Artificial Intelligence Foundation – JC3001

Lecture 20: Quantifying Uncertainty and Reasoning with Probabilities II

Prof. Aladdin Ayesh (aladdin.ayesh@abdn.ac.uk)

Dr. Binod Bhattarai (binod.bhattarai@abdn.ac.uk)

Dr. Gideon Ogunniye, (g.ogunniye@abdn.ac.uk)

September 2025

Material adapted from:

Russell and Norvig (AIMA Book): Chapters 12 and 13

Sebastian Thrun — Stanford University / Udacity

Course Progression

- Part 1: Introduction
  1. Introduction to AI ✓
  2. Agents ✓
- Part 2: Problem-solving
  1. Search 1: Uninformed Search ✓
  2. Search 2: Heuristic Search ✓
  3. Search 3: Local Search ✓
  4. Search 4: Adversarial Search ✓
- Part 3: Reasoning and Uncertainty
  1. Reasoning 1: Constraint Satisfaction ✓
  2. Reasoning 2: Logic and Inference ✓
  3. **Probabilistic Reasoning 1: BNs**
  4. Probabilistic Reasoning 2: HMMs
- Part 4: Planning
  1. Planning 1: Intro and Formalism
  2. Planning 2: Algos and Heuristics
  3. Planning 3: Hierarchical Planning
  4. Planning 4: Stochastic Planning
- Part 5: Learning
  1. Learning 1: Intro to ML
  2. Learning 2: Regression
  3. Learning 3: Neural Networks
  4. Learning 4: Reinforcement Learning
- Part 6: Conclusion
  1. Ethical Issues in AI
  2. Conclusions and Discussion

Objectives

- Probability Theory ✓
- Probabilistic Inference
- Bayes Rule
- Bayesian Networks
  - Graphical Semantics
  - Inference

Outline

- Inference in the Joint Probability Distribution
- Bayes Rule

Inference in the Joint Probability Table

- If you have a toothache, you will go to the dentist, and they will check for cavities by testing whether a probe catches on the tooth.
- It is possible to have a toothache without a cavity, a cavity without the probe catching, etc.

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- $P(cavity \lor toothache) = 0.28$
- $P(cavity) = 0.2$ This is the marginal probability of cavity.

Given a joint distribution over all variables X, we can find the probability of any set of query variables Y by marginalising over the remaining variables $Z = X - Y$:

$$P(Y) = \sum_{z \in Z} P(Y, Z = z)$$

Example:

$$\begin{aligned} P(Cavity) &= P(Cavity, toothache, catch) + P(Cavity, toothache, \neg catch) \\ &+ P(Cavity, \neg toothache, catch) + P(Cavity, \neg toothache, \neg catch) \\ &= \langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle + \langle 0.072, 0.144 \rangle + \langle 0.008, 0.576 \rangle \\ &= \langle 0.2, 0.8 \rangle \end{aligned}$$

We can rewrite the marginalisation rule using the product rule, so that we get:

$$\begin{aligned} P(Y) &= \sum_{z \in Z} P(Y, Z = z) \\ &= \sum_{z \in Z} P(Y \mid z) P(z) \end{aligned}$$

This yields a rule called **conditioning**, which also allows us to compute probabilities, given evidence:

$$P(Y \mid z) = \frac{P(Y, z)}{P(z)}$$

$$P(Cav = true \mid Cat = true) = \frac{P(Cav = true, Cat = true)}{P(Cat = true)}$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(Cav = true \mid Cat = true) = \frac{P(Cav = true, Cat = true)}{P(Cat = true)}$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(Cav = true \mid Cat = true) = \frac{0.18}{P(Cat = true)}$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(Cav = true \mid Cat = true) = \frac{0.18}{P(Cat = true)}$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(Cav = true \mid Cat = true) = \frac{0.18}{0.34}$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(Cav = true \mid Cat = true) = 0.53$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

