

Probability

Alice Gao

Lecture 10

Readings: RN 13.2 - 13.4. PM 8.1 - 8.2.

Outline

Learning Goals

Introduction to Probability Theory

Inferences Using the Joint Distribution

- The Sum Rule

- The Product Rule

Inferences using Prior and Conditional Probabilities

- The Chain Rule

- Bayes' Rule

Revisiting the Learning goals

Learning Goals

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

- ▶ Calculate prior, posterior, and joint probabilities using the sum rule, the product rule, the chain rule and Bayes' rule.

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.
- ▶ Make a decision based on their uncertainty.

Probability

- ▶ Probability is the formal measure of uncertainty.
- ▶ There are two camps: **Frequentists** and **Bayesians**.
- ▶ **Frequentists' view of probability:**
 - ▶ Frequentists view probability as **something objective**.
 - ▶ Compute probabilities by **counting the frequencies of events**.
- ▶ **Bayesians' view of probability:**
 - ▶ Bayesians view probability as something **subjective**.
 - ▶ Probabilities are **degrees of belief**.
 - ▶ We start with **prior** beliefs and **update** beliefs based on new evidence.

Random variable

A random variable

- ▶ Has a **domain** of possible values
- ▶ Has an associated **probability distribution**, which is a function from the domain of the random variable to $[0, 1]$.

Example:

- ▶ random variable: The alarm is going.
- ▶ domain: $\{\text{true}, \text{false}\}$
- ▶ $P(\text{The alarm is going} = \text{true}) = 0.1$
 $P(\text{The alarm is going} = \text{false}) = 0.9$

Shorthand notation

Let A a Boolean random variable.

- ▶ $P(A)$ denotes $P(A = \text{true})$.
- ▶ $P(\neg A)$ denotes $P(A = \text{false})$.

Axioms of Probability

Let A and B be Boolean random variables.

- ▶ Every probability is between 0 and 1.

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

- ▶ Necessarily true propositions have prob 1.
Necessarily false propositions have probability 0.

$$P(\text{true}) = 1, P(\text{false}) = 0$$

- ▶ The inclusion-exclusion principle:

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

These axioms limit the functions that can be considered as probability functions.

Joint Probability Distribution

- ▶ A **probabilistic model** contains a set of random variables.
- ▶ An **atomic event** assigns a value to every random variable in the model.
- ▶ A **joint probability distribution** assigns a probability to every atomic event.

Prior and Posterior Probabilities

$P(X)$:

- ▶ **prior** or **unconditional** probability
- ▶ Likelihood of X in the absence of any other information
- ▶ Based on the background information

$P(X|Y)$

- ▶ **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?

B: A Burglary is happening.

A: The alarm is going.

How many probabilities are there
in the joint probability distribution?

$2^6 = 64$ probabilities.

W: Dr Watson is calling.

G: Mrs Gibbon is calling.

E: Earthquake is happening.

*R: A report of an earthquake
is on the radio.*

Learning Goals

Introduction to Probability Theory

Inferences Using the Joint Distribution

The Sum Rule

The Product Rule

Inferences using Prior and Conditional Probabilities

Unconditional and Conditional Independence

Revisiting the Learning goals

Inference using the Joint Distribution

Given a joint probability distribution,
how do we compute

- ▶ the probability over a subset of the variables?
- ▶ a conditional probability?

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

Probability over a Subset of the Variables

Given a joint probability distribution,
we can compute the probability over a subset of the variables
using the sum rule.

We can sum out every variable that we do not care about.

Given a joint distribution over A , B , and C ,
we can calculate $P(A \wedge B)$ by summing out C :

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

we can calculate $P(A)$ by summing out B and C :

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

CQ: Probability over a subset of the variables

CQ: 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

$$P(\neg A \wedge W) = 0.36$$

$$\begin{aligned} P(\neg A \wedge W) &= P(\neg A \wedge W \wedge G) + P(\neg A \wedge W \wedge \neg G) \\ &= 0.036 + 0.324 = 0.36 \end{aligned}$$

CQ: Probability over a subset of the variables

CQ: 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

$$P(A \wedge \neg G) = 0.06.$$

$$\begin{aligned} P(A \wedge \neg G) &= P(A \wedge \neg G \wedge W) + P(A \wedge \neg G \wedge \neg W) \\ &= 0.048 + 0.012 = 0.06 \end{aligned}$$

