Uncertainty

Chapter 13

Outline

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

Let action $A_t = \text{leave for airport, } t \text{ minutes before flight} \\
Will $A_t$ get me there on time?

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

Hence a purely logical approach either
1. risks falsehood: “$A_{25}$ will get me there on time”, or
2. leads to conclusions that are too weak for decision making:

“$A_{25}$ 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.”

($A_{1440}$ might reasonably be said to get me there on time but I’d have to stay overnight in the airport …)

But…

• A decision must be made!
• No intelligent system can afford to consider all eventualities, wait until all the data is in and complete, or try all possibilities to see what happens
Quick Overview of Reasoning Systems

- **Logic:** True or false, nothing in between. No uncertainty.
- **Non-monotonic logic:** True or false, but new information can change it.
- **Probability:** Degree of belief, but in the end it’s either true or false.
- **Fuzzy:** Degree of belief, allows overlapping of true and false states.

Examples

- **Logic:** All birds fly.
- **Non-monotonic**
  - Tweety flies, since he’s a bird and no evidence he doesn’t fly.
Probability

Probabilistic assertions summarize effects of
- **laziness**: failure to enumerate exceptions, qualifications, etc.
- **ignorance**: lack of relevant facts, initial conditions, etc.

**Subjective probability:**
- Probabilities relate propositions to agent’s own state of knowledge
  e.g., \( P(A_{25} \mid \text{no reported accidents}) = 0.06 \)

These are not assertions about the world

Probabilities of propositions change with new evidence:
  e.g., \( P(A_{25} \mid \text{no reported accidents, 5 a.m.}) = 0.15 \)

Making decisions under uncertainty

Suppose I believe the following:
- \( P(A_{25} \text{ gets me there on time } \mid \ldots) = 0.04 \)
- \( P(A_{90} \text{ gets me there on time } \mid \ldots) = 0.70 \)
- \( P(A_{120} \text{ gets me there on time } \mid \ldots) = 0.95 \)
- \( P(A_{1440} \text{ gets me there on time } \mid \ldots) = 0.9999 \)

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

- Basic element: random variable
- Similar to propositional logic: possible worlds defined by assignment of values to random variables.
  - **Boolean** random variables
    e.g., \textit{Cavity} (do I have a cavity?)
  - **Discrete** random variables
    e.g., \textit{Weather} is one of \{sunny, rainy, cloudy, snow\}
  - Domain values must be exhaustive and mutually exclusive
- Elementary proposition constructed by assignment of a value to a random variable: e.g., \textit{Weather} = sunny, \textit{Cavity} = false (abbreviated as \neg \textit{cavity})
- Complex propositions formed from elementary propositions and standard logical connectives e.g., \textit{Weather} = sunny \lor \textit{Cavity} = false

Atomic event: A complete specification of the state of the world about which the agent is uncertain

E.g., if the world consists of only two Boolean variables \textit{Cavity} and \textit{Toothache}, then there are 4 distinct atomic events:

- \textit{Cavity} = false \land \textit{Toothache} = false
- \textit{Cavity} = false \land \textit{Toothache} = true
- \textit{Cavity} = true \land \textit{Toothache} = false
- \textit{Cavity} = true \land \textit{Toothache} = true

- Atomic events are mutually exclusive and exhaustive
Axioms of probability

- For any propositions $A$, $B$
  - $0 \leq P(A) \leq 1$
  - $P(\text{true}) = 1$ and $P(\text{false}) = 0$
  - $P(A \lor B) = P(A) + P(B) - P(A \land B)$

Prior probability

- Prior or unconditional probabilities of propositions
  e.g., $P(\text{Cavity} = \text{true}) = 0.1$ and $P(\text{Weather} = \text{sunny}) = 0.72$ correspond to belief prior to arrival of any (new) evidence
- Probability distribution gives values for all possible assignments:
  $P(\text{Weather}) = <0.72,0.1,0.08,0.1>$ (normalized, i.e., sums to 1)
- Joint probability distribution for a set of random variables gives the probability of every atomic event on those random variables
  $P(\text{Weather, Cavity}) = a 4 \times 2$ matrix of values:

<table>
<thead>
<tr>
<th>Weather</th>
<th>sunny</th>
<th>rainy</th>
<th>cloudy</th>
<th>snow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity = true</td>
<td>0.144</td>
<td>0.02</td>
<td>0.016</td>
<td>0.02</td>
</tr>
<tr>
<td>Cavity = false</td>
<td>0.576</td>
<td>0.08</td>
<td>0.064</td>
<td>0.08</td>
</tr>
</tbody>
</table>
- Every question about a domain can be answered by the joint distribution
How could we estimate the full joint distribution?

Parameter estimates are provided by expert knowledge, statistics on data samples, or a combination of both.

Suppose you have 20 variables.