We now know that $P(Cav = true \mid Cat = true) = 0.53$

Just to check, we can compute $P(Cav = false \mid Cat = true)$

$$P(Cav = false \mid Cat = true) = \frac{P(Cav = false, Cat = true)}{P(Cat = true)}$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

We now know that $P(Cav = true \mid Cat = true) = 0.53$

Just to check, we can compute $P(Cav = false \mid Cat = true)$

$$P(Cav = false \mid Cat = true) = \frac{P(Cav = false, Cat = true)}{P(Cat = true)}$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

We now know that $P(Cav = true \mid Cat = true) = 0.53$

Just to check, we can compute $P(Cav = false \mid Cat = true)$

$$P(Cav = false \mid Cat = true) = \frac{0.16}{P(Cat = true)}$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

We now know that $P(Cav = true \mid Cat = true) = 0.53$
Just to check, we can compute $P(Cav = false \mid Cat = true)$

$$P(Cav = false \mid Cat = true) = \frac{0.16}{P(Cat = true)}$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

We now know that $P(Cav = true \mid Cat = true) = 0.53$
Just to check, we can compute $P(Cav = false \mid Cat = true)$

$$P(Cav = false \mid Cat = true) = \frac{0.16}{0.34}$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

We now know that $P(Cav = true \mid Cat = true) = 0.53$
Just to check, we can compute $P(Cav = false \mid Cat = true)$

$$P(Cav = false \mid Cat = true) = 0.47$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

We now know that $P(Cav = true \mid Cat = true) = 0.53$
Just to check, we can compute $P(Cav = false \mid Cat = true)$

$$P(Cav = false \mid Cat = true) = 0.47$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Note that the denominator is the same for both $P(cavity|catch)$ and $P(\neg cavity|catch)$

We now know that $P(Cav = true \mid Cat = true) = 0.53$
Just to check, we can compute $P(Cav = false \mid Cat = true)$

$$P(Cav = false \mid Cat = true) = 0.47$$

	toothache		¬ toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Note that the denominator is the same for both $P(cavity|catch)$ and $P(\neg cavity|catch)$
Recall that $\sum_{y \in Y} P(y) = 1$, so $\sum_{y \in Y} P(y \mid z) = 1$

If we have a joint distribution over all variables, then given evidence variables $E = e$, we can find the probability of any query variable $X = x$

$$P(X = x \mid E = e) = \frac{P(X = x, E = e)}{P(E = e)}$$

$$P(X = x, E = e) = \sum_{y \in Y} P(X = x, E = e, Y = y)$$

$$P(E = e) = \sum_{y \in Y} P(E = e, Y = y)$$

If we have a joint distribution over all variables, then given evidence variables $E = e$, we can find the probability of any query variable $X = x$

$$P(X = x \mid E = e) = \frac{P(X = x, E = e)}{P(E = e)}$$

$$P(X = x, E = e) = \sum_{y \in Y} P(X = x, E = e, Y = y)$$ Y are all variables not in X or E

$$P(E = e) = \sum_{y \in Y} P(E = e, Y = y)$$ Y are all variables not in E

If we have a joint distribution over all variables, then given evidence variables $E = e$, we can find the probability of any query variable $X = x$

$$P(X = x \mid E = e) = \frac{P(X = x, E = e)}{P(E = e)}$$

$$P(X = x, E = e) = \sum_{y \in Y} P(X = x, E = e, Y = y) \quad Y \text{ are all variables not in } X \text{ or } E$$

$$P(E = e) = \sum_{y \in Y} P(E = e, Y = y) \quad Y \text{ are all variables not in } E$$

Since we know that $P(X \mid e)$ must sum to 1, we can consider $\frac{1}{P(E=e)}$ to be a **normalising constant** α, so $P(X \mid e) = \alpha P(X, e) = \alpha \sum_{y \in Y} P(X, e, y)$

