CHAPTER 13

UNCERTAINTY
Outline

Uncertainty

Probability

Syntax and Semantics

Inference

Independence and Bayes' Rule

Inference

Probability

Uncertainty
Uncertainty

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

Will $A_t$ get me there on time?

Problems:
1) Partial observability (road state, other drivers' plans, etc.)
2) Noisy sensors (KCBS traffic reports)
3) Uncertainty in action outcomes (that tire, etc.)
4) Immense complexity of modeling and predicting traffic

Hence a purely logical approach either (1) risks falsehood: "$A_t$ will get me there on time"
or (2) leads to conclusions that are too weak for decision making:
"$A_t$ 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."

$A_t$ might reasonably be said to get me there on time ($A_t 1440$)

Chapter 13
Methods for handling uncertainty

Default or nonmonotonic logic:

Assume my car does not have a flat tire:

Given the available evidence, probability

\[ P(\text{AtAirportOnTime}) = 0.3 \]

\[ P(\text{WetGrass}) = 0.2 \]

Issues: Problems with combination, e.g., Sprinkler causes Rain?

\[ \text{WetGrass} \leftarrow 0.0, \text{Rain} \]

\[ \text{Sprinkler} \leftarrow 0.9, \text{WetGrass} \]

\[ \text{A25} \leftarrow 0.3, \text{AtAirportOnTime} \]

Fuzzy logic handles degree of truth. NOT uncertainty.

\[ \text{WetGrass} \text{ is true to degree } 0.2 \]

Mahaviracarya (9th C.), Cardano (1565) theory of gambling

A25 will get me there on time with probability 0.04

Rules with fudge factors:

Issues: What assumptions are reasonable? How to handle contradictions?

Assume A25 works unless contradicted by evidence

Assume my car does not have a flat tire

Mahaviracarya (9th C.), Cardano (1565) theory of gambling

Given the available evidence,
Probabilistic assertions summarize effects of laziness: failure to enumerate exceptions, qualifications, etc. ignorance: lack of relevant facts, initial conditions, etc.

Subjective or Bayesian probability: Probabilistic entailment status $KB \models a$, not truth.

*e.g.* $P(A \not\rightarrow \neg B \mid \neg a \rightarrow B, 5 \text{ a.m.}) = 0.15$

Probabilities of propositions change with new evidence:

*e.g.* $P(A \not\rightarrow \neg B \mid \neg a \rightarrow B, 5 \text{ a.m., 5 a.m.}) = 0.06$

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

Probabilistic assertions summarize effects of
Making decisions under uncertainty

Suppose I believe the following:

\[ P(A_{25} \text{ gets me there on time}) = 0.04 \]
\[ P(A_{90} \text{ gets me there on time}) = 0.70 \]
\[ P(A_{120} \text{ gets me there on time}) = 0.95 \]
\[ P(A_{1440} \text{ gets me there on time}) = 0.9999 \]

Which action to choose?

Which action to choose?

Decision theory = utility theory + probability theory

Utility theory is used to represent and infer preferences

Depends on my preferences for missing flight vs. airport cuisine, etc.

Suppose I believe the following:
Probability basics

Begin with a set $\Omega$—the sample space

An event $A$ is any subset of $\Omega$.

An event $A$ is any subset of $\Omega$.

A probability space or probability model is a sample space

with an assignment of $P$ for every $\omega \in \Omega$.

$P$ is a sample point/possible world/atomic event.

E.g., $6$ possible rolls of a die.

$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$

E.g., $P(\text{die roll} < 4) = P(1) + P(2) + P(3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$
A random variable is a function from sample points to some range, e.g., the reals or Booleans.

\[
\mathbb{P}(\frac{2}{1} = 6/1 + 6/1 + 6/1 = (5) + (3) + (1) = (\text{true} = \text{true} = \text{true}) \) \\
\mathbb{P}(\{x = \text{true} \}) = \mathbb{P}(\{x = X \}) \\
\mathbb{P}(\text{true} = \text{true}) = (\text{true} \) \\
\mathbb{P}(\text{true} = \text{true}) = 1/2
\]
Propositions

Think of a proposition as the event (set of sample points) where the proposition is true.

