COMP111: Artificial Intelligence

Section 8. Reasoning under Uncertainty 1

Frank Wolter

Content

- Why reasoning under uncertainty?
- Basics of probability
- Conditional probability
- Independence
- Bayes' Theorem and basic applications.

Why Reasoning under Uncertainty?

Logic-based KR&R methods mostly assume that knowledge is certain. Often, this is not the case (or it is impossible to list all assumptions that make it certain).

- When going to the airport by car, how early should I start? 45 minutes should be enough from Liverpool to Manchester Airport, but only under the assumption that there are no accidents, no lane closures, that my car does not break down, and so on. This uncertainty is hard to eliminate, but still an agent has to make a decision.
- A dental patient has toothache. Does the patient have a cavity? One might want to capture the relationship between patients having a cavity and patients having toothache by the rule:

  $$\mathsf{Toothache}(x) \rightarrow \mathsf{Cavity}(x)$$

  But this does not work as not every toothache is caused by a cavity. Hard to come up with exhaustive list of reasons:

  $$\mathsf{Toothache}(x) \rightarrow \mathsf{Cavity}(x) \vee \mathsf{GumProblem}(x) \vee \mathsf{Abscess}(x) \vee \cdots$$

Uncertainty

Trying to use exact rules to cope with a domain like medical diagnosis or traffic fails for three main reasons:

- Laziness: it is too much work to list an exceptionless set of rules.
- Theoretical ignorance: Medical science has, in many cases, no strict laws connecting symptoms with diseases.
- Practical ignorance: Even if we have strict laws, we might be uncertain about a particular patient because not all the necessary tests have been or can be run.

Probability as Summary

- Probability provides a way of summarizing the uncertainty that comes from our laziness and ignorance.
- We might not know for sure what disease a particular patient has, but we believe that there is, say, an 80% chance that a patient with toothache has a cavity.
  The 80% summarises those cases in which all the factors needed for a cavity to cause a toothache are present and other cases in which the patient has both toothache and cavity but the two are unconnected.
- The missing 20% summarizes all the other possible causes of toothache that we are too lazy or ignorant to confirm or deny.

Discrete Probability

We represent random experiments using discrete probability spaces (S, P) consisting of:

- the sample space S of all elementary events $x \in S$; members of S are also called outcomes of the experiment.
- a probability distribution P assigning a real number $P(x)$ to every elementary event $x \in S$ such that
  - for every $x \in S$: $0 \leq P(x) \leq 1$;
  - and $\sum_{x \in S} P(x) = 1$.

Recall that if S consists of $x_1, \ldots, x_n$, then

$$\sum_{x \in S} P(x) = P(x_1) + \cdots + P(x_n).$$

Example: Flipping a fair coin

Consider the random experiment of flipping a coin. Then the corresponding probability space (S, P) is given by

- $S = \{H, T\}$;
- $P(H) = P(T) = \frac{1}{2}$.

Consider the random experiment of flipping a coin two times, one after the other. Then the corresponding probability space (S, P) is given as follows:

- $S = \{HH, HT, TH, TT\}$;
- $P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$.

Example: Rolling a fair die

Consider the random experiment of rolling a die. Then the corresponding probability space (S, P) is given by

- $S = \{1, 2, 3, 4, 5, 6\}$;
- For every $x \in S$: $P(x) = \frac{1}{6}$.

Consider the random experiment of rolling a die n times. Then the corresponding probability space (S, P) is given as follows:

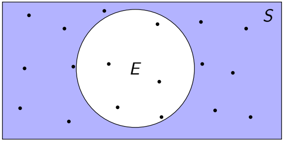
- S is the set of sequences of length n over the alphabet $\{1, \ldots, 6\}$ (sometimes denoted $\{1, \ldots, 6\}^n$).
- $P(x) = \frac{1}{6^n}$ for every elementary event x, since S has 6^n elements.

Uniform Probability Distributions

- A probability distribution is uniform if every outcome is equally likely.
- For uniform probability distributions, the probability of an outcome x is 1 divided by the number $|S|$ of outcomes in S.

$$P(x) = \frac{1}{|S|}$$

Events



- An event is a subset $E \subseteq S$ of the sample space S.
- The probability of the event E is given by

$$P(E) = \sum_{x \in E} P(x)$$

Notice that

- $0 \leq P(E) \leq 1$ for every event E,
- $P(\emptyset) = 0$ and $P(S) = 1$.

Example

If I roll a die three times, the event E of rolling at least one 6 is given by

- the set of sequences of length 3 over $\{1, \ldots, 6\}$ containing at least one 6.
- $P(E)$ is the number of sequences containing at least one 6 divided by $6 \times 6 \times 6 = 216$.

