Dealing with Uncertainty

Paolo Turrini

Department of Computing, Imperial College London

Introduction to Artificial Intelligence
2nd Part

Uncertainty and Probabilities

The main reference

Stuart Russell and Peter Norvig
Artificial Intelligence: a modern approach
Chapter 13

Outline

- Uncertainty
- Probability
- Probability and logic
- Inference

Uncertain outcomes

Uncertain outcomes

I have a lecture on Thursday in the early morning

Uncertain outcomes

I have a lecture on Thursday in the early morning and an alarm clock set for even earlier.

Uncertain outcomes

I have a lecture on Thursday in the early morning and an alarm clock set for even earlier.

Let action S_t = snooze the alarm clock t times

Uncertain outcomes

I have a lecture on Thursday in the early morning
and an alarm clock set for even earlier.

Let action S_t = snooze the alarm clock t times

Will S_t get me there on time?

Uncertain outcomes

I have a lecture on Thursday in the early morning
and an alarm clock set for even earlier.

Let action S_t = snooze the alarm clock t times

Will S_t get me there on time?

Problems:

Uncertain outcomes

I have a lecture on Thursday in the early morning
and an alarm clock set for even earlier.

Let action S_t = snooze the alarm clock t times

Will S_t get me there on time?

Problems:

1. partial observability (planned engineering works, announced strikes, etc.)

Uncertain outcomes

I have a lecture on Thursday in the early morning
and an alarm clock set for even earlier.

Let action S_t = snooze the alarm clock t times

Will S_t get me there on time?

Problems:

1. partial observability (planned engineering works, announced strikes, etc.)
2. noisy sensors (BBC reports, Google maps)

Uncertain outcomes

I have a lecture on Thursday in the early morning
and an alarm clock set for even earlier.

Let action S_t = snooze the alarm clock t times

Will S_t get me there on time?

Problems:

1. partial observability (planned engineering works, announced strikes, etc.)
2. noisy sensors (BBC reports, Google maps)
3. uncertainty in action outcomes (my phone might die, etc.)

Uncertain outcomes

I have a lecture on Thursday in the early morning
and an alarm clock set for even earlier.

Let action S_t = snooze the alarm clock t times

Will S_t get me there on time?

Problems:

1. partial observability (planned engineering works, announced strikes, etc.)
2. noisy sensors (BBC reports, Google maps)
3. uncertainty in action outcomes (my phone might die, etc.)
4. immense complexity of modelling and predicting traffic

Uncertainty

A binary *true-false* approach either:

Uncertainty

A binary *true-false* approach either:

1. might lead to conclusions that are too strong:

Uncertainty

A binary *true-false* approach either:

1. might lead to conclusions that are too strong: "S_{25} will not get me there on time"

Uncertainty

A binary *true-false* approach either:

1. might lead to conclusions that are too strong: "S_{25} will not get me there on time"
2. or too weak:

Uncertainty

A binary *true-false* approach either:

1. might lead to conclusions that are too strong:
   "S_{25} will not get me there on time"
2. or too weak:
   - "S_{25} will not get me there on time unless there's no delay on the District Line and it doesn't rain and I haven't forgotten the keys at home etc."

Methods for handling uncertainty: defaults

default logic handles "normal circumstances":

Methods for handling uncertainty: defaults

default logic handles "normal circumstances":

- Tube normally runs

Methods for handling uncertainty: defaults

default logic handles "normal circumstances":

- Tube normally runs
- Announced strikes normally happen

Methods for handling uncertainty: defaults

default logic handles "normal circumstances":

- Tube normally runs
- Announced strikes normally happen

- Issues:

Methods for handling uncertainty: defaults

default logic handles "normal circumstances":

- Tube normally runs
- Announced strikes normally happen

- Issues:
  - What assumptions are reasonable?

Methods for handling uncertainty: defaults

default logic handles "normal circumstances":

- Tube normally runs
- Announced strikes normally happen

- Issues:
  - What assumptions are reasonable?
  - How to handle contradiction? (e.g., will the tube run?)