Given Boolean random variables $A$ and $B$, 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.

Often in AI applications, the sample points are

event $a \land B$ and event $a \land \neg B$ are

event $a$ where $A = \text{true}$ and event $a$ where $A = \text{false}$ are

event $a$ set of sample points where $A = \text{true}$ and $B$ and

event $a$ set of sample points where $A = \text{false}$

Given Boolean random variables $A$ and $B$, where the proposition is true.

\[
(q \lor a)D + (q\neg \lor a)D + (q \lor a
\neg)D = (q \land a)D
\]

\[
(q \lor a) \land (q\neg \lor a) \land (q \lor a
\neg) = (q \land a)
\]

Example:

- $A = \text{true}$, $B = \text{false}$
- $a \lor \neg q$
The definitions imply that certain logically related events must have related probabilities. Why use probability?
Syntax for propositions

Arbitrary Boolean combinations of basic propositions

e.g., \( \text{Temp} = 21.6; \) also allow, e.g., \( \text{Temp} > 22.0 \).

Continuous random variables (bounded or unbounded)

Values must be exhaustive and mutually exclusive

\( \text{Weather} = \text{rain} \) is a proposition

e.g., \( \text{Weather} \) is one of \( \langle \text{sunny, rain, cloudy, snow} \rangle \)

Discrete random variables (finite or infinite)

\( \text{Quantity} = \text{true} \) is a proposition, also written \( \text{Quantity} \)

e.g., \( \text{Quantity} \) (do I have a quantity?)

Propositional or Boolean random variables
Prior probability

or unconditional probabilities of propositions

\[ P(Cavity = \text{true}) = 0.08 \]
\[ P(Weather = \text{sunny}) = 0.72 \]

Prior to arrival of any (new) evidence:

\[ P(Cavity = \text{true}) = 0.1 \]
\[ P(Cavity = \text{false}) = 0.9 \]

\[ P(Weather = \text{sunny}) = 0.72 \]


correspond to belief prior to arrival of any (new) evidence.

\[ P(Cavity = \text{true}) = 0.1 \]
\[ P(Cavity = \text{false}) = 0.9 \]

Joint probabililty distribution gives values for all possible assignments:

\[ P(Weather, Cavity) = \begin{pmatrix} 0.72, 0.08 \ 0.02, 0.92 \end{pmatrix} \]

(normalized, i.e., sums to 1)
Express distribution as a parameterized function of value:

\[ P(X = x) = \frac{xp}{xp + 20.5} \quad \text{for} \quad 0 < x < 20.5 \]

Here, \( p \) is a density; integrates to 1.

\[ P(20 < X < 20.5) = 0.125 \]

\[ \lim_{dx \to 0} P(20 < X < 20.5 + dx) = dx = 0.125 \]

Express distribution as a parameterized function of value:

Probability for continuous variables
Gaussian density

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

\[ x \in \mathbb{R} \]

Chapter 13
This kind of inference, sanctioned by domain knowledge, is crucial.

\[ P(\text{Cavity} | \text{Toothache}) = 0.8 \]

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

New evidence may be irrelevant, allowing simplification, e.g.,

\[ P(\text{Cavity} | \text{Toothache}) = 1 \]

but is not always useful.

Note: the less specific belief remains valid after more evidence arrives.

If we know more, e.g., cavity is also given, then we have

\[ P(\text{Cavity} | \text{Toothache}) = 2 \text{-element vector of 2-element vectors) \]

(Notation for conditional distributions:

\[ \text{"If Toothache then 80\% chance of Cavity" NOT } \]

\[ \text{"I.e., given that Toothache is all I know} \]

\[ e.g., P(\text{Cavity} | \text{Toothache}) = 0.8 \]

Conditional or posterior probabilities.
Conditional probability

Definition of conditional probability:

\[ P(a \mid b) = \frac{P(a \cap b)}{P(b)} \]

if \( P(b) \neq 0 \)

Product rule gives an alternative formulation:

\[ P(a \cap b) = P(a \mid b) P(b) = P(b \mid a) P(a) \]

