Uncertainty

Chapter 13

Outline

- Uncertainty
- Probability
- Syntax and Semantics
- Inference
- Independence and Bayes' Rule

Uncertainty

- Let's say you want to get to the airport in time for a flight.
- Let action $A_t =$ leave for airport t minutes before flight
- Q: Will A_t get me there on time?

Uncertainty

- Let's say you want to get to the airport in time for a flight.
- Let action A_t = leave for airport t minutes before flight
- Q: Will A_t get me there on time?
- Problems:

  1. partial observability (road state, other drivers' plans, etc.)
  2. noisy sensors (traffic reports, possibly-inaccurate clocks, etc.)
  3. uncertainty in action outcomes (flat tire, etc.)
  4. immense complexity of modelling and predicting traffic

Uncertainty

Hence a purely logical approach either:

1. risks falsehood: "A_{45} will get me there on time"
2. leads to conclusions that are too weak for decision making:
   - "A_{45} will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact etc etc."

Methods for handling uncertainty

Default (or nonmonotonic) logic:

- Assume my car does not have a flat tire
- Assume A_{45} works unless contradicted by evidence
- Issues: What assumptions are reasonable? How do you reason with assumptions?

Methods for handling uncertainty

Default (or nonmonotonic) logic:

- Assume my car does not have a flat tire
- Assume A_{45} works unless contradicted by evidence
- Issues: What assumptions are reasonable? How do you reason with assumptions?

Rules with fudge factors:

- $A_{45} \mapsto_{0.3} AtAirportOnTime$
  $Sprinkler \mapsto_{0.99} WetGrass$; $\quad WetGrass \mapsto_{0.7} Rain$
- Issues: Problems with combination, (*Sprinkler* causes *Rain*??)

Methods for handling uncertainty

Default (or nonmonotonic) logic:

- Assume my car does not have a flat tire
- Assume A_{45} works unless contradicted by evidence
- Issues: What assumptions are reasonable? How do you reason with assumptions?

Rules with fudge factors:

- $A_{45} \mapsto_{0.3} AtAirportOnTime$
  $Sprinkler \mapsto_{0.99} WetGrass$; $\quad WetGrass \mapsto_{0.7} Rain$
- Issues: Problems with combination, (*Sprinkler* causes *Rain*??)

Probability:

- Given available evidence, A_{45} works with probability 0.80
- Aside: *Fuzzy logic* handles *degree of truth* NOT uncertainty
  e.g., *WetGrass* is true to degree 0.2

Probability

- Probabilistic assertions *summarize* effects of
  - *"Laziness"*: Failure to enumerate exceptions, qualifications, etc.
  - *"Ignorance"*: Lack of relevant facts, initial conditions, etc. Lack of a background theory (e.g. medical domains)

Probability

- Probabilistic assertions *summarize* effects of
  - *"Laziness"*: Failure to enumerate exceptions, qualifications, etc.
  - *"Ignorance"*: Lack of relevant facts, initial conditions, etc. Lack of a background theory (e.g. medical domains)
- Approach: *Subjective* or *Bayesian* probability:
  - Probabilities relate propositions to an agent's state of knowledge
  - e.g., $P(A_{45} | \text{no reported accidents}) = 0.90$

Probability

- Probabilistic assertions *summarize* effects of
  - *"Laziness"*: Failure to enumerate exceptions, qualifications, etc.
  - *"Ignorance"*: Lack of relevant facts, initial conditions, etc. Lack of a background theory (e.g. medical domains)
- Approach: *Subjective* or *Bayesian* probability:
  - Probabilities relate propositions to an agent's state of knowledge
  - e.g., $P(A_{45} | \text{no reported accidents}) = 0.90$
- Probabilities of propositions change with new evidence:
  - e.g., $P(A_{45} | \text{no reported accidents, 5 a.m.}) = 0.98$

Making decisions under uncertainty

- Suppose I believe the following:

$$
\begin{aligned}
P(A_{45} \text{ gets me there on time}|\ldots) &= 0.80 \\
P(A_{90} \text{ gets me there on time}|\ldots) &= 0.95 \\
P(A_{120} \text{ gets me there on time}|\ldots) &= 0.99 \\
P(A_{1440} \text{ gets me there on time}|\ldots) &= 0.9999
\end{aligned}
$$

  Which action to choose?