Methods for handling uncertainty: defaults

default logic handles "normal circumstances":

- Tube normally runs
- Announced strikes normally happen

- Issues:
  - What assumptions are reasonable?
  - How to handle contradiction? (e.g., will the tube run?)

Also, fuzzy logic handles **degrees of truth**. It doesn't arguably handle uncertainty e.g., *Asleep* is true to degree 0.2

Rules with fudge factors

e..g, $S_{25} \mapsto_{0.4} AtLectureOnTime$

Rules with fudge factors

e..g, $S_{25} \mapsto_{0.4} AtLectureOnTime$

But...

- $ReadingSteinbeck \mapsto_{0.7} FallAsleep$

Rules with fudge factors

e..g, $S_{25} \mapsto_{0.4}$ $AtLectureOnTime$
But...

- $ReadingSteinbeck \mapsto_{0.7} FallAsleep$
- $FallAsleep \mapsto_{0.99} DarkOutside$

Rules with fudge factors

e..g, $S_{25} \mapsto_{0.4} AtLectureOnTime$
But...

- $ReadingSteinbeck \mapsto_{0.7} FallAsleep$
- $FallAsleep \mapsto_{0.99} DarkOutside$

Problems with combination, e.g., $ReadingSteinbeck \mapsto_{\sim 0.7} DarkOutside$

Rules with fudge factors

e..g, $S_{25} \mapsto_{0.4} AtLectureOnTime$
But...

- $ReadingSteinbeck \mapsto_{0.7} FallAsleep$
- $FallAsleep \mapsto_{0.99} DarkOutside$

Problems with combination, e.g.,
$ReadingSteinbeck \mapsto_{\sim 0.7} DarkOutside$

Causal connections?

Probabilities

Probability

$$P(S_{25} \text{ gets me there on time} | \ldots) = 0.2$$

Given the available evidence, S_{25} will get me there on time with probability 0.2

Probabilities

Probability

$$P(S_{25} \text{ gets me there on time} | \ldots) = 0.2$$

Given the available evidence, S_{25} will get me there on time with probability 0.2

Probabilistic assertions **summarize** effects of:

Probabilities

Probability

$$P(S_{25} \text{ gets me there on time} | \ldots) = 0.2$$

Given the available evidence, S_{25} will get me there on time with probability 0.2

Probabilistic assertions **summarize** effects of:

- laziness: failure to enumerate exceptions, qualifications, etc.

Probabilities

Probability

$$P(S_{25} \text{ gets me there on time} | \ldots) = 0.2$$

Given the available evidence, S_{25} will get me there on time with probability 0.2

Probabilistic assertions **summarize** effects of:

- laziness: failure to enumerate exceptions, qualifications, etc.
- ignorance: lack of relevant facts, initial conditions, etc.

Probabilities

Probability

$$P(S_{25} \text{ gets me there on time} | \ldots) = 0.2$$

Given the available evidence, S_{25} will get me there on time with probability 0.2

Probabilistic assertions **summarize** effects of:

- laziness: failure to enumerate exceptions, qualifications, etc.
- ignorance: lack of relevant facts, initial conditions, etc.

- Subjective/Bayesian view: Probabilities relate propositions to one's own state of knowledge e.g., $P(S_{25} \text{ gets me there on time} | \text{no reported accidents}) = 0.3$

Probability

- These are **not** claims of a "probabilistic tendency" in the current situation (but might be learned from past experience of similar situations)

Probability

- These are **not** claims of a "probabilistic tendency" in the current situation (but might be learned from past experience of similar situations)
- Probabilities of propositions change with new evidence:

Probability

- These are **not** claims of a "probabilistic tendency" in the current situation (but might be learned from past experience of similar situations)
- Probabilities of propositions change with new evidence: e.g., $P(S_{25} | \text{no reported accidents, 5 a.m.}) = 0.8$

Probability

- These are **not** claims of a "probabilistic tendency" in the current situation (but might be learned from past experience of similar situations)
- Probabilities of propositions change with new evidence: e.g., $P(S_{25}|\text{no reported accidents, 5 a.m.}) = 0.8$
- Analogous to logical entailment status $KB \models \alpha$, not truth.