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

\[ P(\text{Weather} \mid \text{Cavity}) P(\text{Cavity}) = P(\text{Weather} \cap \text{Cavity}) \]

Chain rule is derived by successive application of product rule:

\[ (\mathbf{X}^1 \cdots \mathbf{X}^u \mid \mathbf{X}) \mathbf{d}^{\mathbf{1} \mid \mathbf{u}} \]

\[ = (\mathbf{X}^1 \cdots \mathbf{X}^u \mid \mathbf{X}) \mathbf{d} (\mathbf{X}^1 \cdots \mathbf{X}^u \mid \mathbf{X}) \mathbf{d} (\mathbf{X}^1 \cdots \mathbf{X}^u \mid \mathbf{X}) \mathbf{d} = (\mathbf{X}^1 \cdots \mathbf{X}^u \mid \mathbf{X}) \mathbf{d} \]

Product rule gives an alternative formulation:

\[ 0 \neq (q \mid p) \Leftrightarrow \frac{(q \mid p) \mathbf{d}}{(q \lor p) \mathbf{d}} = (q \mid p) \mathbf{d} \]

Definition of conditional probability:
Inference by enumeration

Start with the joint distribution:

<table>
<thead>
<tr>
<th>cavity</th>
<th>catch</th>
<th>toothache</th>
<th></th>
<th>cavity</th>
<th>catch</th>
<th>toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.576</td>
<td>0.144</td>
<td>0.064</td>
<td>0.16</td>
<td>0.072</td>
<td>0.012</td>
<td></td>
</tr>
</tbody>
</table>

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

\[
(\bigwedge_{m}^{m} p_{\phi} = (\phi) p
\]

Chapter 13
Inference by enumeration

Start with the joint distribution:

<table>
<thead>
<tr>
<th></th>
<th>cavity</th>
<th>catch</th>
<th>toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>cavity</td>
<td>0.0576</td>
<td>0.144</td>
<td>0.064</td>
</tr>
<tr>
<td>catch</td>
<td>0.008</td>
<td>0.072</td>
<td>0.108</td>
</tr>
<tr>
<td>toothache</td>
<td></td>
<td>0.12</td>
<td></td>
</tr>
</tbody>
</table>

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

\[
P(\text{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.24\]

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

Start with the joint distribution:

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

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

\[
\sum P(\phi) = P(\text{cavity} \land \text{toothache}) + P(\text{catch} \land \text{toothache})
\]

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

Chapter 13
Inference by enumeration

Start with the joint distribution:

\[
P(\neg \text{cavity}|\text{toothache}) = \frac{P(\text{toothache}) - P(\text{toothache} \land \neg \text{cavity})}{P(\text{toothache})}
\]

Can also compute conditional probabilities:

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

\[
\begin{array}{cccc}
\text{Cavity} & \text{Catch} & \text{Toothache} \\
0.576 & 1.44 & 0.064 \\
0.008 & 0.008 & 0.108 \\
\end{array}
\]

\[
\begin{array}{cccc}
\text{Cavity} & \text{Catch} & \text{Toothache} \\
0.116 & 0.012 & 0.016 \\
\end{array}
\]

\[
\begin{array}{ccc}
\text{Catch} & \text{Toothache} \\
0.012 & 0.016 \\
\end{array}
\]

\[
\begin{array}{c}
\text{Toothache} \\
0.016 \\
\end{array}
\]

Inference by enumeration
Denominator can be viewed as a normalization constant as

\[
\frac{\langle a \rangle}{\langle 0.12', 0.08 \rangle} = \frac{\langle 0.08 \rangle}{\langle 0.012', 0.064 \rangle} = \frac{\langle a \rangle}{\langle Cavity, Toothache, Catch \rangle} = \frac{\langle a \rangle}{\langle Cavity, Toothache \rangle}
\]

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

<table>
<thead>
<tr>
<th>Cavity</th>
<th></th>
<th></th>
<th>Toothache</th>
<th>Catch</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.76</td>
<td>0.144</td>
<td>0.064</td>
<td>0.016</td>
<td>0.012</td>
</tr>
<tr>
<td>0.08</td>
<td>0.72</td>
<td>0.12</td>
<td>0.18</td>
<td>0.12</td>
</tr>
</tbody>
</table>

