CSC442: Introduction to Artificial Intelligence

Probability and Inference

Announcements

- Hoping to grade midterms this week.
- FOL quiz cancelled (no assignment either).
- Next quiz will be take-home (written) given out on Wednesday, subject: probability.
- Next project will also be assigned Wednesday.

Probability

- Sample space: possible worlds
- Probability model: degrees of belief in possible worlds
- Events: subsets of possible worlds

Random Variable

- A function from outcomes to values

- “the number of heads in three flips”
- “the total of the two dice”
- “the dice are matched doubles”
- “a red ball was picked”

Statements

- Elementary (atomic) statement

  RV = value

- Combine with connectives from PL

$$Die_1 = 3 \wedge Total = 7$$

$$Total = 7 \vee doubles$$

$$Die_1 = 3 \wedge \neg doubles$$

Probability Statements

- Assign a probability to a statement

$$P(Die_1 = 3 \wedge Total = 7) = 1/36$$

$$P(Total = 7 \vee doubles) = 12/36$$

$$P(Die_1 = 3 \wedge \neg doubles) = 5/36$$

Probability Distribution

- Assign a probability to every possible value of a random variable

$$Weather : \{sunny, rain, cloudy, snow\}$$

$$P(Weather = sunny) = 0.6$$

$$P(Weather = rain) = 0.1$$

$$P(Weather = cloudy) = 0.29$$

$$P(Weather = snow) = 0.01$$

$$\mathbf{P}(Weather) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$$

Bold

Vector

Joint Distribution

- Distribution over multiple variables
- Gives probabilities of all combinations of the values of the variables

Joint Distributions

$$\mathbf{P}(Weather, Cavity)$$

		Cavity	
		true	false
Weather	sunny		
	rain		
	cloudy		
	snow		

Full Joint Probability Distribution

- Joint probability distribution over all the random variables
- Probabilities for every possible combination of values assigned to random variables
- Probabilities for every possible world

Full Joint Probability Distribution

$$\mathbf{P}(Cavity, Toothache, Weather)$$

	toothache		¬toothache	
	cavity	¬cavity	cavity	¬cavity
sunny				
rain				
cloudy				
snow				

Representing Uncertainty

- Probability: Sample space, probabilities, events
- Random variables, domains
- Language of probability statements
- Probability distributions, joint distributions, full joint distribution

New Stuff

Inference

- Compute what "follows from" our knowledge
- Make implicit knowledge explicit

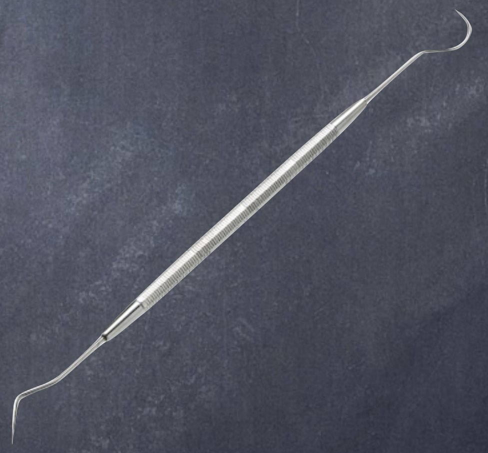
Now with Probabilities!

Probabilistic Inference

- Compute what "follows from" our (uncertain) knowledge
- Make implicit knowledge explicit

Cavity

Toothache



Catch

P(Cavity,Toothache,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

Probability of a Statement

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

Probability of a Statement

$P(\text{cavity} \lor \text{toothache}) = 0.28$

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

Probability of a Statement

$P(cavity \wedge toothache) = 0.12$

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

Background Knowledge + Evidence ⇨ Conclusions

Unconditional (Prior) Probabilities

- Degrees of belief in propositions in the <u>absence</u> of any other information

$$P(\text{doubles}) = 1/3$$

$$P(\text{cavity}) = 0.2$$

Conditional (Posterior) Probability

- Degree of belief in a proposition given some information (evidence)