Making decisions under uncertainty

- Suppose I believe the following:

$$\begin{aligned} P(A_{45} \text{ gets me there on time}|\ldots) &= 0.80 \\ P(A_{90} \text{ gets me there on time}|\ldots) &= 0.95 \\ P(A_{120} \text{ gets me there on time}|\ldots) &= 0.99 \\ P(A_{1440} \text{ gets me there on time}|\ldots) &= 0.9999 \end{aligned}$$

  Which action to choose?
- Depends on my *preferences* for missing flight vs. airport cuisine, etc.
- *Utility theory* is used to represent and infer preferences
- *Decision theory* = probability theory + utility theory

Probability Basics: Syntax

- Basic element: *random variable* that can have a value.
  - Like CSP
    - A variable will have a *domain* of possible values
  - Unlike 1st-order variables.

Probability Basics: Syntax

- Basic element: *random variable* that can have a value.
  - Like CSP
    - A variable will have a *domain* of possible values
  - Unlike 1st-order variables.
- Boolean variables
  - e.g., *Cavity* (do I have a cavity?) – *true*, *false*.

Probability Basics: Syntax

- Basic element: *random variable* that can have a value.
  - Like CSP
    - A variable will have a *domain* of possible values
  - Unlike 1st-order variables.
- Boolean variables
  - e.g., *Cavity* (do I have a cavity?) – *true*, *false*.
- Discrete variables
  - e.g., *Weather* is one of
    $\langle sunny, rainy, cloudy, snow\rangle$

Syntax

- *Atom* = assignment of value to variable.
  - Examples:
    - Weather = sunny
    - Cavity = false.

Syntax

- *Atom* = assignment of value to variable.
  - Examples:
    - Weather = sunny
    - Cavity = false.
- *Sentences* are Boolean combinations of atoms.
  - Same as propositional logic.
  - Examples:
    - Weather = sunny OR Cavity = false.
    - Catch = true AND Tootache = false.

Probabilities and Possible Worlds

- *Possible World*: A complete assignment of values to every variable.

Probabilities and Possible Worlds

- *Possible World*: A complete assignment of values to every variable.
  - Like a model in propositional logic.
  - E.g. [weather=sunny, cavity = false, catch = false, toothache = true]
  - (This is a simplification, but will do for our needs.)

Probabilities and Possible Worlds

- *Possible World*: A complete assignment of values to every variable.
  - Like a model in propositional logic.
  - E.g. [weather=sunny, cavity = false, catch = false,
    toothache = true]
  - (This is a simplification, but will do for our needs.)
- The set of all possible worlds is called the *sample space*, denoted Ω.
- *Event* or *proposition* = a set of possible worlds.
- *Atomic event* = a single possible world.

Propositions

- Think of a *proposition* as the event (set of sample points or possible worlds) where the proposition is true.
- Given Boolean random variables A and B:

Propositions

- Think of a *proposition* as the event (set of sample points or possible worlds) where the proposition is true.
- Given Boolean random variables A and B:
  - event a = set of sample points where $A(\omega) = \mathit{true}$
    (Write a for $A = \mathit{true}$.)

Propositions

- Think of a *proposition* as the event (set of sample points or possible worlds) where the proposition is true.
- Given Boolean random variables A and B:
    - event a = set of sample points where $A(\omega) = true$ (Write a for $A = true$.)
    - event $\neg a$ = set of sample points where $A(\omega) = false$

Propositions

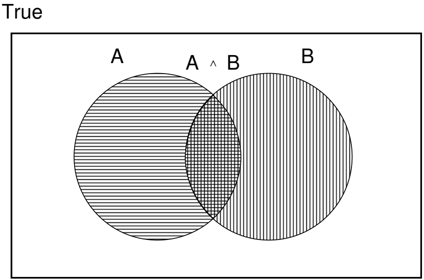
- Think of a *proposition* as the event (set of sample points or possible worlds) where the proposition is true.
- Given Boolean random variables A and B:
  - event a = set of sample points where $A(\omega) = true$
    (Write a for $A = true$.)
  - event $\neg a$ = set of sample points where $A(\omega) = false$
  - event $a \wedge b$ = points where $A(\omega) = true$ and $B(\omega) = true$