**Normalization**
Inference by enumeration, contd.

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

Given specific values $e$ for the evidence variables $E$, the posterior joint distribution of the query variables $Y$ is:

$$
P(Y|E = e) = \sum_{H} P(Y; E = e, H)$$

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

$$
P(Y; E = e) = \sum_{H} P(Y, E, H; P, \lambda)$$

Obvious problems:

1) Worst-case time complexity $O(dn)$ where $d$ is the largest arity
2) Space complexity $O((up)O)$ to store the joint distribution
3) How to find the numbers for the joint entries

The terms in the summation are joint entries because $Y, E, H$ together exhaust the set of random variables.

Ultimately, we want $X$ be all the variables. Typically, we want the posterior joint distribution of the query variables $Y$ given specific values $e$ for the evidence variables $E$.
Independence

A and B are independent iff

\[ P(A \mid B) = P(A) \quad \text{or} \quad P(B \mid A) = P(B) \]

none of which are independent. What to do?

Dentistry is a large field with hundreds of variables,

Absolute independence powerful but rare.

32 entries reduced to 12: for n independent biased coins, 2^n → n

\[ P(Toothache, Catch, Cavity, Weather) = P(Toothache, Catch, Cavity, Weather) \]

Decomposes into

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

Chapter 13
Conditiona

\( p(\text{Toothache}, \text{Catch} | \text{Cavity}) = p(\text{Catch} | \text{Cavity}) \)

\( p(\text{Toothache} | \text{Catch}, \text{Cavity}) = p(\text{Toothache} | \text{Cavity}) \)

\( p(\text{Toothache}, \text{Catch} | \text{Cavity}) = p(\text{Toothache} | \text{Cavity}) \)

Equivalent statements:

Catch is conditionally independent of Toothache given Cavity:

\( p(\text{Catch} | \text{Toothache}, \text{Cavity}) = p(\text{Catch} | \text{Cavity}) \)

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

\( p(\text{Catch} | \text{Toothache}, \text{Cavity}) = p(\text{Catch} | \text{Cavity}) \)

\( p(\text{Catch} | \text{Toothache}, \text{Cavity}) = p(\text{Catch} | \text{Cavity}) \)

\( p(\text{Catch} | \text{Toothache}, \text{Cavity}) = p(\text{Catch} | \text{Cavity}) \)

Conditional independence

\( p(\text{Toothache}, \text{Catch} | \text{Cavity}) = p(\text{Catch} | \text{Cavity}) \)

\( p(\text{Toothache} | \text{Cavity}) \) has \( 2^3 - 1 = 7 \) independent entries.

\( p(\text{Toothache}, \text{Catch} | \text{Cavity}) \) has \( 2^3 - 1 = 7 \) independent entries.
Conditional independence is our most basic and robust form of knowledge about uncertain environments. 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 $n$.

\[
\begin{align*}
P(\text{Toothache}; \text{Catch}; \text{Cavity}) &= P(\text{Toothache}; \text{Catch}| \text{Cavity}) P(\text{Catch}| \text{Cavity}) P(\text{Cavity}) \\
&= P(\text{Toothache}| \text{Cavity}) P(\text{Catch}| \text{Cavity}) P(\text{Cavity}) \\
&= P(\text{Toothache}| \text{Catch}; \text{Cavity}) P(\text{Catch}| \text{Cavity}) P(\text{Cavity}) \\
&= P(\text{Toothache}| \text{Catch}; \text{Cavity}) P(\text{Catch}| \text{Cavity}) P(\text{Cavity})
\end{align*}
\]

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

Write out full joint distribution using chain rule.

