Lecture 9a: Quantifying Uncertainty

CSCI 360

Introduction to Artificial Intelligence

USC

Here is where we are…

	3/1		Project 2 Out	
9	3/4	3/5	Quantifying Uncertainty	[Ch 13.1-13.6]
	3/6	3/7	Bayesian Networks	[Ch 14.1-14.2]
10	3/11	3/12	*(spring break, no class)*	
	3/13	3/14	*(spring break, no class)*	
11	3/18	3/19	Inference in Bayesian Networks	[Ch 14.3-14.4]
	3/20	3/21	Decision Theory	[Ch 16.1-16.3 and 16.5]
	3/23		Project 2 Due	
12	3/25	3/26	Advanced topics (Chao traveling to NSF)	
	3/27	3/28	Advanced topics (Chao traveling to NSF)	
	3/29		Homework 2 Out	
13	4/1	4/2	Markov Decision Processes	[Ch 17.1-17.2]
	4/3	4/4	Decision Tree Learning	[Ch 18.1-18.3]
	4/5		Homework 2 Due	
	4/5		Project 3 Out	
14	4/8	4/9	Perceptron Learning	[Ch 18.7.1-18.7.2]
	4/10	4/11	Neural Network Learning	[Ch 18.7.3-18.7.4]
15	4/15	4/16	Statistical Learning	[Ch 20.2.1-20.2.2]
	4/17	4/18	Reinforcement Learning	[Ch 21.1-21.2]
16	4/22	4/23	Artificial Intelligence Ethics	
	4/24	4/25	Wrap-Up and Final Review	
	4/26		Project 3 Due	
	5/3	5/2	**Final Exam** (2pm-4pm)	

Outline

- What is AI?
- Problem-solving agent (search)
- Knowledge-based agent (logical reasoning)
- **Probabilistic reasoning**
  - **Quantifying Uncertainty**
  - Bayesian Networks
  - Inference in Bayesian Networks
  - Decision Theory
  - Markov Decision Processes
- Machine learning

A little history…

- Early AI researchers largely rejected using probability in their systems
  - “People don’t think that way…”

- However, neither **problem-solving** nor **logical reasoning** agents tolerate approximation well

Agents under uncertainty

- Sources of uncertainty
  - Partial observability
  - Nondeterminism
- Technique used by an agent
  - Keep track of a **belief state** (sets of possible world states) and
  - Generate a **contingency plan** to handle every possible outcome

- Drawbacks
  - Large belief state representations
  - Complex contingency plan
  - Sometimes, no plan can guarantee to achieve the goal!

Uncertainty (example)

- Let **action** A_t = leave for airport t minutes before flight
- Will A_t get me there on time?

- Problems:
  - Partial observability (road state, other drivers, etc.)
  - Noisy sensors
  - Nondeterministic outcomes of actions (flat tire, et.c)
  - Immense complexity in predicting traffic

- Thus, a purely logical approach may risk falsehood
  - A_t will get passenger to airport on time

What if *no plan guarantees to achieve the goal…*

- Example: ***delivering a passenger to the airport on time***
  - A_{90}: leaving home 90 minutes before the flight's departure time
    - The airport is only 5 miles away, but… this is Los Angeles, and nothing can be guaranteed with certainty

"A_{90} will get the passenger to the airport on time, as long as

(1) **the car doesn't break down**, or
(2) **run out of gas**, and
(3) **there are no accidents on the bridge**, and
(4) **the plane doesn't leave early**, and
(5) **no meteorite hits the car**, and
(6) … "

Must compare the merits of plans

- Possible plans
  - A_{30}: leaving home 30 minutes before the flight's departure time
    - More likely to miss flight, but less likely to have a long wait
  - A_{90}: leaving home 90 minutes before the flight's departure time

    This is actually the best plan
  - A_{180}: leaving home 90 minutes before the flight's departure time
    - Less likely to miss flight, but more likely for a long wait
  - …
  - A_{1440}: leaving home 24 hours before the flight's departure time
    - Sleep in the airport?

Making decision under uncertainty

Probability

- P(A_{25} gets me there on time|…) = 0.04
- P(A_{90} gets me there on time|…) = 0.70
- P(A_{120} gets me there on time|…) = 0.95
- P(A_{1440} gets me there on time|…) = 0.9999

Utility is used to represent and infer preferences

preference for missing flight versus airport cuisine, etc.

Which action to choose?