Propositions

- As far as we're concerned, the sample points are *defined* by the values of a set of random variables,
  - I.e., the sample space is the Cartesian product of the ranges of the variables
- With Boolean variables, a sample point = propositional logic model
  - e.g., $A = \mathit{true}$, $B = \mathit{false}$, or $a \wedge \neg b$.
- Probability of a proposition:
  - Can be determined via disjunction of atomic events in which it is true
  - e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$
    So $P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Why use probability?

- The definitions imply that certain logically related events must have related probabilities
- E.g., $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



Axioms of probability

- Basic axioms:
  - $0 \leq P(a) \leq 1$.
  - $P(true) = 1 \qquad P(false) = 0$.
  - $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$
- Obtain:
  - $P(\neg a) = 1 - P(a)$.
- de Finetti (1931): An agent who bets according to probabilities that violate these axioms can be forced to bet so as to lose money regardless of outcome.

Syntax for propositions

More generally, there are three kinds of random variables:

1. *Propositional* or *Boolean* random variables
    - E.g., *Cavity* (do I have a cavity?)
    - $Cavity = true$ is a proposition.
      Recall: write *cavity* for $Cavity = true$.

Syntax for propositions

More generally, there are three kinds of random variables:

1. *Propositional* or *Boolean* random variables
   - E.g., *Cavity* (do I have a cavity?)
   - $Cavity = true$ is a proposition.
     Recall: write *cavity* for $Cavity = true$.
2. *Discrete* random variables (*finite* or *infinite*)
   - E.g., *Weather* is one of $\langle sunny, rain, cloudy, snow \rangle$
   - $Weather = rain$ is a proposition
   - Values are exhaustive and mutually exclusive

Syntax for propositions

More generally, there are three kinds of random variables:

1. *Propositional* or *Boolean* random variables
   - E.g., *Cavity* (do I have a cavity?)
   - $Cavity = true$ is a proposition.
     Recall: write *cavity* for $Cavity = true$.
2. *Discrete* random variables (*finite* or *infinite*)
   - E.g., *Weather* is one of $\langle sunny, rain, cloudy, snow \rangle$
   - $Weather = rain$ is a proposition
   - Values are exhaustive and mutually exclusive
3. *Continuous* random variables (*bounded* or *unbounded*)
   - E.g., $Temp = 21.6$; also allow, e.g., $Temp < 22.0$.
   - We're not going to bother with continuous random variables.

Form complex propositions using logical connectives ($\neg$, $\wedge$, $\vee$, ...).

Prior probability

Prior or unconditional probabilities of propositions

- e.g., $P(Cavity = true) = 0.1$ or $P(Weather = sunny) = 0.72$ specifies belief prior to arrival of any (new) evidence

Prior probability

Prior or unconditional probabilities of propositions

- e.g., $P(Cavity = true) = 0.1$ or $P(Weather = sunny) = 0.72$ specifies belief prior to arrival of any (new) evidence

Probability distribution

- Sometimes we want to talk about the probabilities of all values of a r.v., e.g.:
  $P(Weather = sunny) = 0.6$
  $P(Weather = rain) = 0.1$
  $P(Weather = cloudy) = 0.29$
  $P(Weather = snow) = 0.01$
- The *probability distribution* gives values for all possible assignments:
  - $\mathbf{P}(Weather) = \langle 0.60, 0.10, 0.29, 0.01 \rangle$
  - Assumes ordering on values: $\langle sunny, rain, cloudy, snow \rangle$.

Joint probability distribution

- *Joint probability distribution* for a set of r.v's gives the probability of every atomic event on those r.v's (i.e., every sample point)

  $\mathbf{P}(Weather, Cavity)$ = a 4×2 matrix of values:

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

- Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

Conditional probability

- *Conditional* or *posterior probabilities*:
  $P(\alpha|\beta) =$ prob of α given that β.