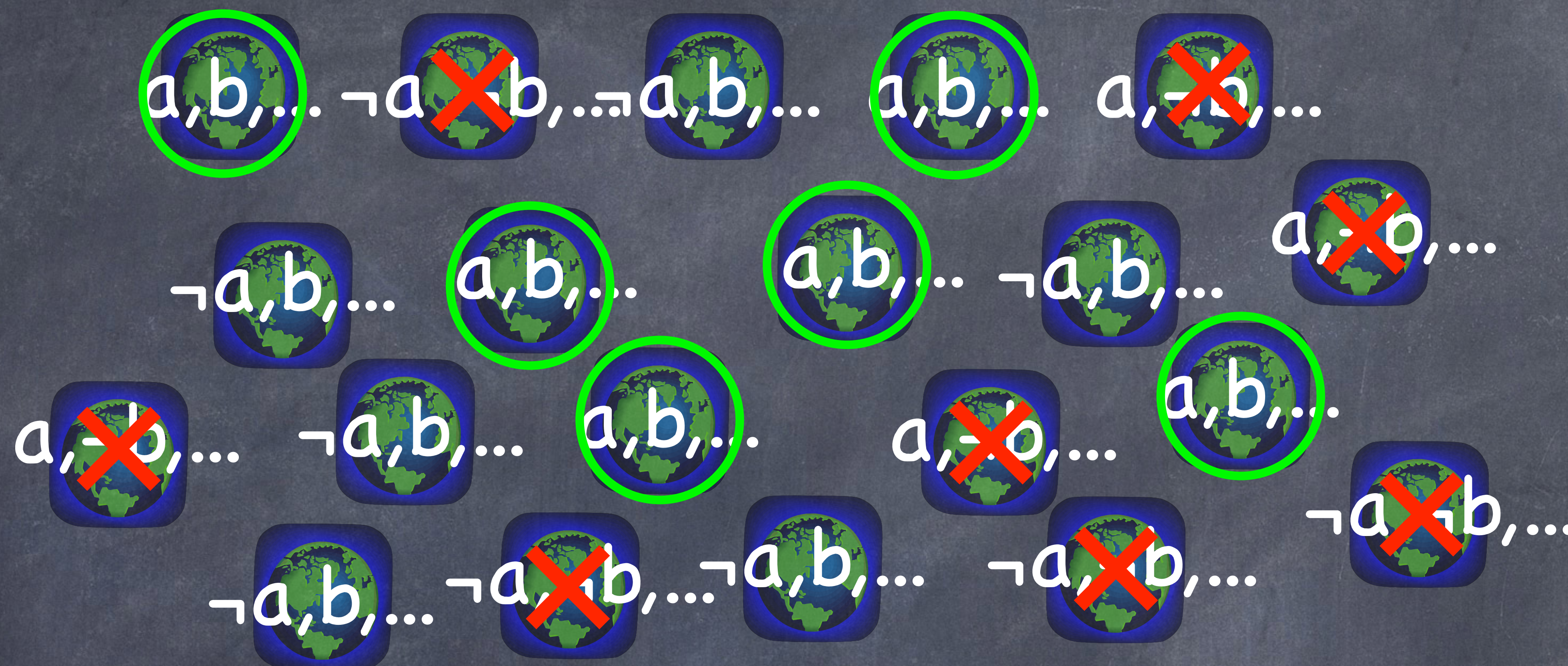
$$P(\text{SnakeEyes} \mid \text{Die}_1 = 1) = 1/6$$

$$P(\text{cavity} \mid \text{toothache}) = 0.6$$

- Whenever evidence is true and we have no further information, conclude probability of proposition