Consider both utility and probability

Another example *(uncertainty reasoning)*

- Diagnosing a dental patient's toothache
  - Toothache ➔ Cavity
- But it may be caused by gum disease, abscess, … and an almost unlimited list of other possible problems
  - Toothache ➔ Cavity ∨ GumProblem ∨ Abscess ∨ …
- Try the causal rule
  - Cavity ➔ Toothache
  - But not right either: cavity does not **always** lead to toothache

Using **logic** to deal with a domain like **medical diagnosis** is difficult… *(similar domains include law, business, design, auto repair, dating, etc.)*

Probability Theory

- Provides a way of summarizing the uncertainty that comes from **laziness** and **ignorance**
  - Laziness
    - Too much work to list the complete set of antecedents (or consequents) needed to ensure a complete rule
  - Ignorance
    - Medical science has no complete theory for the domain
    - Not all necessary tests have been (or can be) run for a particular patient

Making decision

Rational decision depends on

(1) The relative importance of various goals and

(2) likelihood that (and degree to which) they will be reached

Decision theory = Utility theory + Probability theory

Choose the action that yields the ***highest expected utility***, averaged over all the possible outcomes of the action

Decision-theoretic agent

```
function DT-AGENT(percept) returns an action
   persistent: belief_state, probabilistic beliefs about the current state of the world
               action, the agent's action

   update belief_state based on action and percept
   calculate outcome probabilities for actions,
       given action descriptions and current belief_state
   select action with highest expected utility
       given probabilities of outcomes and utility information
   return action
```

Outline

- **Probability Theory**
- Probabilistic Inference using Joint Distribution
- Bayes' Rule

Probability

- Similar to propositional logic: possible worlds defined by assignment of values to **random variables**

- Every **random variable** has a **domain** – the set of values it can take on
  - *Die1* *{1, ..., 6}*
  - *Total* *{2, ..., 12}*
  - *Cavity* *{true, false}*
  - *Age* *{juvenile, teen, adult}*
  - *Weather* *{sunny, rain, cloudy, snow}*

Probability (example)

- Logical expressions are predicates (either true or false)

$$
\begin{aligned}
&P(Weather = sunny) = 0.6 \\
&P(Weather = rain) = 0.1 \\
&P(Weather = cloudy) = 0.29 \\
&P(Weather = snow) = 0.01 \ ,
\end{aligned}
$$

Probability model

- A numerical probability *P(ω)* for each possible world *ω*

$$0 \leq P(\omega) \leq 1 \text{ for every } \omega$$

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

- **Example:** rolling two dice
  - Each possible world (1,1), (1,2), …, (6,6) has probability 1/36
  - *P(Total=11)* = P((5,6)) + P((6,5)) = 1/36 + 1/36 = 1/18

Probability axioms

- A numerical probability *P(ω)* for each possible world

$$0 \leq P(\omega) \leq 1 \text{ for every } \omega$$

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

$$\begin{aligned} P(\neg a) &= \textstyle\sum_{\omega \in \neg a} P(\omega) \\ &= \textstyle\sum_{\omega \in \neg a} P(\omega) + \sum_{\omega \in a} P(\omega) - \sum_{\omega \in a} P(\omega) \\ &= \textstyle\sum_{\omega \in \Omega} P(\omega) - \sum_{\omega \in a} P(\omega) \\ &= 1 - P(a) \end{aligned}$$

Probability axioms

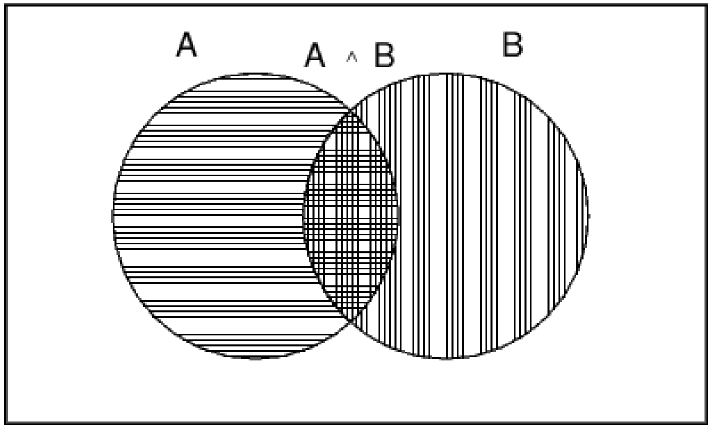
- A numerical probability *P(ω)* for each possible world

$$0 \leq P(\omega) \leq 1 \text{ for every } \omega$$

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

$$P(\neg a) = 1 - P(a)$$

$$\boxed{P(a \vee b) = P(a) + P(b) - P(a \wedge b)}$$



Beliefs needs to be consistent with the axioms…

- The following set of beliefs violates the probability axioms

$$P(a) = 0.4 \qquad P(a \wedge b) = 0.0$$
$$P(b) = 0.3 \qquad P(a \vee b) = 0.8\,.$$

