Uncertainty

AIMA2e Chapter 13
Outline

♦ Uncertainty
♦ Probability
♦ Syntax and Semantics
♦ Inference
♦ Independence and Bayes’ Rule
FOL and Uncertainty

One problem with FOL is that it can not handle uncertainty:

Assume two propositions:

\(A=\text{Leaving for the airport 2hrs before your flight}\)
\(B=\text{Catching the flight}\)

Then, you can reasonably say:

\[A \Rightarrow B\]

But this is not always True.
Uncertainty

Uncertainty is due to either:

◊ Laziness: too much work to list all the antecedents to ensure an exceptionless rule:

\[ A \land \neg Rain \land \neg Traffic \ldots \Rightarrow B \]

◊ Ignorance: We may not have a complete theory for the domain (e.g. medical)

Even if we listed a long precedent list, we may not be able to apply such a rule because we may have only partial information
Uncertainty in FOL

Let action $A_t = \text{leave for airport } t \text{ minutes before flight}$

A purely logical approach either
1) risks falsehood: “$A_{30}$ will get me there on time” or
2) leads to conclusions that are too weak for decision making:
   “$A_{30}$ 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.”
3) be overly pessimistic:
   $A_{1440}$ will get me there on time but I’d have to stay overnight in the airport . . . )
Methods for handling uncertainty

Making assumptions:
  Assume my car does not have a flat tire
  Assume there is no outrageous traffic...
Issues: What assumptions are reasonable? How to handle contradiction?

Rules with fudge factors:
  $A_{25} \leftrightarrow_{0.3} \text{get there on time}$
  $\text{Sprinkler} \leftrightarrow_{0.99} \text{WetGrass}$
  $\text{WetGrass} \leftrightarrow_{0.7} \text{Rain}$
Issues: Problems with combination, e.g., \textit{Sprinkler} causes \textit{Rain}??

Fuzzy logic:
  Handles \textit{degree of truth} NOT uncertainty e.g.,
  $\text{WetGrass}$ is true to degree 0.2

\textbf{Probability}(Mahaviracarya(9th C.), Cardamo(1565) - theory of gambling):
  Given the available evidence,
  $A_{25}$ will get me there on time with probability 0.04
Probability

- Probabilities relate propositions to one’s own state of knowledge, by assigning a numerical degree of belief between 0 and 1, to each event.
  \[ P(A_{25} \text{ gets me there on time}|\text{no reported accidents}) = 0.06 \]

- Probabilities of propositions change with new evidence:
  \[ P(A_{25} \text{ gets me there on time}|\text{no reported accidents, 5 a.m.}) = 0.15 \]

Probabilistic assertions \textit{summarize} effects of
  - laziness: failure to enumerate exceptions, qualifications, etc.
  - ignorance: lack of relevant facts, initial conditions, etc.
Probability

Degree of belief is different than degree of truth:

♦ Fuzzy logic handles degree of truth
  e.g. WetGrass is true to degree 0.2

♦ Probability deals with degree of belief,
  e.g. \( P(\text{rain}) = 0.8 \) reflects our belief that there is an 80% chance of rain.
Making decisions under uncertainty

Suppose then that I believe the following:

\[
\begin{align*}
P(A_{25} \text{ gets me there on time} | \ldots) &= 0.04 \\
P(A_{90} \text{ gets me there on time} | \ldots) &= 0.70 \\
P(A_{120} \text{ gets me there on time} | \ldots) &= 0.95 \\
P(A_{1440} \text{ gets me there on time} | \ldots) &= 0.9999
\end{align*}
\]

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 = utility theory + probability theory
Probability basics

◊ Begin with a set \( \Omega \)—the *sample space*
  e.g., 6 possible rolls of a die.
  \( \omega \in \Omega \) is a sample point/possible world/atomic event

◊ A *probability space* or *probability model* is a sample space
  with an assignment \( P(\omega) \) for every \( \omega \in \Omega \) s.t.
  \[
  0 \leq P(\omega) \leq 1 \\
  \sum_\omega P(\omega) = 1
  \]
  e.g., \( P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6} \).