Conditional (Posterior) and Unconditional (Prior) Probability

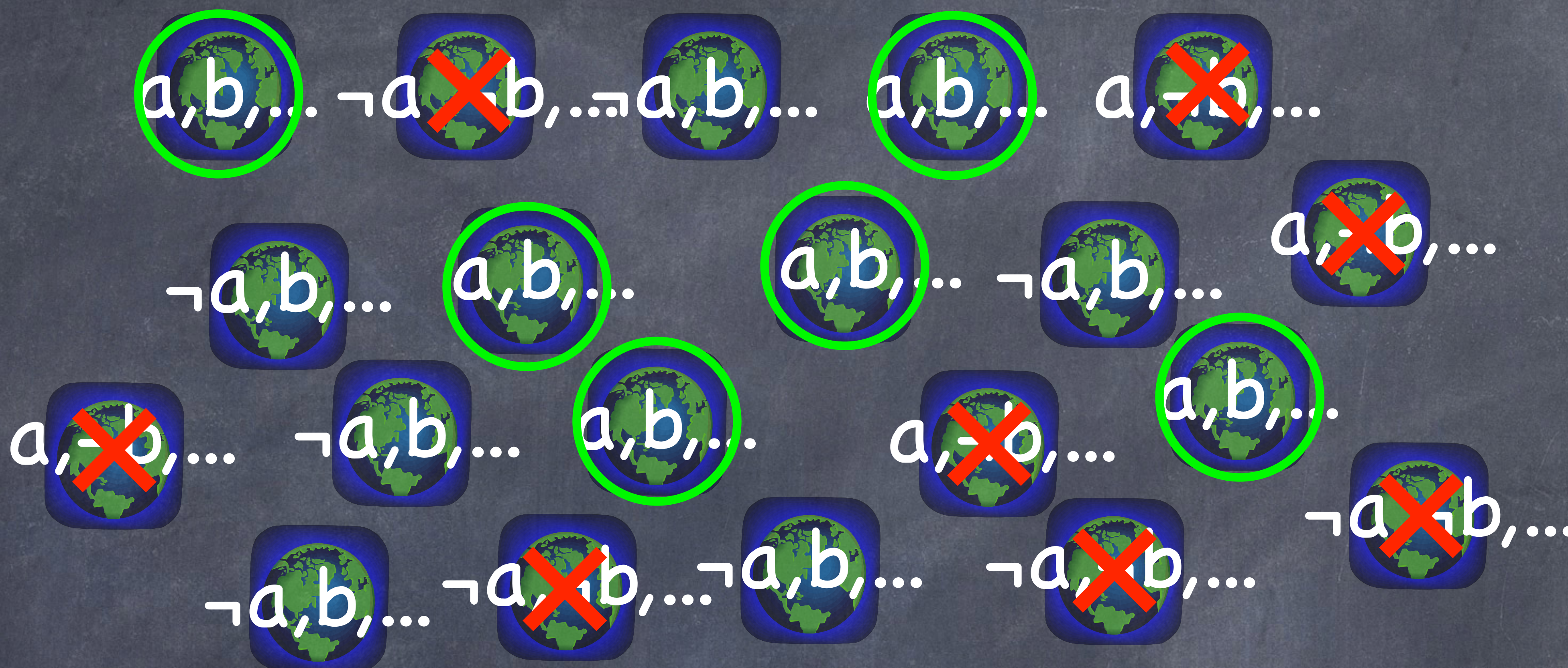
$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} \text{ when } P(b) > 0$$



20 equally likely possible worlds

12 possible worlds where b is true $P(b) = 12/20$

a is true in 6 of those 12 $P(a \mid b) = 6/12 = 1/2$



20 equally likely possible worlds

12 possible worlds where b is true $\qquad P(b) = 12/20$

6 worlds where a∧b is true: $\qquad P(a \wedge b) = 6/20$

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} = \frac{6/20}{12/20} = 1/2$$

Conditional (Posterior) and Unconditional (Prior) Probability

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} \text{ when } P(b) > 0$$

Conditional (Posterior) and Unconditional (Prior) Probability

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} \text{ when } P(b) > 0$$

Product Rule:

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

Background Knowledge	+ Evidence	⇨	Conclusions
Prior Probabilities			Posterior (Conditional) Probabilities given evidence

Probabilistic Inference

- Computing posterior probabilities for statements given prior probabilities and observed evidence
- Given priors and evidence, compute probabilities given evidence (posteriors)