CQ: Probability over a subset of the variables

CQ: 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

$$P(\neg A) = 0.9$$

$$\begin{aligned} P(\neg A) &= P(\neg A \wedge W \wedge G) + P(\neg A \wedge W \wedge \neg G) \\ &\quad + P(\neg A \wedge \neg W \wedge G) + P(\neg A \wedge \neg W \wedge \neg G) \\ &= 0.036 + 0.324 + 0.054 + 0.486 = 0.9 \end{aligned}$$

A Conditional Probability

Given a joint probability distribution,
how do we compute the probability one variable A
conditioned on knowing the value of another variable B?

We can use the product rule.

For example, how do we calculate $P(A|B)$
given a joint distribution over A, B, C?

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

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

CQ: A conditional probability

CQ: 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

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

$$P(\neg A \wedge W) = 0.36,$$

$$P(A \wedge \neg G) = 0.06,$$

$$P(\neg A) = 0.9.$$

$$P(W|\neg A) = 0.4$$

$$\begin{aligned} P(W|\neg A) &= \frac{P(W \wedge \neg A)}{P(\neg A)} \\ &= \frac{0.36}{0.9} = 0.4 \end{aligned}$$

CQ: A conditional probability

CQ: 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

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

$$P(\neg A \wedge W) = 0.36,$$

$$P(A \wedge \neg G) = 0.06,$$

$$P(\neg A) = 0.9.$$

$$\begin{aligned} P(\neg G|A) &= \frac{P(\neg G \wedge A)}{P(A)} \\ &= \frac{0.06}{1 - 0.9} = \frac{0.06}{0.1} = 0.6 \end{aligned}$$

$$P(\neg G|A) = 0.6$$

Learning Goals

Introduction to Probability Theory

Inferences Using the Joint Distribution

Inferences using Prior and Conditional Probabilities

The Chain Rule

Bayes' Rule

Unconditional and Conditional Independence

Revisiting the Learning goals

Inference Using the Prior and Conditional Probabilities

How do we

- ▶ calculate a probability over a subset of the variables?
- ▶ calculate a conditional probability?

The prior probabilities

$$P(A) = 0.1$$

The conditional probabilities

$$P(W|A) = 0.9$$

$$P(W|\neg A) = 0.4$$

$$P(G|A) = 0.3$$

$$P(G|\neg A) = 0.1$$

$$P(W|A \wedge G) = 0.9$$

$$P(W|A \wedge \neg G) = 0.9$$

$$P(W|\neg A \wedge G) = 0.4$$

$$P(W|\neg A \wedge \neg G) = 0.4$$

$$P(G|A \wedge W) = 0.3$$

$$P(G|A \wedge \neg W) = 0.3$$

$$P(G|\neg A \wedge W) = 0.1$$

$$P(G|\neg A \wedge \neg W) = 0.1$$

Calculate a Joint Probability Using the Chain Rule

For two variables (a.k.a. the product rule):

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

For three variables:

$$P(A \wedge B \wedge C) = P(A|B \wedge C) * P(B|C) * P(C)$$

For any number of variables:

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

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

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

CQ: Calculate a joint probability

$$\begin{aligned}P(A \wedge W \wedge \neg G) &= P(A) P(W|A) P(\neg G|A \wedge W) \\&= 0.1 * 0.9 * (1 - 0.3) = 0.063\end{aligned}$$

order: $A \rightarrow W \rightarrow \neg G$

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

(A) 0.060

(B) 0.061

(C) 0.062

(D) 0.063

(E) 0.064

$$P(A) = 0.1$$

$$P(W|A) = 0.9$$

$$P(W|A \wedge \neg G) = 0.9$$

$$P(G|A) = 0.3$$

$$P(G|A \wedge W) = 0.3$$

$$P(G|A \wedge W)$$

$$+ P(\neg G|A \wedge W)$$

$$= 1$$

$$\begin{aligned}P(A \wedge W \wedge \neg G) &= P(A) P(\neg G|A) P(W|A \wedge \neg G) \\&= 0.1 * (1 - 0.3) * 0.9 = 0.063\end{aligned}$$

