CS761 Artificial Intelligence

16. Reasoning with Uncertainty: Quantifying Uncertainty

Jiamou Liu
The University of Auckland

Reason with Uncertainty

Agents's plans are contingent on events in the world:

- What will be the weather in Auckland tomorrow?
- Would a COVID lockdown take place next week?
- Would the All Blacks win their next Test match?
- Would I get head or tail if I toss a coin?

Uncertainty is an important aspect of an agent's decision making:

- Agents only have partial information about the environment.
- Agents' knowledge of the truth of a statement is uncertain.
- The state of the world is inherently uncertain.

The agent's knowledge can at best provide only a degree of belief.

(Subjective) Probability

- Probability theory is a calculus of belief

  E.g. To say that "There is a 80% probability of raining" expresses a belief based on past experience.

- Probability theory is the study of how knowledge affects belief.
- The probability of a hypothesis α is a scale of the agent's belief in α in the range $[0, 1]$.

Next we will link propositional logic framework with probability to capture the above intuitions.

- Let $X_1, \ldots, X_d$ be atomic propositions.

  E.g. Sunny means outlook will be sunny, Hot means high temperature.

- The sample space Ω over these atomic propositions contains all possible interpretations. Each interpretation is also called a sample.

  E.g. Samples over Sunny, Hot:

interpretations	Sunny	Hot
e_1	true	true
e_2	true	false
e_3	false	true
e_4	false	false

- A proposition describes a constraint on atoms.

  E.g.

  1. α_1: $\neg$Hot
     $e_2 \models \alpha_1, e_4 \models \alpha_1$
  2. α_2: $\neg$Sunny $\vee$ $\neg$Hot
     $e_2 \models \alpha_2, e_3 \models \alpha_2, e_4 \models \alpha_2$

Definition [Belief measure]

- For a sample $s \in \Omega$ whose features are unknown, a belief measure over Ω is a function $\mu : 2^{\Omega} \to [0, 1]$ such that for any $S \subseteq \Omega$, $\mu(S)$ expresses the amount of belief in the fact that s is an element of S.
- Any belief measure must satisfy the following properties:
  1. (Unit Measure) s must be a sample in Ω, i.e.,
  $$\mu(\Omega) = 1$$
  2. (Additivity) Suppose S_1 and S_2 are disjoint subsets of Ω. Then our belief that $s \in S_1 \cup S_2$ is the sum of our belief that $s \in S_1$ and our belief that $s \in S_2$, i.e.,
  $$\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$$

Definition [Probability]

Given a belief measure μ over Ω, the probability of a proposition α is $P(\alpha) = \mu(\{w \in \Omega : w \models \alpha\})$, i.e., it is the strength of belief that α will hold.

Example. Consider the atoms Sunny and Hot.

- Belief measure 1:

Sunny	*true*	*true*	*false*	*false*
Hot	*true*	*false*	*true*	*false*
Belief μ	0	1	0	0

  This corresponds to a single interpretation in propositional logic without uncertainty.

- Belief measure 2:

Sunny	*true*	*true*	*false*	*false*
Hot	*true*	*false*	*true*	*false*
Belief μ	0.35	0.19	0.12	0.34

$$
\begin{aligned}
& P(\mathsf{Sunny}) \\
=& \mu(\{\mathsf{Sunny} \wedge \mathsf{Hot}, \mathsf{Sunny} \wedge \neg \mathsf{Hot}\}) \\
=& 0.35 + 0.19 = 0.54
\end{aligned}
$$

Conditional Probability

- The measure of belief in hypothesis h based on proposition e is called the conditional probability of h given e, written $P(h \mid e)$.
- The proposition e represents certain given experience.
- The probability $P(h)$ is the prior probability of h and is the same as $P(h \mid true)$. It is the probability of h without any given experience.
- The conditional probability $P(h \mid e) = \frac{P(h \wedge e)}{P(e)}$ is the agent's posterior probability of h.

