# Uncertainty

## Chapter 13

# Outline

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

# Making decisions under uncertainty

Problem: choosing between actions with uncertain outcomes

Assumption: we have utility of each outcome + probability of each outcome

Rational choice: choose action that maximizes expected utility

Needed:

a) method for obtaining probabilities $\Rightarrow$ probability theory
b) method for inferring utility values $\Rightarrow$ utility theory

Maximum expected utility principle is part of decision theory:

Decision theory = utility theory + probability theory

We will look at probability theory first.

# Probability theory: Syntax

Basic element: random variable

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

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

Continuous random variables (bounded or unbounded)
e.g., $Temp = 21.6$; also allow, e.g., $Temp < 22.0$.

Elementary proposition formed by assigning value to random variable.

Complex propositions can be formed with standard logical connectives.

# Atomic event

Atomic event: A complete specification of the state of the world about which we are uncertain.

E.g. if the world consists of only two Boolean variables *Cavity* and *Toothache*, then there are four distinct atomic evens

Atomic events are mutually exclusive and exhaustive.

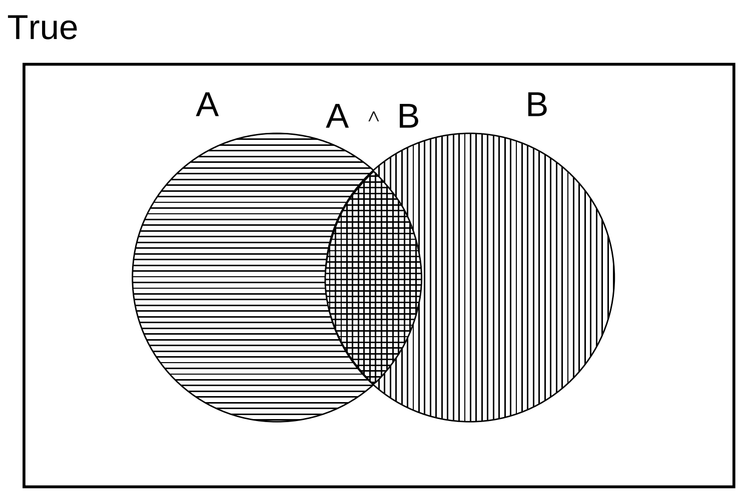
# Axioms of probability

For any propositions *A*, *B*:

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

$P(true) = 1$ and $P(false) = 0$

$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



# Prior probability

Prior or 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 $1$)

Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s
$\mathbf{P}(Weather, Cavity) =$ a $4 \times 2$ matrix of values:

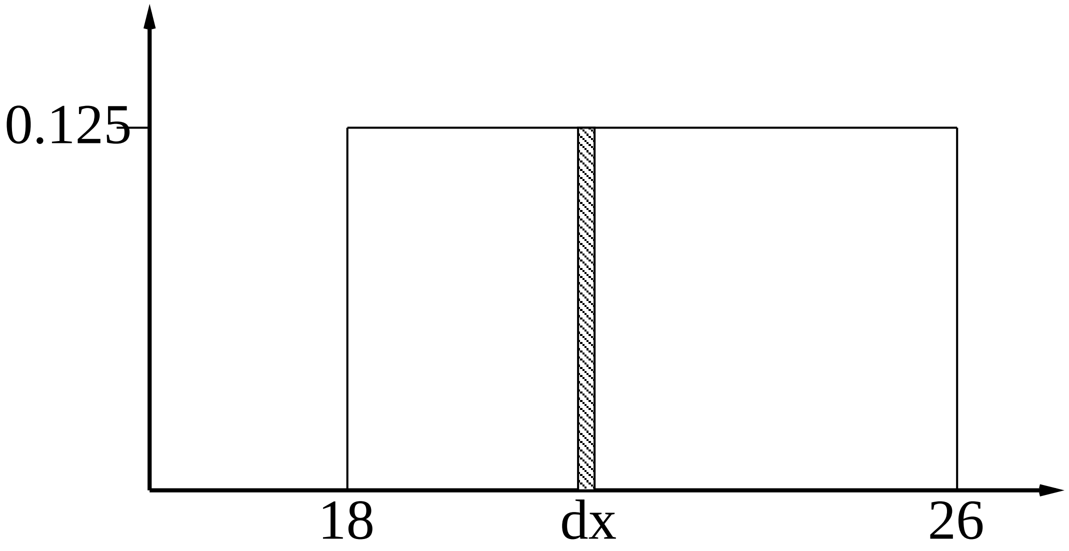
| $Weather =$ | $sunny$ | $rain$ | $cloudy$ | $snow$ |
|---|---|---|---|---|
| $Cavity = true$ | $0.144$ | $0.02$ | $0.016$ | $0.02$ |
| $Cavity = false$ | $0.576$ | $0.08$ | $0.064$ | $0.08$ |

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

# Probability for continuous variables

Express distribution as a parameterized function of value:

$P(X = x) = U[18, 26](x)$ = uniform density between $18$ and $26$



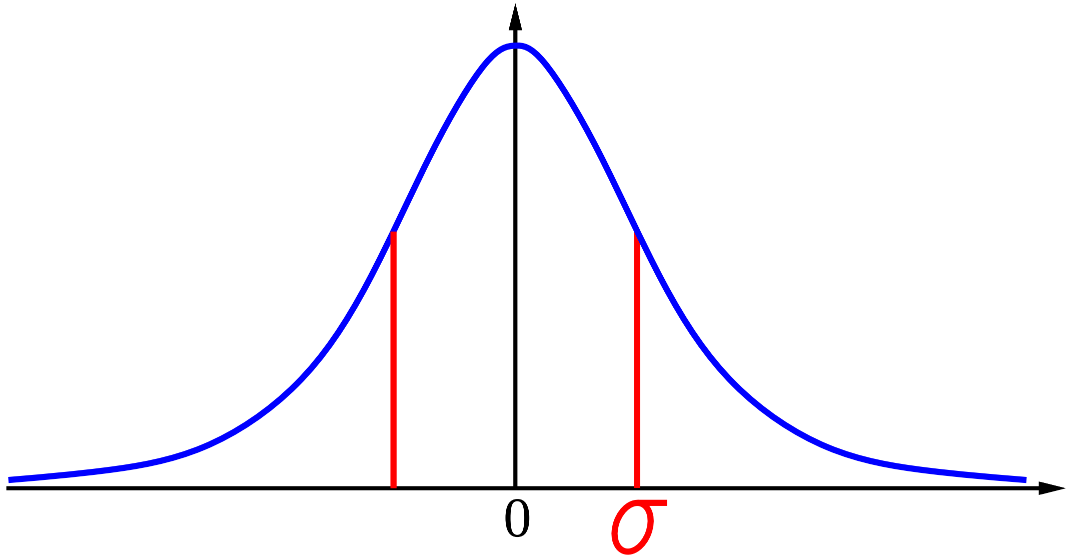

Here $P$ is a density; integrates to 1.

$P(X = 20.5) = 0.125$ really means

$$\lim_{dx \to 0} P(20.5 \leq X \leq 20.5 + dx)/dx = 0.125$$

# Gaussian density

$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$



# 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)

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

New evidence may be irrelevant, allowing simplification, e.g.,
$P(cavity|toothache, CrusadersWin) = P(cavity|toothache) = 0.8$

# 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 $4 \times 2$ set of equations, **not** matrix mult.)

Chain rule is derived by successive application of product rule:

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

# 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 $\phi$, 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 $\phi$, 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* | | ¬ *toothache* | |
|---|---|---|---|---|
| | *catch* | ¬ *catch* | *catch* | ¬ *catch* |
| *cavity* | **.108** | **.012** | **.072** | **.008** |
| ¬ *cavity* | **.016** | **.064** | **.144** | **.576** |

For any proposition $\phi$, sum the atomic events where it is true:

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

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

# Inference by enumeration

Start with the joint distribution:

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

Can also compute conditional probabilities:

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

# Normalization and summing out

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

Denominator can be viewed as a normalization constant.

In general, inference is done by summing out and then normalizing:

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

General idea: compute distribution on query variable
by fixing evidence variables and summing over hidden 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 $\mathbf{E}$

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:

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

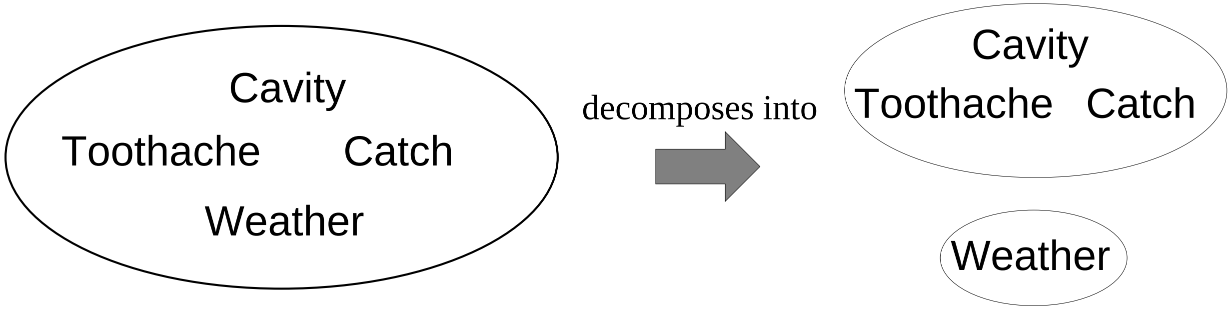
The terms in the summation are joint entries because $\mathbf{Y}$, $\mathbf{E}$, and $\mathbf{H}$ together exhaust the set of random variables

Obvious problems:
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???

# Independence

$A$ and $B$ are independent iff

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




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

32 entries reduced to 12; for $n$ independent biased coins, $2^n \rightarrow n$

Absolute independence powerful but rare

Dentistry is a large field with hundreds of variables,
none of which are independent. What to do?

# 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:

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

I.e., $2 + 2 + 1 = 5$ independent numbers (equations 1 and 2 remove 2)

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 our most basic and robust form of knowledge about uncertain environments.**

# Bayes' Rule

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

$\Rightarrow$ Bayes' rule $P(a|b) = \dfrac{P(b|a)P(a)}{P(b)}$

or in distribution form

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

Useful for assessing diagnostic probability from causal probability:

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

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

$$\begin{aligned}&\mathbf{P}(Cavity|toothache \land catch)\\ &= \alpha\,\mathbf{P}(toothache \land catch|Cavity)\mathbf{P}(Cavity)\\ &= \alpha\,\mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)\end{aligned}$$

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