CSC 384
Introduction to Artificial Intelligence

Uncertainty 1
Introduction and Probability Rules

Alice Gao and Randy Hickey

Winter 2023

Learning Goals

By the end of this lecture, you should be able to

1. Calculate a prior/marginal/joint probability using the sum rule, the product rule, the chain rule, and the Bayes' rule.
2. Explain which rule is useful for which type of task.

Outline

- Motivation and Probability Review
- Sum Rule and Product Rule
- Chain Rule and Bayes Rule

Questions

- Why does an agent need to handle uncertainty?
- What are the Frequentists v.s. Bayesian views of probability?
- Sum rule:

$$P(A \wedge B) = P(A \wedge B \wedge c) + P(A \wedge B \wedge \neg c)$$

- Product rule:

$$P(A|B) = P(A \wedge B)/P(B)$$

- Chain rule:

$$P(X_3 \wedge X_2 \wedge X_1) = P(X_3|X_2 \wedge X_1)P(X_2|X_1)P(X_1)$$

- Bayes rule:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

MOTIVATION AND PROBABILITY REVIEW

Why handle uncertainty?

- Why does an agent need to handle uncertainty?
  - An agent may not observe everything in the world.
  - An action may not have its intended consequences.

- An agent needs to
  - Reason about their uncertainty.
  - Decide based on their uncertainty.

Probability – the formal measure of uncertainty

Frequentists	Bayesians
• Probability is **objective**.	• Probability is **subjective**, i.e. degrees of belief
• Compute probabilities by counting frequency of events.	• Start with **prior** beliefs and **update** beliefs based on new evidence.
• E.g. Probs of heads for coin = probs of heads in history	• E.g. Probs of heads for coin = probs of heads in agent's previous experience.

Random Variable

A random variable

- Has a **domain** of possible values, and
- Has an associated **probability distribution** – a function from the domain to [0,1].

Example:

- Random variable: The alarm is sounding.
- Domain: {true, false}.
- P(The alarm is sounding = true) = 0.1.
- P(The alarm is sounding = false) = 0.9.

Notation for Binary Random Variables

- Capital letter denotes an unassigned random variable.
- Small letter denotes a variable and its value.

Example:

- Let X be a Boolean random variable.
- $P(x)$ denotes $P(X = true)$.
- $P(\neg x)$ denotes $P(X = false)$.
- $P(X)$ denotes the distribution $\{P(x), P(\neg x)\}$.

Axioms of Probability

Let A and B be Boolean random variables.

1. Every probability is between 0 and 1.

$$0 \leq P(A) \leq 1$$

The [0,1] interval is an arbitrary choice.

2. Necessarily true propositions have probability 1.

Necessarily false propositions have probability 0.

$$P(true) = 1, P(false) = 0$$

3. The inclusion-exclusion principle:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

Prior and Posterior Probabilities

$$P(X)$$

- **Prior** or **unconditional** probability. (marginal)
- Likelihood of X in the absence of any information.
- Based on background info or prior experience.

$$P(X|Y)$$

- **Posterior** or **conditional** probability.
- Likelihood of X given Y.
- Based on Y as evidence.

The Holmes Scenario

Mr. Holmes lives in a high crime area and therefore has installed a burglar alarm. He relies on his neighbors to phone him when they hear the alarm sound. Mr. Holmes has two neighbors, Dr. Watson and Mrs. Gibbon.

Unfortunately, his neighbors are not entirely reliable. Dr. Watson is known to be a tasteless practical joker and Mrs. Gibbon, while more reliable in general, has occasional drinking problems.

Mr. Holmes also knows from reading the instruction manual of his alarm system that the device is sensitive to earthquakes and can be triggered by one accidentally. He realizes that if an earthquake has occurred, it would surely be on the radio news.

Modeling the Holmes Scenario

What are the random variables?

How many probabilities are there in the joint probability distribution?

Modeling the Holmes Scenario

What are the random variables?

- B: A burglary is happening.
- A: The alarm is ringing.
- W: Dr. Watson is calling.
- G: Mrs. Gibbon is calling.
- E: An earthquake is happening.
- R: A report of an earthquake is on the news.

