Probability Basics

Bayes' Theorem

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

likelihood of *t* given a given features

Generalized Bayes' Theorem

prior

$$P(t = l|\mathbf{q}_1, \ldots, \mathbf{q}_m) = \frac{P(\mathbf{q}_1, \ldots, \mathbf{q}_m|t = l)P(t = l)}{P(\mathbf{q}_1, \ldots, \mathbf{q}_m)}$$

Normalising

posterior probability: degree of belief after knowing $(\mathbf{q}_1, \ldots, \mathbf{q}_m)$

Example

Table: A simple dataset for MENINGITIS diagnosis with descriptive features that describe the presence or absence of three common symptoms of the disease: HEADACHE, FEVER, and VOMITING.

ID	HEADACHE	FEVER	VOMITING	MENINGITIS
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

Meningitis is a serious disease

$$
\mathbf{P}(H, F, V, M) = \begin{bmatrix}
P(h, f, v, m), & P(\neg h, f, v, m) \\
P(h, f, v, \neg m), & P(\neg h, f, v, \neg m) \\
P(h, f, \neg v, m), & P(\neg h, f, \neg v, m) \\
P(h, f, \neg v, \neg m), & P(\neg h, f, \neg v, \neg m) \\
P(h, \neg f, v, m), & P(\neg h, \neg f, v, m) \\
P(h, \neg f, v, \neg m), & P(\neg h, \neg f, v, \neg m) \\
P(h, \neg f, \neg v, m), & P(\neg h, \neg f, \neg v, m) \\
P(h, \neg f, \neg v, \neg m), & P(\neg h, \neg f, \neg v, \neg m)
\end{bmatrix}
$$

$$P(M|h, \neg f, v) = ?$$

- In the terms of Bayes' Theorem this problem can be stated as:

$$P(M|h, \neg f, v) = \frac{P(h, \neg f, v|M) \times P(M)}{P(h, \neg f, v)}$$

- There are two values in the domain of the MENINGITIS feature, *'true'* and *'false'*, so we have to do this calculation twice.

- We will do the calculation for m first
- To carry out this calculation we need to know the following probabilities: $P(m)$, $P(h, \neg f, v)$ and $P(h, \neg f, v \mid m)$.

ID	HEADACHE	FEVER	VOMITING	MENINGITIS
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

$h, \neg f, v$

m

- We can calculate the required probabilities directly from the data. For example, we can calculate $P(m)$ and $P(h, \neg f, v)$ as follows:

$$P(m) = \frac{|\{\mathbf{d}_5, \mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \mathbf{d}_4, \mathbf{d}_5, \mathbf{d}_6, \mathbf{d}_7, \mathbf{d}_8, \mathbf{d}_9, \mathbf{d}_{10}\}|} = \frac{3}{10} = 0.3$$

$$P(h, \neg f, v) = \frac{|\{\mathbf{d}_3, \mathbf{d}_4, \mathbf{d}_6, \mathbf{d}_7, \mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \mathbf{d}_4, \mathbf{d}_5, \mathbf{d}_6, \mathbf{d}_7, \mathbf{d}_8, \mathbf{d}_9, \mathbf{d}_{10}\}|} = \frac{6}{10} = 0.6$$

- However, as an exercise we will use the chain rule calculate:

$$P(h, \neg f, v \mid m) = ?$$

ID	HEADACHE	FEVER	VOMITING	MENINGITIS
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

- Using the chain rule calculate:

$$P(h, \neg f, v \mid m) = P(h \mid m) \times P(\neg f \mid h, m) \times P(v \mid \neg f, h, m)$$

ID	HEADACHE	FEVER	VOMITING	MENINGITIS
5	false	true	false	true
8	true	false	true	true
10	true	false	true	true

- Using the chain rule calculate:

$$P(h, \neg f, v \mid m) = P(h \mid m) \times P(\neg f \mid h, m) \times P(v \mid \neg f, h, m)$$
$$= \frac{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_5, \mathbf{d}_8, \mathbf{d}_{10}\}|} \times \frac{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|} \times \frac{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|}$$

ID	HEADACHE	FEVER	VOMITING	MENINGITIS
5	false	true	false	true
8	true	false	true	true
10	true	false	true	true

- Using the chain rule calculate:

$$
\begin{aligned}
P(h, \neg f, v \mid m) &= P(h \mid m) \times P(\neg f \mid h, m) \times P(v \mid \neg f, h, m) \\
&= \frac{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_5, \mathbf{d}_8, \mathbf{d}_{10}\}|} \times \frac{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|} \times \frac{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|} \\
&= \frac{2}{3} \times \frac{2}{2} \times \frac{2}{2} = 0.6666
\end{aligned}
$$