◊ An *event* \( A \) is any subset of \( \Omega \)
  \[
  P(A) = \sum_{\{\omega \in A\}} P(\omega)
  \]
  E.g., \( P(\text{die roll} < 4) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \)
Random variables and Probability

A *random variable* is a function that maps outcomes of some random experiment to some range, e.g., Real or Boolean values

◊ A random variable *DiceValue*, can be used to describe the outcome of rolling a fair die to the possible outcomes 1, 2, 3, 4, 5, 6.

◊ A random variable *OddDice*, can be used to describe the evenness-oddness of the rolled dice with possible outcomes True,False.

◊ $P$ induces a *probability distribution* for any random variable $X$:

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

where $\omega \in \Omega$ is the samples in the *sample space*

  e.g., $P(OddDice = true) = 1/6 + 1/6 + 1/6 = 1/2$
  i.e., the event is the random variable taking certain values.
Given Boolean random variables $A$ and $B$:

- event $a = \text{set of sample points } \omega \text{ where } A(\omega) = \text{true}$
- event $\neg a = \text{set of sample points } \omega \text{ where } A(\omega) = \text{false}$
- event $a \land b = \text{set of sample points } \omega \text{ where } A(\omega) = \text{true}$ and $B(\omega) = \text{true}$

◊ Often in AI applications, 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

e.g., $A = \text{true}$, $B = \text{false}$, or $a \land \neg b$.

◊ Proposition = disjunction of atomic events in which it is true

e.g., $(a \lor b) \equiv (\neg a \land b) \lor (a \land \neg b) \lor (a \land b)$

$\Rightarrow P(a \lor b) = P(\neg a \land b) + P(a \land \neg b) + P(a \land b)$
Why use probability?

The definitions imply that certain logically related events must have related probabilities

E.g., \( P(a \lor b) = P(a) + P(b) - P(a \land b) \)

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

◊ **Propositional** or **Boolean** random variables
e.g., *Cavity* (do I have a cavity?)

◊ **Discrete** random variables (**finite** or **infinite**)
e.g., *Weather* is one of \{sunny, rain, cloudy, snow\}
Values must be exhaustive and mutually exclusive

◊ **Continuous** random variables (**bounded** or **unbounded**)
e.g., *Temp* ∈ ℝ

◊ **Arbitrary Boolean combinations of basic propositions**
*Weather = rain* is a proposition
*Temp < 22.0* is a proposition
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:
\[
P(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle \quad \text{(normalized, i.e., sums to 1)}
\]

Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s (i.e., every sample point)
\[
P(Weather, Cavity) = \begin{bmatrix}
    \text{sunny} & \text{rain} & \text{cloudy} & \text{snow} \\
    \text{Cavity = true} & 0.144 & 0.02 & 0.016 & 0.02 \\
    \text{Cavity = false} & 0.576 & 0.08 & 0.064 & 0.08
\end{bmatrix}
\]

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:

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

Here $P$ is a *density*, integrates to 1.

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

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

\[ P(x) = \frac{1}{\sqrt{2\pi \sigma}} e^{-(x-\mu)^2/2\sigma^2} \]
Joint probability distribution for a set of variables gives values for each possible assignment to all the variables

<table>
<thead>
<tr>
<th></th>
<th>Toothache</th>
<th>¬Toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>¬Cavity</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Note that the entries sum up to 1 (mutually exclusive and exhaustive)
## Joint Distribution

<table>
<thead>
<tr>
<th></th>
<th>Toothache</th>
<th>¬Toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>¬Cavity</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Adding across rows or columns give unconditional probability of a variable:

\[
P(Cavity) = \]

\[
P(Cavity \lor Toothache) = \]

\[
P(Cavity \land Toothache) = \]
### Joint Distribution

<table>
<thead>
<tr>
<th></th>
<th>Toothache</th>
<th>¬Toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>¬Cavity</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Adding across rows or columns gives unconditional probability of a variable:

\[
P(Cavity) = 0.04 + 0.06 = 0.10
\]

\[
P(Cavity \lor Toothache) = 0.04 + 0.06 + 0.01 = 0.11
\]