E.g. Prior probability $P(\mathsf{Sunny}) = 0.54$

Evidence e: $\neg\mathsf{Hot}$

Other knowledge:

$$\begin{aligned} P(\mathsf{Sunny} \wedge \neg\mathsf{Hot}) &= 0.19 \\ P(\neg\mathsf{Hot}) &= 0.53 \end{aligned}$$

Posterior probability

$$P(\mathsf{Sunny} \mid \neg\mathsf{Hot}) = 0.19/0.53 \approx 0.36$$

Theorem (Properties of conditional probability)

The following hold for all propositions a and b and e:

- $P(e \mid e) = 1$
- If $a \wedge b$ is a contradiction, $P(a \mid e) + P(b \mid e) = P(a \vee b \mid e)$
- $P(\neg a \mid e) = 1 - P(a \mid e)$
- If a and b are logically equivalent, then $P(a \mid e) = P(b \mid e)$
- $P(a \mid e) = P(a \wedge b \mid e) + P(a \wedge \neg b \mid e)$
- $P(a \vee b \mid e) = P(a \mid e) + P(b \mid e) - P(a \wedge b \mid e)$
- Chain rule: $P(a \wedge b) = P(a)P(b \mid a)$
- Law of total probability: $P(a) = P(a \mid b)P(b) + P(a \mid \neg b)P(\neg b)$
- Baye's rule: $P(a \mid b) = \frac{P(b|a) \times P(a)}{P(b)}$

Probability Distribution

Suppose $S = \{X_1, \ldots, X_m\}$ is a set of atomic propositions, we use Ω_S to denote the set of samples on S, i.e.,

$$\Omega_S = \{(\ell_1, \ldots, \ell_m) \mid \text{ each } \ell_i = X_i \text{ or } \ell_i = \neg X_i \text{ for } 1 \leq i \leq m\}$$

Definition. [probability distribution]

- A probability distribution of a set of atoms S, denoted as $\mathbf{P}_S$, is a function from Ω_S into $[0, 1]$ such that

$$\sum_{\omega \in \Omega_S} \mathbf{P}_S(\omega) = 1.$$

  We express it as $\mathbf{P}(X_1, \ldots, X_m)$.

- A probability distribution of a set of atoms S conditioned on atoms $Y_1, \ldots, Y_\ell$, denoted as $\mathbf{P}_{S|Y_1,\ldots,Y_\ell}$, is a function from $\Omega_{\{Y_1,\ldots,Y_\ell\} \cup S}$ to $[0, 1]$ such that for any $\tau \in \Omega_{\{Y_1,\ldots,Y_m\}}$

$$\sum_{\omega \in \Omega_S} \mathbf{P}_{S|Y_1,\ldots,Y_\ell}(\omega, \tau) = 1.$$

  We express it as $\mathbf{P}(X_1, \ldots, X_m \mid Y_1, \ldots, Y_\ell)$.

Compact Representation of Probability Distribution

Conditional probability table (CPT)

A probability distribution can be represented as a probability table:

Sunny	**P**(Sunny)
0	0.46
1	0.54

and

Hot	**P**(Hot)
0	0.53
1	0.47

We can represent conditional probability distribution as a conditional probability table (CPT).

Sunny	Hot	**P**(Hot|Sunny)
0	0	$0.34/0.46 \approx 0.74$
0	1	$0.12/0.46 \approx 0.26$
1	0	$0.19/0.54 \approx 0.35$
1	1	$0.35/0.54 \approx 0.65$

Summary of notations:

- **Atomic propositions:** $X_1, X_2, X_3, \ldots, Y_1, Y_2, Y_3, \ldots, Z_1, \ldots$
- **Sample space:** Ω_S where $S = \{X_1, \ldots, X_m\}$.
- **Probability of proposition h:** $P(h)$
- **Probability of proposition h conditioned on e:** $P(h \mid e)$
- **Probability distribution of S:** $\mathbf{P}_S = \mathbf{P}(X_1, \ldots, X_m)$
- **Probability distribution of S conditioned on $Y_1, \ldots, Y_\ell$:**

$$\mathbf{P}(X_1, \ldots, X_m \mid Y_1, \ldots, Y_\ell)$$

- **CPT for $\mathbf{P}(X_1, \ldots, X_m \mid Y_1, \ldots, Y_\ell)$:**

Y_1	$\cdots$	Y_ℓ	X_1	$\cdots$	X_m	$\mathbf{P}_{X_1,\ldots,X_m}(Y_1, \ldots, Y_\ell)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

Note: Green columns denote variables for which (joint) probability is measured.