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- We can compute the probability of any set of variables Y by summing the probabilities of all possible combinations of unrelated variables Z:
  $P(Y) = \sum_{z \in Z} P(Y, z)$ — **Marginalisation**
  Example: $P(cavity) = \sum_{z \in \{Catch, Toothache\}} P(Cavity, z)$
- $P(Y) = \sum_{z} P(Y \mid z) P(z)$ — **Conditioning**

- Inference in the Joint Probability Distribution
- Bayes Rule

The Cookie Problem

- Suppose there are two bowls of cookies:
  - **Bowl 1:** 30 **vanilla** cookies and 10 **chocolate** cookies
  - **Bowl 2:** 20 **vanilla** cookies and 20 **chocolate** cookies

The Cookie Problem

- Suppose there are two bowls of cookies:
  - **Bowl 1:** 30 **vanilla** cookies and 10 **chocolate** cookies
  - **Bowl 2:** 20 **vanilla** cookies and 20 **chocolate** cookies
- Suppose you pick a random bowl and cookie (without looking)
  The cookie is vanilla.
  What is the **probability** that it came from **Bowl 1**?

- Suppose there are two bowls of cookies:
  - **Bowl 1:** 30 **vanilla** cookies and 10 **chocolate** cookies
  - **Bowl 2:** 20 **vanilla** cookies and 20 **chocolate** cookies
- Suppose you pick a random bowl and cookie (without looking)
  The cookie is vanilla.
  What is the **probability** that it came from **Bowl 1**?
- This is conditional probability; we want $P(Bowl1 \mid vanilla)$
  but this is not obvious with what we know.

- Suppose there are two bowls of cookies:
  - **Bowl 1:** 30 **vanilla** cookies and 10 **chocolate** cookies
  - **Bowl 2:** 20 **vanilla** cookies and 20 **chocolate** cookies
- Suppose you pick a random bowl and cookie (without looking)
  The cookie is vanilla.
  What is the **probability** that it came from **Bowl 1**?
- This is conditional probability; we want $P(Bowl1 \mid vanilla)$
  but this is not obvious with what we know.
- If we wanted to know $P(vanilla \mid Bowl1) = 3/4$
  Sadly $P(a \mid b)$ **is not** the same as $P(b \mid a)$

Since $P(a \wedge b) = P(a \mid b)P(b)$, Since $P(a \wedge b) = P(a \mid b)P(b)$, $P(a \wedge b) = P(b \mid a)P(a)$

Since $P(a \wedge b) = P(a \mid b)P(b)$, $P(a \wedge b) = P(b \mid a)P(a)$ then

$$P(b \mid a)P(a) = P(a \mid b)P(b)$$

Since $P(a \wedge b) = P(a \mid b)P(b)$, $P(a \wedge b) = P(b \mid a)P(a)$ then

$$P(b \mid a)P(a) = P(a \mid b)P(b)$$

and

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

Since $P(a \wedge b) = P(a \mid b)P(b)$, $P(a \wedge b) = P(b \mid a)P(a)$ then

$$P(b \mid a)P(a) = P(a \mid b)P(b)$$

and

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

so

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

And that's Bayes's Rule!

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

- Back to the cookie example, consider B_1 for the hypothesis that the cookie came from Bowl 1, and V for vanilla cookie.
  Plugging in the values into the theorem, we get:

$$P(B_1|V) = \frac{P(V \mid B_1)P(B_1)}{P(V)}$$

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

- Back to the cookie example, consider B_1 for the hypothesis that the cookie came from Bowl 1, and V for vanilla cookie.
  Plugging in the values into the theorem, we get:

$$P(B_1|V) = \frac{P(V \mid B_1)P(B_1)}{P(V)}$$

- The terms on the right are:
  - $P(B_1)$: Unconditional probability of choosing Bowl 1, so $P(B_1) = 1/2$
  - $P(V \mid B_1)$: Probability of getting vanilla from Bowl 1, which is $3/4$
  - $P(V)$: Probability of drawing a vanilla cookie from either bowl, or $5/8$