If we roll a fair die, then the event E of rolling an odd number is given by

- the set $E = \{1, 3, 5\}$
- $P(E) = P(1) + P(3) + P(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

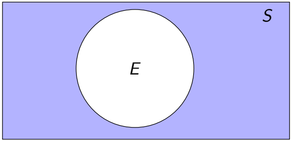
The probability of composed events

We show how the probability of

- the complement of an event can be computed from the probability of the event;
- the union of events can be computed from the probabilities of the individual events.

The probability of the complement of an event

Let $\neg E = S \setminus E$. Then $P(\neg E) = 1 - P(E)$



Side remark:

$S = \neg E \cup E$

Proof.

$$1 = \sum_{x \in S} P(x) = \sum_{x \in E} P(x) + \sum_{x \in \neg E} P(x)$$

Thus,

$$\sum_{x \in \neg E} P(x) = 1 - \sum_{x \in E} P(x)$$

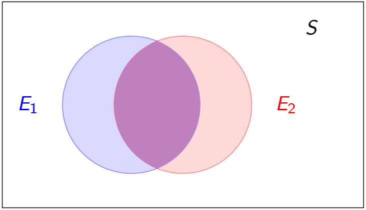
Example

What is the probability that at least one bit in a randomly generated sequence of 10 bits is 0?

- $S = \{0, 1\}^{10}$ = all sequences of 0 and 1 of length 10.
- For every $x \in S$, $P(x) = (\frac{1}{2})^{10} = \frac{1}{2^{10}}$.
- E = all sequences of 0 and 1 of length 10 containing at least one 0.
- $\neg E = \{1111111111\}$.
- $P(\neg E) = \frac{1}{2^{10}}$.
- $P(E) = 1 - \frac{1}{2^{10}}$.

The probability of the union of two events

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$



Side remark:

$$|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2|$$

Proof of $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

Recall:

- $P(E_1) = \sum_{x \in E_1} P(x)$;
- $P(E_2) = \sum_{x \in E_2} P(x)$;
- $P(E_1 \cup E_2) = \sum_{x \in E_1 \cup E_2} P(x)$

Thus,

$$\begin{aligned} P(E_1 \cup E_2) &= \sum_{x \in E_1 \cup E_2} P(x) \\ &= \sum_{x \in E_1} P(x) + \sum_{x \in E_2} P(x) - \sum_{x \in E_1 \cap E_2} P(x) \\ &= P(E_1) + P(E_2) - P(E_1 \cap E_2) \end{aligned}$$

Example

Suppose I have a jar of 30 sweets, as follows.

	red	blue	green
circular	2	4	3
square	6	7	8

The sample space S has 30 elements and if one chooses a sweet uniformly at random, then

$$P(x) = \frac{1}{30}$$

for all $x \in S$. What is the probability of choosing a red or circular sweet?

- The probability that it is red is $\frac{2+6}{30} = \frac{8}{30}$ ($P(R) = \frac{8}{30}$).
- The probability that it is circular is $\frac{2+4+3}{30} = \frac{9}{30}$ ($P(C) = \frac{9}{30}$).

Then $P(R \cup C)$ is is the probability that the sweet is red or circular. Thus

$$P(R \cup C) = P(R) + P(C) - P(R \cap C) = \frac{8}{30} + \frac{9}{30} - \frac{2}{30} = \frac{15}{30} = \frac{1}{2}$$

Disjoint events

Assume that $E_1, \ldots, E_n$ are mutually disjoint events. So $E_i \cap E_j = \emptyset$ whenever $i \neq j$.

Then

$$P(\bigcup_{1 \leq i \leq n} E_i) = \sum_{1 \leq i \leq n} P(E_i)$$

Example: three dice

Suppose that I roll a fair die three times. Then

- S is the set of sequences of length three over $\{1, \ldots, 6\}$ (or $\{1, \ldots, 6\}^3$).
- $P(x) = \frac{1}{6 \times 6 \times 6} = \frac{1}{216}$ for all $x \in S$.

What is the probability that I roll at least one 6?

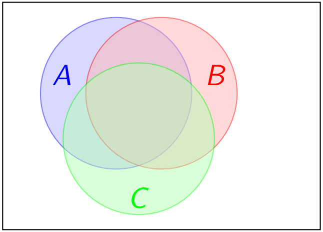
Let E_1: event that 1st roll is a 6, E_2: event that 2nd roll is a 6; E_3: event that 3rd roll is a 6.

