Dealing with Uncertainty

- We want to get to the point where we can reason with uncertainty
- This will require using probability e.g. probability that it will rain today is 0.99
- We will review the fundamentals of probability

Outline

1. Random variables
2. Probability

Random Variables

- The basic element of probability is the random variable
- Think of the random variable as an event with some degree of uncertainty as to whether that event occurs
- Random variables have a domain of values it can take on

Random Variables

3 types of random variables:
1. Boolean random variables
2. Discrete random variables
3. Continuous random variables

Example:
- ProfLate is a random variable for whether your prof will be late to class or not
- The domain of ProfLate is \{true, false\}
  - ProfLate = true: proposition that prof will be late to class
  - ProfLate = false: proposition that prof will not be late to class
Random Variables

Example:
• \textit{ProfLate} is a random variable for whether your prof will be late to class or not
• The domain of \textit{ProfLate} is \textless true, false\textgreater
  – \textit{ProfLate} = true: proposition that prof will be late to class
  – \textit{ProfLate} = false: proposition that prof will not be late to class

You can assign some degree of belief to this proposition e.g.
\[P(\text{ProfLate} = \text{true}) = 0.9\]

Random Variables

Example:
• \textit{ProfLate} is a random variable for whether your prof will be late to class or not
• The domain of \textit{ProfLate} is \textless true, false\textgreater
  – \textit{ProfLate} = true: proposition that prof will be late to class
  – \textit{ProfLate} = false: proposition that prof will not be late to class

And to this one e.g.
\[P(\text{ProfLate} = \text{false}) = 0.1\]

Random Variables

• We will refer to random variables with capitalized names e.g. \textit{X, Y, ProfLate}
• We will refer to names of values with lower case names e.g. \textit{x, y, proflate}
• This means you may see a statement like \textit{ProfLate} = \textit{proflate}
  – This means the random variable \textit{ProfLate} takes the value \textit{proflate} (which can be \textit{true} or \textit{false})
• Shorthand notation:
  \[\text{ProfLate} = \text{true} \text{ is the same as } \text{proflate} \text{ and } \text{ProfLate} = \text{false} \text{ is the same as } \neg \text{proflate}\]

Boolean Random Variables

• Take the values \textit{true} or \textit{false}
• E.g. Let \textit{A} be a Boolean random variable
  – \[P(\text{A} = \text{false}) = 0.9\]
  – \[P(\text{A} = \text{true}) = 0.1\]

Discrete Random Variables

Allowed to taken on a finite number of values e.g.
• \[P(\text{DrinkSize}=\text{small}) = 0.1\]
• \[P(\text{DrinkSize}=\text{medium}) = 0.2\]
• \[P(\text{DrinkSize}=\text{large}) = 0.7\]

Discrete Random Variables

Values of the domain must be:
• Mutually Exclusive i.e. \[P(\text{A} = v_i \text{ AND } A = v_j) = 0\]
  if \(i \neq j\)
  This means, for instance, that you can’t have a drink that is both \textit{small} and \textit{medium}
• Exhaustive i.e. \[P(\text{A} = v_1 \text{ OR } A = v_2 \text{ OR ... OR } A = v_k) = 1\]
  This means that a drink can only be either \textit{small}, \textit{medium} or \textit{large}. There isn’t an \textit{extra large}. 
Discrete Random Variables

Values of the domain must be:
• Mutually Exclusive i.e. \( P( A = v_i \text{ AND } A = v_j ) = 0 \) if \( i \neq j \)
  This means, for instance, that you can't have a drink that is both Small and Medium
• Exhaustive i.e. \( P( A = v_1 \text{ OR } A = v_2 \text{ OR } ... \text{ OR } A = v_k ) = 1 \)
  This means that a drink can only be either small, medium, or large.
  The AND here means intersection i.e. \((A = v_i) \cap (A = v_j)\)
  The OR here means union i.e. \((A = v_i) \cup (A = v_j) \cup ... \cup (A = v_k)\)

Continuous Random Variables

• Can take values from the real numbers
• E.g. They can take values from \([0, 1]\)
• Note: We will primarily be dealing with discrete random variables
• (The next slide is just to provide a little bit of information about continuous random variables)

