UNCERTAINTY

Outline

- Acting under uncertainty
- Basic probability notation
- The axioms of probability
- Bayes' rule and its use
- Where do probabilities come from?

Based on the textbook by S. Russell & P. Norvig:

Artificial Intelligene, A Modern Approach, Chapter 14

ACTING UNDER UNCERTAINTY

- Agents almost never have access to the whole truth about their environment and have thus act under **uncertainty**.
- **Qualification problem:** how to define the circumstances under which a given action is *guaranteed* to work.

  It is typical that there are too many conditions (or exceptions to conditions) to be explicitly enumerated.
- The right thing to do, the **rational decision**, depends on the relative importance of the various goals and the likelihood that, and degree to which, they will be achieved.

Example. Suppose that our taxi-driving agent wants to drive someone to an airport 15 miles away to catch a flight.

- Plan A_{90} involves leaving 90 minutes before the flight.
- Plan A_{90} is successful given that
  1. the car does not break or run out of gas,
  2. the agent does not get into an accident,
  3. the plane does not leave early, and so on ...
- Performance measure: getting to the airport on time, avoiding unproductive, long waits as well as speeding tickets.
- Other plans, such as A_{120}, increases the likelihood of getting to the airport on time, but also the likelihood of a long wait.

Handling Uncertain Knowledge

Example. Consider formalizing some diagnostic principles:

$$\forall p(Symptom(p, Toothache) \rightarrow Disease(p, Cavity))$$

$$\forall p(Symptom(p, Toothache) \rightarrow Disease(p, Cavity) \lor Disease(p, ImpactedWisdom) \lor Disease(p, GumDisease) \lor \cdots)$$

$$\forall p(Disease(p, Cavity) \land \cdots \rightarrow Symptom(p, Toothache))$$

Difficulties with formalizations using sentences of first-order logic:

1. **Laziness:** completing antecedents/consequents is very laborious.
2. **Theoreticalignorance:** the domain lacks a comprehensive theory.
3. **Practicalignorance:** applicability to a patient is not guaranteed.

- Agent's knowledge on the environment can at best provide only a **degree of belief** in relevant sentences.
- **Probability theory** assigns a degree of belief $P(\phi)$ (a real number from the interval [0,1]) to a sentence ϕ.
- Individual sentences ϕ are considered to be either true or false.
  - — $P(\phi) = 0$ means that ϕ is false in all circumstances
  - — $P(\phi) = 1$ means that ϕ is true in all circumstances.
- Probabilities provide *a way of summarizing the uncertainty*.

Example. A patient has a cavity with a probability of 0.8 if (s)he has a toothache. The remaining probability mass (0.2) summarizes all other explanations for toothache.

Probability Theory vs. Fuzzy Logic

- Degrees of belief (as in probability theory) are different from **degrees of truth** (as in fuzzy logic).

Example. Consider an atomic sentence A stating "the door is closed".

- $P(A) = 0.99$ means that the door is closed almost for sure.
- In contrast to this, a degree of truth $V(A) = 0.99$ would mean that the door is almost completely closed.

On The Role of Evidence

- The probability that an agent assigns to a sentence ϕ depends of the percepts $\phi_1, \ldots, \phi_n$ (evidence) obtained so far.
- Analogous to logical consequence $\{\phi_1, \ldots, \phi_n\} \models \phi$.
- *Prior/unconditional probability* $P(\phi)$ is the probability of ϕ without evidence.
- *Posterior/conditional probability* $P(\phi \mid \phi_1 \land \cdots \land \phi_n)$ is the probability of ϕ after obtaining pieces of evidence $\phi_1, \ldots, \phi_n$.

Example. Consider a shuffled standard pack of 52 playing cards. Let A mean "the card drawn from the pack is the ace of spades".

