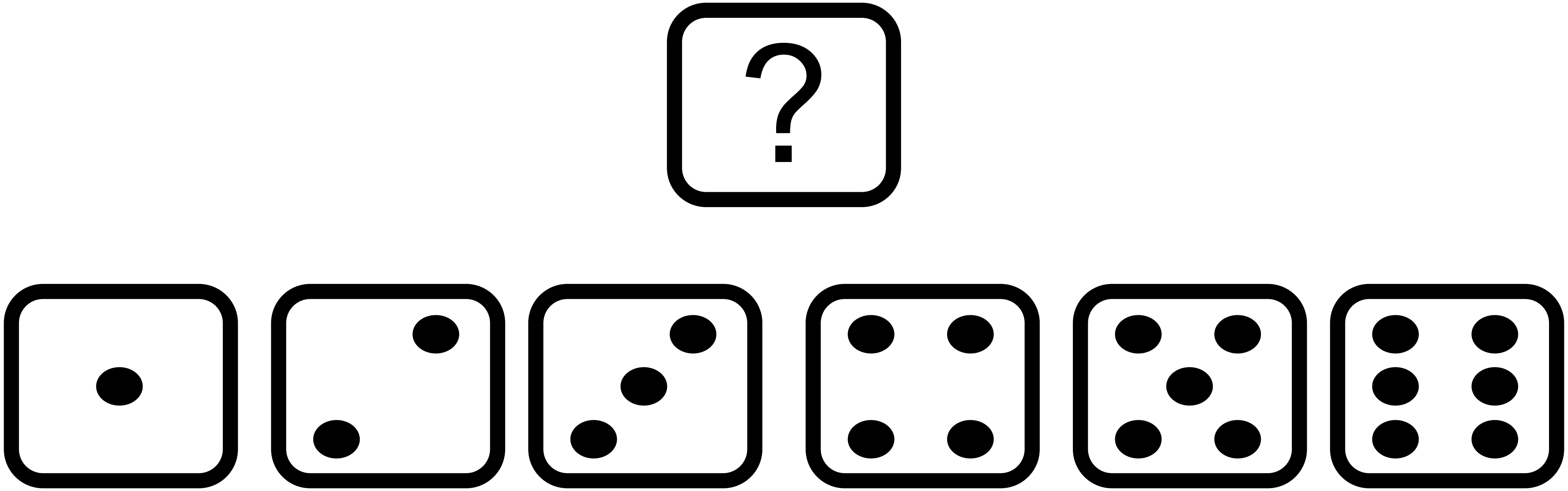
$$P(\omega)$$

Probability Axioms

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

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

Example: Rolling Fair Dice



$0 \leq \frac{1}{6} \leq 1$ $\quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6}$

sum = 1

Random Variable

A variable in probability theory with a domain of possible values it can take on.

Random Variable

Dice

1, 2, 3, 4, 5, 6

Random Variable

Weather

sun, rain, cloud, snow

Random Variables

- A **random variable** is some aspect of the world about which we have uncertainty, and their names begin with **an uppercase letter**
  - **D**ice
  - **R** = Is it raining?
  - **T** = Is it hot or cold? temperature
  - **W** = How's the weather?
- Random variables have domains and **names for values** are always **lowercase**
  - Dice in {1, ..., 6}
  - R in {true, false}
  - T in {hot, cold}
  - W in {sun, rain, cloud, snow}

Shorthand notation:

- **R = true**, abbreviated as **r** real
- **R = false**, abbreviated as ¬**r**
- **T = cold**, abbreviated as **cold**

Probability Distributions

- A probability distribution for the random variable
  - Associate a probability with each value
- Example
  - Weather
    - P(W = sun) = 0.6
    - P(W = rain) = 0.1
    - P(W = cloud) = 0.29
    - P(W = snow) = 0.01

Shorthand notation:
P(sun) = 0.6
P(rain) = 0.1
P(cloud) = 0.29
P(snow) = 0.01

Must have:
1) $\forall x \; P(X = x) \geq 0$
2) $\Sigma_x P(X = x) = 1$

an abbreviation

P(W) = <0.6, 0.1, 0.29, 0.01> (vector)
where we assume a predefined ordering <sun, rain, cloud, snow>

Joint Distributions

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

$$P(X_1 = x_1, \ X_2 = x_2, \ldots, X_n = x_n) \text{ or } P(x_1, x_2, \ldots, x_n)$$