---

Conditional independence contd.
Bayes' Rule

$$P(a^b) = P(a|b)P(b) = P(b|a)P(a)$$

or in distribution form

$$(X)d(X|X)d = (X)d = (X|X)d$$

Bayes' Rule

E.g., let meningitis, be stiff neck:

$$(s)d = (s|m)d = (s|m)d$$

Note: posterior probability of meningitis still very small!

$$8000 = \frac{0.1}{1000} \times 8.0 = \frac{(s)d}{(m)d} = (s|m)d$$

Useful for assessing diagnostic probability from causal probability:

$$(s)d = (s|m)d = (s|m)d$$

Bayes' Rule $\iff$

$$(s)d = (s|m)d = (s|m)d$$

Product Rule

$$(s)d = (s|m)d = (s|m)d$$

Bayes' Rule
Bayes' Rule and conditional independence

Total number of parameters is linear in $n$

This is an example of a naive Bayes model:

\[ p(Cause, Effect_1, \ldots, Effect_n) = p(Cause) \prod_{i=1}^{n} p(Effect_i | Cause) \]

\[ p(Cause) = p(toothache | Cavity) p(Cavity) \]

\[ p(toothache \land catch | Cavity) = p(toothache | catch \land Cavity) p(Cavity) \]
\begin{itemize}
  \item $P_{ij} = \text{true}$ iff $[i,j]$ contains a pit.
  \item $B_{ij} = \text{true}$ iff $[i,j]$ is breezy.
\end{itemize}

Include only $B_{1,1}, B_{1,2}, B_{2,1}$ in the probability model.

\begin{tabular}{|c|c|c|c|c|}
  \hline
  \textbf{i} & 1 & 2 & 3 & 4 \\
  \hline
  \textbf{j} & 1 & 2 & 3 & 4 \\
  \hline
  1 & 2 & 3 & 2 & B \textbf{OK} \\
  \hline
  2 & 3 & 2 & 3 & B \textbf{OK} \\
  \hline
  3 & 2 & 2 & 3 & \textbf{OK} \\
  \hline
  4 & 2 & 1 & 3 & \textbf{OK} \\
  \hline
\end{tabular}
Specifying the probability model

The full joint distribution is

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

Apply product rule:

\[
P(B_{1,1}, B_{1,2}, B_{2,1}, B_{2,2} | P_{1,1}, \ldots, P_{4,4}, P_{1,1}, \ldots, P_{4,4}, P_{1,1}) = P(B_{1,1}, B_{1,2} | P_{1,1}, \ldots, P_{4,4})P(B_{2,1}, B_{2,2} | P_{1,1}, \ldots, P_{4,4})
\]

First term: If pits are adjacent to breezes, 0 otherwise.

Second term: Pits are placed randomly, probability 0.2 per square.

Do it this way to get \(P(E|F, E, \text{cause})\).

For \(n\) pits:

\[
u \in [0, 0.816 - n] = \left(\sum_{i=1}^{n} P_i\right) \prod_{i=1}^{4} P_{i,1} P_{i,2} = \left(\sum_{i=1}^{4} P_{i,1} \cdot \cdot \cdot P_{i,1}\right) \prod_{i=1}^{4} P_{i,1} P_{i,2}
\]

Specifying the probability model
We know the following facts:

\[ p_{1,3}^{\text{known}, \text{unknown}} = q \]

For inference by enumeration, we have

\[ p_{1,3}^{\text{known}, \text{unknown}} = p_{1,3}^{\text{known}, \text{unknown}} = q \]

We know the following facts:
Using conditional independence

Basic insight: observations are conditionally independent given neighboring hidden squares.

Define \( \text{Unknown} = \text{Fringe} \cap \text{Other} \)

Manipulate query into a form where we can use this:

\[
\Pr(b_j \mid P_{1,3}; \text{Known}; \text{Unknown}) = \Pr(b_j \mid P_{1,3}; \text{Known}; \text{Fringe})
\]

Unknown = Fringe \Cap Other

Using conditional independence
Using conditional independence contd.

\[ P(P_1; P_3 | \text{known}; b) = \begin{cases} \text{known} & \text{known} \\ \text{other} & \text{other} \end{cases} \]
Using conditional independence contd.
Summary

Probability is a rigorous formalism for uncertain knowledge. Joint probability distributions specify probabilities of every atomic event.

Queries can be answered by summing over atomic events.

Independence and conditional independence provide the tools for nontrivial domains, we must find a way to reduce the joint size.

Probabilility is a rigorous formalism for uncertain knowledge.