ID	HEADACHE	FEVER	VOMITING	MENINGITIS
5	false	true	false	true
8	true	false	true	true
10	true	false	true	true

- So the calculation of $P(m|h, \neg f, v)$ is:

$$P(m|h, \neg f, v) = \frac{\begin{pmatrix} P(h|m) \times P(\neg f|h, m) \\ \times P(v|\neg f, h, m) \times P(m) \end{pmatrix}}{P(h, \neg f, v)}$$

- So the calculation of $P(m|h, \neg f, v)$ is:

$$P(m|h, \neg f, v) = \frac{\begin{pmatrix} P(h|m) \times P(\neg f|h, m) \\ \times P(v|\neg f, h, m) \times P(m) \end{pmatrix}}{P(h, \neg f, v)}$$

$$= \frac{0.6666 \times 0.3}{0.6} = 0.3333$$

- The corresponding calculation for $P(\neg m|h, \neg f, v)$ is:

$$P(\neg m \mid h, \neg f, v) = \frac{P(h, \neg f, v \mid \neg m) \times P(\neg m)}{P(h, \neg f, v)}$$

$$= \frac{\begin{pmatrix} P(h|\neg m) \times P(\neg f \mid h, \neg m) \\ \times P(v|\neg f, h, \neg m) \times P(\neg m) \end{pmatrix}}{P(h, \neg f, v)}$$

H.W.

$$P(m|h, \neg f, v) = 0.3333$$
$$P(\neg m|h, \neg f, v) = 0.6667$$

- These calculations tell us that it is twice as probable that the patient does not have meningitis than it is that they do even though the patient is suffering from a headache and is vomiting!

Uncertainty

Acting Under Uncertainty

- Agents often make decisions based on **incomplete** information
  - partial observability
  - nondeterministic actions

Acting Under Uncertainty

- Agents often make decisions based on **incomplete** information
  - partial observability
  - nondeterministic actions
- **Partial solution** (previous discussion): maintain **belief states**
  - represent the set of all possible world states the agent might be in
  - generating a **contingency plan** handling every possible eventuality
- **Several drawbacks**:

Acting Under Uncertainty

- Agents often make decisions based on **incomplete** information
  - partial observability
  - nondeterministic actions
- **Partial solution** (previous discussion): maintain **belief states**
  - represent the set of all possible world states the agent might be in
  - generating a contingency plan handling every possible eventuality
- **Several drawbacks**:
  - must consider **every possible** explanation for the observation
  - contingent plans handling every eventuality grow **arbitrarily large**
  - sometimes there is no plan that is **guaranteed** to achieve the goal

Logic and Uncertainty

Major problem with logical-agent approaches:

> *Agents almost never have access to the whole truth about their environments*

- Very often, even in simple worlds, there are important questions for which there is no yes/no answer
- In that case, an agent must reason under uncertainty
- Uncertainty also arises because of an agent's incomplete or incorrect understanding of its environment

Uncertainty Example

Let action A_t = "leave for airport t minutes before flight"

Will A_t get me there on time?

Problems (Too many sources of uncertainty)

- partial observability (road state, other drivers' plans, etc.)
- noisy sensors (unreliable traffic reports)
- uncertainty in action outcomes (flat tire, etc.)
- immense complexity of modeling and predicting traffic

Goal: deliver a passenger to the airport on time

Uncertainty Example

Let action A_t = "leave for airport t minutes before flight"

Will A_t get me there on time?

A purely logical approach either

1. risks falsehood ("A_{25} will get me there on time"), or
2. leads to conclusions that are too weak for decision making ("A_{25} will get me there on time if there's no accident on the way, it doesn't rain, my tires remain intact, ...")

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport ...)

Uncertainty Example

Let action A_t = "leave for airport t minutes before flight"

Will A_t get me there on time?

A purely logical approach either

1. risks falsehood ("A_{25} will get me there on time"), or
2. leads to conclusions that are too weak for decision making ("A_{25} will get me there on time if there's no accident on the way, it doesn't rain, my tires remain intact, ...")

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport ...)

