COMP304: Knowledge Representation and Reasoning

Lecture 17: Probability Theory (1)

Today

- Introduction to Probability Theory
- Some basic definitions
  - Sample spaces
  - Events
  - Combining events and computing their probability
  - Conditional probability

Probability Theory - Introduction

- So far we have looked at logical aspects of KR&R; there are many different varieties of logic that contribute towards the adequate and efficient representation of particular forms of knowledge for computer manipulation.
- We now we turn our attention to a different aspect of KR&R, based on Probability and Decision Theory, to help represent and reason about certain things that logic cannot deal with.
- The purpose of this is to enable us to design and build agents that will operate in a wide range of complex environments.
- Suppose that we want a robot agent to perform an action, say picking up an object.
  - What happens if the environment is stochastic, so picking up the item is not reliable?
  - What happens if the environment is partially observable?

Probability Theory - Introduction

- We have already considered approaches that distinguish between and represent 'possible' and 'necessary' truths.

- Another approach is to use probability theory.

- Thus, we can say things like: 90% of the time a *Pickup* will leave the item in the robot's hand, 10% of the time it will leave the item on the table.

- We can then perform computations to find out how likely a plan is to achieve a given goal.

- But this is only part of what we need to do: an agent also needs to figure out what to do.

Probability Theory - Introduction

- We want to build agents which “do the right thing”, which clearly this needs a notion of “rightness”.
- One way to deal with “rightness” is to define it in terms of what is in an agent’s best interests.
- Given a set of actions that can be performed, we can look at the situations which result from the actions and choose the one which is best.
- This requires a way of quantifying “best” from the perspective the agent.
- In addition, since the effects of actions are *uncertain*, we need to factor in a way of reasoning about how likely the situations which result from actions are.

Some Basic Definitions

- **Definition 1:** *A* sample space *of an experiment is a set S of elements* $E_1, E_2, \ldots, E_k$ *such that any outcome of the experiment corresponds to exactly one element in the set.*

- **Definition 2:** *Given an experiment with a sample space S, the elements* $E_1, E_2, \ldots, E_k$ *of S are called* sample points*.*

- **Definition 3:** *An* event *is a subset of the sample space S.*

  We say that the event *E* has occurred if the outcome of the experiment corresponds to an element in the set *E*.

- **Definition 4**: *The sample space* $S = \{E_1, E_2, \ldots, E_k\}$ *and the probabilities* $\Pr(E_i)$, $i = 1, \ldots, k$ *determine the* probability model *for a random experiment.*

Definition 5: *The* probability of an event *E* *is the sum of the probabilities of the simple events which constitute E.*

Example - Sample Spaces

- Consider tossing a coin - there are two possible outcomes, a head (H) and a tail (T).
- The sample space for the experiment is $S = \{H, T\}$
- In a more complex experiment, such as tossing a dime and a nickel, there are a number of ways of defining the sample space.
- We could count the total number of heads giving: $S = \{0, 1, 2\}$
- Alternatively we could distinguish exactly what the outcome was, for instance using: $S = \{H_n H_d, H_n T_d, T_n H_d, T_n T_d\}$

  where H_n indicates that the nickel came up heads.
- The second sample space (but not the first) allows us, for instance, to identify whether a particular coin came up heads.

Events

- But what if we want to talk about situations in which "at least one head appears"?

- Considering the nickel and dime example, there are particular outcomes which might be interesting.

- For instance:
  - U: Exactly one head appears.
  - V: Exactly two heads appear.
  - W: At least one head appears.
  - X: A head appears on the dime.
  - Y: A head appears on the nickel.
  - Z: No head appears.

- These outcomes are "events".

Events

- Each event corresponds to a set of sample points:
  - $U = \{H_n T_d, T_n H_d\}$
  - $V = \{H_n H_d\}$
  - $W = \{H_n H_d, H_n T_d, T_n H_d\}$
  - $X = \{H_n H_d, T_n H_d\}$
  - $Y = \{H_n H_d, H_n T_d\}$
  - $Z = \{T_n T_d\}$

- It is clear that V and Z are different from the rest since they contain a single sample point; we call events which contain a single sample point *simple events*.