- Prior probabilities before looking the card:

$$P(A) = \tfrac{1}{52} \text{ and } P(\neg A) = \tfrac{51}{52}.$$

- Posterior probabilities after looking the card:

$$P(A \mid A) = \frac{P(A \land A)}{P(A)} = \frac{P(A)}{P(A)} = 1 \text{ and}$$

$$P(A \mid \neg A) = \frac{P(A \land \neg A)}{P(\neg A)} = \frac{0}{P(\neg A)} = 0.$$

Note: all pieces of evidence have to be taken into account when the posterior probabilities of sentences are determined.

Uncertainty and Rational Decisions

Example. Regarding the airport example, suppose that

1. P("Plan A_{90} succeeds.") $= 0.95$,
2. P("Plan A_{120} succeeds.") $= 0.98$, and
3. P("Plan A_{1440} succeeds.") $= 0.9999$.

– Which plan should be selected for execution?

– What kind of criteria could be used for making such a decision?

☞ In addition to estimating the success rates of plans/actions, we have to specify preferences on the possible outcomes.

➤ By **utility theory** every state has a *degree of usefulness*, or *utility*, to an agent and the agent prefers states with higher utility.

➤ An agent may freely define its preferences that may appear even irrational from the point of view of other agents.

➤ Utility theory allows for altruism (unselfishness).

➤ **Decision theory = probability theory + utility theory**

The principle of Maximum Expected Utility (MEU):

"*an agent is* ***rational*** *if and only if it chooses an action that yields the highest expected utility, averaged over all the possible outcomes of the action*".

Design for a Decision-theoretic Agent

➤ An abstract algorithm for a decision-theoretic agent that selects rational actions is the following:

```
function DT-AGENT(percept) returns an action
  static: a set probabilistic beliefs about the state of the world

  calculate updated probabilities for current state based on
      available evidence including current percept and previous action
  calculate outcome probabilities for actions,
      given action descriptions and probabilities of current states
  select action with highest expected utility
      given probabilities of outcomes and utility information
  return action
```

➤ The steps of the algorithm will be refined in the sequel.

BASIC PROBABILITY NOTATION

➤ A formal language is needed for representing and reasoning with uncertain knowledge.

➤ An extension of the language of propositional logic is used.

➤ Degrees of belief are expressed as probabilities $P(\phi)$ that are assigned to sentences ϕ of the language.

➤ The dependence on evidence/experience $\phi_1, \ldots, \phi_n$ is expressed in terms of conditional probability statements $P(\phi \mid \phi_1, \ldots, \phi_n)$.

Prior/Unconditional Probabilities

- Unconditional probabilities are applied when no other information (evidence) is available.

  Example. Let $Cavity$ mean that "a patient has a cavity".

  Then $P(Cavity) = 0.1$ means that *in the absence of any other information* the patient has a cavity with a probability of 0.1.

- This probability may change if new information becomes available.
- To enrich propositional language, also **random variables** X that range over particular domains $\langle x_1, \ldots, x_n \rangle$ are used in the sequel.
- Any propositional atom A can be viewed as a Boolean random variable ranging over the domain $\langle true, false \rangle$.

- An expression $X = x_i$ (denoting that X has the value x_i) can be interpreted as a propositional atom.

Example. Consider a random variable $Weather$ that ranges over weather conditions $Sunny$, $Rain$, $Cloudy$, and $Snow$.

Then we may assign probabilities to particular values of $Weather$:

$$P(Weather = Sunny) = 0.7$$
$$P(Weather = Rain) = 0.2$$
$$P(Weather = Cloudy) = 0.08$$
$$P(Weather = Snow) = 0.02$$

- A **probability distribution** $\mathbf{P}$ assigns probabilities to all value combinations of the random variables involved.

Example. In the example above, $\mathbf{P}(Weather) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$.

The probability distribution $\mathbf{P}(Weather, Cavity)$ is two-dimensional.

Posterior/Conditional Probabilities

- If new evidence is acquired, conditional probabilities have to be used instead of unconditional ones.
- Conditional probabilities can be defined in terms of unconditional ones. When $P(B) > 0$ we have that

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

Example. Suppose that $Cavity$ and $Toothache$ mean that "the patient has a cavity" and "the patient has a toothache", respectively.