Inference with Probability

Recall: Baye's rule

$$P(a \mid b) = \frac{P(b \mid a) \times P(a)}{P(b)}$$

This means that $\mathbf{P}(a \mid b)$ is proportional to $\mathbf{P}(b \mid a) \times \mathbf{P}(a)$, written as

$$\mathbf{P}(a \mid b) \propto \mathbf{P}(b \mid a) \times \mathbf{P}(a)$$

or equivalently

$$\begin{pmatrix} P(a \mid b) \\ P(\neg a \mid b) \end{pmatrix} = \alpha \begin{pmatrix} P(b \mid a) \times P(a) \\ P(b \mid \neg a) \times P(\neg a) \end{pmatrix} \text{ for constant } \alpha$$

Inference with Probability

Recall: Baye's rule

$$P(a \mid b) = \frac{P(b \mid a) \times P(a)}{P(b)}$$

This means that $\mathbf{P}(a \mid b)$ is proportional to $\mathbf{P}(b \mid a) \times \mathbf{P}(a)$, written as

$$\overbrace{\mathbf{P}(\underbrace{a}_{\text{hypothesis}} \mid \underbrace{b}_{\text{evidence}})}^{\text{posterior probability}} \propto \underbrace{\mathbf{P}(b \mid a)}_{\text{likelihood}} \times \overbrace{\mathbf{P}(a)}^{\text{prior probability}}$$

or equivalently

$$\begin{pmatrix} P(a \mid b) \\ P(\neg a \mid b) \end{pmatrix} = \alpha \begin{pmatrix} P(b \mid a) \times P(a) \\ P(b \mid \neg a) \times P(\neg a) \end{pmatrix} \text{ for constant } \alpha$$

"Posterior_probability ∝ Likelihood × Prior_Probability"

Example. [infection] Suppose a test for a certain viral infection is 95% reliable for infected patients and 99% reliable for the healthy one. Suppose also that we have prior belief that 4% of patients have the virus. How to decide if the patient is infected?

Atoms:

- $Test$: the test outcome is positive.
- Inf: the patient is infected

Prior knowledge:

- Probability of infection $P(Inf) = 0.04$
- The test is 95% reliable for infected patients: $P(Test \mid Inf) = 0.95$
- The test is 99% reliable for healthy patients: $P(\neg Test \mid \neg Inf) = 0.99$

Task: Compute $P(Inf \mid Test)$.

- Write down the CPT:

Inf	$\mathbf{P}(Inf)$
0	0.96
1	0.04

and

Inf	Test	$\mathbf{P}(Test \mid Inf)$
0	0	0.99
0	1	0.01
1	0	0.05
1	1	0.95

- By Bayes' rule:

$$\mathbf{P}(Inf \mid Test) \propto \mathbf{P}(Test \mid Inf)\mathbf{P}(Inf)$$

- Thus

$$\begin{pmatrix} P(Inf \mid Test) \\ P(\neg Inf \mid Test) \end{pmatrix} = \alpha \begin{pmatrix} 0.95 \times 0.04 \\ 0.01 \times 0.96 \end{pmatrix} = \alpha \begin{pmatrix} 0.038 \\ 0.0096 \end{pmatrix}$$

- This means that $P(Inf \mid Test) = \frac{0.038}{0.0476} \approx 79.83\%$.

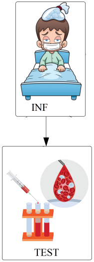
Example. [infection] The simple diagnostic problem above involves only two atoms: Inf and $Test$. Inf can be seen as a causal factor for $Test$:

- Infected patients are most likely to be tested positive.
- Uninfected patients are less likely to be tested positive.

We may represent such a causal relation with a clause: $Test \leftarrow Inf$

Note. The clause $Inf \leftarrow Test$ would not be accurate as

- testing positive does not guarantee infection; and more importantly
- even if the patient is infected, the fact that a patient is tested positive is not the reason for the infection of the patient.



Bayes’ rule tells us how to update the agent’s belief in hypothesis h as new evidence e arrives, given existing knowledge k.

Bayes’ Rule with Existing Evidence

As long as $P(e \mid k) \neq 0$

Hypothesis Likelihood Prior probability

$$P(h \mid e \wedge k) = \frac{P(e \mid h \wedge k) \times P(h \mid k)}{P(e \mid k)}$$

Posterior probability New observation Existing evidence

This means that

$$\mathbf{P}(h \mid e, k) \propto \mathbf{P}(e \mid h, k) \times \mathbf{P}(h \mid k)$$

Example. [infection] Suppose a patient has been tested positive for the viral infection. Experience tells that 70% of infected patients and 30% of uninfected patients develop a fever. We then observe that the patient has developed a fever. How likely is the patient infected with the virus then?