Probabilistic Inference

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

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

Probabilistic Inference

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = \underline{0.12}$$

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	$\boxed{0.108}$	$\boxed{0.012}$	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Probabilistic Inference

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = \frac{0.12}{0.2}$$

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

Probabilistic Inference

$$P(cavity \mid toothache) = \frac{P(cavity \land toothache)}{P(toothache)} = 0.6$$

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

Probabilistic Inference

$$P(cavity \mid toothache) = \frac{P(cavity \land toothache)}{P(toothache)} = 0.6$$

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

Probabilistic Inference

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = 0.4$$

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

Probabilistic Inference

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = 0.4$$

$$\mathbf{P}(Cavity \mid toothache) = \langle 0.6, 0.4 \rangle$$

Probabilistic Inference

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = 0.4$$

$$\mathbf{P}(Cavity \mid toothache) = \langle 0.6, 0.4 \rangle$$

Probabilistic Inference

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = 0.4$$

Probabilistic Inference

$$P(cavity \mid toothache) = \alpha \, P(cavity \wedge toothache) = \alpha \, 0.12$$

$$P(\neg cavity \mid toothache) = \alpha \, P(\neg cavity \wedge toothache) = \alpha \, 0.08$$

Probabilistic Inference

$$P(cavity \mid toothache) = \alpha\, P(cavity \wedge toothache) = \alpha\, 0.12$$

$$P(\neg cavity \mid toothache) = \alpha\, P(\neg cavity \wedge toothache) = \alpha\, 0.08$$

$$\mathbf{P}(Cavity \mid toothache) = \alpha\, \langle 0.12, 0.08 \rangle$$

Probabilistic Inference

$$P(cavity \mid toothache) = \alpha\, P(cavity \land toothache) = \alpha\, 0.12$$

$$P(\neg cavity \mid toothache) = \alpha\, P(\neg cavity \land toothache) = \alpha\, 0.08$$

$$\mathbf{P}(Cavity \mid toothache) = \frac{1}{0.12 + 0.08}\, \langle 0.12, 0.08 \rangle$$

Normalization

$$P(cavity \mid toothache) = \alpha\, P(cavity \wedge toothache) = \alpha\, 0.12$$

$$P(\neg cavity \mid toothache) = \alpha\, P(\neg cavity \wedge toothache) = \alpha\, 0.08$$

$$\mathbf{P}(Cavity \mid toothache) = \langle 0.6, 0.4 \rangle$$

Normalization constant

We didn't know P(toothache)!

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

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

$$P(cavity) = \sum_{y_1 \in Toothache} \sum_{y_2 \in Catch} P(cavity, y_1, y_2)$$

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

Marginalization

$$\mathbf{P}(X) = \sum_{\mathbf{y} \in \mathbf{Y}} \mathbf{P}(X, \mathbf{y})$$

X : Cavity

Y : { Toothache, Catch }

P(Cavity) = $\displaystyle\sum_{y_1 \in Toothache} \sum_{y_2 \in Catch} \mathbf{P}(Cavity, y_1, y_2)$

Probabilistic Inference (Single Variable)

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

Query variable $X : Domain(X) = \{x_1, \ldots, x_m\}$

Evidence variables $\mathbf{E} : \{E_1, \ldots, E_k\}$

Observations $\mathbf{e} : \{e_1, \ldots, e_k\}$ s.t. $E_i = e_i$

Unobserved variables $\mathbf{Y} : \{Y_1, \ldots, Y_l\}$

$Domain(Y_i) = \{y_{i,1}, \ldots, y_{i,n_i}\}$

Probabilistic Inference (Single Variable)

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

Query variable $X : Domain(X) = \{x_1, \ldots, x_m\}$

Evidence variables $\mathbf{E} : \{E_1, \ldots, E_k\}$

Observations $\mathbf{e} : \{e_1, \ldots, e_k\}$ s.t. $E_i = e_i$

Unobserved variables $\mathbf{Y} : \{Y_1, \ldots, Y_l\}$

$Domain(Y_i) = \{y_{i,1}, \ldots, y_{i,n_i}\}$

Probabilistic Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Query variable $X : Domain(X) = \{x_1, \ldots, x_m\}$