Alternatively: \[1 - P(\neg Cavity \land \neg Toothache) = 1 - 0.89\]

\[
P(Cavity \land Toothache) = 0.04
\]
### Inference by enumeration

Start with the joint distribution:

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>(\neg) toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>0.108</td>
<td>0.012</td>
</tr>
<tr>
<td>(\neg) catch</td>
<td>0.072</td>
<td>0.008</td>
</tr>
<tr>
<td>cavity</td>
<td>0.016</td>
<td>0.064</td>
</tr>
<tr>
<td>(\neg) cavity</td>
<td>0.144</td>
<td>0.576</td>
</tr>
</tbody>
</table>

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

\[
P(\phi) = \sum_{\omega: \omega \models \phi} P(\omega)
\]
Inference by enumeration

Start with the joint distribution:

<table>
<thead>
<tr>
<th>cavity</th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>.108</td>
<td>.012</td>
</tr>
<tr>
<td>¬ catch</td>
<td>.016</td>
<td>.064</td>
</tr>
<tr>
<td>catch</td>
<td>.072</td>
<td>.008</td>
</tr>
<tr>
<td>¬ catch</td>
<td>.144</td>
<td>.576</td>
</tr>
</tbody>
</table>

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

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

$$P(\text{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$
Inference by enumeration

Start with the joint distribution:

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>0.108</td>
<td>0.012</td>
</tr>
<tr>
<td>¬ catch</td>
<td>0.072</td>
<td>0.008</td>
</tr>
<tr>
<td>cavity</td>
<td>0.016</td>
<td>0.064</td>
</tr>
<tr>
<td>¬ cavity</td>
<td>0.144</td>
<td>0.576</td>
</tr>
</tbody>
</table>

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

\[
P(\phi) = \sum_{\omega: \omega \models \phi} P(\omega)
\]

\[
P(\text{cavity} \lor \text{toothache}) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28
\]
Conditional probability

Definition of conditional (or posterior) probability:

\[ P(A|B) = \frac{P(A \land B)}{P(B)} \text{ if } P(B) \neq 0 \]

e.g., \( P(Cavity|Toothache) = 0.8 \)

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

\[ P(Weather, Cavity) = P(Weather|Cavity)P(Cavity) \]
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.,

\[ P(Cavity|Toothache, GSWon) = P(Cavity|Toothache) = 0.8 \]

This kind of inference, sanctioned by domain knowledge, is crucial
Bayes’ Rule

Product rule \( P(A \land B) = P(A|B)P(B) = P(B|A)P(A) \)

\[ \Rightarrow \text{Bayes’ rule} \quad P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

Why is this 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!
Chain Rule

Product rule $P(A \land B) = P(A|B)P(B) = P(B|A)P(A)$

$\Rightarrow$ Bayes’ rule $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

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

$P(Weather, Cavity) = P(Weather|Cavity)P(Cavity)$

(View as a $4 \times 2$ set of equations, not matrix mult.)

Chain rule is derived by successive application of product rule:

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

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

$= \ldots$

$= \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1})$
Back to our example:

<table>
<thead>
<tr>
<th></th>
<th>Toothache</th>
<th>¬Toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>¬Cavity</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

We can compute conditional probabilities from joint probabilities:

\[
P(Cavity|Toothache) = ?
\]
Inference by enumeration

<table>
<thead>
<tr>
<th></th>
<th>Toothache</th>
<th>¬Toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>¬Cavity</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

We can compute conditional probabilities from joint probabilities:

\[
P(Cavity|Toothache) = \frac{P(Cavity \land Toothache)}{P(Toothache)} = \frac{0.04}{0.04 + 0.01} = 0.8
\]
Inference by enumeration - 3 variables

Start with the joint distribution:

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>.108</td>
<td>.012</td>
</tr>
<tr>
<td>¬ catch</td>
<td>.016</td>
<td>.064</td>
</tr>
<tr>
<td>catch</td>
<td>.072</td>
<td>.008</td>
</tr>
<tr>
<td>¬ catch</td>
<td>.144</td>
<td>.576</td>
</tr>
</tbody>
</table>

\[
P(\neg \text{cavity}|\text{toothache}) = \frac{P(\neg \text{cavity} \land \text{toothache})}{P(\text{toothache})} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4
\]
Normalization

Suppose we wish to compute the posterior probability of Meningitis given StiffNeck.

\[ P(m|S) = P(s|m)P(m)/P(s) \]

If we are also entertaining the possibility that the patient may be suffering from a Whiplash, so we consider:

\[ P(w|S) = P(s|w)P(w)/P(s) \]

\[\checkmark\] Now we can use the relative likelihoods of Meningitis and Whiplash, \textit{without} computing \( P(s) \) (divide the above two equations two cancel out \( P(s) \)), to find which is the more likely cause.

\[\checkmark\] If we still want to be able to compute \( P(M|S) \), we need normalization...
Consider:
\[ P(m|S) = P(s|m)P(m)/P(s) \text{ and } P(\neg m|S) = P(s|\neg m)P(\neg m)/P(s) \]

Since the above two terms sum to 1 (there are only two possibilities given \( s \)), we obtain:
\[ P(M|s) + P(\neg m|s) = P(s|m)P(m)/P(s) + P(s|\neg m)P(\neg m) = 1 \]

From which we can find that:
\[ P(s|m)P(m) + P(s|\neg m)P(\neg m) = P(s) \]

So once we have:
\[ P(M|s) = \langle P(s|m)P(m)/P(s), P(s|\neg m)P(\neg m)/P(s) \rangle \]
\[ P(M|s) = \alpha \langle P(s|m)P(m), P(s|\neg m)P(\neg m) \rangle \]

we can sum the terms of the vector \( (P(s|m)P(m) \text{ and } P(s|\neg m)P(\neg m)) \) to find \( \alpha \) using the above equality and normalize the terms by that.
Normalized Example

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>catch</td>
<td>¬ catch</td>
</tr>
<tr>
<td>cavity</td>
<td>.108</td>
<td>.012</td>
</tr>
<tr>
<td>¬ cavity</td>
<td>.016</td>
<td>.064</td>
</tr>
</tbody>
</table>

Denominator can be viewed as a **normalization constant**
\[ \alpha = 1/P(\text{toothache}) \] for this case:

\[
P(Cavity | \text{toothache}) = \alpha P(Cavity, \text{toothache})
\]
\[
= \alpha (P(Cavity, \text{toothache}, \text{catch}) + P(Cavity, \text{toothache}, \neg \text{catch}))
\]
\[
= \alpha (\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle)
\]
\[
= \alpha \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle
\]
Note: Vector indicators ⟨ and ⟩ and proper Uppercase/lowercase usage, for random variable (Cavity) or given fixed evidence (toothache).

General idea: compute distribution on query variable by fixing evidence variables and summing over hidden variables.
Suppose we wish to compute $P(A|B = b)$ and suppose $A$ has possible values $a_1 \ldots a_m$

We can apply Bayes’ rule for each value of $A$:

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

$$\ldots$$

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

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

$$P(B = b) = \sum_i P(B = b|A = a_i)P(A = a_i)$$

This is the normalization factor $1/\alpha$:

$$P(A|B = b) = \alpha P(B = b|A)P(A)$$

Typically compute an unnormalized distribution, normalize at end

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

then $P(A|B = b) = \alpha \langle 0.4, 0.2, 0.2 \rangle = \frac{\langle 0.4, 0.2, 0.2 \rangle}{0.4 + 0.2 + 0.2} = \langle 0.5, 0.25, 0.25 \rangle$
Inference by enumeration, contd.