Examples

1 $\quad = P(\text{true})$

$=$

Examples

$$
\begin{aligned}
1 &= P(true) \\
&= P(A \vee \neg A) \\
&= P(A) + P(\neg A) - P(A \wedge \neg A) \\
&= P(A) + P(\neg A) - P(false) \\
&= P(A) + P(\neg A) - 0 \\
&= P(A) + P(\neg A)
\end{aligned}
$$

$$P(\neg A) = 1 - P(A)$$

Examples (cont'd)

$P(B) = P((A \wedge B) \vee (\neg A \wedge B))$

$=$

Examples (cont'd)

$$
\begin{aligned}
P(B) &= P(\,(A \wedge B) \vee (\neg A \wedge B)\,) \\
&= P(A \wedge B) + P(\neg A \wedge B) - P(A \wedge B \wedge \neg A \wedge B) \\
&= P(A \wedge B) + P(\neg A \wedge B) - P(\text{false}) \\
&= P(A \wedge B) + P(\neg A \wedge B) - 0 \\
&= P(A \wedge B) + P(\neg A \wedge B)
\end{aligned}
$$

Unconditional (or prior) probability

- Probability **in the absence of any other information**

$$P(cavity) = 0.2$$

- **Conditional (or posterior)** probability

$$P(cavity \mid toothache) = 0.6$$

Conditional (or posterior) probability

- For any propositions *a* and *b*, we have

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

- Example:

$$P(doubles \mid Die_1 = 5) = \frac{P(doubles \wedge Die_1 = 5)}{P(Die_1 = 5)}$$

Product rule *(of conditional probability)*

- For *a* and *b* to be true
  - we need *b* to be true, and
  - we also need *a* to be true given *b*

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

Probability distribution

- Probabilities of all possible values of a random variable

$$
\begin{aligned}
P(Weather = sunny) &= 0.6 \\
P(Weather = rain) &= 0.1 \\
P(Weather = cloudy) &= 0.29 \\
P(Weather = snow) &= 0.01\ ,
\end{aligned}
$$

- In a vector format

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

Joint probability distribution

- Probabilities of all possible values of **multiple** random variables

$$
\begin{aligned}
P(W = sunny \land C = true) &= P(W = sunny | C = true)\, P(C = true) \\
P(W = rain \land C = true) &= P(W = rain | C = true)\, P(C = true) \\
P(W = cloudy \land C = true) &= P(W = cloudy | C = true)\, P(C = true) \\
P(W = snow \land C = true) &= P(W = snow | C = true)\, P(C = true) \\
P(W = sunny \land C = false) &= P(W = sunny | C = false)\, P(C = false) \\
P(W = rain \land C = false) &= P(W = rain | C = false)\, P(C = false) \\
P(W = cloudy \land C = false) &= P(W = cloudy | C = false)\, P(C = false) \\
P(W = snow \land C = false) &= P(W = snow | C = false)\, P(C = false)\ .
\end{aligned}
$$

- In a vector format

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

Outline

- Probability Theory
- **Probabilistic Inference using Joint Distribution**
- Bayes’ Rule

Probabilistic inference

- It's the computation of posterior probabilities for **query propositions**, given observed evidence.

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Example

$$P(cavity \vee toothache) =$$

Probabilistic inference

- It's the computation of posterior probabilities for **query propositions**, given observed evidence.