If you snooze you lose?

Suppose I believe the following:

If you snooze you lose?

Suppose I believe the following:

$$P(S_0 \text{ gets me there on time} | \ldots) = 0.99$$

If you snooze you lose?

Suppose I believe the following:

$$
\begin{aligned}
P(S_0 \text{ gets me there on time} | \ldots) &= 0.99 \\
P(S_1 \text{ gets me there on time} | \ldots) &= 0.90
\end{aligned}
$$

If you snooze you lose?

Suppose I believe the following:

$$
\begin{aligned}
P(S_0 \text{ gets me there on time} | \ldots) &= 0.99 \\
P(S_1 \text{ gets me there on time} | \ldots) &= 0.90 \\
P(S_{10} \text{ gets me there on time} | \ldots) &= 0.6
\end{aligned}
$$

If you snooze you lose?

Suppose I believe the following:

$$
\begin{aligned}
P(S_0 \text{ gets me there on time} | \ldots) &= 0.99 \\
P(S_1 \text{ gets me there on time} | \ldots) &= 0.90 \\
P(S_{10} \text{ gets me there on time} | \ldots) &= 0.6 \\
P(S_{25} \text{ gets me there on time} | \ldots) &= 0.1
\end{aligned}
$$

If you snooze you lose?

Suppose I believe the following:

$$
\begin{aligned}
P(S_0 \text{ gets me there on time} | \ldots) &= 0.99 \\
P(S_1 \text{ gets me there on time} | \ldots) &= 0.90 \\
P(S_{10} \text{ gets me there on time} | \ldots) &= 0.6 \\
P(S_{25} \text{ gets me there on time} | \ldots) &= 0.1
\end{aligned}
$$

Which action should I choose?

If you snooze you lose?

Suppose I believe the following:

$$
\begin{aligned}
P(S_0 \text{ gets me there on time} | \ldots) &= 0.99 \\
P(S_1 \text{ gets me there on time} | \ldots) &= 0.90 \\
P(S_{10} \text{ gets me there on time} | \ldots) &= 0.6 \\
P(S_{25} \text{ gets me there on time} | \ldots) &= 0.1
\end{aligned}
$$

Which action should I choose?

IT DEPENDS

If you snooze you lose?

Suppose I believe the following:

$$
\begin{aligned}
P(S_0 \text{ gets me there on time} | \ldots) &= 0.99 \\
P(S_1 \text{ gets me there on time} | \ldots) &= 0.90 \\
P(S_{10} \text{ gets me there on time} | \ldots) &= 0.6 \\
P(S_{25} \text{ gets me there on time} | \ldots) &= 0.1
\end{aligned}
$$

Which action should I choose?

IT DEPENDS on my preferences

If you snooze you lose?

Suppose I believe the following:

$$
\begin{aligned}
P(S_0 \text{ gets me there on time} | \ldots) &= 0.99 \\
P(S_1 \text{ gets me there on time} | \ldots) &= 0.90 \\
P(S_{10} \text{ gets me there on time} | \ldots) &= 0.6 \\
P(S_{25} \text{ gets me there on time} | \ldots) &= 0.1
\end{aligned}
$$

Which action should I choose?

IT DEPENDS on my preferences

e.g., missing class vs. sleeping

If you snooze you lose?

Suppose I believe the following:

$$
\begin{aligned}
P(S_0 \text{ gets me there on time} | \ldots) &= 0.99 \\
P(S_1 \text{ gets me there on time} | \ldots) &= 0.90 \\
P(S_{10} \text{ gets me there on time} | \ldots) &= 0.6 \\
P(S_{25} \text{ gets me there on time} | \ldots) &= 0.1
\end{aligned}
$$

Which action should I choose?

IT DEPENDS on my preferences

e.g., missing class vs. sleeping

S_0: ages in the Huxley building, therefore feeling miserable.