Typically, we are interested in
the posterior joint distribution of the query variables $Y$
given specific values $e$ for the evidence variables $E$

Let the hidden variables be $H = X - Y - E$

Then the required summation of joint entries is done by summing out the hidden variables:

$$P(Y|E=e) = \alpha P(Y, E=e) = \alpha \sum_h P(Y, E=e, H=h)$$

The terms in the summation are joint entries because $Y$, $E$, and $H$ 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

$P(A|B) = P(A)$ or $P(B|A) = P(B)$ or $P(A, B) = P(A)P(B)$

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

◊ Required parameters is greatly reduced: in this case, 32 entries reduced to 12. ($2^3 + 4$ for weather)

◊ For $n$ independent biased coins, required parameters: $2^n \rightarrow n$

◊ Absolute independence powerful but rare. What to do? (e.g. dentistry is a large field with hundreds of variables, none of which are independent.)
**Conditional independence**

\[ P(\text{Toothache}, \text{Cavity}, \text{Catch}) \] 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) \[ P(\text{catch} | \text{toothache}, \text{cavity}) = P(\text{catch} | \text{cavity}) \]

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

(2) \[ P(\text{catch} | \text{toothache}, \neg \text{cavity}) = P(\text{catch} | \neg \text{cavity}) \]

We say that **Catch is conditionally independent** of **Toothache** given **Cavity**: 

\[ P(\text{Catch} | \text{Toothache}, \text{Cavity}) = P(\text{Catch} | \text{Cavity}) \]

Equivalent statements (you may use one or the other as suitable):

\[ P(\text{Toothache} | \text{Catch}, \text{Cavity}) = P(\text{Toothache} | \text{Cavity}) \]

\[ P(\text{Toothache}, \text{Catch} | \text{Cavity}) = P(\text{Toothache} | \text{Cavity})P(\text{Catch} | \text{Cavity}) \]
Conditional independence contd.

Write out full joint distribution using chain rule:
\[ P(\text{Toothache}, \text{Catch}, \text{Cavity}) \]
\[ = P(\text{Toothache}|\text{Catch}, \text{Cavity})P(\text{Catch}, \text{Cavity}) \]
\[ = P(\text{Toothache}|\text{Catch}, \text{Cavity})P(\text{Catch}|\text{Cavity})P(\text{Cavity}) \]
\[ = P(\text{Toothache}|\text{Cavity})P(\text{Catch}|\text{Cavity})P(\text{Cavity}) \]

Note that the bold \( P \)s indicate full probability distributions, weather joint or conditional, and we multiply the appropriate terms.
- e.g. \( 2 \times 2 \) for the first and second and \( 2 \times 1 \) for the last
- \( \rightarrow 2 + 2 + 1 = 5 \) 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 and Conditional independence

Using the Bayes rule:

\[ P(Cavity|\text{toothache, catch}) = \alpha P(\text{toothache, catch, Cavity}) = \alpha P(\text{toothache, catch}|Cavity)P(Cavity) = \alpha P(\text{toothache}|Cavity)P(\text{catch}|Cavity)P(Cavity) \]

This is an example of the *naive Bayes* model which makes a simplifying assumption of conditional independence between different attributes/effects/symptoms given the same cause:

\[ P(Cause, Effect_1, \ldots, Effect_n) = P(Cause) \prod_i P(Effect_i|Cause) \]
Total number of parameters is *linear* in $n$.
The knowledgebase contains the following:

\[ P_{ij} = \text{true} \text{ iff } [i, j] \text{ contains a pit} \]

\[ B_{ij} = \text{true} \text{ iff } [i, j] \text{ is breezy} \]

\[ b = \neg b_{1,1} \land b_{1,2} \land b_{2,1} \text{ and } \text{known} = \neg p_{1,1} \land \neg p_{1,2} \land \neg p_{2,1} \]

Assume pits are placed randomly, probability 0.2 per square.
Specifying the probability model

The full joint distribution is involving only pits and \( b_{1,1}, \neg b_{1,2}, \neg b_{2,1} \) is

\[
P(P_{1,1}, \ldots, P_{4,4}, b_{1,1}, \neg b_{1,2}, \neg b_{2,1})
\]