The prior probability $P(Cavity) = 0.1$ has to be replaced by a conditional one $P(Cavity \mid Toothache) = 0.8$ in case of a toothache.

- Note: the conditional probability $P(Cavity \mid Toothache) = 0.8$ does not mean that $P(Cavity) = 0.8$ when $Toothache$ is true!
- The preceding definition can be rewritten as **product rule**:

$$P(A \wedge B) = P(A \mid B)P(B), \text{ or alternatively}$$
$$P(A \wedge B) = P(B \mid A)P(A).$$

- Conditional probabilities and the product rule can be generalized for probability distributions of random variables as follows:

$$\mathbf{P}(X \mid Y) = \frac{\mathbf{P}(X \wedge Y)}{\mathbf{P}(Y)} \text{ and } \mathbf{P}(X \wedge Y) = \mathbf{P}(X \mid Y)\mathbf{P}(Y).$$

- These have to be interpreted with respect to particular values of the random variables X and Y involved. For instance,

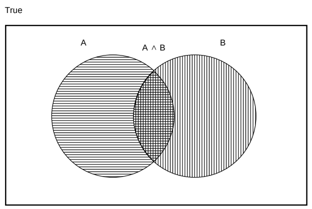
$$P(X = x_1 \wedge Y = y_2) = P(X = x_1 \mid Y = y_2)P(Y = y_2).$$

THE AXIOMS OF PROBABILITY

➤ Probabilities associated with sentences are axiomatized as follows:

For all ϕ and ψ:
A1. $0 \leq P(\phi) \leq 1$,
A2. $P(\phi) = 0$ if ϕ is unsatisfiable,
A3. $P(\phi) = 1$ if ϕ is valid, and
A4. $P(\phi \vee \psi) = P(\phi) + P(\psi) - P(\phi \wedge \psi)$.

➤ The last axiom is easily verified from a Venn diagram:



Lemma. If ϕ and ψ are logically equivalent, then $P(\phi) = P(\psi)$.

Proof. Suppose that ϕ and ψ are logically equivalent, i.e., $\models \phi \leftrightarrow \psi$.

1. $\models \psi \vee \neg\psi$ and $\models \phi \vee \neg\psi$.
2. Both $\psi \wedge \neg\psi$ and $\phi \wedge \neg\psi$ are unsatisfiable.
3. Using A4 we obtain

$$P(\psi \vee \neg\psi) = P(\psi) + P(\neg\psi) - P(\psi \wedge \neg\psi)$$
$$\Longrightarrow P(\neg\psi) = 1 - P(\psi)$$

and

$$P(\phi \vee \neg\psi) = P(\phi) + P(\neg\psi) - P(\phi \wedge \neg\psi)$$
$$\Longrightarrow 1 = P(\phi) + 1 - P(\psi) - 0$$
$$\Longrightarrow P(\phi) = P(\psi).$$

➤ Other propositional connectives are covered as follows:

1. $P(\phi \wedge \psi) = P(\phi) + P(\psi) - P(\phi \vee \psi)$ (A4)

2. $P(\neg\phi) = 1 - P(\phi)$

3. $P(\phi \rightarrow \psi) = P(\neg\phi \vee \psi) = P(\neg\phi \vee (\phi \wedge \psi))$ (Lemma)
$= P(\neg\phi) + P(\phi \wedge \psi) - P(\neg\phi \wedge \phi \wedge \psi)$ (A4)
$= 1 - P(\phi) + P(\phi \wedge \psi) - 0$
$= 1 - P(\phi) + P(\psi \mid \phi)P(\phi)$ (Def. of $P(\psi \mid \phi)$)

4. $P(\phi \leftrightarrow \psi) = P((\neg\phi \vee \psi) \wedge (\neg\psi \vee \phi))$ (Lemma)
$= 1 - P(\phi) + 1 - P(\psi) + 2 \cdot P(\phi \wedge \psi) - 1$ (A4,A3)
$= 1 - P(\phi) - P(\psi) + 2 \cdot P(\phi \wedge \psi)$