We would like to know

$$P(E_1 \cup E_2 \cup E_3)$$

Computing the probability of a union of three sets

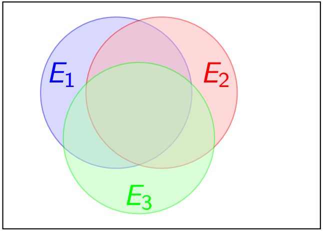
$$
\begin{aligned}
P(A \cup B \cup C) = &\, P(A) + P(B) + P(C) \\
&- P(A \cap B) - P(A \cap C) - P(B \cap C) \\
&+ P(A \cap B \cap C).
\end{aligned}
$$



$$
\begin{aligned}
|A \cup B \cup C| \quad &= |A| + |B| + |C| \\
&-|A \cap B| - |A \cap C| - |B \cap C| \\
&+|A \cap B \cap C|.
\end{aligned}
$$

Three dice example continued

E_1: event that 1st roll is a 6.
E_2: event that 2nd roll is a 6.
E_3: event that 3rd roll is a 6.



$$\begin{aligned}
P(E_1 \cup E_2 \cup E_3) &= P(E_1) + P(E_2) + P(E_3) \\
&- P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) \\
&+ P(E_1 \cap E_2 \cap E_3) \\
&= \tfrac{36}{216} + \tfrac{36}{216} + \tfrac{36}{216} \\
&- \tfrac{6}{216} - \tfrac{6}{216} - \tfrac{6}{216} \\
&+ \tfrac{1}{216} \\
&= \tfrac{91}{216} \approx 0.42
\end{aligned}$$

Conditional Probability

- Often, we are interested in just part of the sample space.
- Conditional probability gives us a means of handling this situation.
- Consider a family 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 $S = \{BB, GB, BG, GG\}$.
- It is reasonable to assume that $P(x) = \frac{1}{4}$ for all $x \in S$. Thus $P(BB) = \frac{1}{4}$.

Conditional Probability

- Now you learn that the families were selected from those who have one child at a boys' school. Does this change probabilities?
- The new sample space (denoted S') is

$$S' = \{BB, GB, BG\}$$

  and we are now looking for

$$P(BB \mid \text{ at least one boy }) = P(BB \mid S')$$

  where the vertical line is read "given that".
- How do we assign probabilities to the events in S'?

Normalisation

- S' is a subset of S, so every outcome x in S' is also in S. Its probability $P(x)$ in S we can determine.
- However, if we just take the sum of these probabilities, they will sum to less than 1. (In the example $P(BB) + P(GB) + P(BG) = \frac{3}{4}$.) This violates our assumptions for probability spaces.
- We therefore normalise by dividing the probability $P(x)$ of the outcome x in S by the probability $P(S')$ of S' in S. Thus

$$P(BB \mid \text{ at least one boy }) = P(BB \mid S') = \frac{P(BB)}{P(S')} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Conditioning

- Assume now that events A and B are given.
- Assume we know that B happens. So we want to condition on B. Thus, we want to know

$$P(A \mid B),$$

the probability that A occurs given that B is known to occur.
- So we want to know the probability $P(A \cap B)$ (since we know that B occurs) after the conditioning on B.
- Once again, we can't take $P(A \cap B)$ itself but have to normalise by dividing by the probability of the new sample space $P(B)$. Thus

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

Conditional Probability

Let A and B be events, with $P(B) > 0$.
The conditional probability $P(A|B)$ of A given B is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Example

I choose a sweet uniformly at random from a jar of 30 sweets.

	red	blue	green
circular	2	4	3
square	6	7	8

- The probability that it is red conditioned on that it is circular is

$$\frac{2}{30} \div \frac{2+4+3}{30} = \frac{2}{30} \times \frac{30}{2+4+3} = \frac{2}{2+4+3} = \frac{2}{9}, \text{ which is}$$

  less than the unconditional probability that it is red, $\frac{2+6}{30} = \frac{8}{30}$.

- The probability that it is circular conditioned on the event that it is red is

$$\frac{2}{30} \div \frac{2+6}{30} = \frac{2}{6+2} = \frac{1}{4}$$

Multiplication Rule

- The conditional probability $P(A|B)$ of A given B is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- We can rewrite this as

  $$P(A \cap B) = P(A \mid B)P(B) \quad (\text{ also: } P(A \cap B) = P(B \mid A)P(A))$$

  This is knows as the multiplication rule.