With purely-logical approach it is difficult to anticipate everything that can go wrong

Reasoning under Uncertainty

A *rational* agent is one that makes rational decisions — in order to maximize its performance measure)

A rational decision depends on

- the relative importance of various goals
- the likelihood they will be achieved
- the degree to which they will be achieved

Reasoning under Uncertainty

A *rational* agent is one that makes rational decisions — in order to maximize its performance measure)

A rational decision depends on

- the relative importance of various goals
- the likelihood they will be achieved
- the degree to which they will be achieved

Probability theory offers a clean way to quantify likelihood

Example

Given the symptoms (toothache) infer the cause (cavity)

How to encode this relation in logic?

- diagnostic rules:

- causal rules:

Example

Given the symptoms (toothache) infer the cause (cavity)

How to encode this relation in logic?

- diagnostic rules:

    Toothache$\longrightarrow$ Cavity (wrong)

    Toothache$\longrightarrow$ (Cavity V Gum-Problem V Abscess V ...)

    (too many possible causes, some very unlikely)

Example

Given the symptoms (toothache) infer the cause (cavity)

How to encode this relation in logic?

- diagnostic rules:

  Toothache$\longrightarrow$ Cavity (wrong)

  Toothache$\longrightarrow$ (Cavity V Gum-Problem V Abscess V ...)

  (too many possible causes, some very unlikely)

- causal rules:

  Cavity $\longrightarrow$ Toothache (wrong)

  (Cavity ...) $\longrightarrow$ Toothache (many possible (con)causes)

Handling Uncertain Knowledge

Reasons FOL-based approaches fail to cope with domains like, for instance, medical diagnosis:

- Laziness: too much work to write complete axioms, or too hard to work with the enormous sentences that result
- Theoretical Ignorance: The available knowledge of the domain is incomplete
- Practical Ignorance: The theoretical knowledge of the domain is complete but some evidential facts are missing

Degrees of Belief and probability

In several real-world domains the agent's knowledge can only provide a degree of belief in the relevant sentences

The agent cannot say whether a sentence is true, but only that is true $x\%$ of the times

The main tool for handling degrees of belief is *Probability Theory*

The use of *probability summarizes the uncertainty that stems from our laziness or ignorance about the domain*

Probability theory

- Probability can be derived from
  - statistical data (ex: 80% of toothache patients so far had cavities)
  - some knowledge (ex: 80% of toothache patients has cavities)
  - their combination

Probability theory

- Probability can be derived from
  - statistical data (ex: 80% of toothache patients so far had cavities)
  - some knowledge (ex: 80% of toothache patients has cavities)
  - their combination
- Probability statements are made with respect to a state of knowledge (aka evidence)
  - e.g., "The probability that the patient has a cavity, given that she has a toothache, is 0.8":
  - P(HasCavity (patient) | hasToothAche(patient)) = 0.8

Probability theory

- Probabilities of **propositions** change with new evidence:
  - "The probability that the patient has a cavity, given that she has a toothache and a history of gum disease, is 0.4":
  - $P(HasCavity\ (patient)\ |\ hasToothAche(patient)\ \wedge\ HistoryOfGum(patient)) = 0.4$

Making Decisions Under Uncertainty

Ex: Suppose I believe:

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

Making Decisions Under Uncertainty

Ex: Suppose I believe:

- 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

Which action to choose?

- Depends on tradeoffs among preferences:
  - missing flight vs. costs (airport cuisine, sleep overnight in airport)

Making Decisions Under Uncertainty

Ex: Suppose I believe:

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

Which action to choose?

- Depends on tradeoffs among preferences:
  - missing flight vs. costs (airport cuisine, sleep overnight in airport)

When there are conflicting goals the agent may express preferences among them by means of a **utility function**.

Making Decisions Under Uncertainty

- Utilities are combined with probabilities in the general theory of rational decisions, aka **decision theory**:
- **Decision theory = Probability theory + Utility theory**
- **Maximum Expected Utility (MEU)**: an agent is **rational** if and only if it chooses the action that yields the **maximum expected utility**, averaged over all the possible outcomes of the action.

Probabilistic inference

Probability Theory

Probability Theory makes the same ontological commitments as First-order Logic:

Every sentence φ *is either true or false*

The *degree of belief* that φ is true is a number P between 0 and 1

$$
\begin{array}{lcl}
P(\varphi) = 1 & \longrightarrow & \varphi \text{ is certainly true} \\
P(\varphi) = 0 & \longrightarrow & \varphi \text{ is certainly not true} \\
P(\varphi) = 0.65 & \longrightarrow & \varphi \text{ is true with a } 65\% \text{ chance}
\end{array}
$$

Syntax for propositions

Propositional or Boolean random variables

- e.g., $Cavity$ (do I have a cavity?)
- $Cavity = true$ is a proposition, also written $cavity$

Discrete random variables (finite or infinite)

- e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow \rangle$
- $Weather = rain$ is a proposition
- Values must be exhaustive and mutually exclusive

Continuous random variables (bounded or unbounded)

- e.g., $Temp = 21.6$; also allow, e.g., $Temp < 22.0$.