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Example

$$P(cavity \lor toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

Marginal probability

- Extracting the distribution **over some subset** of variables, or a single variable, from the full joint distribution

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Example

$$P(cavity) =$$

Marginal probability

- Extracting the distribution **over some subset** of variables, or a single variable, from the full joint distribution

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Example

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

Normalization

- The probability of **cavity**, or **no cavity**, given **toothache**

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Example

$$P(cavity \mid toothache) \; =$$

$$P(\neg cavity \mid toothache) \; =$$

Normalization

- The probability of **cavity**, or **no cavity**, given **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

- Example

$$P(cavity \mid toothache) = \frac{P(cavity \land toothache)}{P(toothache)} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \land toothache)}{P(toothache)} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Sum is always 1.0

Normalization

- The probability of **cavity**, or **no cavity**, given **toothache**

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Example

No need to compute *P (toothache)* any more

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

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

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

$$= \frac{0.016 + 0.064}{\cancel{0.108 + 0.012 + 0.016 + 0.064}} = 0.4$$

Normalization

- The probability of **cavity**, or **no cavity**, given **toothache**

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Example

Assume that
$\alpha = 1/P(toothache)$

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

Normalization

- The probability of **cavity**, or **no cavity**, given **toothache**

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Example

Assume that
α = 1/P (toothache) = 1 / (0.12+0.08) = 1/0.2 = 5

$$\mathbf{P}(Cavity \mid toothache) = \alpha\, \mathbf{P}(Cavity, toothache)$$
$$= \alpha\, [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]$$
$$= \alpha\, [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] = \alpha\, \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle\ .$$

Normalization

Suppose we wish to compute a posterior distribution over A given $B=b$, and suppose A has possible values $a_1 \ldots a_m$

Normalization

Suppose we wish to compute a posterior distribution over A given $B=b$, and suppose A has possible values $a_1 \ldots a_m$

We can apply Bayes' rule for each value of A:

$$P(A=a_1|B=b) = P(B=b|A=a_1)P(A=a_1)/P(B=b)$$

$$\ldots$$

$$P(A=a_m|B=b) = P(B=b|A=a_m)P(A=a_m)/P(B=b)$$

Adding these up, and noting that $\Sigma_i P(A=a_i|B=b) = 1$:

$$1/P(B=b) = \boxed{1/\Sigma_i P(B=b|A=a_i)P(A=a_i)}$$

This is the <u>normalization factor</u>, constant w.r.t. i, denoted α:

$$\mathbf{P}(A|B=b) = \alpha \mathbf{P}(B=b|A)\mathbf{P}(A)$$

Normalization

Suppose we wish to compute a posterior distribution over A given $B=b$, and suppose A has possible values $a_1 \ldots a_m$

We can apply Bayes' rule for each value of A:

$$P(A=a_1|B=b) = P(B=b|A=a_1)P(A=a_1)/P(B=b)$$
$$\ldots$$
$$P(A=a_m|B=b) = P(B=b|A=a_m)P(A=a_m)/P(B=b)$$

Adding these up, and noting that $\Sigma_i P(A=a_i|B=b) = 1$:

$$1/P(B=b) = \boxed{1/\Sigma_i P(B=b|A=a_i)P(A=a_i)}$$

This is the <u>normalization factor</u>, constant w.r.t. i, denoted α:

$$\mathbf{P}(A|B=b) = \alpha \mathbf{P}(B=b|A)\mathbf{P}(A)$$

Typically compute an unnormalized distribution, normalize at end

e.g., suppose $\mathbf{P}(B=b|A)\mathbf{P}(A) = \langle 0.4, 0.2, 0.2 \rangle$

then $\mathbf{P}(A|B=b) = \alpha\langle 0.4, 0.2, 0.2 \rangle = \frac{\langle 0.4, 0.2, 0.2 \rangle}{\boxed{0.4+0.2+0.2}} = \langle 0.5, 0.25, 0.25 \rangle$

Exponential blowup

- Given **full joint probability distribution**, we can answer any probabilistic queries for discrete variables

E.g., suppose $Toothache$ and $Cavity$ are the random variables:

	$Toothache = true$	$Toothache = false$
$Cavity = true$	0.04	0.06
$Cavity = false$	0.01	0.89

Possible worlds are mutually exclusive $\Rightarrow$ $P(w_1 \wedge w_2) = 0$
Possible worlds are exhaustive $\Rightarrow$ $w_1 \vee \cdots \vee w_n$ is $True$
hence $\Sigma_i P(w_i) = 1$

- However, the (full joint distribution) table is exponential in the number of random variables
  - For **(n>100)**, the complexity **$O(2^n)$** becomes impractical

Outline

- Probability Theory
- **Probabilistic Inference using Joint Distribution**
  - Basic procedure
  - **Independence**
- Bayes' Rule

Independence *to the rescue…*

- Consider ***P****(Toothache, Catch, Cavity, Weather)*, which has 32 entries in the **full joint distribution** table

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

- Applying the product rule

$$P(toothache, catch, cavity, cloudy)$$
$$= P(cloudy \mid toothache, catch, cavity) P(toothache, catch, cavity)$$