  - Must have
    - $P(x_1, x_2, \ldots, x_n) \geq 0$ and
    - $\sum_{(x1,\ x2,\ \ldots,\ xn)} P(x_1, x_2, \ldots, x_n) = 1$

Example

Given two variables

- Weather W in {sun, rain, cloud, snow}
- Cavity C (do I have a cavity?) in {true, false}

The joint probability distribution of Weather and Cavity is

$$P(W = sun \wedge C = true) \qquad P(W = sun \wedge C = false)$$
$$P(W = rain \wedge C = true) \qquad P(W = rain \wedge C = false)$$
$$P(W = cloud \wedge C = true) \qquad P(W = cloud \wedge C = false)$$
$$P(W = snow \wedge C = true) \qquad P(W = snow \wedge C = false)$$

$4 \times 2 = 8$

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 (or outcomes) are likely
  - Normalized: sum to 1

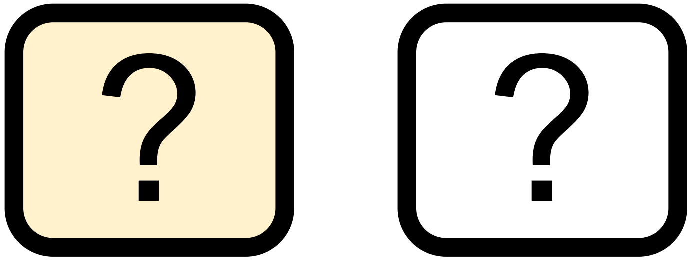
命题.

Proposition (Event)

- Proposition (or Event) is a set of worlds (or outcomes)
- For each proposition, the corresponding set contains just those possible worlds in which the proposition holds
- The probability associated with a proposition Φ, is defined to be

$$P(\Phi) = \Sigma_{\omega \in \Phi} P(\omega)$$

Example: Rolling Two Fair Dice



1	1	2	1	3	1	4	1	5	1	6	1
1	2	2	2	3	2	4	2	5	2	6	2
1	3	2	3	3	3	4	3	5	3	6	3
1	4	2	4	3	4	4	4	5	4	6	4
1	5	2	5	3	5	4	5	5	5	6	5
1	6	2	6	3	6	4	6	5	6	6	6

Example: Rolling Two Fair Dice

- Φ: The two dice add up to 11 (Total=11)
- P(Total=11) = ?

1	1	2	1	3	1	4	1	5	1	6	1
1	2	2	2	3	2	4	2	5	2	6	2
1	3	2	3	3	3	4	3	5	3	6	3
1	4	2	4	3	4	4	4	5	4	6	4
1	5	2	5	3	5	4	5	5	5	6	5
1	6	2	6	3	6	4	6	5	6	6	6

Example: Rolling Two Fair Dice

- Φ: The two dice add up to 11 (Total=11)
- P(Total=11) = 1/36 + 1/36 = 1/18

1	1	2	1	3	1	4	1	5	1	6	1
1	2	2	2	3	2	4	2	5	2	6	2
1	3	2	3	3	3	4	3	5	3	6	3
1	4	2	4	3	4	4	4	5	4	6	4
1	5	2	5	3	5	4	5	5	5	6 **11**	5
1	6	2	6	3	6	4	6	5 **11**	6	6	6

Full Joint Probability Distribution

- A probability model is completely determined by the joint distribution for all of the random variables
- Example
  - Boolean variables: Cavity, Toothache, Catch*
  - Full joint probability distribution, **P**(Cavity, Toothache,Catch) where the joint distribution is represented by

bold P: distribution
non-bold P: probability

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

|**Knowledge Base**| $= 2^3 = 8$

*the dentist's nasty steel probe catches in my tooth

Probabilistic Inference

- Compute a desired probability from other known probabilities
- Example
  - Variables: Toothache, Cavity, Catch*

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

P(cavity V toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28

P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2

- called the **marginal probability** of cavity

*the dentist's nasty steel probe catches in my tooth

Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables
- **Marginalization (summing out)**
  - Combine collapsed rows by adding
    - Extract the distribution over some subset of variables or a single variable
    - Sum up the probabilities for each possible value of the other variables
- General **marginalization rule** for any sets of variables Y and Z:

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y}, \mathbf{Z} = \mathbf{z})$$

Example 1

- Variables
  - R: is it raining?
  - T: is it hot or cold?
- **P**(R) = ? **P**(T) =?