Arbitrary Boolean combinations of basic propositions

Conditional probability

Definition of conditional probability:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ if } P(b) \neq 0$$

Product rule gives an alternative formulation:

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

A general version holds for whole distributions, e.g.,

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

(View as a 4×2 set of equations, **not** matrix mult.)

Inference by enumeration

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$

Inference by enumeration

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$$

$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Can also compute conditional probabilities:

$$P(\neg cavity | toothache) = ?$$

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Can also compute conditional probabilities:

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

Inference by enumeration

Start with the joint distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Can also compute conditional probabilities:

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

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

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

Denominator can be viewed as a normalization constant α

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

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

Denominator can be viewed as a normalization constant α

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

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

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

Denominator can be viewed as a normalization constant α

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

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

$$= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

Denominator can be viewed as a normalization constant α

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

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

$$= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$

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

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Denominator can be viewed as a normalization constant α

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

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

$$= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$

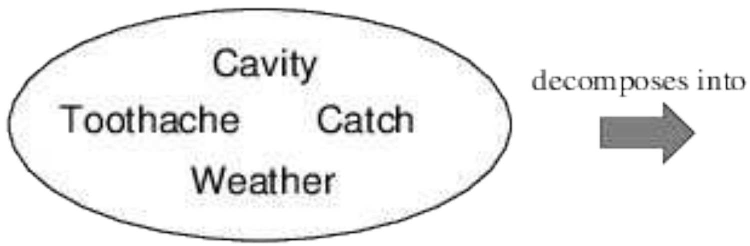
$$= \alpha\,\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle$$

General idea: compute distribution on query variable
by fixing evidence variables and summing over hidden variables

Independence

A and B are independent iff

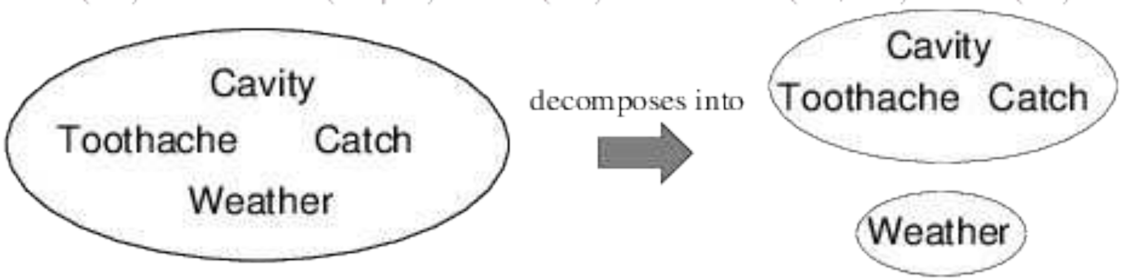
$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$



Independence

A and B are independent iff

$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$



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

Conditional independence

$\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ independent entries

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

(1) $P(catch|toothache, cavity) = P(catch|cavity)$

Conditional independence

$\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ independent entries

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

(1) $P(catch|toothache, cavity) = P(catch|cavity)$

The same independence holds if I haven't got a cavity:

(2) $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

Conditional independence

$\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ independent entries

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
(1) $P(catch|toothache, cavity) = P(catch|cavity)$

The same independence holds if I haven't got a cavity:
(2) $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

$Catch$ is conditionally independent of $Toothache$ given $Cavity$:
$\mathbf{P}(Catch|Toothache, Cavity) = \mathbf{P}(Catch|Cavity)$

Conditional independence

$\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ independent entries

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

(1) $P(catch|toothache, cavity) = P(catch|cavity)$

The same independence holds if I haven't got a cavity:

(2) $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

$Catch$ is conditionally independent of $Toothache$ given $Cavity$:

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

Equivalent statements:

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

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

Conditional independence contd.

Write out full joint distribution using chain rule:

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

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

Conditional independence contd.

Write out full joint distribution using chain rule:

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

Conditional independence contd.

Write out full joint distribution using chain rule:

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

How many independent numbers?

Conditional independence contd.

Write out full joint distribution using chain rule:

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

$2 + 2 + 1 = 5$ independent numbers

In many cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in n to linear in n.

Conditional independence is our most basic and robust form of knowledge about uncertain environments.