Chances and Utility

Utility theory is used to represent and infer preferences

Chances and Utility

Utility theory is used to represent and infer preferences

Decision theory = utility theory + probability theory

Probability basics

Begin with a set Ω—the sample space

Probability basics

Begin with a set Ω—the sample space
e.g., 6 possible rolls of a dice.

Probability basics

Begin with a set Ω—the sample space
e.g., 6 possible rolls of a dice.

$w \in \Omega$ is a sample point/possible world/atomic event

Probability basics

A probability space or probability model is a sample space Ω with an assignment $P(w)$ for every $w \in \Omega$ s.t.

Probability basics

A probability space or probability model is a sample space Ω with an assignment $P(w)$ for every $w \in \Omega$ s.t.

$0 \leq P(w) \leq 1$

Probability basics

A probability space or probability model is a sample space Ω with an assignment $P(w)$ for every $w \in \Omega$ s.t.

$0 \leq P(w) \leq 1$

$\sum_w P(w) = 1$

Probability basics

A probability space or probability model is a sample space Ω with an assignment $P(w)$ for every $w \in \Omega$ s.t.

$0 \leq P(w) \leq 1$

$\Sigma_w P(w) = 1$

e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

Events

An event A is any subset of Ω

Events

An event A is any subset of Ω

$$P(A) = \sum_{\{w \in A\}} P(w)$$

Events

An event A is any subset of Ω

$$P(A) = \sum_{\{w \in A\}} P(w)$$

E.g.,
$P(\text{dice roll} < 4) = P(1) + P(2) + P(3) = 1/6 + 1/6 + 1/6 = 1/2$

Random variables

A random variable is a function from sample points to some range, e.g., $\mathbb{R}$, $[0, 1]$, $\{true, false\}$ ...

Random variables

A random variable is a function from sample points to some range, e.g., $\mathbb{R}$, $[0,1]$,$\{true, false\}$ ...

e.g., $Odd(1) = true$.

Random variables

A random variable is a function from sample points to some range, e.g., $\mathbb{R}$, $[0,1]$,$\{true, false\}$ ...

e.g., $Odd(1) = true$.

P induces a probability distribution for any random variable X:

$$P(X = x_i) = \sum_{\{w:X(w)=x_i\}} P(w)$$

Random variables

A random variable is a function from sample points to some range, e.g., $\mathbb{R}$, $[0,1]$, $\{true, false\}$ ...

e.g., $Odd(1) = true$.

P induces a probability distribution for any random variable X:

$$P(X = x_i) = \sum_{\{w:X(w)=x_i\}} P(w)$$

e.g.,
$P(Odd = true) = P(1) + P(3) + P(5) = 1/6 + 1/6 + 1/6 = 1/2$

Propositions

A proposition can be seen as an event (set of sample points) where the proposition is true

Propositions

A proposition can be seen as an event (set of sample points) where the proposition is true

Given Boolean random variables A and B:

Propositions

A proposition can be seen as an event (set of sample points) where the proposition is true

Given Boolean random variables A and B:

event a = set of sample points where $A(w) = true$

Propositions

A proposition can be seen as an event (set of sample points) where the proposition is true

Given Boolean random variables A and B:

event a = set of sample points where $A(w) = true$

event $\neg a$ = set of sample points where $A(w) = false$

Propositions

A proposition can be seen as an event (set of sample points) where the proposition is true

Given Boolean random variables A and B:

event a = set of sample points where $A(w) = true$

event $\neg a$ = set of sample points where $A(w) = false$

event $a \wedge b$ = points where $A(w) = true$ and $B(w) = true$

Events and Propositional Logic

Proposition = disjunction of atomic events in which it is true

Events and Propositional Logic

Proposition = disjunction of atomic events in which it is true

e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$

Events and Propositional Logic

Proposition = disjunction of atomic events in which it is true

e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$

$\Rightarrow P(a \vee b)$

Events and Propositional Logic

Proposition = disjunction of atomic events in which it is true

e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$

$\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Events and Propositional Logic

Proposition = disjunction of atomic events in which it is true

e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$

$$\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$$

$$= P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b) + P(a \wedge b) - P(a \wedge b)$$

Events and Propositional Logic

Proposition = disjunction of atomic events in which it is true

e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$

$\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

$= P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b) + P(a \wedge b) - P(a \wedge b)$

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

Probabilities are logical

Theorem (De Finetti 1931)

An agent who bets according to "illogical" probabilities can be tricked into a bet that loses money regardless of outcome.