P(R,T)

R	T	P
true	hot	0.1
true	cold	0.3
false	hot	0.4
flase	cold	0.2

P(R)

R	P
true	0.4
false	0.6

P(T)

T	P
hot	0.5
cold	0.5

Example 2

bold → distribution

uppercase → random variable

- The distribution **P**(Cavity) =<?,?>

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$
\begin{aligned}
\mathbf{P}(Cavity) &= \mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch) \\
&+ \mathbf{P}(Cavity, \neg toothache, catch) + \mathbf{P}(Cavity, \neg toothache, \neg catch) \\
&= \langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle + \langle 0.072, 0.144 \rangle + \langle 0.008, 0.576 \rangle \\
&= \langle 0.2, 0.8 \rangle \,.
\end{aligned}
$$

Unconditional Probability

(Prior Probability, Priors)

Degree of belief in a proposition in the absence of any other evidence

Conditional Probability

Degree of belief in a proposition given some evidence that has already been revealed

Bayes' Theorem:

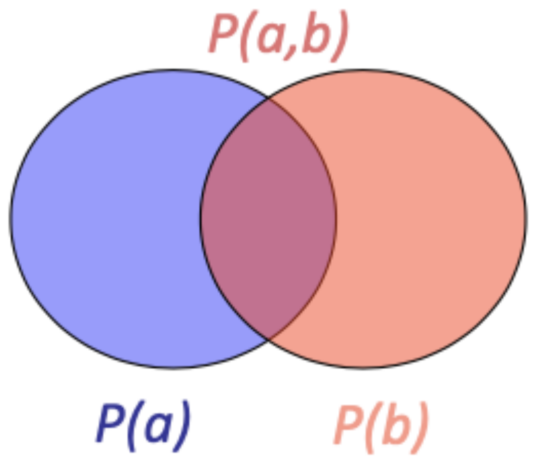
$$\underbrace{P(A|B)}_{\text{posterior}} = \frac{\overbrace{P(B|A)}^{\text{likelihood}}\ \overbrace{P(A)}^{\text{prior}}}{\underbrace{P(B)}_{\text{evidence}}}$$

Conditional Probability

- For any propositions a and b, we have

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)}$$

which holds whenever P(b) > 0



- Product rule

$$P(a \wedge b) = P(a \mid b) P(b).$$

Example 1

- Variables
  - R: is it raining?
  - T: is it hot or cold?
- P(R=false | T=cold) = ?

P(R,T)

R	T	P
true	hot	0.1
true	cold	0.3
false	hot	0.4
flase	cold	0.2

P(R = false | T = cold)
= P(R = false, T = cold) / P(T = cold)
= 0.2/0.5
= 0.4

Example 2

↙ non bold ⇒ probability

- P(cavity | toothache)=?

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

Example 2

- P(cavity | toothache)=?

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6.$$

Conditional Distributions

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

Example 1

- Variables
  - R: is it raining?
  - T: is it hot or cold?
- The distribution **P**(R|cold) =<?,?> <0.6, 0.4>

P(R,T)

R	T	P
true	hot	0.1
true	cold	0.3
false	hot	0.4
flase	cold	0.2

Evidence →

P(R, cold)

R	T	P
true	cold	0.3
flase	cold	0.2

Normalization →

P(R|cold)

R	P
true	0.6
false	0.4

Example 2

- The distribution **P**(Cavity | toothache)=<?,?>

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Select evidence:

$$[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]$$

$$= \quad [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] = \quad \langle 0.12, 0.08 \rangle$$

Normalization: (α: a constant)

$$\alpha \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle$$

Independence

- The knowledge that one proposition occurs does not affect the probability of the other proposition
- Independence between propositions a and b can be written as

$$P(a \mid b) = P(a)$$

$$\text{or} \quad P(b \mid a) = P(b)$$

$$\text{or} \quad P(a \wedge b) = P(a)P(b)$$

$$P(a|b) = P(a) \xRightarrow[\text{rule}]{\text{product}} \frac{P(ab)}{P(b)} = P(a) \Rightarrow P(ab) = P(a)P(b)$$

Bayes' Rule (Bayes' Law or Bayes' Theorem)

For any two propositions a and b,

posterior likelihood prior

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

evidence

Update Factor

Prof. By product rule, we have

$$P(a \wedge b) = P(a \mid b)P(b) \qquad \text{and} \qquad P(a \wedge b) = P(b \mid a)P(a)$$

$$\Rightarrow P(a \mid b)P(b) = P(b \mid a)P(a) \;\; (= P(ab))$$

$$\Rightarrow P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$

Update Factor = $\frac{\text{likelihood}}{\text{evidence}} = \frac{P(a|b)}{P(a)}$: It indicates how the occurrence of b changes the occurence of a.