Conditional probability

- *Conditional* or *posterior probabilities*:
  $P(\alpha|\beta)$ = prob of α given that β.
- E.g., $P(\mathit{cavity}|\mathit{toothache}) = 0.8$
  - Probablility of cavity is 0.8 *given that* *toothache* *is all I know*
  - NOT "if *toothache* then 80% chance of *cavity*"

Conditional probability

- *Conditional* or *posterior probabilities*:
  $P(\alpha|\beta)$ = prob of α given that β.
- E.g., $P(\mathit{cavity}|\mathit{toothache}) = 0.8$
  - Probablility of cavity is 0.8 *given that toothache is all I know*
  - NOT "if *toothache* then 80% chance of *cavity*"
- If we know more (e.g., *cavity* is also given) then we have
  $P(\mathit{cavity}|\mathit{toothache}, \mathit{cavity}) = 1$
  (or: $P(\mathit{cavity}|\mathit{toothache} \wedge \mathit{cavity}) = 1$)

Conditional probability

- *Conditional* or *posterior probabilities*:

  $P(\alpha|\beta)$ = prob of α given that β.
- E.g., $P(\mathit{cavity}|\mathit{toothache}) = 0.8$
  - Probablility of cavity is 0.8 *given that toothache is all I know*
  - NOT "if *toothache* then 80% chance of *cavity*"
- If we know more (e.g., *cavity* is also given) then we have

  $P(\mathit{cavity}|\mathit{toothache}, \mathit{cavity}) = 1$

  (or: $P(\mathit{cavity}|\mathit{toothache} \wedge \mathit{cavity}) = 1$)
- Note: the less specific belief *remains valid* after more evidence arrives, but is not always *useful*
- New evidence may be irrelevant, allowing simplification, e.g., $P(\mathit{cavity}|\mathit{toothache}, \mathit{LeafsWin}) = P(\mathit{cavity}|\mathit{toothache}) = 0.8$

Conditional probability

- Definition of conditional probability:

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

Conditional probability

- Definition of conditional probability:

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

- This can be rewritten as follows (called the *product rule*):
  $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$
- 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 multiplication.
  - I.e. $\mathbf{P}(X, Y) = \mathbf{P}(X|Y)\mathbf{P}(Y)$ stands for

$$\begin{aligned} P(X = x_1 \wedge Y = y_1) &= P(X = x_1|Y = y_1)P(y = y_1) \\ P(X = x_1 \wedge Y = y_2) &= P(X = x_1|Y = y_2)P(y = y_2) \\ \ldots \end{aligned}$$

Conditional probability

- The *chain rule* is derived by application of the product rule:

$$\begin{aligned}
&\mathbf{P}(X_1, \ldots, X_n) = \\
&\quad = \mathbf{P}(X_1, \ldots, X_{n-1})\ \mathbf{P}(X_n | X_1, \ldots, X_{n-1}) \\
&\quad = \mathbf{P}(X_1, \ldots, X_{n-2})\ \mathbf{P}(X_{n-1} | X_1, \ldots, X_{n-2}) \\
&\qquad\qquad \mathbf{P}(X_n | X_1, \ldots, X_{n-1}) \\
&\quad = \ldots \\
&\quad = \Pi_{i=1}^{n} \mathbf{P}(X_i | X_1, \ldots, X_{i-1})
\end{aligned}$$

- I.e.

$$\begin{aligned}
&\mathbf{P}(X_1, \ldots, X_n) = \\
&\qquad \mathbf{P}(X_n | X_1, \ldots, X_{n-1})\ \mathbf{P}(X_{n-1} | X_1, \ldots, X_{n-2})\ \ldots \\
&\qquad\qquad \mathbf{P}(X_2 | X_1)\ \mathbf{P}(X_1)
\end{aligned}$$

- E.g. $\mathbf{P}(X_1, X_2, X_3) = \mathbf{P}(X_3 | X_1, X_2) \mathbf{P}(X_2 | X_1) \mathbf{P}(X_1)$

Inference by enumeration

- Start with the joint distribution
- E.g. consider a domain with 3 Boolean variables: *Toothache*, *Catch*, *Cavity*.