Why the Axioms of Probability Are Reasonable

➤ Bruno de Finetti, 1931:

"If Agent 1 expresses a set of degrees of belief that violate the axioms of probability theory then ***there is a betting strategy*** *for Agent 2 that guarantees that Agent 1 will lose money."*

Example. Consider the following betting scenario:

Agent 1		Agent 2		Outcome for Agent 1			
Proposition	Belief	Bet	Stakes	$A \wedge B$	$A \wedge \neg B$	$\neg A \wedge B$	$\neg A \wedge \neg B$
A	0.4	A	4 to 6	-6	-6	4	4
B	0.3	B	3 to 7	-7	3	-7	3
$A \vee B$	0.8	$\neg(A \vee B)$	2 to 8	2	2	2	-8
				-11	-1	-1	-1

☞ Choices made by Agent 2 guarantee that Agent 1 loses money.

The Joint Probability Distribution

- Consider a system of n random variables $X_1, \dots, X_n$ that may range over different domains.
- An **atomic event** $X_1 = x_1 \wedge \cdots \wedge X_n = x_n$ is an assignment of particular values $x_1, \dots, x_n$ to the variables $X_1, \dots, X_n$.
- The joint probability distribution $\mathbf{P}(X_1, \dots, X_n)$ assigns probabilities to all possible atomic events.

Example. For the Boolean random variables Cavity and $\mathit{Toothache}$:

	$\mathit{Toothache}$	$\neg\mathit{Toothache}$
Cavity	0.04	0.06
$\neg\mathit{Cavity}$	0.01	0.89

- For Boolean random variables, atomic events correspond to conjunctions of *literals* (propositional atoms or their negations).
- Atomic events are *mutually exclusive*: any conjunction of atomic events is necessarily false.
- The disjunction of all atomic events is necessarily true: entries in the joint probability distribution sum to 1.
- Probabilities provided by the joint probability distribution can be used for computing probabilities of arbitrary sentences ϕ:

  $P(\phi)$ is the sum of probabilities
  assigned to atomic events satisfying ϕ.

- Also conditional probabilities $P(\phi \mid \phi_1, \dots, \phi_n)$ can be determined by the relationship $P(\phi \mid \phi_1, \dots, \phi_n) = \frac{P(\phi \wedge \phi_1 \wedge \cdots \wedge \phi_n)}{P(\phi_1 \wedge \cdots \wedge \phi_n)}$.

- The joint probability distribution grows rapidly with respect to the number of variables (e.g., 2^n entries for n Boolean variables).

☞ It is infeasible to specify/store the whole distribution.

Example. For the preceding joint probability distribution:

1. $\mathit{Cavity} \wedge \neg\mathit{Toothache}$ is one of the atomic events,
2. $P(\mathit{Cavity}) = P(\mathit{Cavity} \wedge \mathit{Tootache}) + P(\mathit{Cavity} \wedge \neg\mathit{Toothache})$
   $= 0.04 + 0.06 = 0.10,$
3. $P(\mathit{Cavity} \vee \mathit{Toothache}) = 1 - P(\neg\mathit{Cavity} \wedge \neg\mathit{Toothache})$
   $= 1 - 0.89 = 0.11,$
4. $P(\mathit{Cavity} \mid \mathit{Toothache}) = \dfrac{P(\mathit{Cavity} \wedge \mathit{Toothache})}{P(\mathit{Toothache})} = \dfrac{0.04}{0.04 + 0.01}$
   $= 0.80.$

BAYES' RULE AND ITS USE

- Bayes' rule (or Bayes' theorem) is derived from the product rule:

  $P(A \mid B)P(B) = P(A \wedge B) = P(B \mid A)P(A)$

  $\implies P(B \mid A) = \dfrac{P(A \mid B)P(B)}{P(A)}$ given that $P(A) > 0$.