- It can be extended to more events, for example:

  $$P(A \cap B \cap C) = P(C \mid A \cap B)P(A \cap B) = P(C \mid A \cap B)P(B \mid A)P(A)$$

An Application of the Multiplication Rule

- Consider choosing a family from a set of families with just one pair of twins (and thus no other children).
- What is the probability $P(BB)$ that both twins are boys?
- Recall that twins are either identical (I) or fraternal (F). We know that a third of human twins are identical. Thus

$$P(I) = \frac{1}{3}, \quad P(F) = \frac{2}{3}$$

  and

$$P(BB) = P(I \cap BB) + P(F \cap BB)$$

- By multiplication rule

$$P(I \cap BB) = P(BB \mid I)P(I), \quad P(F \cap BB) = P(BB \mid F)P(F)$$

An Application of the Multiplication Rule

- The probability of being a girl or boy for fraternal twins will be the same as for any other two-child family. For the identical twins, the outcomes BG and GB are no longer possible. Thus:

$$P(BB \mid I) = \frac{1}{2}, \quad P(BB \mid F) = \frac{1}{4}$$

- We obtain:

$$\begin{aligned} P(BB) &= P(I \cap BB) + P(F \cap BB) \\ &= P(BB \mid I)P(I) + P(BB \mid F)P(F) \\ &= \frac{1}{2} \times \frac{1}{3} + \frac{1}{4} \times \frac{2}{3} \\ &= \frac{1}{3} \end{aligned}$$

Independence

- In everyday language we refer to events that "have nothing to do with each other" as being independent.
- A similar notion of independence is useful in probability theory as it helps to structure probabilistic knowledge and reduce complexity.

Events A and B are independent if

$$P(A \cap B) = P(A) \times P(B)$$

If $P(A) \neq 0$ and $P(B) \neq 0$, then the following are equivalent:

- A and B are independent;
- $P(B) = P(B|A)$; (recall that $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$);
- $P(A) = P(A|B)$; (recall that $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$).

Example 1: Independence

If you roll two dice, then the rolls are independent. So the event A of rolling a 1 with the first die is independent of the event B of rolling a 1 with the second die. So the probability that both of these happen is

$$P(A \cap B) = P(A) \times P(B) = \tfrac{1}{6} \times \tfrac{1}{6} = \tfrac{1}{36}$$

This holds for all $ab \in \{1, 2, 3, 4, 5, 6\}^2$:

$$P(ab) = \frac{1}{36} = P(a) \times P(b)$$

Example continued

- Let A be the event of rolling a 1 with the first die.
- Let C be the event of rolling a 1 with the first or second die.
- Let B be the event of rolling an even sum.

We show that A and B are independent but C and B are not.

- As
$$A = \{11, 12, 13, 14, 15, 16\}.$$
we have $P(A) = \frac{6}{36} = \frac{1}{6}$.
- As
$$B = \{11, 13, 15, 22, 24, 26, 31, 33, 35, 42, 44, 46, 51, 53, 55, 62, 64, 66\}$$
we have $P(B) = \frac{18}{36} = \frac{1}{2}$.
- We have $A \cap B = \{11, 13, 15\}$. So
$$P(A \cap B) = \frac{3}{36} = \frac{1}{12} = P(A)P(B)$$
and conclude that A and B are independent.

Example continued

- As
$$C = \{11, 12, 13, 14, 15, 16, 21, 31, 41, 51, 61\}$$
we have $P(C) = \frac{11}{36}$.
- As
$$B \cap C = \{11, 13, 15, 31, 51\}$$
we have $P(B \cap C) = \frac{5}{36}$.
- So
$$P(B \cap C) = \tfrac{5}{36} \neq \frac{5.5}{36} = P(B)P(C)$$
and we conclude that B and C are not independent.

Example 2

- Consider the random experiment of drawing two balls, one after the other, from a basket containing a red, a blue, and a green ball.
- Let R_1 be the event that the first ball is red.
- Let R_2 be the event that the second ball is blue.

We show that R_1 and R_2 are not independent.
Let

- $S = \{RB, RG, BR, BG, GR, GB\}$;
- $P(RB) = P(RG) = P(BR) = P(BG) = P(GR) = P(GB) = \frac{1}{6}$.
- Then $R_1 = \{RB, RG\}$ and $R_2 = \{RB, GB\}$.
- $P(R_1) = P(RB) + P(RG) = \frac{1}{3}$ and $P(R_2) = P(RB) + P(GB) = \frac{1}{3}$.
- Thus,

$$P(R_1 \cap R_2) = P(RB) = \frac{1}{6} \neq P(R_1) \times P(R_2) = \frac{1}{9}$$

Independence for more than two events

- Consider a finite set of events

$$A_1, \ldots, A_n$$

- $A_1, \ldots, A_n$ are pairwise independent if every pair of events is independent: for all distinct k, m

$$P(A_m \cap A_k) = P(A_m)P(A_k)$$

- $A_1, \ldots, A_n$ are mutually independent if every event is independent of any intersection of the other events: for all distinct $k_1, \ldots, k_m$:

$$P(A_{k_1}) \times \cdots \times P(A_{k_m}) = P(A_{k_1} \cap \cdots \cap A_{k_m})$$

- Pairwise independence is not a particularly important notion.