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

- Back to the cookie example, consider B_1 for the hypothesis that the cookie came from Bowl 1, and V for vanilla cookie.
  Plugging in the values into the theorem, we get:

$$P(B_1|V) = \frac{P(V \mid B_1)P(B_1)}{P(V)} = \frac{(1/2)(3/4)}{5/8}$$

- The terms on the right are:
  - $P(B_1)$: Unconditional probability of choosing Bowl 1, so $P(B_1) = 1/2$
  - $P(V \mid B_1)$: Probability of getting vanilla from Bowl 1, which is $3/4$
  - $P(V)$: Probability of drawing a vanilla cookie from either bowl, or $5/8$

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

- Back to the cookie example, consider B_1 for the hypothesis that the cookie came from Bowl 1, and V for vanilla cookie.
  Plugging in the values into the theorem, we get:

$$P(B_1|V) = \frac{P(V \mid B_1)P(B_1)}{P(V)} = \frac{(1/2)(3/4)}{5/8} = \frac{3}{5}$$

- The terms on the right are:
  - $P(B_1)$: Unconditional probability of choosing Bowl 1, so $P(B_1) = 1/2$
  - $P(V \mid B_1)$: Probability of getting vanilla from Bowl 1, which is $3/4$
  - $P(V)$: Probability of drawing a vanilla cookie from either bowl, or $5/8$

- Another way of thinking about Bayes's Theorem:
  Update the probability of a hypothesis H, in light of new data D

Diachronic Interpretation

- Another way of thinking about Bayes's Theorem:
  Update the probability of a hypothesis H, in light of new data D
- This is called "diachronic interpretation"
  (i.e., probability changing over time, as we see data)

$$P(H|D) = \frac{P(H)P(D \mid H)}{P(D)}$$

- In this interpretation, each term has a name:
  - $P(H)$ is the probability of the hypothesis, before any data, the **prior** probability;
  - $P(H \mid D)$ is what we want to compute, i.e., the hypothesis probability after the data, or **posterior**;
  - $P(D \mid H)$ is the probability of the data under the hypothesis, or **likelihood**; and
  - $P(D)$ is the probability of the data under any hypothesis, called the **normalizing constant**.

- The prior can be computed from background information (when it is not subjective)
- The likelihood is usually the easiest part to compute (e.g., we found the probability of a vanilla cookie by counting)
- The normalizing constant is often tricky, since it means the independent probability of the data

- The prior can be computed from background information (when it is not subjective)
- The likelihood is usually the easiest part to compute (e.g., we found the probability of a vanilla cookie by counting)
- The normalizing constant is often tricky, since it means the independent probability of the data
  so, in most cases we simplify by specifying only hypotheses that are:

- The prior can be computed from background information (when it is not subjective)
- The likelihood is usually the easiest part to compute (e.g., we found the probability of a vanilla cookie by counting)
- The normalizing constant is often tricky, since it means the independent probability of the data

  so, in most cases we simplify by specifying only hypotheses that are:
  - **Mutually exclusive:** at most one hypothesis can be true at a time; and

- The prior can be computed from background information (when it is not subjective)
- The likelihood is usually the easiest part to compute (e.g., we found the probability of a vanilla cookie by counting)
- The normalizing constant is often tricky, since it means the independent probability of the data

  so, in most cases we simplify by specifying only hypotheses that are:
  - **Mutually exclusive:** at most one hypothesis can be true at a time; and
  - **Collectively exhaustive:** there are no other possibilities, and at least one must be true.