- But weather is not influenced by dentistry!

$$P(cloudy \mid toothache, catch, cavity) = P(cloudy)$$

$$\boxed{P(toothache, catch, cavity, cloudy) = P(cloudy) P(toothache, catch, cavity)}$$

Independence *to the rescue…*

- Consider $\mathbf{P}(Toothache, Catch, Cavity, Weather)$, which has 32 entries in the **full joint distribution** table

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

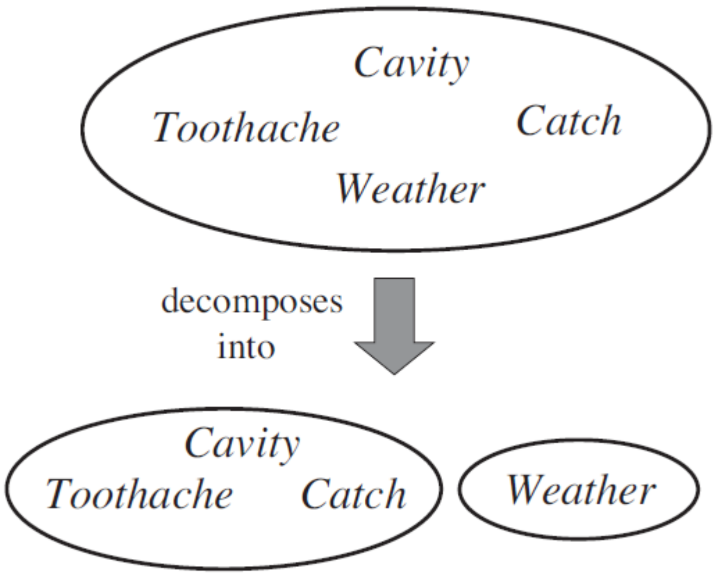
	toothache		¬toothache		toothache		¬toothache	
	catch	¬catch	catch	¬catch	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576	0.016	0.064	0.144	0.576

- The 32-element table can be reduced to a 8-element table and a 4-element table

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

Factoring the large joint distribution

Leveraging the (absolute) independence

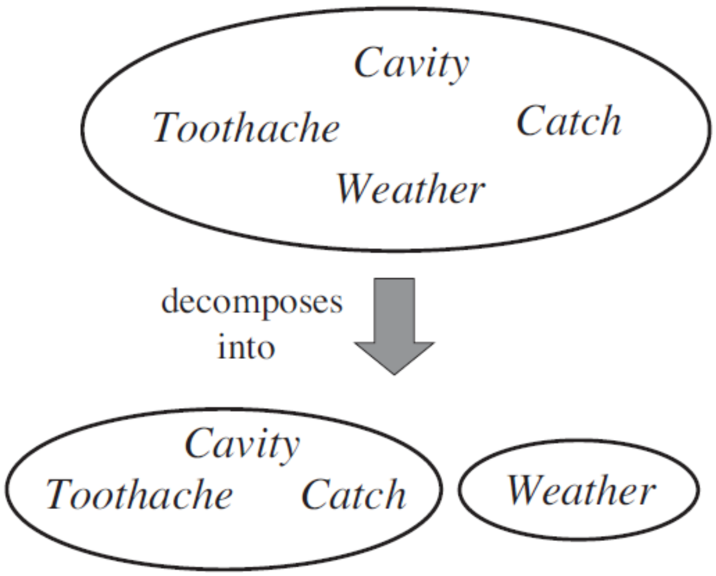


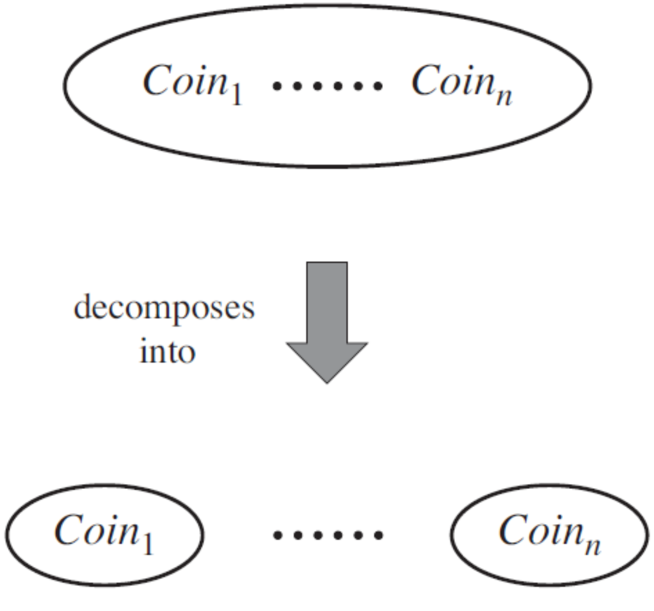
Weather and dentistry are independent

Coin flips are independent

Factoring the large joint distribution

Leveraging the (absolute) independence





Weather and dentistry are independent

Coin flips are independent

Outline

- Probability Theory
- Probabilistic Inference using Joint Distribution
- **Bayes' Rule**