- Bayes' rule can be used for *diagnostic inference*, i.e., computing $P(D \mid S)$ on the basis of other three probabilities:
  - $P(D)$ for a disease D,
  - $P(S)$ for a symptom S, and
  - $P(S \mid D)$ for the *causal relationship* of S and D.
- Generalization for arbitrary sentences and random variables:

$$P(\phi \mid \psi) = \frac{P(\psi \mid \phi)P(\phi)}{P(\psi)} \text{ and } \mathbf{P}(Y \mid X) = \frac{\mathbf{P}(X \mid Y)\mathbf{P}(Y)}{\mathbf{P}(X)}.$$

- Bayes' rule can be further generalized by conditioning:

$$P(\phi \mid \psi \wedge \chi) = \frac{P(\phi \wedge \psi \wedge \chi)}{P(\psi \wedge \chi)}$$

$$= \frac{P(\phi \wedge \psi \wedge \chi)}{P(\phi \wedge \chi)} \cdot \frac{P(\phi \wedge \chi)}{P(\chi)} \cdot \frac{P(\chi)}{P(\psi \wedge \chi)}$$

$$= \frac{P(\psi \mid \phi \wedge \chi) P(\phi \mid \chi)}{P(\psi \mid \chi)}.$$

Here the sentence χ stands for any background evidence.

- For random variables and a background evidence E this becomes

$$\mathbf{P}(Y \mid X, E) = \frac{\mathbf{P}(X \mid Y, E)\mathbf{P}(Y \mid E)}{\mathbf{P}(X \mid E)}.$$

Applying Bayes' Rule: the Simple Case

Example. Let the atomic propositions S and M mean that "the patient has a stiff neck" and "the patient has meningitis", respectively.

- Given the probabilities $P(S \mid M) = 1/2$, $P(M) = 1/50000$, and $P(S) = 1/20$, we may apply Bayes' rule to compute

$$P(M \mid S) = \frac{P(S \mid M)P(M)}{P(S)} = \frac{\frac{1}{2} \cdot \frac{1}{50000}}{\frac{1}{20}} = \frac{1}{5000}.$$

- Diagnostic knowledge is often more tenuous than causal one: an epidemic increases $P(M)$ and $P(M \mid S)$ but not $P(S \mid M)$.

Normalization

Example. Suppose we are interested in another condition of the patient: W means that "the patient has a whiplash injury".

- The relative likelihood of meningitis and whiplash can be assessed without knowing the prior probability $P(S)$ of the symptom.

$$\frac{P(M \mid S)}{P(W \mid S)} = \frac{P(S \mid M)P(M)}{P(S \mid W)P(W)} = \frac{\frac{1}{2} \cdot \frac{1}{50000}}{\frac{4}{5} \cdot \frac{1}{1000}} = \frac{1}{80}$$

- This kind of comparison may be enough for decision making.
- Would it be possible to compute the value of $P(M \mid S)$ without assessing the prior probability $P(S)$ directly?

- One possibility is to consider *an exhaustive set of cases*: By combining $P(M \mid S) + P(\neg M \mid S) = 1$ with products

$$P(M \mid S)P(S) = P(S \mid M)P(M) \text{ and}$$

$$P(\neg M \mid S)P(S) = P(S \mid \neg M)P(\neg M)$$

we obtain $P(S) = P(S \mid M)P(M) + P(S \mid \neg M)P(\neg M)$.

- Thus $P(M \mid S) = \alpha P(S \mid M)P(M)$ and $P(\neg M \mid S) = \alpha P(S \mid \neg M)P(\neg M)$ where $\alpha = 1/P(S)$.
- Thus α is a *normalizing constant* that scales the products $P(S \mid M)P(M)$ and $P(S \mid \neg M)P(\neg M)$ so that they sum to 1.
- Generalizing for arbitrary random variables X and Y:

$$\mathbf{P}(Y \mid X) = \alpha \mathbf{P}(X \mid Y)\mathbf{P}(Y)$$

where α makes the entries in $\mathbf{P}(Y \mid X)$ sum to 1.