How many probabilities are there in the joint probability distribution?

Modeling the Holmes Scenario

What are the random variables?

- B: A burglary is happening.
- A: The alarm is ringing.
- W: Dr. Watson is calling.
- G: Mrs. Gibbon is calling.
- E: An earthquake is happening.
- R: A report of an earthquake is on the news.

How many probabilities are there in the joint probability distribution?

- $2^6 = 64$ probabilities.

SUM RULE AND PRODUCT RULE

Inference using the Joint Distribution

Given a joint distribution, how do we compute

- The probability over a subset of the variables?
- A conditional probability?

Two useful rules:

- The sum rule
- The product rule

Example of a Joint Distribution

	a			$\neg a$	
	g	$\neg g$		g	$\neg g$
w	0.032	0.048	w	0.036	0.324
$\neg w$	0.008	0.012	$\neg w$	0.054	0.486

- A: The alarm is ringing.
- W: Dr. Watson is calling.
- G: Mrs. Gibbon is calling.

The Sum Rule

Q: Given $P(A \wedge B \wedge C)$,

how can we calculate $P(A \wedge B)$ and $P(A)$?

A: Sum out every variable that we do not care about.

Calculate $P(A \wedge B)$ by summing out C:

$$P(A \wedge B) = P(A \wedge B \wedge c) + P(A \wedge B \wedge \neg c)$$

Calculate $P(A)$ by further summing out B:

$$P(A) = P(A \wedge b) + P(A \wedge \neg b)$$

Let's Apply the Sum Rule

Q1: What is probability that
the alarm is NOT going and Dr. Watson is calling?

(A) 0.36 (B) 0.46 (C) 0.56 (D) 0.66 (E) 0.76

	a			$\neg a$	
	g	$\neg g$		g	$\neg g$
w	0.032	0.048	w	0.036	0.324
$\neg w$	0.008	0.012	$\neg w$	0.054	0.486

Let's Apply the Sum Rule

Q1: What is probability that

the alarm is NOT going and Dr. Watson is calling?

(A) 0.36 (B) 0.46 (C) 0.56 (D) 0.66 (E) 0.76

A: $P(\neg a \wedge w) = P(\neg a \wedge w \wedge g) + P(\neg a \wedge w \wedge \neg g)$

$$= 0.036 + 0.324 = 0.36$$

	a			$\neg a$	
	g	$\neg g$		g	$\neg g$
w	0.032	0.048	w	0.036	0.324
$\neg w$	0.008	0.012	$\neg w$	0.054	0.486

Let's Apply the Sum Rule

Q2: What is probability that
the alarm is going and Mrs. Gibbon is NOT calling?

(A) 0.05 (B) 0.06 (C) 0.07 (D) 0.08 (E) 0.09

	a			$\neg a$	
	g	$\neg g$		g	$\neg g$
w	0.032	0.048	w	0.036	0.324
$\neg w$	0.008	0.012	$\neg w$	0.054	0.486

Let's Apply the Sum Rule

Q2: What is probability that

the alarm is going and Mrs. Gibbon is NOT calling?

(A) 0.05 **(B) 0.06** (C) 0.07 (D) 0.08 (E) 0.09

A: $P(a \land \neg g) = P(a \land w \land \neg g) + P(a \land \neg w \land \neg g)$

$$= 0.048 + 0.012 = 0.06$$

	a			$\neg a$	
	g	$\neg g$		g	$\neg g$
w	0.032	0.048	w	0.036	0.324
$\neg w$	0.008	0.012	$\neg w$	0.054	0.486

Let's Apply the Sum Rule

Q3: What is probability that **the alarm is NOT going**?

(A) 0.1 (B) 0.3 (C) 0.5 (D) 0.7 (E) 0.9

	a			$\neg a$	
	g	$\neg g$		g	$\neg g$
w	0.032	0.048	w	0.036	0.324
$\neg w$	0.008	0.012	$\neg w$	0.054	0.486

Let's Apply the Sum Rule

Q3: What is probability that **the alarm is NOT going**?