Knowledge:

- Let *Fev* be the proposition "The person has a fever"
- Inf is a causal factor of both *Fev* and *Test*
- $P(Fev \mid Inf) = 0.70$, $P(Fev \mid \neg Inf) = 0.3$
- By Baye's rule,

$$P(Inf \mid Fev \wedge Test) \propto P(Fev \mid Test \wedge Inf) P(Inf \mid Test).$$

We need extra knowledge for $P(Fev \mid Test \wedge Inf)$.

Definition

An atom X is independent of another atom Y conditioned on a set of atoms $Z_1, \ldots, Z_m$ if for any $y \in \{0, 1\}$

$$\mathbf{P}(X \mid Z_1, \ldots, Z_m) \times \mathbf{P}(Y \mid Z_1, \ldots, Z_m) = \mathbf{P}(X, Y \mid Z_1, \ldots, Z_m)$$

Example. [exam] Consider a domain that consists of students and exams. There are three atoms:

$$Smart, WorkHard, GoodAnswer$$

We may have the following samples

Smart	WorkHard	GoodAnswer	$\mathbf{P}(Smart, WorkHard, GoodAnswer)$
0	0	0	0.55
0	0	1	0.01
0	1	0	0.04
0	1	1	0.2
1	0	0	0.08
1	0	1	0.06
1	1	0	0.01
1	1	1	0.05

Then $P(Smart) = 0.2$, $P(WorkHard) = 0.3$.

$\mathbf{P}(Smart) \times \mathbf{P}(WorkHard) =$

Smart	WorkHard	**P**
0	0	0.56
0	1	0.24
1	0	0.14
1	1	0.06

$= \mathbf{P}(Smart, WorkHard)$.

- So *Smart* and *WorkHard* are independent.
- But *Smart* and *WorkHard* are dependent conditioned on *GoodAnswer*.
- Suppose there is a fourth atom *GoodGrade*. *GoodGrade* only depends on the answer of the students. *GoodGrade* and *Smart* are not independent. But conditioned on *GoodAnswer*, *GoodGrade* and *Smart* are independent.

Example. [infection] Suppose we have an extra knowledge that *Fev* and *Test* are independent conditioned on Inf,i.e.,

- Inf is a causal factor of *Fev*, and *Test*
- *Fev* and *Test* are independent conditioned on Inf

We can represent this by clauses $Test \leftarrow Inf$ and $Fev \leftarrow Inf$.

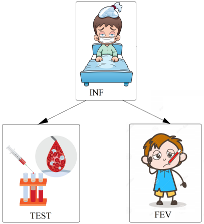
Inf	$\mathbf{P}(Inf)$
0	0.96
1	0.04

,

Inf	Test	$\mathbf{P}(Test \mid Inf)$
0	0	0.99
0	1	0.01
1	0	0.05
1	1	0.95

and

Inf	Fev	$\mathbf{P}(Fev \mid Inf)$
0	0	0.7
0	1	0.3
1	0	0.3
1	1	0.7



By conditional independence,

- $P(Fev \mid Test \wedge Inf) = P(Fev \mid Inf) = 0.70.$
- $P(Fev \mid Test \wedge \neg Inf) = P(Fev \mid \neg Inf) = 0.30$

By Baye's theorem,

$$P(Inf \mid Fev \wedge Test) = \frac{P(Fev \mid Inf)P(Inf \mid Test)}{P(Fev \mid Test)} \approx 0.5588/P(Fev \mid Test)$$
$$P(\neg Inf \mid Fev \wedge Test) = \frac{P(Fev \mid \neg Inf)P(\neg Inf \mid Test)}{P(Fev \mid Test)} \approx 0.0605/P(Fev \mid Test)$$

Thus

$$P(Inf \mid Fev \wedge Test) : P(\neg Inf \mid Fev \wedge Test) \approx 0.5588/0.0605.$$

Furthermore, $P(Inf \wedge Fev \mid Test) + P(\neg Inf \mid Fev \wedge Test) = 1.$
Therefore,
$P(Inf \mid Fev \wedge Test) = (0.7 \times 0.7983)/(0.7 \times 0.7983 + 0.3 \times 0.2017) \approx 90\%.$

Example. [wumpus world] Suppose the agent senses a breeze at both squares $(1,2)$ and $(2,1)$. Then all $(1,3),(2,2),(3,1)$ may have a pit.