- Events made up of more than one sample point will be called *composite events*.
  - Clearly composite events can always be decomposed into simple events.

- Because experiments are random, listing the outcomes does not fully describe the experiment - we need probabilities as well.

Probabilities

- Given a sample space: $S = \{E_1, E_2, \ldots, E_k\}$
  we want to determine: $\Pr(E_i)$ for $i = 1, \ldots, k$

- There are three ways in which this might be done:

  (1) Assume that every outcome is equally likely: $\Pr(E_i) = \frac{1}{k}$ for $i = 1, \ldots, k$.
  This can be justified when nothing is known about the likelihood of the various outcomes.

  (2) Use observed relative frequencies obtained from a series of experiments:
  $\Pr(E_i) =$ *relative frequency of the event* E_i

  (3) Assign the probabilities based on our belief about the likelihood of the events.

- In all cases we assign probabilities so that:

$$0 \leq \Pr(E_i) \leq 1$$

$$\sum_{i=1}^{k} \Pr(E_i) = 1$$

Tautologies and Contradictions

- Since all possible outcomes are enumerated in the sample space

  $$\Pr(S) = 1$$

  This might be translated as "some event in S must occur".

- The event S is sometimes known as the *tautology* - the event which is always true - and is written as T:

  $$\Pr(\mathrm{T}) = 1$$

- If an event is not a possible outcome of the experiment, then there are no corresponding sample points in S. The probability of such an event is zero.

- This event is sometimes known as the *contradiction*, and is written as $\perp$. It may also be written as Ø.

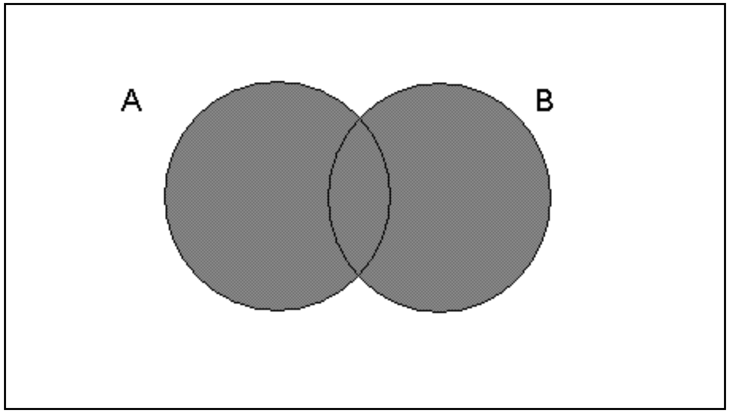
$$\Pr(\perp) = \Pr(\emptyset) = 0$$

Combining Events

- As we have defined them, events are sets of outcomes. We thus combine events set-theoretically.
- In the following, we will consider events A and B in some sample space S.
- The event "*A or B*" is the union of A and B. Thus:

$$A \text{ or } B = A \cup B$$
$$= \{x \mid x \in A \text{ or } x \in B\}$$

Depicted as a Venn Diagram:



Combining Events

- The event "*A* and *B*" is the intersection of *A* and *B*. Thus:

$$A \text{ and } B = A \cap B$$
$$= \{x \mid x \in A \text{ and } x \in B\}$$

- We can also write *A*,*B* or *AB* in place of $A \cap B$.

- The event "not *A*", also written as $\neg A$ or $\sim A$, is the set of all events in *S* which are not in *A*:

$$\neg A = \{x \in S \mid x \notin A\}$$

- The difference of *A* and *B*, also called the relative complement of *B* with respect to *A*, is the set of points in *A* which belong to *A* and not *B*.

$$A - B = A \cap \neg B$$
$$= \{x \mid x \in A \text{ and } x \notin B\}$$

Partitions

- Clearly the events A and $\neg A$ cannot occur at the same time; we say such sets of events are *mutually exclusive*.

- If A and B are mutually exclusive, they have no points in common, so:
$$A \cap B = \emptyset$$

- The set of events $A_1, \ldots, A_n$ are mutually exclusive if:
$$A_i \cap A_j = \emptyset$$

  for all i, j such that $i \neq j$.