(A) 0.1 (B) 0.3 (C) 0.5 (D) 0.7 **(E) 0.9**

Answer: $P(\neg a) = 0.036 + 0.324 + 0.054 + 0.486 = 0.9$

	a			$\neg a$	
	g	$\neg g$		g	$\neg g$
w	0.032	0.048	w	0.036	0.324
$\neg w$	0.008	0.012	$\neg w$	0.054	0.486

The Product Rule

Given $P(A \wedge B)$ and $P(B)$,

how can we calculate $P(A|B)$?

The product rule:

$$P(A \wedge B) = P(A|B)P(B), \text{ or}$$
$$P(A|B) = P(A \wedge B)/P(B)$$

Convert between joint and conditional probabilities.

Let's Apply the Product Rule

Q1: What is probability that

Dr. Watson is calling given that the alarm is NOT going?

(A) 0.2 (B) 0.4 (C) 0.6 (D) 0.8 (E) 1.0

$$P(\neg a \wedge w) = 0.36, P(a \wedge \neg g) = 0.06, P(\neg a) = 0.9$$

	a			$\neg a$	
	g	$\neg g$		g	$\neg g$
w	0.032	0.048	w	0.036	0.324
$\neg w$	0.008	0.012	$\neg w$	0.054	0.486

Let's Apply the Product Rule

Q1: What is probability that

Dr. Watson is calling given that the alarm is NOT going?

(A) 0.2 **(B) 0.4** (C) 0.6 (D) 0.8 (E) 1.0

Answer: $P(\neg a \land w) / P(\neg a) = 0.36/0.9 = 0.4$

$$P(\neg a \land w) = 0.36, P(a \land \neg g) = 0.06, P(\neg a) = 0.9$$

	a			$\neg a$	
	g	$\neg g$		g	$\neg g$
w	0.032	0.048	w	0.036	0.324
$\neg w$	0.008	0.012	$\neg w$	0.054	0.486

Let's Apply the Product Rule

Q2: What is probability that

Mrs. Gibbon is NOT calling given that the alarm is going?

(A) 0.2 (B) 0.4 (C) 0.6 (D) 0.8 (E) 1.0

$$P(\neg a \land w) = 0.36, P(a \land \neg g) = 0.06, P(\neg a) = 0.9$$

	a			$\neg a$	
	g	$\neg g$		g	$\neg g$
w	0.032	0.048	w	0.036	0.324
$\neg w$	0.008	0.012	$\neg w$	0.054	0.486

Let's Apply the Product Rule

Q2: What is probability that

Mrs. Gibbon is NOT calling given that the alarm is going?

(A) 0.2 (B) 0.4 **(C) 0.6** (D) 0.8 (E) 1.0

Answer:

$$P(\neg g|a) = P(a \wedge \neg g)/P(a) = 0.06/(1-0.9) = 0.6$$

$$P(\neg a \wedge w) = 0.36, P(a \wedge \neg g) = 0.06, P(\neg a) = 0.9$$

	a			$\neg a$	
	g	$\neg g$		g	$\neg g$
w	0.032	0.048	w	0.036	0.324
$\neg w$	0.008	0.012	$\neg w$	0.054	0.486

CHAIN RULE AND BAYES RULE

Probabilities for the Holmes Scenario

$$P(a) = 0.1, P(\neg a) = 0.9$$

$$P(w|a) = 0.8, P(w|\neg a) = 0.4$$
$$P(g|a) = 0.4, P(g|\neg a) = 0.1$$

$$P(w|a \wedge g) = 0.8, P(w|a \wedge \neg g) = 0.8$$
$$P(w|\neg a \wedge g) = 0.4, P(w|\neg a \wedge \neg g) = 0.4$$

$$P(g|a \wedge w) = 0.4, P(g|a \wedge \neg w) = 0.4$$
$$P(g|\neg a \wedge w) = 0.1, P(g|\neg a \wedge \neg w) = 0.1$$

The Chain Rule

- Calculate a joint probability given the marginal and the conditional probabilities.

  Any order of variables leads to the same result.