To get to the form of \( P(Effect | Cause) \), apply the product rule:

\[
P(b_{1,1}, \neg b_{1,2}, \neg b_{2,1} | P_{1,1}, \ldots, P_{4,4}) P(P_{1,1}, \ldots, P_{4,4})
\]

First term: 1 if pits are adjacent to breezes, 0 otherwise
Second term: pits are placed randomly, probability 0.2 per square:

\[
P(P_{1,1}, \ldots, P_{4,4}) = \prod_{i,j=1,1}^{4,4} P(P_{i,j}) = 0.2^n \times 0.8^{16-n} \text{ for } n \text{ pits.}
\]
Observations and query

Now let's go to our actual problem:

◊ We know the following facts:
  \[ b = \neg b_{1,1} \land b_{1,2} \land b_{2,1} \]
  \[ known = \neg p_{1,1} \land \neg p_{1,2} \land \neg p_{2,1} \]

◊ Query is \( P(P_{1,3}|known, b) \)

◊ Define \textit{Unknown} = \( P_{ij} \)s other than \( P_{1,3} \) and \( Known \)

◊ Using enumeration, we have

\[
P(P_{1,3}|known, b) = \alpha \sum_{unknown} P(P_{1,3}, unknown, known, b)
\]

Note: The terms in this grows exponentially with the number of squares!
Using conditional independence

Basic insight: observations are conditionally independent of other hidden squares given neighbouring hidden squares

Define $Unknown = Fringe \cup Other$

$P(b|P_{1,3}, Known, Unknown) = P(b|P_{1,3}, Known, Fringe)$

Manipulate query into a form where we can use this!
Using conditional independence contd.

\[ P(P_{1,3}|\text{known}, b) = \alpha \sum_{\text{unknown}} P(P_{1,3}, \text{unknown}, \text{known}, b) \]
\[ = \alpha \sum_{\text{unknown}} P(b|P_{1,3}, \text{known}, \text{unknown})P(P_{1,3}, \text{known}, \text{unknown}) \]
\[ = \alpha \sum_{\text{fringe, other}} P(b|\text{known}, P_{1,3}, \text{fringe}, \text{other})P(P_{1,3}, \text{known}, \text{fringe, other}) \]
\[ = \alpha \sum_{\text{fringe}} P(b|\text{known}, P_{1,3}, \text{fringe}) \sum_{\text{other}} P(P_{1,3}, \text{known}, \text{fringe, other}) \]
\[ = \alpha \sum_{\text{fringe}} P(b|\text{known}, P_{1,3}, \text{fringe}) \sum_{\text{other}} P(P_{1,3})P(\text{known})P(\text{fringe})P(\text{other}) \]
\[ = \alpha P(\text{known})P(P_{1,3}) \sum_{\text{fringe}} P(b|\text{known}, P_{1,3}, \text{fringe})P(\text{fringe}) \sum_{\text{other}} P(\text{other}) \]
\[ = \alpha' P(P_{1,3}) \sum_{\text{fringe}} P(b|\text{known}, P_{1,3}, \text{fringe})P(\text{fringe}) \]
Using conditional independence contd.

\[ P(P_{1,3}|\text{known}, b) = \alpha' P(P_{1,3}) \sum_{\text{fringe}} P(b|\text{known}, P_{1,3}, \text{fringe})P(\text{fringe}) \]
\[ = \alpha' \langle 0.2, 0.8 \rangle \sum_{\text{compatible fringe}} P(\text{fringe}) \]
\[ = \alpha' \langle 0.2, 0.8 \rangle \langle (0.04 + 0.16 + 0.16), (0.04 + 0.16) \rangle \]
\[ = \alpha' \langle 0.2 \times 0.36, 0.8 \times 0.20 \rangle = \alpha' \langle 0.072, 0.16 \rangle \]
\[ \approx \langle 0.31, 0.69 \rangle \]

\[ P(P_{2,2}|\text{known}, b) \approx \langle 0.86, 0.14 \rangle \]

The agent infers that it is safer to go to \([1,3]\).
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.