Evidence variables $\mathbf{E} : \{E_1, \ldots, E_k\}$

Observations $\mathbf{e} : \{e_1, \ldots, e_k\} \text{ s.t. } E_i = e_i$

Unobserved variables $\mathbf{Y} : \{Y_1, \ldots, Y_l\}$

$$Domain(Y_i) = \{y_{i,1}, \ldots, y_{i,n_i}\}$$

Add numerical example of querying/evidence

Probabilistic Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Given values e for evidence variables E:

For each possible value x_i for query X

  For each possible combination of values y for Y

    Sum up $P(x_i, e, y)$

Result: $P(X|e) = \langle P(x_i, e, y) \rangle$

Probabilistic Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Given values e for evidence variables E:

For each possible value x_i for query X

  For each possible combination of values y for Y

    Sum up $P(x_i, e, y) = P(x_i, e_1, \ldots, e_k, y_{1,i_1}, \ldots, y_{l,i_l})$

Result: $P(X|e) = \langle P(x_i, e, y) \rangle$

Probabilistic Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Given values e for evidence variables E:

For each possible value x_i for query X

For each possible combination of values y for Y

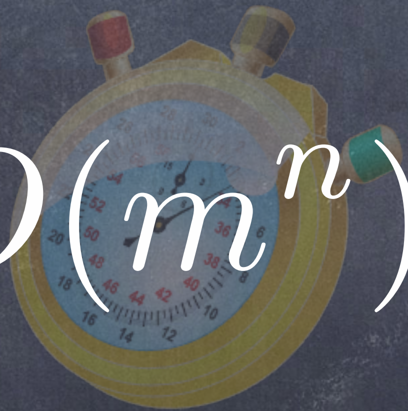
Sum up $P(x_i, e, y) = P(x_i, e_1, \ldots, e_k, y_{1,i_1}, \ldots, y_{l,i_l})$

$= P(X=x_i, E_1=e_1, \ldots, E_k=e_k, Y_1=y_{1,i_1}, \ldots, Y_l=y_{l,i_l})$

Result: $P(X|e) = \langle P(x_i, e, y) \rangle$

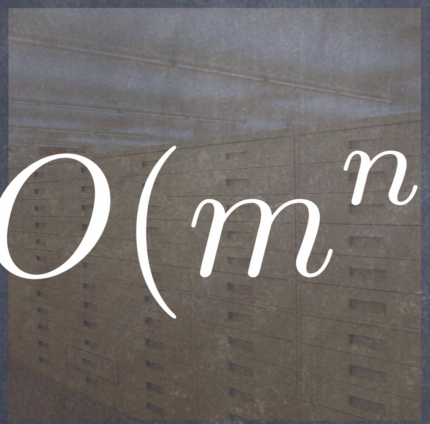
Probabilistic Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$