- For three variables: $= P(X_2 | X_1 \wedge X_3) * P(X_1 | X_3) * P(X_3)$

  $$P(X_3 \wedge X_2 \wedge X_1) = P(X_3 | X_2 \wedge X_1) P(X_2 | X_1) P(X_1)$$

- For any number of variables:

$$P(X_n \wedge X_{n-1} \wedge \cdots \wedge X_2 \wedge X_1)$$

$$= \prod_{i=1}^{n} P(X_i | X_{i-1} \wedge \cdots \wedge X_2 \wedge X_1)$$

$$= P(X_n | X_{n-1} \wedge \cdots X_2 \wedge X_1) \ldots P(X_2 | X_1) P(X_1)$$

Let's Apply the Chain Rule

Q1: What is probability that **the alarm is going, Dr. Watson is calling and Mrs. Gibbon is NOT calling**?

(A) 0.048

(B) 0.058

(C) 0.068

(D) 0.078

(E) 0.088

$$P(a) = 0.1, P(\neg a) = 0.9$$

$$P(w|a) = 0.8, P(w|\neg a) = 0.4$$
$$P(g|a) = 0.4, P(g|\neg a) = 0.1$$

$$P(w|a \wedge g) = 0.8, P(w|a \wedge \neg g) = 0.8$$
$$P(w|\neg a \wedge g) = 0.4, P(w|\neg a \wedge \neg g) = 0.4$$

$$P(g|a \wedge w) = 0.4, P(g|a \wedge \neg w) = 0.4$$
$$P(g|\neg a \wedge w) = 0.1, P(g|\neg a \wedge \neg w) = 0.1$$

Let's Apply the Chain Rule

Q1: What is probability that **the alarm is going, Dr. Watson is calling and Mrs. Gibbon is NOT calling**?

(A) 0.048

(B) 0.058

(C) 0.068

(D) 0.078

(E) 0.088

Answer:

$$P(a) * P(w|a) * P(\neg g|a \wedge w) = 0.1 * 0.8 * 0.6 = 0.048$$

$$P(a) = 0.1, P(\neg a) = 0.9$$

$$P(w|a) = 0.8, P(w|\neg a) = 0.4$$
$$P(g|a) = 0.4, P(g|\neg a) = 0.1$$

$$P(w|a \wedge g) = 0.8, P(w|a \wedge \neg g) = 0.8$$
$$P(w|\neg a \wedge g) = 0.4, P(w|\neg a \wedge \neg g) = 0.4$$

$$P(g|a \wedge w) = 0.4, P(g|a \wedge \neg w) = 0.4$$
$$P(g|\neg a \wedge w) = 0.1, P(g|\neg a \wedge \neg w) = 0.1$$

Let's Apply the Chain Rule

Q2: What is probability that **the alarm is NOT going, Dr. Watson is NOT calling and Mrs. Gibbon is NOT calling**?

(A) 0.476

(B) 0.486

(C) 0.496

(D) 0.506

(E) 0.606

$$P(a) = 0.1, P(\neg a) = 0.9$$

$$P(w|a) = 0.8, P(w|\neg a) = 0.4$$
$$P(g|a) = 0.4, P(g|\neg a) = 0.1$$

$$P(w|a \wedge g) = 0.8, P(w|a \wedge \neg g) = 0.8$$
$$P(w|\neg a \wedge g) = 0.4, P(w|\neg a \wedge \neg g) = 0.4$$

$$P(g|a \wedge w) = 0.4, P(g|a \wedge \neg w) = 0.4$$
$$P(g|\neg a \wedge w) = 0.1, P(g|\neg a \wedge \neg w) = 0.1$$

Let's Apply the Chain Rule

Q2: What is probability that **the alarm is NOT going, Dr. Watson is NOT calling and Mrs. Gibbon is NOT calling**?