order: $A \rightarrow \neg G \rightarrow W$

CQ: Calculate a joint probability

$$\begin{aligned}P(\neg A \wedge \neg W \wedge \neg G) &= P(\neg A) P(\neg W | \neg A) P(\neg G | \neg A \wedge \neg W) \\&= (1-0.1) * (1-0.4) * (1-0.1) = 0.486\end{aligned}$$

order : $\neg A \rightarrow \neg W \rightarrow \neg G$

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

(A) 0.486

(B) 0.586

(C) 0.686

(D) 0.786

(E) 0.886

$$P(A) = 0.1$$

$$P(W | \neg A) = 0.4$$

$$P(W | \neg A \wedge \neg G) = 0.4$$

$$P(G | \neg A) = 0.1$$

$$P(G | \neg A \wedge \neg W) = 0.1$$

$$\begin{aligned}P(\neg A \wedge \neg W \wedge \neg G) &= P(\neg A) P(\neg G | \neg A) P(\neg W | \neg A \wedge \neg G) \\&= (1-0.1) * (1-0.1) * (1-0.4) = 0.486\end{aligned}$$

order : $\neg A \rightarrow \neg G \rightarrow \neg W$

Flipping a Conditional Probability

Often you have causal knowledge:

- ▶ $P(\text{symptom} \mid \text{disease})$
- ▶ $P(\text{alarm} \mid \text{fire})$

...and you want to do evidential reasoning:

- ▶ $P(\text{disease} \mid \text{symptom})$
- ▶ $P(\text{fire} \mid \text{alarm})$.

Flipping a Conditional Probability using the Bayes' Rule

$$P(X \wedge Y) = \underbrace{P(X|Y) * P(Y)}_{\text{joint probability}} = \underbrace{P(Y|X) * P(X)}_{\text{joint probability}}$$

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

Definition (Bayes' rule)

$$P(X|Y) = \frac{P(Y|X) * P(X)}{P(Y)} \rightarrow \text{a normalization constant.}$$

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

① calculate $P(Y|X) P(X)$ and $P(Y|\neg X) P(\neg X)$

② normalize the two values so that they sum to 1.

CQ: Flipping a conditional probability

$$P(\neg A|W) = \frac{P(W|\neg A) P(\neg A)}{P(W)} = \frac{0.4 * (1 - 0.1)}{0.45} = 0.8$$

$$P(W) = P(W|A) P(A) + P(W|\neg A) P(\neg A) = 0.9 * 0.1$$

CQ: What is the probability that the alarm is NOT going given that Dr. Watson is calling? $+ 0.4 * 0.9 = 0.09 + 0.36 = 0.45$

(A) 0.6

(B) 0.7

(C) 0.8

(D) 0.9

(E) 1.0

$$P(A) = 0.1$$

$$P(W|A) = 0.9$$

$$P(W|\neg A) = 0.4$$

$$P(W|A) P(A) = 0.9 * 0.1 = 0.09$$

$$0.09 + 0.36 = 0.45$$

$$P(W|\neg A) P(\neg A) = 0.4 * 0.9 = 0.36$$

$$P(A|W) = \frac{0.09}{0.45} = 0.2$$

$$P(\neg A|W) = \frac{0.36}{0.45} = 0.8$$

CQ: Flipping a conditional probability

$$P(A|\neg G) = \frac{P(\neg G|A)P(A)}{P(\neg G|A)P(A) + P(\neg G|\neg A)P(\neg A)} = \frac{(1-0.3)0.1}{(1-0.3)0.1 + (1-0.1)(1-0.1)}$$

CQ: What is the probability that the alarm is going given that Mrs. Gibbon is NOT calling? $= 0.07/0.88 = 0.08$

(A) 0.04

(B) 0.05

(C) 0.06

(D) 0.07

(E) 0.08

$$P(A) = 0.1$$

$$P(G|A) = 0.3$$

$$P(G|\neg A) = 0.1$$

$$P(\neg G|A)P(A) = (1-0.3)0.1 = 0.07 \quad (\text{for } P(A|\neg G))$$

$$P(\neg G|\neg A)P(\neg A) = (1-0.1)(1-0.1) = 0.81 \quad (\text{for } P(\neg A|\neg G))$$

$$0.07 + 0.81 = 0.88 \quad P(A|\neg G) = \frac{0.07}{0.88} = 0.08$$

Revisiting the Learning Goals

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

- ▶ Calculate prior, posterior, and joint probabilities using the sum rule, the product rule, the chain rule and Bayes' rule.