$$O(m^n)$$

Time Complexity



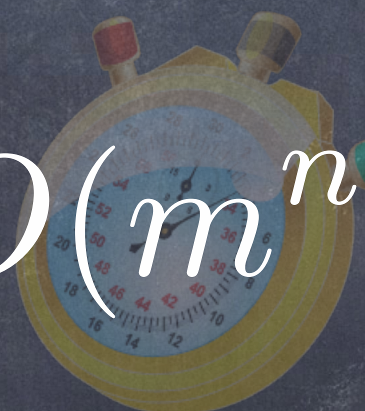
$$O(m^n)$$

Space Complexity

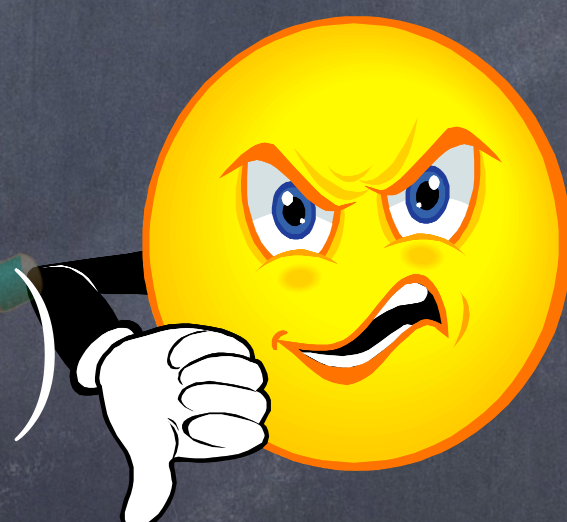
For variables with at most m values

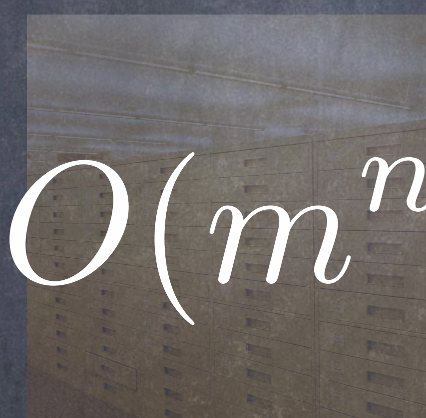
Probabilistic Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$



$$O(m^n)$$





$$O(m^n)$$

Time Complexity   Space Complexity

Intractable!

Bayesian Reasoning

Conditional Probability

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} \text{ when } P(b) > 0$$

Product Rule

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

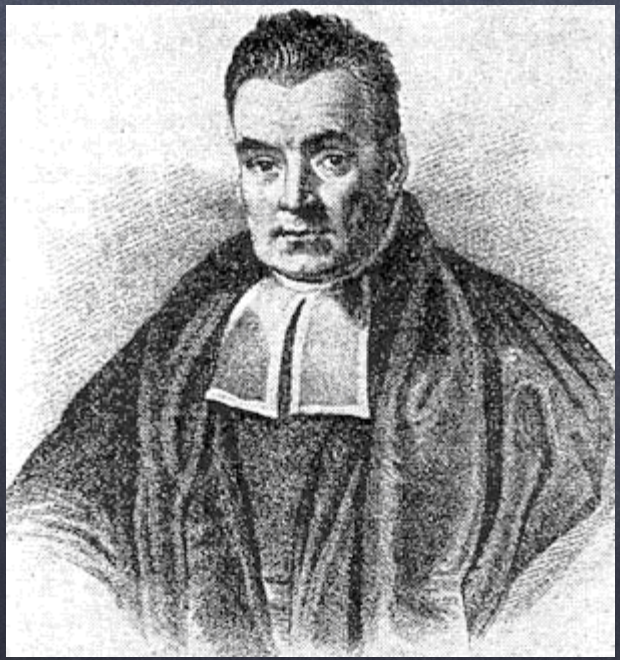
$$P(a \mid b)P(b) = P(b \mid a)P(a)$$

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

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

Bayes' Rule

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



Thomas Bayes

(c. 1702 – 1761)

Bayes' Rule

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

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

Causal and Diagnostic Knowledge

Causal knowledge: $P(\mathit{effect} \mid \mathit{cause})$

Diagnostic knowledge: $P(\mathit{cause} \mid \mathit{effect})$

Causal and Diagnostic Knowledge

Causal knowledge: $P(symptom \mid disease)$

Diagnostic knowledge: $P(disease \mid symptom)$

Bayesian Diagnosis

$$P(disease \mid symptom) = \frac{P(symptom \mid disease)P(disease)}{P(symptom)}$$