(a) ① $=1$: b does not change the likelihood of a ⇒ independent

② >1 : positive association between a & b, suggesting a is more likely when b occurs than otherwise.

③ <1 : negative association, suggesting a is less likely when A occurs than otherwise

(b) ① $\gg 1$ ($P(a|b) \gg P(a)$) : b tells us significant amount of a ⇒ high information

② $\ll 1$ ($P(a|b) \ll P(a)$) : a is very unlikely given b ⇒ high information

③ ≈ 1 ($P(a|b) \approx P(a)$) : low information

Bayes' Rule (Bayes' Law or Bayes' Theorem)

- General case for multivalued variables can be written as follows:

$$\mathbf{P}(Y \mid X) = \frac{\mathbf{P}(X \mid Y)\mathbf{P}(Y)}{\mathbf{P}(X)}$$

- General version conditionalized on some background evidence e:

$$\mathbf{P}(Y \mid X, \mathbf{e}) = \frac{\mathbf{P}(X \mid Y, \mathbf{e})\mathbf{P}(Y \mid \mathbf{e})}{\mathbf{P}(X \mid \mathbf{e})}$$

Bayes' Rule (Bayes' Law or Bayes' Theorem)

- General form of Bayes' rule with normalization is → normalization constant

$$\mathbf{P}(Y \mid X) = \alpha\, \mathbf{P}(X \mid Y)\mathbf{P}(Y) \qquad = \frac{1}{P(x)}$$

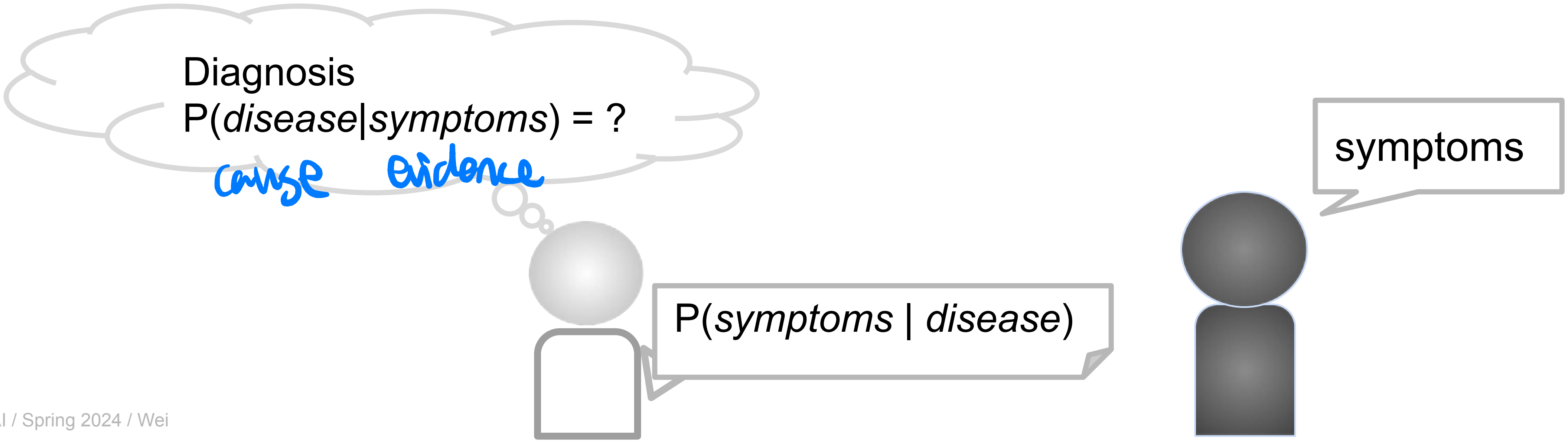
$P(XY)$

where α is the normalization constant needed to make the entries in $\mathbf{P}(Y \mid X)$ sum to 1

Inference: Applying Bayes' Rule

Often, we perceive as evidence the ***effect*** of some unknown ***cause***, and we would like to determine that *cause*

- Doctor



Inference: Applying Bayes' Rule (Cont.)