Syntax for propositions

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

Syntax for propositions

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

Discrete e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow\rangle$.
$Weather = rain$ is a proposition.

Syntax for propositions

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

Discrete e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow \rangle$.
$Weather = rain$ is a proposition.
Important: exhaustive and mutually exclusive

Syntax for propositions

Propositional	e.g., $Cavity$ (do I have a cavity?) $Cavity = true$ is a proposition, also written $cavity$
Discrete	e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow \rangle$. $Weather = rain$ is a proposition. **Important**: exhaustive and mutually exclusive
Continuous	e.g., $Temp = 21.6$; $Temp < 22.0$.

Probabilities

- Unconditional probabilities
- Conditional probabilities

Prior probability

Prior/unconditional probabilities of propositions:

Prior probability

Prior/unconditional probabilities of propositions: e.g.,
$P(Cavity = true) = 0.1$ and
$P(Weather = sunny) = 0.72$, correspond to belief
prior to arrival of any (new) evidence

Prior probability

Prior/unconditional probabilities of propositions: e.g.,
$P(Cavity = true) = 0.1$ and
$P(Weather = sunny) = 0.72$, correspond to belief prior to arrival of any (new) evidence

Probability distribution gives values for all possible assignments:
$\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (normalized, i.e., sums to 1)

Prior probability cont.

Joint probability distribution probability of every sample point

Prior probability cont.

Joint probability distribution probability of every sample point

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

Prior probability cont.

Joint probability distribution probability of 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

Prior probability cont.

Joint probability distribution probability of 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

Conditional probability

Conditional or posterior probabilities

e.g., $P(cavity|toothache) = 0.8$

Conditional probability

Conditional or posterior probabilities

e.g., $P(cavity|toothache) = 0.8$ i.e., **given that** *toothache* **is all I know NOT** "if *toothache* then 80% chance of *cavity*"

Conditional probability

Conditional or posterior probabilities

e.g., $P(cavity|toothache) = 0.8$ i.e., **given that** *toothache* **is all I know NOT** "if *toothache* then 80% chance of *cavity*"

(Notation for conditional distributions: $\mathbf{P}(Cavity|Toothache) =$ 2-element vector of 2-element vectors)

Conditional probability

If we know more, e.g., *cavity* is also given, then we have
$P(cavity|toothache, cavity) = \ldots$

Conditional probability

If we know more, e.g., *cavity* is also given, then we have $P(cavity|toothache, cavity) = ... = 1$

Conditional probability

If we know more, e.g., *cavity* is also given, then we have $P(cavity|toothache, cavity) = ... = 1$

- Note: the less specific belief **remains valid** after more evidence arrives, but is not always **useful**

Conditional probability

If we know more, e.g., *cavity* is also given, then we have
$P(cavity|toothache, cavity) = ... = 1$

- Note: the less specific belief **remains valid** after more evidence arrives, but is not always **useful**
- New evidence may be irrelevant, allowing simplification

Conditional probability

If we know more, e.g., $cavity$ is also given, then we have
$P(cavity|toothache, cavity) = ... = 1$

- 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., $P(cavity|toothache) =$
  $P(cavity|toothache, \text{Cristiano Ronaldo scores}) = 0.8$

Conditional probability

If we know more, e.g., *cavity* is also given, then we have
$P(cavity|toothache, cavity) = ... = 1$

- 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.,
  $P(cavity|toothache) =$
  $P(cavity|toothache, \text{Cristiano Ronaldo scores}) = 0.8$
  This kind of inference is crucial!