Combining Evidence

Example. Recall the dentist example (propositional atoms *Cavity* and *Toothache*) and a further atom *Catch* meaning that "a cavity is detected with a steel probe".

- Suppose that we know the probabilities

$$P(Cavity \mid Toothache) = 0.8 \text{ and } P(Cavity \mid Catch) = 0.95.$$

- What if both *Toothache* and *Catch* are known?
- We know by Bayes' rule that $P(Cavity \mid Catch \wedge Toothache) =$

$$\frac{P(Catch \wedge Toothache \mid Cavity)P(Cavity)}{P(Catch \wedge Toothache)}.$$

- Many (nontrivial) probabilities have to be known!

Bayesian Updating

- The idea is to incorporate pieces of evidence one at a time.

1. $P(Cavity \mid Toothache) = P(Cavity)\dfrac{P(Toothache \mid Cavity)}{P(Toothache)}$
2. Using *Toothache* as conditioning context:

$$P(Cavity \mid Toothache \wedge Catch) =$$

$$P(Cavity \mid Toothache)\frac{P(Catch \mid Toothache \wedge Cavity)}{P(Catch \mid Toothache)} =$$

$$P(Cavity)\frac{P(Toothache \mid Cavity)}{P(Toothache)}\frac{P(Catch \mid Toothache \wedge Cavity)}{P(Catch \mid Toothache)}.$$

☞ Still many probabilities have to be specified!

- Bayesian updating is order-independent.
- Further simplification can be achieved by identifying *conditional independence relations* between variables.
- For instance, Boolean variables *Tootache* and *Catch* are **conditionally independent** given *Cavity* $\iff$

$$P(Catch \mid Toothache \wedge Cavity) = P(Catch \mid Cavity) \text{ and}$$

$$P(Toothache \mid Catch \wedge Cavity) = P(Toothache \mid Cavity).$$

- Using thes we obtain $P(Cavity \mid Toothache \wedge Catch) =$

$$P(Cavity)\frac{P(Toothache \mid Cavity)}{P(Toothache)}\frac{P(Catch \mid Cavity)}{P(Catch \mid Toothache)}$$

- Finally, the product $P(Toothache)P(Catch \mid Toothache)$ in the denominator can be eliminated by normalization:

$$\mathbf{P}(Z \mid X, Y) = \alpha\,\mathbf{P}(Z)\mathbf{P}(X \mid Z)\mathbf{P}(Y \mid Z).$$

WHERE DO PROBABILITIES COME FROM?

- **Frequentist view**: probabilities come from experiments.
  If 10 out of 100 people have a cavity, then $P(Cavity) = 0.10$.
- **Objectivist view**: probabilities are real aspects of the universe that are approximated by the probabilities obtained with experiments.
- **Subjectivist view:** an analyst tries to estimate probabilities.
- **Reference class problem:** the more evidence is taken into account, the smaler becomes the reference class from which colect experimental data. This setting suggests the following:
  1. Minimizing the number of probabilities that need assessment.
  2. Maximizing the number of cases available for each assessment.

SUMMARY

- Uncertainty arises because of both laziness and ignorance.
- Probabilities provide a way of summarizing the agent's beliefs.
- **Bayes' rule/theorem** allows unknown probabilities to be computed from known, stable ones.
- The **joint probability distribution** specifies the probability of each complete assignment of values to random variables.
- The joint is typically far too large to create or use.
- **Conditional independence** relations can make **Bayesian updating** effective even with multiple pieces of evidence.

QUESTIONS

Reconsider soccer playing agents:

- Which factors cause uncertainty in this domain?

  In particular, consider factors that are related with

  1. the environment of agents,
  2. perceptual information, and
  3. outcomes of actions.
- Is it possible to deal with these factors using probabilities?
- What are the ways for determining the probabilities involved?