- The prior can be computed from background information (when it is not subjective)
- The likelihood is usually the easiest part to compute (e.g., we found the probability of a vanilla cookie by counting)
- The normalizing constant is often tricky, since it means the independent probability of the data
  so, in most cases we simplify by specifying only hypotheses that are:
  - **Mutually exclusive:** at most one hypothesis can be true at a time; and
  - **Collectively exhaustive:** there are no other possibilities, and at least one must be true.
- In this case, we can compute $P(D)$ using the law of **total probability**
  If there are two exclusive hypotheses, you can add up the probabilities:

$$P(D) = P(H_1)P(D|H_1) + P(H_2)P(D|H_2)$$

- The prior can be computed from background information (when it is not subjective)
- The likelihood is usually the easiest part to compute (e.g., we found the probability of a vanilla cookie by counting)
- The normalizing constant is often tricky, since it means the independent probability of the data
  so, in most cases we simplify by specifying only hypotheses that are:
  - **Mutually exclusive:** at most one hypothesis can be true at a time; and
  - **Collectively exhaustive:** there are no other possibilities, and at least one must be true.
- In this case, we can compute $P(D)$ using the law of **total probability**
  If there are two exclusive hypotheses, you can add up the probabilities:

$$P(D) = P(1/2)P(3/4) + P(1/2)P(1/2) = 5/8$$

Prior Probability Likelihood

$$P(H|D) = \frac{P(H)P(D|H)}{P(D)}$$

Posterior Probability Evidence

- We can use Bayes rule to reduce the size of our joint distribution table by relying on conditional independence

Conditional Independence and Simplification

- We can work out the probability of any proposition from the joint probability distribution table.
- Given n variables, we have 2^n entries in the table.
- Independence allows us to reduce this table size.

$$P(toothache, catch, cavity, coin = heads) = P(toothache, catch, cavity)P(coin = heads)$$

- Generally, if a and b are independent, $P(a \mid b) = P(a)$ and $P(a \wedge b) = P(a)P(b)$

Chain Rule

- Given the product rule: $P(a \wedge b) = P(a \mid b)P(b)$
- We can express an arbitrary joint probability via the chain rule:

$$P(X_1, \ldots X_n) = \prod_{i=1}^{n} P(X_i \mid X_{i-1}, \ldots X_1)$$

$$P(X_1, X_2, X_3, X_4) = P(X_4 \mid X_1, X_2, X_3)P(X_1, X_2, X_3)$$

- Note that $P(X_1, X_2) = P(X_1 \wedge X_2)$

- Given the product rule: $P(a \wedge b) = P(a \mid b)P(b)$
- We can express an arbitrary joint probability via the chain rule:

$$P(X_1, \dots X_n) = \prod_{i=1}^{n} P(X_i \mid X_{i-1}, \dots X_1)$$

$$P(X_1, X_2, X_3, X_4) = P(X_4 \mid X_1, X_2, X_3)P(X_3 \mid X_1, X_2)P(X_1, X_2)$$

- Note that $P(X_1, X_2) = P(X_1 \wedge X_2)$

- Given the product rule: $P(a \wedge b) = P(a \mid b)P(b)$
- We can express an arbitrary joint probability via the chain rule:

$$P(X_1, \dots X_n) = \prod_{i=1}^{n} P(X_i \mid X_{i-1}, \dots X_1)$$

$$P(X_1, X_2, X_3, X_4) = P(X_4 \mid X_1, X_2, X_3)P(X_3 \mid X_1, X_2)P(X_2 \mid X_1)P(X_1)$$

- Note that $P(X_1, X_2) = P(X_1 \wedge X_2)$

- Given that a cavity is known, there is some probability that a toothache will exist.
- Similarly, there is some probability that the probe will catch.
- The absence or presence of one does not affect the other.

$$\begin{aligned} P(Toothache, Catch, Cavity) &= P(Cavity) * P(Toothache, Catch \mid Cavity) \\ &= P(Cavity) * P(Catch \mid Cavity) * P(Toothache \mid Cavity) \end{aligned}$$

- So rather than having 7 independent elements, we have 1+2+2=5 elements.
- For n conditional independent elements, we have a total of $O(n)$ rather than $O(2^n)$ entries in the table. Common pattern:

$$P(Cause, Effect_1, \ldots, Effect_n) = P(Cause) \prod_i P(Effect_i \mid Cause)$$

To continue in the next session.