Uncertainty

Russell & Norvig Chapter 13

Uncertainty

Let $A_t$ be the action of leaving for the airport $t$ minutes before your flight.
Will $A_t$ get you there on time?

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

Let $A_t$ be the action of leaving for the airport $t$ minutes before your flight. Will $A_t$ get you there on time?

A purely logical approach either
1. risks falsehood: “$A_{120}$ will get me there on time”, or
2. leads to conclusions that are too weak for decision making:
   “$A_{120}$ will get me there on time if there's no accident and it doesn't rain and my tires remain intact etc.”

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

- How to represent uncertainty in knowledge?
- How to perform inference with uncertain knowledge?
- Which action to choose under uncertainty?
Dealing with uncertainty

- Implicit
  - Ignore what you are uncertain of when you can
  - Build procedures that are robust to uncertainty

- Explicit
  - Build a model of the world that describes uncertainty about its state, dynamics, and observations
  - Reason about the effect of actions given the model
Methods for handling uncertainty

- **Default Reasoning:**
  - Assume the car does not have a flat tire
  - Assume $A_{120}$ works unless contradicted by evidence

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

- **Worst case reasoning** (the world behaves according to Murphy’s law).

- **Probability**
  - Model agent's degree of belief
  - Given the available evidence, $A_{120}$ will get me there on time with probability 0.95
Probability

- Probabilities relate propositions to agent's own state of knowledge
  
  e.g., \( P(A_{120} \mid \text{no reported accidents}) = 0.96 \)

- Probabilities of propositions change with new evidence:
  
  e.g., \( P(A_{120} \mid \text{no reported accidents, 5 a.m.}) = 0.99 \)
Making decisions under uncertainty

Suppose I believe the following:

- \( P(A_{60} \text{ gets me there on time} | \ldots) = 0.001 \)
- \( P(A_{90} \text{ gets me there on time} | \ldots) = 0.70 \)
- \( P(A_{120} \text{ gets me there on time} | \ldots) = 0.95 \)
- \( P(A_{150} \text{ gets me there on time} | \ldots) = 0.99 \)
- \( P(A_{1440} \text{ gets me there on time} | \ldots) = 0.9999 \)

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
Axioms of probability

For any events $A$, $B$ in a space of events $\Omega$

- $0 \leq P(A) \leq 1$
- $P(\Omega) = 1$ and $P(\phi) = 0$
- $P(A \lor B) = P(A) + P(B) - P(A \land B)$
Axioms of probability

- $0 \leq P(\omega) \leq 1 \quad \sum_{\omega \in \Omega} P(\omega) = 1$

- $P(A \lor B) = P(A) + P(B) - P(A \land B)$
  [inclusion-exclusion principle]
Example

You draw a card from a deck of cards (52 cards). What is the probability of each of the following events:

- A king
- A face card
- A spade
- A face card or a red suit
- A card
Where do probabilities come from

Two camps:
- Frequentist interpretation
- Bayesian interpretation
Frequentist interpretation

- Draw a ball from an urn containing \( n \) balls of the same size; \( r \) are red, the rest black.
- The probability of the event “the ball is red” corresponds to the relative frequency with which we expect to draw a red ball 
  \[ P(\text{red}) = ? \]
Subjective probabilities

There are many situations in which there is no objective frequency interpretation:

- E.g. the probability that you will get to the airport in time.
- There are theoretical justifications for subjective probabilities!
The Bayesian viewpoint

- Probability is "degree-of-belief".
- To the Bayesian, probability lies subjectively in the mind, and can be different for people with different information.
- In contrast, to the frequentist, probability lies objectively in the external world.
Random Variables

A random variable can be thought of as an unknown value that may change every time it is inspected.

Suppose that a coin is tossed three times and the sequence of heads and tails is noted. The event space for this experiment is:

\[ S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \]

\( X \) - the number of heads in three coin tosses. \( X \) assigns each outcome in \( S \) a number from the set \{0, 1, 2, 3\}.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>HHH</th>
<th>HHT</th>
<th>HTH</th>
<th>THH</th>
<th>HTT</th>
<th>THT</th>
<th>TTH</th>
<th>TTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

We can now ask the question – what is the probability for observing a particular value for \( X \) (the distribution of \( X \)).
Random Variables

- **Boolean** random variables
e.g., *Cavity* (do I have a cavity?)
Distribution characterized by a number *p*.