Meningitis causes a stiff neck 70% of the time

$$P(stiffneck \mid meningitis) = 0.7$$

Meningitis causes a stiff neck 70% of the time

$$P(stiffneck \mid meningitis) = 0.7$$

Prior probability of meningitis $P(meningitis) = 0.00002$

Prior probability of stiff neck $P(stiffneck) = 0.01$

Meningitis causes a stiff neck 70% of the time

$$P(stiffneck \mid meningitis) = 0.7$$

Prior probability of meningitis $P(meningitis) = 0.00002$

Prior probability of stiff neck $P(stiffneck) = 0.01$

$$P(meningitis \mid stiffneck) =$$

Meningitis causes a stiff neck 70% of the time

$$P(stiffneck \mid meningitis) = 0.7$$

Prior probability of meningitis $P(meningitis) = 0.00002$

Prior probability of stiff neck $P(stiffneck) = 0.01$

$$P(meningitis \mid stiffneck) = \frac{P(stiffneck \mid meningitis)P(meningitis)}{P(stiffneck)}$$

Meningitis causes a stiff neck 70% of the time

$$P(stiffneck \mid meningitis) = 0.7$$

Prior probability of meningitis $P(meningitis) = 0.00002$

Prior probability of stiff neck $P(stiffneck) = 0.01$

$$P(meningitis \mid stiffneck) = \frac{P(stiffneck \mid meningitis)P(meningitis)}{P(stiffneck)}$$

$$= \frac{0.7 \times 0.00002}{0.01}$$

= 0.0014

Bayesian Diagnosis

$$P(disease \mid symptom) = \frac{P(symptom \mid disease)P(disease)}{P(symptom)}$$

toothache (Toothache = true)

catch (Catch = true)

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

Combining Evidence

$P(Cavity \mid toothache \wedge catch)$

Combining Evidence

$$\mathbf{P}(Cavity \mid toothache \land catch) \\ = \alpha \langle 0.180, 0.016 \rangle \approx \langle 0.871, 0.129 \rangle$$

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

Combining Evidence

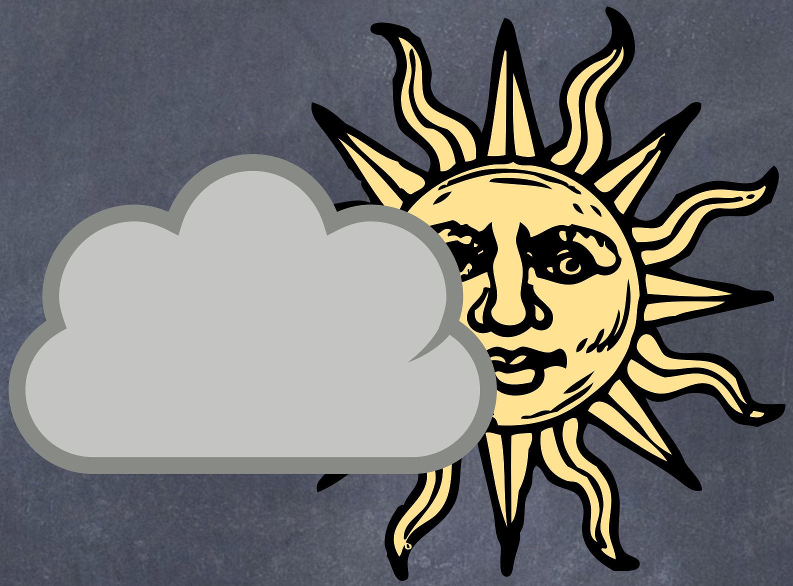
$P(Cavity \mid toothache \wedge catch)$

$= \alpha\, P(toothache \wedge catch \mid Cavity)\, P(Cavity)$

Combining Evidence

- In general, if there are n evidence variables, then there are $O(2^n)$ possible combinations of observed values for which we would need to know the conditional probabilities