Conditional probability

Definition of conditional probability:

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

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:

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

Conditional probability

A general version holds for whole distributions, e.g.,
$\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity)\mathbf{P}(Cavity)$

Conditional probability

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)

Conditional probability

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)

Chain rule is derived by successive application of product rule:

Conditional probability

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)

Chain rule is derived by successive application of product rule:

$$\mathbf{P}(X_1, \ldots, X_n) = \mathbf{P}(X_1, \ldots, X_{n-1})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1})$$

Conditional probability

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)

Chain rule is derived by successive application of product rule:

$$\mathbf{P}(X_1, \ldots, X_n) = \mathbf{P}(X_1, \ldots, X_{n-1})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1})$$
$$= \mathbf{P}(X_1, \ldots, X_{n-2})\ \mathbf{P}(X_{n-1}|X_1, \ldots, X_{n-2})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1})$$

Conditional probability

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)

Chain rule is derived by successive application of product rule:

$$\mathbf{P}(X_1, \ldots, X_n) = \mathbf{P}(X_1, \ldots, X_{n-1})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1})$$
$$= \mathbf{P}(X_1, \ldots, X_{n-2})\ \mathbf{P}(X_{n-1}|X_1, \ldots, X_{n-2})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1})$$
$$= \ldots$$
$$= \prod_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1})$$

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:

$$P(\varphi) = \sum_{w:w\models\varphi} P(w)$$

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:

$$P(\varphi) = \sum_{w:w\models\varphi} P(w)$$

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

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:

$$P(\varphi) = \sum_{w:w\models\varphi} P(w)$$

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

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:

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

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:

$$P(cavity|toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.108 + 0.12}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

Normalization

Start with the joint distribution:

	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

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

Inference by enumeration, contd.

Let $\mathbf{X}$ be all the variables.

Inference by enumeration, contd.

Let $\mathbf{X}$ be all the variables.

- Typically, we want the posterior joint distribution of the query variables $\mathbf{Y}$ given specific values $\mathbf{e}$ for the evidence variables $\mathbf{E}$

Inference by enumeration, contd.

Let $\mathbf{X}$ be all the variables.

- Typically, we want the posterior joint distribution of the query variables $\mathbf{Y}$ given specific values $\mathbf{e}$ for the evidence variables $\mathbf{E}$
- Let the hidden variables be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

Inference by enumeration, contd.

Let $\mathbf{X}$ be all the variables.

- Typically, we want the posterior joint distribution of the query variables $\mathbf{Y}$ given specific values $\mathbf{e}$ for the evidence variables $\mathbf{E}$
- Let the hidden variables be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$
  Then the required summation of joint entries is done by summing out the hidden variables:

$$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e}) = \alpha\sum_{\mathbf{h}}\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e},\mathbf{H}=\mathbf{h})$$

Inference by enumeration, contd.

Let $\mathbf{X}$ be all the variables.

- Typically, we want the posterior joint distribution of the query variables $\mathbf{Y}$ given specific values $\mathbf{e}$ for the evidence variables $\mathbf{E}$
- Let the hidden variables be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$
  Then the required summation of joint entries is done by summing out the hidden variables:
  $$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e}) = \alpha\Sigma_{\mathbf{h}}\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e},\mathbf{H}=\mathbf{h})$$
  The terms in the summation are joint entries because $\mathbf{Y}$, $\mathbf{E}$, and $\mathbf{H}$ together exhaust the set of random variables.

Inference by enumeration, contd.

- Obvious problems: with n variables...

Inference by enumeration, contd.

- Obvious problems: with *n* variables...
  1. Worst-case time complexity $O(d^n)$ where d is the largest arity

Inference by enumeration, contd.

- Obvious problems: with *n* variables...
  1. Worst-case time complexity $O(d^n)$ where d is the largest arity
  2. Space complexity $O(d^n)$ to store the joint distribution

Inference by enumeration, contd.

- Obvious problems: with n variables...
  1. Worst-case time complexity $O(d^n)$ where d is the largest arity
  2. Space complexity $O(d^n)$ to store the joint distribution
  3. How to find the numbers for $O(d^n)$ entries?

Inference by enumeration, contd.