(A) 0.476

(B) 0.486

(C) 0.496

(D) 0.506

(E) 0.606

Answer:

$$P(\neg a) * P(\neg w \mid \neg a) * P(\neg g \mid \neg a \land \neg w) = 0.9 * 0.6 * 0.9 = 0.486$$

$$P(a) = 0.1, P(\neg a) = 0.9$$

$$P(w|a) = 0.8, P(w|\neg a) = 0.4$$
$$P(g|a) = 0.4, P(g|\neg a) = 0.1$$

$$P(w|a \land g) = 0.8, P(w|a \land \neg g) = 0.8$$
$$P(w|\neg a \land g) = 0.4, P(w|\neg a \land \neg g) = 0.4$$

$$P(g|a \land w) = 0.4, P(g|a \land \neg w) = 0.4$$
$$P(g|\neg a \land w) = 0.1, P(g|\neg a \land \neg w) = 0.1$$

Bayes' Rule

- Flipping a conditional probability
- Example: Medical Diagnosis
  - Suppose that we know P(symptom | disease),
  - But we want P(disease | symptom)

- Bayes' Rule

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Bayes' Rule

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

- Could you derive the Bayes' rule using the product rule?
- Do we need to know $P(Y)$?

Bayes' Rule

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

$$P(X|Y) = \underbrace{\alpha}_{\text{normalization constant.}} P(Y|X)P(X) \text{ where } \alpha = \frac{1}{P(Y)}.$$

- Could you derive the Bayes' rule using the product rule?

$$P(X|Y)P(Y) = P(X \wedge Y) = P(Y|X)P(X)$$

- Do we need to know $P(Y)$?
  - No.
  - $P(Y) = P(Y|x)P(x) + P(Y|\neg x)P(\neg x)$
  - (1) Calculate $P(Y|x)P(x)$ and $P(Y|\neg x)P(\neg x)$,
  - (2) Normalize them (divide by the sum of the two terms).

Let's Apply the Bayes' Rule

Q1: What is the probability that

the alarm is NOT going given that Dr. Watson is calling?

$\neg a$ w

(A) 0.518 (B) 0.618 (C) 0.718 (D) 0.818 (E) 0.918

$$P(\neg a | w) = \frac{P(\neg a \wedge w)}{P(w)} = \frac{P(\neg a \wedge w)}{P(a \wedge w) + P(\neg a \wedge w)}$$

$$= \frac{P(w|\neg a)\, P(\neg a)}{P(w|a)\, P(a) + P(w|\neg a)\, P(\neg a)}$$

$$P(a) = 0.1, P(\neg a) = 0.9$$
$$P(w|a) = 0.8, P(w|\neg a) = 0.4$$

Let's Apply the Bayes' Rule

Q1: What is the probability that

the alarm is NOT going given that Dr. Watson is calling?

(A) 0.518 (B) 0.618 (C) 0.718 **(D) 0.818** (E) 0.918

A: $P(\neg a|w) = P(w|\neg a)P(\neg a)/P(w)$

$$= 0.4 * 0.9/(0.4 * 0.9 + 0.1 * 0.8) = 0.818$$

$$P(a) = 0.1, P(\neg a) = 0.9$$
$$P(w|a) = 0.8, P(w|\neg a) = 0.4$$

Let's Apply the Bayes' Rule

Q2: What is the probability that

the alarm is going given that Mrs. Gibbon is NOT calling?

(A) 0.049 (B) 0.059 (C) 0.069 (D) 0.079 (E) 0.089

$$P(a) = 0.1, P(\neg a) = 0.9$$
$$P(g|a) = 0.4, P(g|\neg a) = 0.1$$

Let's Apply the Bayes' Rule

Q2: What is the probability that

the alarm is going given that Mrs. Gibbon is NOT calling?

(A) 0.049 (B) 0.059 **(C) 0.069** (D) 0.079 (E) 0.089

A: $P(a \mid \neg g) = P(\neg g \mid a)P(a)/P(\neg g)$

$$= 0.1 * 0.6/(0.1 * 0.6 + 0.9 * 0.9) = 0.069$$

$$P(a) = 0.1, P(\neg a) = 0.9$$
$$P(g|a) = 0.4, P(g|\neg a) = 0.1$$

Summary

- Why does an agent need to handle uncertainty?
- What are the Frequentists v.s. Bayesian views of probability?
- What probability can we calculate with
  - Sum rule?
  - Product rule?
  - Chain rule?
  - Bayes' rule?