- **Discrete** random variables
e.g., *Weather* is one of <sunny, rainy, cloudy, snow>
Domain values must be exhaustive and mutually exclusive

- The *(probability) distribution* of a random variable *X* with *m* values
  \(x_1, x_2, \ldots, x_n\) is:

  \[(p_1, p_2, \ldots, p_m)\]

  with \(P(X=x_i) = p_i\) and \(\Sigma_i p_i = 1\)
Joint Distribution

- Given \( n \) random variables \( X_1, \ldots, X_n \)
- The **joint distribution** of these variables is a table in which each entry gives the probability of one combination of values of \( X_1, \ldots, X_n \)
- Example:

<table>
<thead>
<tr>
<th></th>
<th>Toothache</th>
<th>( \neg \text{Toothache} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>( \neg \text{Cavity} )</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

\[ P(\neg \text{Cavity} \land \text{Toothache}) \quad P(\text{Cavity} \land \neg \text{Toothache}) \]
It’s all in the joint

<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>

- \( P(\text{Toothache}) = P((\text{Toothache} \land \text{Cavity}) \lor (\text{Toothache} \land \neg \text{Cavity})) \)
  \[= P(\text{Toothache} \land \text{Cavity}) + P(\text{Toothache} \land \neg \text{Cavity}) \]
  \[= 0.04 + 0.01 = 0.05 \]

  We summed over all values of Cavity: **marginalization**

- \( P(\text{Toothache} \lor \text{Cavity}) = P((\text{Toothache} \land \text{Cavity}) \lor (\text{Toothache} \land \neg \text{Cavity})) \)
  \[= P(\text{Toothache} \land \text{Cavity}) \lor (\neg \text{Toothache} \land \text{Cavity}) \]
  \[= 0.04 + 0.01 + 0.06 = 0.11 \]

  These are examples of **inference by enumeration**
Conditional Probability

- **Definition:**
  \[ P(A|B) = \frac{P(A \land B)}{P(B)} \quad (if \ P(B) > 0) \]

- **Read:** probability of A given B
  - Example: \( P(\text{snow}) = 0.03 \) but \( P(\text{snow} | \text{winter}) = 0.06 \), \( P(\text{snow} | \text{summer}) = 1e-4 \)

- can also write this as:
  \[ P(A \land B) = P(A|B) \ P(B) \]

- called the **product rule**
Example

<table>
<thead>
<tr>
<th></th>
<th>Toothache</th>
<th>(\neg)Toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>(\neg)Cavity</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

\[
P(\text{Cavity}|\text{Toothache}) = \frac{P(\text{Cavity} \land \text{Toothache})}{P(\text{Toothache})} = \frac{0.04}{0.05} = 0.8
\]
Events A and B are independent if

\[ P(A \mid B) = P(A) \quad \text{which is equivalent to:} \]

\[ P(A \land B) = P(A) \cdot P(B) \]

Example: the outcomes of rolling two dice are independent.
Bayes’ Rule

\[ P(A \land B) = P(A|B) \, P(B) = P(B|A) \, P(A) \]

\[ P(B|A) = \frac{P(A|B) \, P(B)}{P(A)} \]

Image from: http://commons.wikimedia.org/wiki/File:Thomas_Bayes.gif
Example

- **Given:**
  - \( P(\text{Cavity}) = 0.1 \)
  - \( P(\text{Toothache}) = 0.05 \)
  - \( P(\text{Cavity}|\text{Toothache}) = 0.8 \)

- **Using Bayes’ rule:**
  \[
P(\text{Toothache}|\text{Cavity}) = \frac{(0.8 \times 0.05)}{0.1} = 0.4
\]
Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

source: http://en.wikipedia.org/wiki/Monty_Hall_problem
Solution

\[ P(C2|O3) = \frac{P(O3|C2) P(C2)}{P(O3)} = \frac{1 \times 1/3}{1/2} = \frac{2}{3} \]

\[ P(C1|O3) = \frac{P(O3|C1) P(C1)}{P(O3)} = \frac{1/2 \times 1/3}{1/2} = \frac{1}{3} \]