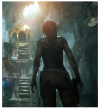
Propositions:

- $\text{Pit}_{i,j}$, $\text{Brz}_{i,j}$ for $(i,j) \in \{1,2,3,4\}^2$
- $\mathit{Known} = \neg\text{Pit}_{1,1} \wedge \neg\text{Pit}_{1,2} \wedge \neg\text{Pit}_{2,1}$
- $\mathit{Percept} = \neg\text{Brz}_{1,1} \wedge \text{Brz}_{1,2} \wedge \text{Brz}_{2,1}$.

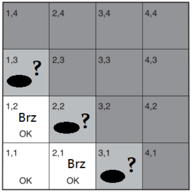
Assumptions:

- A pit causes breezes in all adjacent squares, e.g., $P(\text{Brz}_{1,2} \mid \text{Pit}_{2,2}) = 1$.
- Prior: $P(\text{Pit}_{i,j}) = 0.2$ for any $(i,j) \in \{1,2,3,4\}^2$.
- Independence: $P(\text{Pit}_{i,j}) \times P(\text{Pit}_{i',j'}) = P(\text{Pit}_{i,j} \wedge \text{Pit}_{i',j'})$ for $(i',j') \neq (i,j)$.

Query: What is $P(\text{Pit}_{1,3} \mid \mathit{Known}, \mathit{Percept})$?



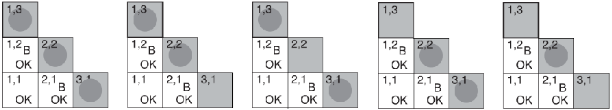




Let $\mathit{Frontier} = \mathrm{Pit}_{2,2} \wedge \mathrm{Pit}_{3,1}$.

$$
\begin{aligned}
\mathbf{P}(\mathrm{Pit}_{1,3} \mid \mathit{Known}, \mathit{Percept}) &\propto \mathbf{P}(\mathit{Percept} \mid \mathrm{Pit}_{1,3}, \mathit{Known})\mathbf{P}(\mathrm{Pit}_{1,3}, \mathit{Known}) \\
&\propto \sum_{\mathit{Frontier}} \mathbf{P}(\mathit{Percept} \mid \mathrm{Pit}_{1,3}, \mathit{Known}, \mathit{Frontier})\mathbf{P}(\mathrm{Pit}_{1,3}, \mathit{Known}, \mathit{Frontier}) \\
&\propto \sum_{\mathit{Frontier}} \mathbf{P}(\mathit{Percept} \mid \mathrm{Pit}_{1,3}, \mathit{Known}, \mathit{Frontier})\mathbf{P}(\mathrm{Pit}_{1,3})P(\mathit{Known})P(\mathit{Frontier}) \\
&\propto P(\mathit{Known})\mathbf{P}(P_{1,3}) \sum_{\mathit{Frontier}} \mathbf{P}(\mathit{Percept} \mid \mathrm{Pit}_{1,3}, \mathit{Known}, \mathit{Frontier})P(\mathit{Frontier}) \\
&\propto \mathbf{P}(P_{1,3}) \sum_{\mathit{Frontier}} \mathbf{P}(\mathit{Percept} \mid \mathrm{Pit}_{1,3}, \mathit{Known}, \mathit{Frontier})P(\mathit{Frontier})
\end{aligned}
$$

To evaluate $\mathbf{P}(\mathit{Percept} \mid \mathrm{Pit}_{1,3}, \mathit{Known}, \mathit{Frontier})$ and $P(\mathit{Frontier})$:



$\mathbf{P}(\mathit{Percept} \mid \mathrm{Pit}_{1,3}, \mathit{Known}, \mathit{Frontier})$	**1**	**1**	**1**	**1**	**1**
$P(\mathit{Frontier})$	0.2 x 0.2 = 0.04	0.2 x 0.8 = 0.16	0.8 x 0.2 = 0.16	0.2 x 0.2 = 0.04	0.2 x 0.8 = 0.16

Therefore
$\mathbf{P}(\mathrm{Pit}_{1,3} \mid \mathit{Known}, \mathit{Percept}) = (0.2(0.04 + 0.16 + 0.16), 0.8(0.04 + 0.16)) \approx (0.31, 0.69)$.

Summary of The Topic

The following are the main knowledge points covered:

- Probability theory is a calculus of belief.
- Sample space and belief measure
- Probability
- Conditional probability and its properties
  - Chain rule
  - Law of total probability
  - Baye's rule
- Probability distribution and CPT.
- Posterior probability $\propto$ Likelihood $\times$ Prior Probability
- Independence and conditional independence