Often, we perceive as evidence the ***effect*** of some unknown ***cause***, and we would like to determine that *cause*

$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}$$

Diagnostic

Example: Applying Bayes' Rule

- Doctor
  - The disease covid-19 causes a patient to have a running nose, say, 70% of the time
  - c: Any patient has covid-19 is 1/50000
  - R: Any patient has a running nose is 1%
- Question: P(covid-19 | running nose) = P(c|r)?

P(r|c) = 0.7 likelihood

P(c) = 1/50000 prior

P(r) = 0.01 evidence

$$P(c \mid r) = \frac{P(r \mid c)P(c)}{P(r)} = \frac{0.7 \times 1/50000}{0.01} = 0.0014.$$

posterior

Naive Bayes' Model (Bayesian Classifier)

- A single cause directly influences a number of **effects**, all of which are **conditionally independent**, given the cause
- The full joint distribution can be written as

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i \mid Cause).$$

Conditional Independence

- Definition
  - **Conditional independence** of two variables X and Y, given a third variable Z, i.e., $X \perp\!\!\!\perp Y \mid Z$, is

$$\mathbf{P}(X,Y \mid Z) = \mathbf{P}(X \mid Z)\mathbf{P}(Y \mid Z)$$

Example

$\mathbf{P}(\text{Cavity} \mid \text{toothache} \wedge \text{catch})$

$= \alpha\, \mathbf{P}(\text{toothache} \wedge \text{catch} \mid \text{Cavity})\, \mathbf{P}(\text{Cavity})$ (By Bayes' Rule)

$= \alpha\, \mathbf{P}(\text{toothache} \mid \text{Cavity})\, \mathbf{P}(\text{catch} \mid \text{Cavity})\, \mathbf{P}(\text{Cavity})$

(If Toothache and Catch are conditional independence given Cavity)

Inference: Applying Naive Bases Models

- Given some **observed** effects **E** = **e** while the remaining effect variables **Y** are **unobserved**, the probability of the cause is

$$\mathbf{P}(Cause \mid \mathbf{e}) = \alpha\, \mathbf{P}(Cause) \prod_j \mathbf{P}(e_j \mid Cause)$$

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\prod_i \mathbf{P}(Effect_i \mid Cause).$$

Prof.

$$\mathbf{P}(Cause \mid \mathbf{e})$$

$$= \alpha\, \mathbf{P}(Cause, e)$$

$$= \alpha \sum_{\mathbf{y}} \mathbf{P}(Cause, \mathbf{e}, \mathbf{y}).$$

unobserved observed

$$= \alpha \sum_{\mathbf{y}} \mathbf{P}(Cause)\, \mathbf{P}(\mathbf{y} | Cause) \left(\prod_j \mathbf{P}(e_j \mid Cause) \right)$$ (by Naive Bayes)

$$= \alpha\, \mathbf{P}(Cause) \left(\prod_j \mathbf{P}(e_j | Cause) \right) \underbrace{\sum_{\mathbf{y}} \mathbf{P}(\mathbf{y} | Cause)}_{1}$$

$$= \alpha\, \mathbf{P}(Cause) \prod_j \mathbf{P}(e_j \mid Cause)$$

Example: Text Classification with Naive Bayes

Newspaper articles:

- Stocks rallied on Monday, with major indexes gaining 1% as optimism persisted over the first quarter earnings season.
- Heavy rain continued to pound much of the east coast on Monday, with flood warnings issued in New York City and other locations.

Task:

Classify each sentence into a Category:

news, sports, business, weather, or *entertainment*

Example: Text Classification with Naive Bayes (Cont.)

To categorize a new document, we check which key words appear in the document and then apply

$$\mathbf{P}(Cause \mid \mathbf{e}) = \alpha\, \mathbf{P}(Cause) \prod_j \mathbf{P}(e_j \mid Cause)$$

(Cause) Category

(Effect) Keywords - $HasWord_i$

If 9% of articles are about weather,

P(Category=weather) = 0.09

If 37% of articles about weather contain word 6, “rain”,

P($HasWord_6$=true|Category=weather) = 0.37

Then take the one with the highest probability

Probabilistic Models

- Reason about unknown variables, given evidence
- Applications
  - Explanation (diagnostic reasoning)
  - Prediction (causal reasoning)