- Obvious problems: with n variables...
  1. Worst-case time complexity $O(d^n)$ where d is the largest arity
  2. Space complexity $O(d^n)$ to store the joint distribution
  3. How to find the numbers for $O(d^n)$ entries?

Summary

Summary

- Probability is a rigorous formalism for uncertain knowledge

Summary

- Probability is a rigorous formalism for uncertain knowledge
- Joint probability distribution specifies probability of every atomic event

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

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 size

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 size
- Independence and conditional independence provide the tools.

What's next?

- Bayes' rule
- Conditional and unconditional independence
- (hopefully) Bayesian Networks

Appendix:Independence

A and B are independent iff

Appendix:Independence

A and B are independent iff

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

Appendix:Independence

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

$P(cavity|\text{Cristiano Ronaldo scores}) = P(cavity)$

Appendix:Independence

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

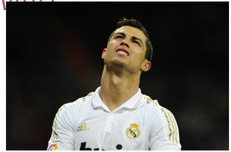
$P(cavity|\text{Cristiano Ronaldo scores}) = P(cavity)$

$P(\text{Cristiano Ronaldo scores}|cavity) =$
$P(\text{Cristiano Ronaldo scores}|\neg cavity) =$
$P(\text{Cristiano Ronaldo scores})$

Appendix:Independence

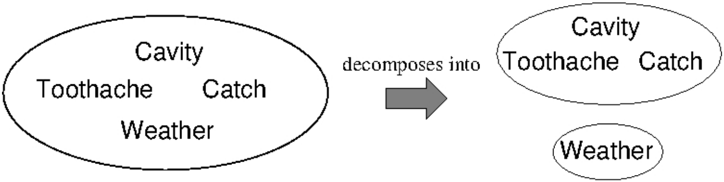
A and B are independent iff
$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$

$P(cavity|\text{Cristiano Ronaldo scores}) = P(cavity)$

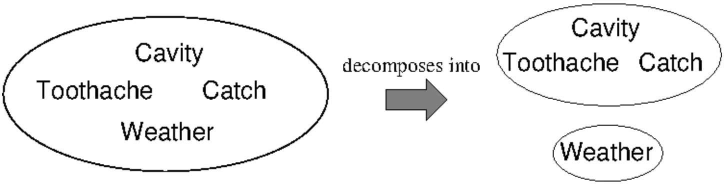


$P(\text{Cristiano Ronaldo scores}|cavity) =$
$P(\text{Cristiano Ronaldo scores}|\neg cavity) =$
$P(\text{Cristiano Ronaldo scores})$

Appendix: Independence

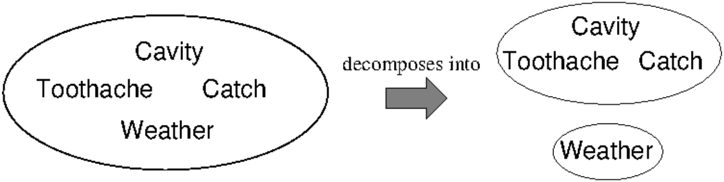


Appendix: Independence



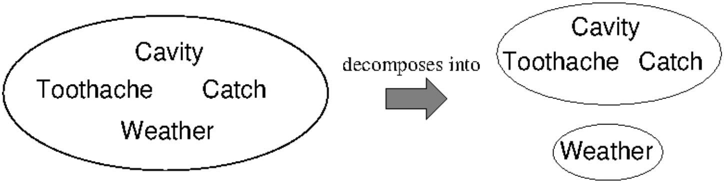
$\mathbf{P}(Toothache, Catch, Cavity, Weather)$
$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$

Appendix: Independence



$\mathbf{P}(Toothache, Catch, Cavity, Weather)$
$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$
32 entries reduced to 12; for n independent biased coins, $2^n \rightarrow n$

Appendix: Independence



$\mathbf{P}(Toothache, Catch, Cavity, Weather)$
$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$
32 entries reduced to 12; for n independent biased coins, $2^n \rightarrow n$

Absolute independence powerful but rare