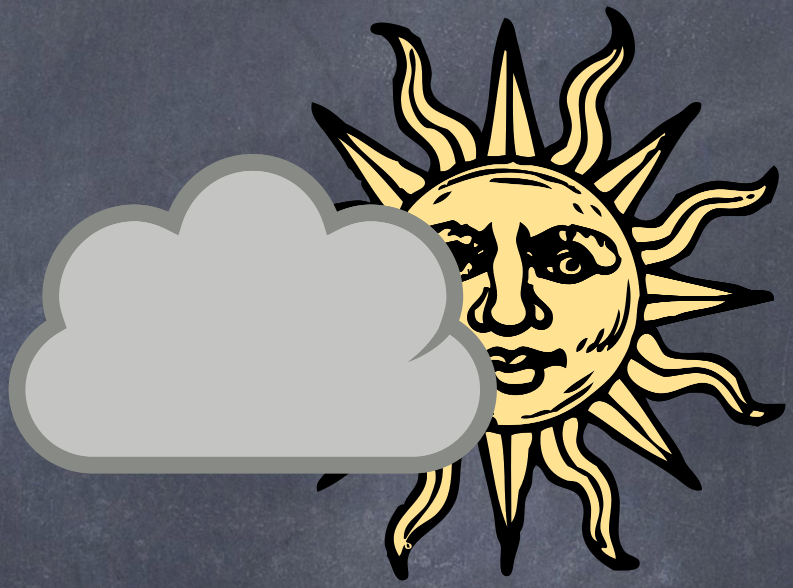
Cavity



Weather

Independence

Cavity



Weather

Independence

$$P(cavity \mid sunny) = P(cavity)$$

$$P(rain \mid cavity) = P(rain)$$

$$\mathbf{P}(Cavity \mid Weather) = \mathbf{P}(Cavity)$$

$$\mathbf{P}(Weather \mid Cavity) = \mathbf{P}(Weather)$$

Independence

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

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

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

Independence

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

Independence

$$\mathbf{P}(X) = \langle p_{x_1}, p_{x_2}, \ldots, p_{x_n} \rangle$$

$$\mathbf{P}(Y) = \langle p_{y_1}, p_{y_2}, \ldots, p_{y_m} \rangle$$

P(X,Y)	x_1	x_2	...	x_n
y_1	$p_{x_1}p_{y_1}$	$p_{x_2}p_{y_1}$	...	$p_{x_n}p_{y_1}$
y_2	$p_{x_1}p_{y_2}$	$p_{x_2}p_{y_2}$	...	$p_{x_n}p_{y_2}$
...	...	...	...	...
y_m	$p_{x_1}p_{y_m}$	$p_{x_2}p_{y_m}$	...	$p_{x_n}p_{y_m}$

Independence

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

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

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

Can compute n×m probabilities

from n+m probabilities

(if random variables are independent)

toothache catch

(Toothache = true) (Catch = true)

$P(Cavity \mid toothache \land catch)$

toothache   catch

(Toothache = true)   (Catch = true)

$$P(Cavity \mid toothache \wedge catch)$$

$$= \alpha\, P(toothache \wedge catch \mid Cavity)\, P(Cavity)$$

toothache (Toothache = true)

catch (Catch = true)

Independent?

$$P(Cavity \mid toothache \land catch)$$

$$= \alpha\, P(toothache \land catch \mid Cavity)\, P(Cavity)$$

Conditional Independence

- Both toothache and catch are caused by a cavity, but neither has a direct effect on the other
- The variables are independent given the presence or absence of a cavity

Conditional Independence

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

Conditional Independence

$$\mathbf{P}(Toothache, Catch \mid Cavity) = \mathbf{P}(Toothache \mid Cavity)\mathbf{P}(Catch \mid Cavity)$$

Combining Evidence

$P(Cavity \mid toothache \wedge catch)$

$= \alpha\, P(toothache \wedge catch \mid Cavity)\, P(Cavity)$

$= \alpha\, P(toothache \mid Cavity)\, P(catch \mid Cavity)\, P(Cavity)$

Combining Evidence

- For n symptoms (e.g., Toothache, Catch) that are all conditionally independent given a disease (e.g., Cavity), we need $O(n)$ probabilities rather than $O(2^n)$
  - Representation scales to larger problems
  - Conditional probabilities more likely to be available than absolute independence assumptions

Probabilistic Inference

- Computing posterior distribution given evidence
  - Normalization, Marginalization
- Full joint distribution: intractable as problem grows
- Bayes' Rule
- Independence assumptions
  - Factor FJPD into smaller distributions

Next Up:

Bayesian Networks