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *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**

- For any proposition ϕ, sum the atomic events where it is true:
  - $P(\phi) = \Sigma_{\omega:\omega\models\phi} P(\omega)$
  - $P(cavity \vee toothache) =$
    $0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$

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:

$$\begin{aligned} P(\neg cavity \mid toothache) &= \frac{P(\neg cavity \wedge toothache)}{P(toothache)} \\ &= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} \\ &= 0.4 \end{aligned}$$

Inference by enumeration, contd.

- Let $\mathbf{X}$ be the query variables.
  - Typically, we want the posterior joint distribution of $\mathbf{X}$ given specific values $\mathbf{e}$ for the *evidence variables* $\mathbf{E}$
  - The remaining (unobserved) variables are sometimes called the *hidden variables*; let the hidden variables be $\mathbf{Y}$

Inference by enumeration, contd.

- Let **X** be the query variables.
  - Typically, we want the posterior joint distribution of **X** given specific values **e** for the *evidence variables* **E**
  - The remaining (unobserved) variables are sometimes called the *hidden variables*; let the hidden variables be **Y**
- The required summation of joint entries is done by *summing out* the hidden variables (i.e. **y** ranges over all combinations of values of **Y**):

$$\mathbf{P}(\mathbf{X}|\mathbf{e}) = \mathbf{P}(\mathbf{X},\mathbf{e})/\mathbf{P}(\mathbf{e}) = \Sigma_{\mathbf{y}}\mathbf{P}(\mathbf{X},\mathbf{e},\mathbf{y})/\mathbf{P}(\mathbf{e})$$

- The terms in the summation are joint entries because **X**, **E**, and **Y** together exhaust the set of random variables.

Inference by enumeration, contd.

- Let $\mathbf{X}$ be the query variables.
  - Typically, we want the posterior joint distribution of $\mathbf{X}$ given specific values $\mathbf{e}$ for the *evidence variables* $\mathbf{E}$
  - The remaining (unobserved) variables are sometimes called the *hidden variables*; let the hidden variables be $\mathbf{Y}$
- The required summation of joint entries is done by *summing out* the hidden variables (i.e. $\mathbf{y}$ ranges over all combinations of values of $\mathbf{Y}$):

$$\mathbf{P}(\mathbf{X}|\mathbf{e}) = \mathbf{P}(\mathbf{X},\mathbf{e})/\mathbf{P}(\mathbf{e}) = \Sigma_{\mathbf{y}}\mathbf{P}(\mathbf{X},\mathbf{e},\mathbf{y})/\mathbf{P}(\mathbf{e})$$

- The terms in the summation are joint entries because $\mathbf{X}$, $\mathbf{E}$, and $\mathbf{Y}$ together exhaust the set of random variables.
- Problem: For n Boolean variables need a table of size $O(2^n)$.
- Also: *Where do you get all the probabilities?*

Aside: 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

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

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

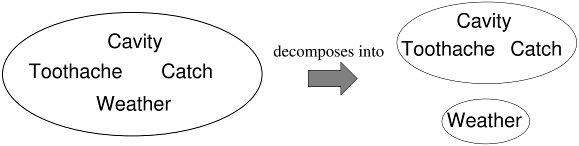
- General idea: Compute the distribution on query variable by fixing *evidence variables* and summing over *hidden variables*

Independence

- A and B are *independent* iff $\quad \mathbf{P}(A|B) = \mathbf{P}(A) \quad$ or
  $$\mathbf{P}(B|A) = \mathbf{P}(B) \quad \text{or} \quad \mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$$

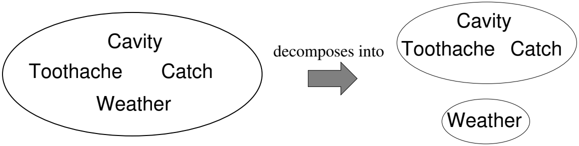
Independence

- A and B are *independent* iff $\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$
- E.g.:



Independence

- A and B are *independent* iff $\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$
- E.g.:



- So: $\mathbf{P}(Toothache, Catch, Cavity, Weather)$
$$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$$
- 32 entries reduce to 12.
- For n independent binary variables, $2^n \rightarrow n$

Independence

- Absolute independence is powerful, but rare
- E.g. dentistry is a large field with hundreds of variables, none of which are independent.
- What to do?