Your pick

Host opens

Should you pick this one instead?
Solution

\[ P(C_2|O_3) = \frac{P(O_3|C_2)P(C_2)}{P(O_3)} = \frac{1 \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3} \]

\[ P(O_3) = P(O_3|C_1)P(C_1) + P(O_3|C_2)P(C_2) + P(O_3|C_3)P(C_3) = \]
\[ = \frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3} = \frac{1}{2} \]
Probabilities in the wumpus world

<table>
<thead>
<tr>
<th></th>
<th>1,4</th>
<th>2,4</th>
<th>3,4</th>
<th>4,4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,3</td>
<td>2,3</td>
<td>3,3</td>
<td>4,3</td>
<td></td>
</tr>
<tr>
<td>1,2</td>
<td>2,2</td>
<td>3,2</td>
<td>4,2</td>
<td></td>
</tr>
<tr>
<td>1,1</td>
<td>2,1</td>
<td>3,1</td>
<td>4,1</td>
<td></td>
</tr>
</tbody>
</table>

There is no safe choice at this point! But are there squares that are less likely to contain a pit?
Probabilities in the *wumpus* world

<table>
<thead>
<tr>
<th></th>
<th>1,1</th>
<th>1,2</th>
<th>1,3</th>
<th>1,4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,1</td>
<td>B</td>
<td>OK</td>
<td>2,1</td>
<td>OK</td>
</tr>
<tr>
<td>1,2</td>
<td>B</td>
<td>OK</td>
<td>2,2</td>
<td>OK</td>
</tr>
<tr>
<td>1,3</td>
<td>B</td>
<td>OK</td>
<td>2,3</td>
<td>OK</td>
</tr>
<tr>
<td>1,4</td>
<td>B</td>
<td>OK</td>
<td>2,4</td>
<td>OK</td>
</tr>
</tbody>
</table>

There is no safe choice at this point! But are there squares that are less likely to contain a pit?

\[
0.2 \times 0.2 = 0.04 \\
0.2 \times 0.8 = 0.16 \\
0.8 \times 0.2 = 0.16
\]
Generalization of Bayes’ rule

\[ P(A \land B \land C) = P(A \land B | C) \ P(C) \]
\[ = P(A | B, C) \ P(B | C) \ P(C) \]
\[ P(A \land B \land C) = P(A \land B | C) \ P(C) \]
\[ = P(B | A, C) \ P(A | C) \ P(C) \]

\[ P(B | A, C) = \frac{P(A | B, C) \ P(B | C)}{P(A | C)} \]
It’s all in the joint but…

- The naïve representation runs into problems.
- Example:
  - Patients in a hospital are described by attributes such as:
    - Background: age, gender, history of diseases, …
    - Symptoms: fever, blood pressure, headache, …
    - Diseases: pneumonia, heart attack, …
  - A probability distribution needs to assign a number to each combination of values of these attributes
    - Size of table is exponential in number of attributes
Bayesian Networks

- Provide an efficient representation that relies on independence relations between variables.
Product rule

\[ P(A \land B \land C) = P(A|B,C) \cdot P(B|C) \cdot P(C) \]
Bayesian Networks

<table>
<thead>
<tr>
<th>Event</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burglary</td>
<td>0.001</td>
</tr>
<tr>
<td>Earthquake</td>
<td>0.002</td>
</tr>
</tbody>
</table>

| Alarm | P(A | B, E) |
|-------|--------|
|       | T T   0.95 |
|       | T F   0.94 |
|       | F T   0.29 |
|       | F F   0.001 |

| JohnCalls | P(J | A) |
|------------|-----|
|            | T 0.90 |
|            | F 0.05 |

| MaryCalls | P(M | A) |
|-----------|-----|
|           | T 0.70 |
|           | F 0.01 |
Bayesian Networks

\[ P(X_1, X_2, \ldots, X_n) = \prod_{i=1,\ldots,n} P(X_i|\text{Pa}(X_i)) \]