- The set of events $A_1, \ldots, A_n$ is said to be exhaustive if:
$$A_1 \cup \ldots \cup A_n = S$$

- A set of mutually exclusive and exhaustive events forms a *partition* of S.

- The simple events which describe a sample space are always a partition.

Partitions

- For any event A, A and $\neg A$ form a partition of S since they are mutually exclusive and:

$$A \cup \neg A = S$$

- If $A_1, \ldots, A_n$ form a partition of S, then any event F can be written as:

$$F = (A_1 \cap F) \cup (A_2 \cap F) \cup \ldots \cup (A_n \cap F)$$

- This allows us to decompose F into the union of a set of mutually exclusive events.

- We can now consider how to establish the probabilities of combined events.

Probabilities of Combined Events

- In our nickel and dime example we could establish the probabilities of the simple events, and so we could calculate the probability of any composite event.

- This is not always possible, so we will develop methods for dealing with composite events without having to know the probabilities of simple events.

- We start from the following axioms:

$$0 \leq \Pr(A) \leq 1 \tag{1}$$

$$\Pr(S) = 1 \tag{2}$$

$$\Pr(A_1 \cup A_2) = \Pr(A_1) + \Pr(A_2) \tag{3}$$

where A_1 and A_2 are mutually exclusive.

Addition Law

- We can now show the following (called the addition law):

  Theorem 1: *If A and B are any two events in S, then the probability of the event "A or B" is given by:*

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

- We subtract the intersection of A and B to ensure that events in the intersection are not counted twice, when A and B are not mutually exclusive.

- In the nickel and dime example, consider the event of a head on the nickel or the dime, that is $X \cup Y$.

- Applying the addition law we find the probability of this event is:

$$\begin{aligned}\Pr(X \cup Y) &= \Pr(X) + \Pr(Y) - \Pr(X \cap Y) \\ &= \Pr(X) + \Pr(Y) - \Pr(V)\end{aligned}$$

  where the probabilities of X, Y and V can be established, using axiom (3), from the probabilities of the original sample points.

Probabilities of Combined Events

- Often we need to deal with more than two events, and for such situations we have:

- **Theorem 2:** *If $A_1, \ldots, A_n$ are a set of mutually exclusive events in S, the probability of their union is:*

$$\Pr(A_1 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \Pr(A_i)$$

- Of course, if $A_1, \ldots, A_n$ forms a partition of S, then:

$$A_1 \cup \ldots \cup A_n = S$$

and

$$\Pr(A_1 \cup \ldots \cup A_n) = 1$$

- We also have:

  Theorem 3: *If $\neg A$ is the complement of the event A in S, then*

$$\Pr(\neg A) = 1 - \Pr(A)$$

Probabilities of Combined Events

- Another useful result is the extension of the addition law to more than two events which are not mutually exclusive.

- In this case we have:

  Theorem 4: *If A, B and C are any three events in S, then the probability of their union is:*

$$
\begin{aligned}
\Pr(A \cup B \cup C) = \Pr(A) + \Pr(B) + \Pr(C) & \\
- \Pr(A \cap B) & \\
- \Pr(A \cap C) & \\
- \Pr(B \cap C) & \\
+ \Pr(A \cap B \cap C) &
\end{aligned}
$$

Conditional Probability

- Frequently we are interested in just part of the sample space.
  - e.g. given a sample space of "English children" we may be interested in the sub-population with blue eyes.

- Conditional probability gives us a means of handling this kind of problem.

- Consider a family is chosen at random from a set of families having two children (but not having twins).

- What is the probability that both children are boys?

- A suitable sample space is:

$$S = \{(B, B), (G, B), (B, G), (G, G)\}$$

- It is reasonable to assume that each of the sample points is equally likely, so that:

$$\Pr(\textit{Both children are boys}) = \tfrac{1}{4}$$

Conditional Probability

- Now you learn that the families were selected from those who have one child at a boys' school, how does this affect the probability?

- The new sample space (denoted by S^*) is:

$$S^* = \{(B, B), (G, B), (B, G)\}$$

and we are now looking for:

$$\Pr(\textit{Two boys} \mid \textit{At least one boy})$$

where the vertical line is read "given that".

