Probability

Alice Gao

Lecture 10

Readings: RN 13.2 - 13.3. PM 8.1.

Outline

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: {true, false}
- P(The alarm is going $=$ true) $= 0.1$
  P(The alarm is going $=$ false) $= 0.9$

Shorthand notation

Let A a Boolean random variable.

- $P(A)$ denotes $P(A = true)$.
- $P(\neg A)$ denotes $P(A = false)$.

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 probability 1.
  Necessarily false propositions have probability 0.

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

- 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.

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?

How many probabilities are there
in the joint probability distribution?

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.

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

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

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

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?

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

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

$P(\neg A \wedge W) = 0.36$,
$P(A \wedge \neg G) = 0.06$,
$P(\neg A) = 0.9$.

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

$P(\neg A \wedge W) = 0.36$,
$P(A \wedge \neg G) = 0.06$,
$P(\neg A) = 0.9$.

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

$$P(W|A) = 0.9 \qquad P(G|A) = 0.3$$
$$P(W|\neg A) = 0.4 \qquad P(G|\neg A) = 0.1$$

$$P(W|A \wedge G) = 0.9 \qquad P(G|A \wedge W) = 0.3$$
$$P(W|A \wedge \neg G) = 0.9 \qquad P(G|A \wedge \neg W) = 0.3$$
$$P(W|\neg A \wedge G) = 0.4 \qquad P(G|\neg A \wedge W) = 0.1$$
$$P(W|\neg A \wedge \neg G) = 0.4 \qquad P(G|\neg A \wedge \neg W) = 0.1$$

Calculate a Joint Probability Using the Chain Rule

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

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

For three variables:

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

For any number of variables:

$$\begin{aligned}
&P(X_n \land X_{n-1} \land \cdots \land X_2 \land X_1) \\
&= \prod_{i=1}^{n} P(X_i | X_{i-1} \land \cdots \land X_1) \\
&= P(X_n | X_{n-1} \land \cdots \land X_2 \land X_1) * ... * P(X_2|X_1) * P(X_1)
\end{aligned}$$

CQ: Calculate a joint probability

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

$$P(A) = 0.1$$
$$P(W|A) = 0.9$$
$$P(W|A \land \neg G) = 0.9$$
$$P(G|A) = 0.3$$
$$P(G|A \land W) = 0.3$$

CQ: Calculate a joint probability

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

$$P(A) = 0.1$$
$$P(W|\neg A) = 0.4$$
$$P(W|\neg A \wedge \neg G) = 0.4$$
$$P(G|\neg A) = 0.1$$
$$P(G|\neg A \wedge \neg W) = 0.1$$

Flipping a Conditional Probability

Often you have causal knowledge:

- $P(symptom \mid disease)$
- $P(alarm \mid fire)$

...and you want to do evidential reasoning:

- $P(disease \mid symptom)$
- $P(fire \mid alarm)$.

Flipping a Conditional Probability using the Bayes' Rule

Definition (Bayes' rule)

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

CQ: Flipping a conditional probability

CQ: What is the probability that
the alarm is NOT going given that Dr. Watson is calling?

(A) 0.6
(B) 0.7
(C) 0.8
(D) 0.9
(E) 1.0

$$P(A) = 0.1$$
$$P(W|A) = 0.9$$
$$P(W|\neg A) = 0.4$$

CQ: Flipping a conditional probability

CQ: What is the probability that
the alarm is going given that Mrs. Gibbon is NOT calling?

(A) 0.04
(B) 0.05
(C) 0.06
(D) 0.07
(E) 0.08

$$P(A) = 0.1$$
$$P(G|A) = 0.3$$
$$P(G|\neg A) = 0.1$$

Some example problems

How do we calculate a (conditional) probability given several probabilities?

Example 1: Calculate $P(A \land E)$
given $P(A)$, $P(B|A)$ and $P(E|A \land B)$.

Example 2: Calculate $P(A|E)$
given $P(A)$, $P(B|A)$ and $P(E|A \land B)$.

A universal approach

1. To calculate a conditional probability, convert it into a fraction of two joint probabilities using the product rule in reverse.

2. To calculate a joint probability (not involving all the variables), write it as a summation of joint probabilities (involving all the variables) by introducing the other variables using the sum rule in reverse.

3. Calculate every joint probability (involving all the variables) using the chain rule.

Step 1

Convert a conditional probability
to joint probabilities.

Step 2

Change each joint probability
to involve all the variables.

Step 3

Calculate a joint probability using the chain rule.

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.