Conditional independence

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

Conditional independence

- $\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ indep entries
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  $P(catch|toothache, cavity) = P(catch|cavity)$

Conditional independence

- $\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ indep entries
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  $P(catch|toothache, cavity) = P(catch|cavity)$
- The same independence holds if I haven't got a cavity:
  $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

Conditional independence

- $\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ indep entries
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  $P(catch|toothache, cavity) = P(catch|cavity)$
- The same independence holds if I haven't got a cavity:
  $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$
- *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 $2^3 - 1 = 7$ indep entries
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  $P(catch|toothache, cavity) = P(catch|cavity)$
- The same independence holds if I haven't got a cavity:
  $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$
- *Catch* is *conditionally independent* of *Toothache* given *Cavity*:
  $\mathbf{P}(Catch|Toothache, Cavity) = \mathbf{P}(Catch|Cavity)$
- Statements equivalent to the last one:
  $\mathbf{P}(Toothache|Catch, Cavity) = \mathbf{P}(Toothache|Cavity)$
  $\mathbf{P}(Toothache, Catch|Cavity) =$
  $\qquad \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)$

Conditional independence contd.

- Write out the full joint distribution using the chain rule:

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

- I.e., $2 + 2 + 1 = 5$ independent numbers instead of 8.

Conditional independence contd.

- Write out the full joint distribution using the chain rule:

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

- I.e., $2 + 2 + 1 = 5$ independent numbers instead of 8.
- In most 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 contd.

- Write out the full joint distribution using the chain rule:
  $$\begin{aligned}&\mathbf{P}(Toothache, Catch, Cavity)\\&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch, Cavity)\\&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)\\&= \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)\end{aligned}$$
- I.e., $2 + 2 + 1 = 5$ independent numbers instead of 8.
- In most 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 the most basic and robust form of knowledge about uncertain environments.*

Conditional independence contd.

- Write out the full joint distribution using the chain rule:

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

- I.e., $2 + 2 + 1 = 5$ independent numbers instead of 8.
- In most 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 the most basic and robust form of knowledge about uncertain environments.*
- Problem: How to systematically figure out conditional independence relations?

Bayes' Rule

- Product rule $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$
- *Bayes' rule*:

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

- Or in joint distribution form :

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

Bayes' Rule

- Useful for assessing *diagnostic* probability from *causal* probability:

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

- Typically $P(Effect|Cause)$ is easier to determine than $P(Cause|Effect)$
- E.g., let M be meningitis, S be stiff neck:

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

  - Note: posterior probability of meningitis still very small!

Bayes' Rule and conditional independence

- Recall:

$$\begin{aligned} &\mathbf{P}(\mathit{Cavity} \wedge \mathit{Toothache} \wedge \mathit{Catch}) \\ &\quad = \alpha\,\mathbf{P}(\mathit{Toothache} \wedge \mathit{Catch} | \mathit{Cavity})\mathbf{P}(\mathit{Cavity}) \\ &\quad = \alpha\,\mathbf{P}(\mathit{Toothache} | \mathit{Cavity})\mathbf{P}(\mathit{Catch} | \mathit{Cavity})\mathbf{P}(\mathit{Cavity}) \end{aligned}$$

(Recall: α is a normalization constant, st. entries sum to 1.)

Bayes' Rule and conditional independence

- Recall:

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

  (Recall: α is a normalization constant, st. entries sum to 1.)

- This is an example of a *naive Bayes* model:
  $\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i \mathbf{P}(Effect_i | Cause)$

- Total number of parameters is *linear* in n

Summary

- Probability is a rigorous formalism for uncertain knowledge
- *Joint probability distribution* specifies probability of every *atomic event*
- Queries can be answered by summing over atomic events
- For nontrivial domains, we must find a way to reduce the joint size
- *Independence* and *conditional independence* provide the tools