- How do we assign probabilities to the events in S^*?

- The answer comes from considering how S^* relates to S.

Normalisation

- *S** is a subset of *S*, so every sample point in *S** is a sample point in *S* and therefore has a probability we can determine.

- However, if we sum these probabilities they will sum to less than 1 (because the sum of the probabilities of all the sample points in *S* is 1) in violation of axiom (2) given earlier.

- We therefore *normalise* by dividing the probability of the sample point, calculated from *S,* by the sum of the probabilities of all the sample points in *S** (which is the probability of the event *S** in the sample space *S*):

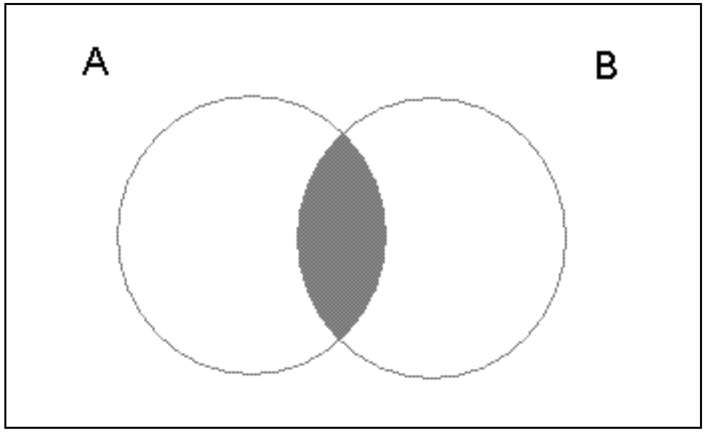
$$\Pr(\textit{Two boys} \mid \textit{At least one boy}) = \frac{1/4}{3/4}$$

$$= 1/3$$

Conditioning

- We can generalise this. Consider a sample space *S* with two events *A* and *B* with a non-empty intersection $A \cap B$.

- We can picture this as:

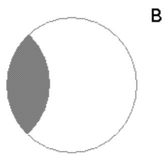


S

- Now, consider we want to *condition on* event *B*, so that we are interested in discovering $\Pr(A \mid B)$, the probability that *A* will occur given that *B* is known to occur.

Conditioning

- The situation in which we are interested is thus the situation in which event *B* has occurred, a situation we can picture as:



- By comparing the two diagrams, it is clear that the sample points in *A* after the conditioning on *B* are exactly those which were in $\Pr(A \cap B)$ before the conditioning.

- Once again though we have to normalise, dividing by the probability of all the sample points in the new sample space:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Conditional Probability

- **Definition 6:** *The conditional probability of A given B, is denoted by* $\Pr(A \mid B)$ *and is defined by:*

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

for $\Pr(B) \neq 0$.

- We can re-write this as:

$$\Pr(A \cap B) = \Pr(A \mid B)\,\Pr(B)$$

- Note: we can also write this the other way round:
$\Pr(A \cap B) = \Pr(B \mid A)\,\Pr(A)$

- This is known as the *multiplication rule* or *product law*, and is useful is establishing joint probabilities.

- The rule can easily be extended for more events, thus for three events A, B and C we have:

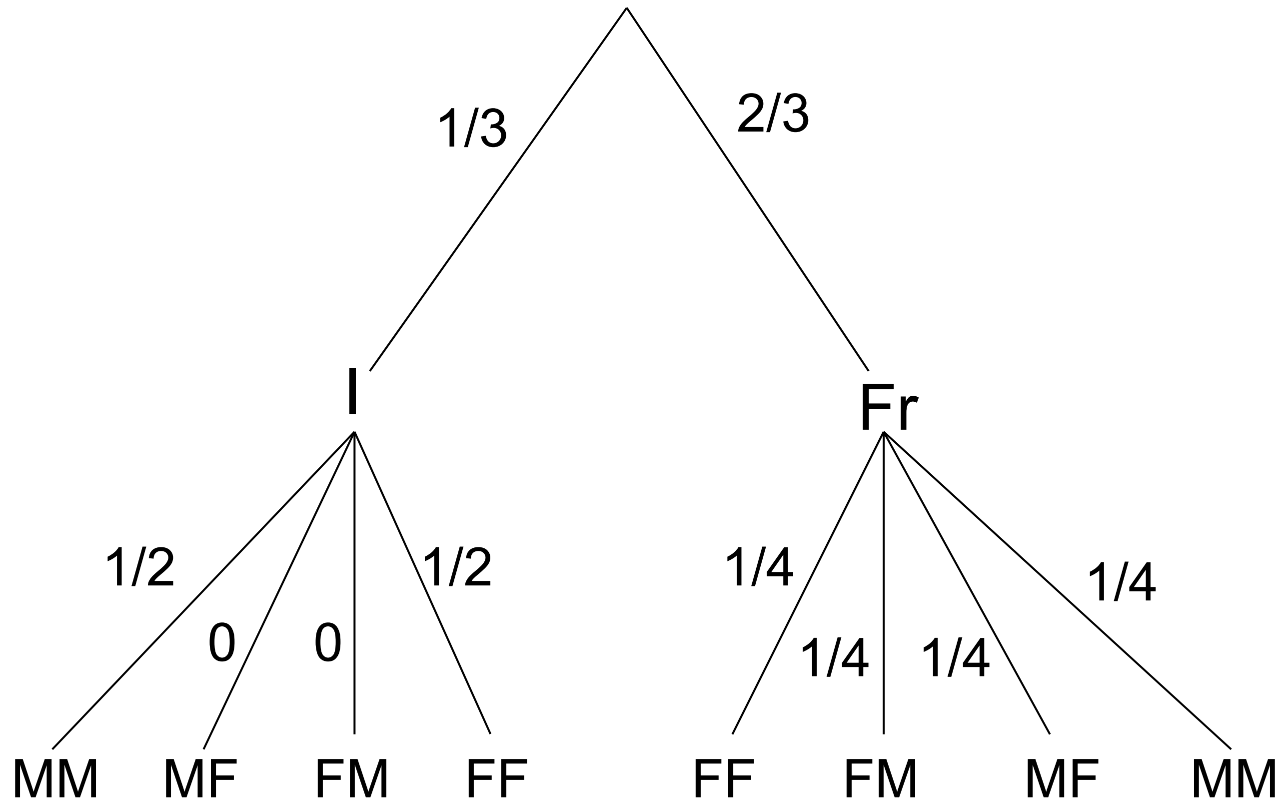
$$\Pr(A \cap B \cap C) = \Pr(C \mid A \cap B)\,\Pr(B \mid A)\,\Pr(A)$$

Multiplication Rule

- Consider choosing one family from a set of families with just one pair of twins (and thus no other children).
- What is the probability that both twins are boys?
- We can think of the choice of twins as occurring in two stages.
  - First we select whether the twins are identical, *I*, or fraternal, *Fr*. (Note that about one third of human twins are identical.)
  - Then we select the sex of the twins (*MM*, *MF*, *FM*, *FF*).
- The probabilities of the sexes of the fraternal twins will be the same as for any other two-child family. For the identical twins, the outcomes *MF* and *FM* are no longer possible.
- Essentially we are conditioning on the event "same sex" in families with two children.

Tree Representation

- We can represent this two-stage experiment as a tree:



Joint Probabilities

- Since the probabilities for the second stage are conditional, we can write:

- $\Pr(\textit{Twin boys}) = \Pr(I \cap MM)$
  $+ \Pr(Fr \cap MM)$

  $= \Pr(MM \mid I) \Pr(I)$
  $+ \Pr(MM \mid Fr) \Pr(Fr)$

  $= 1/2 \cdot 1/3 + 1/4 \cdot 2/3$

  $= 1/3$

Conditioning

- This example also illustrates the fact that it is often easier to obtain information in the form of conditional probabilities such as:

$$\Pr(MM \mid I)$$

than joint probabilities such as:

$$\Pr(MM \cap I)$$

- There is no real explanation for why this should be the case, but it is one of the reasons behind the use of Bayesian networks (which we will see later).

Summary

- Motivation for using Probability Theory in KR&R
- Sample spaces
- Events
- Combining events
- Probabilities of combined events
- Addition law
- Conditional probabilities
- Product law