Expert knowledge:
\[ P(X_1=0, X_2=0, ..., X_{13}=1, ..., X_{20}=0) \text{ vs. } P(X_1=0, X_2=0, ..., X_{13}=0, ..., X_{20}=0) \]

Data Samples: practically speaking, we don’t typically have enough data

---

Conditional probability

- Conditional or posterior probabilities
  - e.g., \( P(\text{cavity} \mid \text{toothache}) = 0.8 \)
  - i.e., given that \( \text{toothache} \) is all I know

- (Notation for conditional distributions:
  \[ P(Cavity \mid Toothache) = \text{2-element vector of 2-element vectors} \]

- If we know more, e.g., \( \text{cavity} \) is also given, then we have
  \[ P(\text{cavity} \mid \text{toothache}, \text{cavity}) = 1 \]

- New evidence may be irrelevant, allowing simplification, e.g.,
  \[ P(\text{cavity} \mid \text{toothache}, \text{sunny}) = P(\text{cavity} \mid \text{toothache}) = 0.8 \]

- This kind of inference, sanctioned by domain knowledge, is crucial
More on Conditional Probabilities

- $P(\text{CarWontStart} \mid \text{NoGas})$
  - This predicts a symptom based on an underlying cause
  - These can be generated empirically (Drain N gastanks, see how many cars start) or using expert knowledge
- $P(\text{NoGas} \mid \text{CarWontStart})$
  - Diagnosis. We have a symptom and want to predict the cause. This is what the system wants to determine

Conditional probability

- Definition of conditional probability:
  \[ P(a \mid b) = \frac{P(a \land b)}{P(b)} \text{ if } P(b) > 0 \]
- Product rule gives an alternative formulation:
  \[ P(a \land b) = P(a \mid b) P(b) = P(b \mid a) P(a) \]
- A general version holds for whole distributions, e.g.,
  \[ P(\text{Weather}, \text{Cavity}) = P(\text{Weather} \mid \text{Cavity}) P(\text{Cavity}) \]
  - (View as a set of $4 \times 2$ equations, not matrix mult.)
- Chain rule is derived by successive application of product rule:
  \[
P(X_1, \ldots, X_n) = P(X_1, \ldots, X_{n-1}) P(X_n \mid X_1, \ldots, X_{n-1})
  = P(X_1, \ldots, X_{n-2}) P(X_{n-1} \mid X_1, \ldots, X_{n-2}) P(X_n \mid X_1, \ldots, X_{n-1})
  = \ldots
  = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})
\]
Inference by enumeration

• Start with the joint probability 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>0.016</td>
<td>0.064</td>
</tr>
<tr>
<td>cavity</td>
<td>0.072</td>
<td>0.008</td>
</tr>
<tr>
<td>¬cavity</td>
<td>0.144</td>
<td>0.576</td>
</tr>
</tbody>
</table>

• For any proposition φ, sum the atomic events where it is true: \( P(φ) = \sum_{ω: ω╞ φ} P(ω) \)

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

- Start with the joint probability distribution:

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

- Can also compute conditional probabilities:

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

Normalization

- Denominator can be viewed as a normalization constant \( \alpha \)

\[
\mathbf{P}(\text{Cavity} \mid \text{toothache}) = \alpha, \quad \mathbf{P}(\text{Cavity}, \text{toothache}) = \alpha, [\mathbf{P}(\text{Cavity}, \text{toothache}, \text{catch}) + \mathbf{P}(\text{Cavity}, \text{toothache}, \neg \text{catch})] = \alpha, [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] = \alpha, <0.12, 0.08> = <0.6, 0.4>
Independence

- A and B are independent iff
  \[ P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B) \quad \text{or} \quad P(A, B) = P(A) \cdot P(B) \]

- 32 entries reduced to 12; for \( n \) independent biased coins, \( O(2^n) \rightarrow O(n) \)
- Absolute independence powerful but rare
- Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

Conditional independence

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  1. \( P(\text{catch} \mid \text{toothache}, \text{cavity}) = P(\text{catch} \mid \text{cavity}) \)
- The same independence holds if I haven't got a cavity:
  2. \( P(\text{catch} \mid \text{toothache}, \neg \text{cavity}) = P(\text{catch} \mid \neg \text{cavity}) \)
- \textit{Catch} is \textit{conditionally independent} of \textit{Toothache} given \textit{Cavity}:
  \( P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity}) \)
- Equivalent statements:
  \[
  P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) \\
  P(\text{Toothache, Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) \cdot P(\text{Catch} \mid \text{Cavity})
  \]
Conditional independence contd.

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

- 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 \(P(a \land b) = P(a | b) P(b) = P(b | a) P(a)\)
  \(\Rightarrow\) Bayes' rule: \(P(a | b) = P(b | a) P(a) / P(b)\)

- or in distribution form
  \[P(Y|X) = P(X|Y) P(Y) / P(X) = \alpha P(X|Y) P(Y)\]

- Useful for assessing diagnostic probability from causal probability:
  - \(P(\text{Cause}|\text{Effect}) = P(\text{Effect}|\text{Cause}) P(\text{Cause}) / P(\text{Effect})\)
  - E.g., let \(M\) be meningitis, \(S\) be stiff neck:
    \[P(m|s) = P(s|m) P(m) / P(s) = 0.8 \times 0.0001 / 0.1 = 0.0008\]
  - Note: posterior probability of meningitis still very small!
Bayes' Rule and conditional independence

\[
P(Cavity \mid toothache \land catch) = \alpha P(toothache \land catch \mid Cavity) P(Cavity) = \alpha P(toothache \mid Cavity) P(catch \mid Cavity) P(Cavity)
\]

- This is an example of a naïve Bayes model:
  \[
P(Cause, Effect_1, \ldots, Effect_n) = P(Cause) \prod_i P(Effect_i \mid Cause)
\]

- Total number of parameters is \textit{linear} in \(n\)
- All features/symptoms/effects conditionally independent of each other given the class/diagnosis/cause

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