Bayes' Rule

- Derive **Bayes' rule** from the **product rule** of conditional probability

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

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

- Equating the right-hand sides and dividing by $P(a)$

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

Bayes' Rule

- Derive **Bayes' rule** from the **product rule** of conditional probability

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

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

- Equating the right-hand sides and dividing by $P(a)$

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

This equation underlies most **modern AI systems** for **probabilistic inference…**

Applying Bayes' rule (simple case)

- **Question**: Why would anyone want to compute a single term $P(b|a)$ using three terms: $P(a|b)$, $P(b)$, and $P(a)$?

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

- **Answer**: whenever you have $P(a|b)$, $P(b)$, $P(a)$ but not $P(b|a)$ ---- for example, in medical diagnosis

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

Applying Bayes' rule (simple case)

- Assume that the doctor knows some unconditional facts:
  - Prior probability that a patient has meningitis $P(m)=1/50000$
  - Prior probability that a patient has stiff neck $P(s) = 0.01$
  - Meningitis causes patient to have stiff neck $P(s|m) = 0.7$

- Now, a patient has a stiff neck; what is the probability that this particular patient has meningitis?

$$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)}$$

Applying Bayes' rule (simple case)

- Assume that the doctor knows some unconditional facts:
  - Prior probability that a patient has meningitis $P(m)=1/50000$
  - Prior probability that a patient has stiff neck $P(s) = 0.01$
  - Meningitis causes patient to have stiff neck $P(s|m) = 0.7$
- Now, a patient has a stiff neck; what is the probability that this particular patient has meningitis?

$$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.7 \times 1/50000}{0.01} = 0.0014$$

Applying Bayes' rule (simple case)

- Assume that the doctor knows some unconditional facts:
  - Prior probability that a patient has meningitis $P(m)=1/50000$
  - ~~Prior probability that a patient has stiff neck $P(s) = 0.01$~~
  - Meningitis causes patient to have stiff neck $P(s|m) = 0.7$
  - **Non-meningitis causes patient to have stiff neck** $P(s|\neg m) = 0.0\ldots$

- The probability that a stiff-neck patient has meningitis?

$$P(s \mid \neg m) P(\neg m)$$

$$P(m \mid s) = \frac{P(s \mid m) P(m)}{\quad} = \frac{0.7 \times 1/50000}{\quad}$$

Using Bayes' rule (combining evidence)

- **Question**: What if we have two or more evidences?

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

- *Toothache* and *Catch* are **not independent**. But they will be independent **given the presence or absence of a cavity**

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

$$\boxed{\begin{aligned}&\mathbf{P}(Cavity \mid toothache \wedge catch)\\ &\quad = \alpha\, \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)\, \mathbf{P}(Cavity)\end{aligned}}$$

Conditional independence

- Two variables X and Y are conditional independent, given a third variable Z

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

- Alternatively, we have

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

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

Conditional independence (example)

- Two variables X and Y are conditional independent, given a third variable Z

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

- Example:

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

Conditional independence (example)

- Two variables X and Y are conditional independent, given a third variable Z

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

- Example:

$$\begin{aligned}
&\mathbf{P}(Toothache, Catch, Cavity) \\
&\quad = \mathbf{P}(Toothache, Catch \mid Cavity)\mathbf{P}(Cavity) \\
&\quad = \mathbf{P}(Toothache \mid Cavity)\mathbf{P}(Catch \mid Cavity)\mathbf{P}(Cavity)
\end{aligned}$$

(0.108+0.012)/0.2 =0.6

(0.108+0.072)/0.2 =0.9

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

Separation

- For *n* symptoms that are conditionally independent given *Cavity*, the **full joint distribution** table size grows as *O(n)* instead of $O(2^n)$

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

Outline

- Probability Theory
- Probabilistic Inference using Joint Distribution
- Bayes' Rule