Probability Density Functions

Discrete random variables have probability distributions:

Continuous random variables have probability density functions e.g:

Probabilities

• We will write \( P( A = \text{true} ) \) as “the fraction of possible worlds in which \( A \) is true”
• We can debate the philosophical implications of this for the next 4 hours
• But we won’t
Visualizing A

The Event space of all possible worlds

- World's in which A is true
  - $P(a) = $ Area of reddish oval
- World's in which A is false

Interpreting the axioms

- $0 \leq P(a) \leq 1$
- $P(\text{true}) = 1$
- $P(\text{false}) = 0$
- $P(a \text{ OR } b) = P(a) + P(b) - P(a \text{ AND } b)$

These Axioms are Not to be Trifled With

- There have been attempts to do different methodologies for uncertainty
  - Fuzzy Logic
  - Three-valued logic
  - Dempster-Shafer
  - Non-monotonic reasoning

- But the axioms of probability are the only system with this property:
  
  If you gamble using them you can’t be unfairly exploited by an opponent using some other system [de Finetti 1931]
Prior Probability

- We can consider \( P(A) \) as the unconditional or prior probability
  - E.g. \( P(\text{ProfLate} = \text{true}) = 1.0 \)
- It is the probability of event \( A \) in the absence of any other information
- If we get new information that affects \( A \), we can reason with the conditional probability of \( A \) given the new information.

Conditional Probability

- \( P(A \mid B) \) = Fraction of worlds in which \( B \) is true that also have \( A \) true
- Read this as: “Probability of \( A \) conditioned on \( B \)”
- Prior probability \( P(A) \) is a special case of the conditional probability \( P(A \mid \) ) conditioned on no evidence

Conditional Probability Example

\( H = \) "Have a headache"
\( F = \) "Coming down with Flu"

- \( P(H) = 1/10 \)
- \( P(F) = 1/40 \)
- \( P(H \mid F) = 1/2 \)

"Headaches are rare and flu is rarer, but if you're coming down with 'flu there's a 50-50 chance you'll have a headache."

Definition of Conditional Probability

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

Corollary: The Chain Rule (aka The Product Rule)

\[
P(A, B) = P(A \mid B)P(B)
\]

Important Note

\[
P(A \mid B) + P(\neg A \mid B) = 1
\]

But:

\[
P(A \mid B) + P(A \mid \neg B) \text{ does not always } = 1
\]
The Joint Probability Distribution

• P(A, B) is called the joint probability distribution of A and B
• It captures the probabilities of all combinations of the values of a set of random variables

The Joint Probability Distribution

• For example, if A and B are Boolean random variables, then P(A,B) could be specified as:

<table>
<thead>
<tr>
<th>Event</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A=false, B=false</td>
<td>0.25</td>
</tr>
<tr>
<td>A=false, B=true</td>
<td>0.25</td>
</tr>
<tr>
<td>A=true, B=false</td>
<td>0.25</td>
</tr>
<tr>
<td>A=true, B=true</td>
<td>0.25</td>
</tr>
</tbody>
</table>

The Joint Probability Distribution

• Now suppose we have the random variables:
  – Drink = {coke, sprite}
  – Size = {small, medium, large}
• The joint probability distribution for P(Drink, Size) could look like:

<table>
<thead>
<tr>
<th>Event</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drink=coke, Size=small</td>
<td>0.1</td>
</tr>
<tr>
<td>Drink=coke, Size=medium</td>
<td>0.1</td>
</tr>
<tr>
<td>Drink=coke, Size=large</td>
<td>0.3</td>
</tr>
<tr>
<td>Drink=sprite, Size=small</td>
<td>0.1</td>
</tr>
<tr>
<td>Drink=sprite, Size=medium</td>
<td>0.2</td>
</tr>
<tr>
<td>Drink=sprite, Size=large</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Full Joint Probability Distribution

• Suppose you have the complete set of random variables used to describe the world
• A joint probability distribution that covers this complete set is called the full joint probability distribution
• Is a complete specification of one’s uncertainty about the world in question
• Very powerful: Can be used to answer any probabilistic query