Pairwise independence does not imply mutual independence

- Consider the random experiment of flipping a coin two times, one after the other. Then $S = \{HH, HT, TH, TT\}$ and $P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$.
- Let $H_1 = \{HT, HH\}$. Then $P(H_1) = \frac{1}{2}$;
- Let $H_2 = \{HH, TH\}$. Then $P(H_2) = \frac{1}{2}$;
- Let $H^* = \{HH, TT\}$. Then $P(H^*) = \frac{1}{2}$.
- H_1, H_2, H^* are pairwise independent:

$$H_1 \cap H_2 = H_1 \cap H^* = H_2 \cap H^* = \{HH\}, \quad P(HH) = \frac{1}{4}$$

- H_1, H_2, H^* are not mutually independent:

$$P(H_1 \cap H_2 \cap H^*) = \frac{1}{4} \neq \frac{1}{8} = P(H_1)P(H_2)P(H^*)$$

Example

- Consider the random experiment of rolling a die n times. Recall that probability space (S, P) is given as follows:
  - S is the set of sequences of length n over the alphabet $\{1, \ldots, 6\}$ (sometimes denoted $\{1, \ldots, 6\}^n$).
  - $P(x) = \frac{1}{6^n}$ since S has 6^n elements.
- Let E_i be the event of getting a 6 the ith time the die is rolled, for $1 \leq i \leq n$.
- Then $E_1, \ldots, E_n$ are mutually independent.

Bayes' Theorem (first form)

If $P(A) > 0$, then

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A)}$$

Proof. We have

- $P(A \cap B) = P(A|B) \times P(B)$ and
- $P(A \cap B) = P(B|A) \times P(A)$.

Thus,

$$P(A|B) \times P(B) = P(B|A) \times P(A)$$

By dividing by $P(A)$ we get

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A)}$$

Application: Diagnosis

Assume a patient walks into a doctor's office complaining of a stiff neck. The doctor knows:

- meningitis may cause a patient to have a stiff neck 50% of the time (causal knowledge);
- the probability of having meningitis if $\frac{1}{50000}$;
- the probability of having a stiff neck is $\frac{1}{20}$.

Problem: What is the probability that the patient has meningitis?

Let A be the event that the patient has a stiff neck and B the event that he has meningitis. Then

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)} = \frac{1/2 \times 1/50000}{1/20} = \frac{1}{5000}$$

Comments

- We can interpret the fact that the patient has a stiff neck as a new observation.
- Given this observation, we want to classify the patient as either having meningitis or not having meningitis.
- We have prior knowledge about the unconditional probability of having a stiff neck and having meningitis.
- We have causal knowledge about the number of times in which meningitis causes a stiff neck.
- We can then compute the diagnostic probabilities using

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

Bayes' Theorem (alternative form)

If $P(A) > 0$, then

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A|B) \times P(B) + P(A|\neg B) \times P(\neg B)}$$

It suffices to show

$$P(A) = P(A|B) \times P(B) + P(A|\neg B) \times P(\neg B)$$

But this follows from

$$\begin{aligned} P(A) &= P((A \cap B) \cup (A \cap \neg B)) \\ &= P(A \cap B) + P(A \cap \neg B) \\ &= P(A|B) \times P(B) + P(A|\neg B) \times P(\neg B) \end{aligned}$$

Application: Diagnosis

Assume a drug test is

- positive for users 99% of the time;
- negative for non-users 99% of the time.

Assume that 0.5% take the drug.

Problem: What is the probability that a person whose test is positive (event A) takes the drug (event B)?

We have $P(A|B) = \frac{99}{100}$, $P(\neg A|\neg B) = \frac{99}{100}$, and $P(B) = \frac{1}{200}$.

Thus, $P(A|\neg B) = \frac{1}{100}$ and $P(\neg B) = \frac{199}{200}$.

Thus,

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A|B) \times P(B) + P(A|\neg B) \times P(\neg B)} = 0.33$$