

Reasoning Under Uncertainty: Conditional Probability

CPSC 322 – Uncertainty 2

Textbook §10.1

Lecture Overview

- 1 Recap
- 2 Probability Distributions
- 3 Conditional Probability
- 4 Bayes' Theorem

Possible World Semantics

Probability is a formal measure of uncertainty.

- A **random variable** is a variable that is randomly assigned one of a number of different values.
- The **domain** of a variable X , written $dom(X)$, is the set of values X can take.
- A **possible world** specifies an assignment of one value to each random variable.
- $w \models X = x$ means variable X is assigned value x in world w .
- Let Ω be the set of all possible worlds.
- Define a nonnegative **measure** $\mu(w)$ to each world w so that the measures of the possible worlds sum to 1.
- The **probability** of proposition f is defined by:

$$P(f) = \sum_{w \models f} \mu(w).$$

Lecture Overview

- 1 Recap
- 2 Probability Distributions**
- 3 Conditional Probability
- 4 Bayes' Theorem

Probability Distributions

Definition (probability distribution)

A **probability distribution** P on a random variable X is a function $dom(X) \rightarrow [0, 1]$ such that

$$x \mapsto P(X = x).$$

- When $dom(X)$ is infinite we need a **probability density function**.

Joint Distribution and Marginalization

- When there are multiple random variables, their **joint distribution** is a probability distribution over the variables' Cartesian product
 - E.g., $P(X, Y, Z)$ means $P(\langle X, Y, Z \rangle)$.
 - Think of a joint distribution over n variables as an n -dimensional table
 - Each entry, indexed by $X_1 = x_1, \dots, X_n = x_n$, corresponds to $P(X_1 = x_1 \wedge \dots \wedge X_n = x_n)$.
 - The sum of entries across the whole table is 1.
- Given the joint distribution, we can compute distributions over smaller sets of variables through **marginalization**:
 - E.g., $P(X, Y) = \sum_{z \in \text{dom}(Z)} P(X, Y, Z = z)$.
 - This corresponds to summing out a dimension in the table.
 - The new table still sums to 1.

Lecture Overview

- 1 Recap
- 2 Probability Distributions
- 3 Conditional Probability**
- 4 Bayes' Theorem

Conditioning

- Probabilistic conditioning specifies how to revise beliefs based on new information.
- You build a probabilistic model taking all background information into account. This gives the **prior probability**.
- All other information must be conditioned on.
- If **evidence** e is all of the information obtained subsequently, the **conditional probability** $P(h|e)$ of h given e is the **posterior probability** of h .

Semantics of Conditional Probability

- Evidence e rules out possible worlds **incompatible** with e .
- We can represent this using a new measure, μ_e , over possible worlds

$$\mu_e(\omega) = \begin{cases} \frac{1}{P(e)} \times \mu(\omega) & \text{if } \omega \models e \\ 0 & \text{if } \omega \not\models e \end{cases}$$

Definition

The **conditional probability of formula h given evidence e** is

$$\begin{aligned} P(h|e) &= \sum_{\omega \models h} \mu_e(\omega) \\ &= \frac{P(h \wedge e)}{P(e)} \end{aligned}$$

Chain Rule

Definition (Chain Rule)

$$\begin{aligned} & P(f_1 \wedge f_2 \wedge \dots \wedge f_n) \\ &= P(f_n | f_1 \wedge \dots \wedge f_{n-1}) \times P(f_1 \wedge \dots \wedge f_{n-1}) \\ &= P(f_n | f_1 \wedge \dots \wedge f_{n-1}) \times P(f_{n-1} | f_1 \wedge \dots \wedge f_{n-2}) \times \\ &\quad P(f_1 \wedge \dots \wedge f_{n-2}) \\ &= P(f_n | f_1 \wedge \dots \wedge f_{n-1}) \times P(f_{n-1} | f_1 \wedge \dots \wedge f_{n-2}) \\ &\quad \times \dots \times P(f_3 | f_1 \wedge f_2) \times P(f_2 | f_1) \times P(f_1) \\ &= \prod_{i=1}^n P(f_i | f_1 \wedge \dots \wedge f_{i-1}) \end{aligned}$$

Lecture Overview

- 1 Recap
- 2 Probability Distributions
- 3 Conditional Probability
- 4 Bayes' Theorem**

Bayes' theorem

The chain rule and commutativity of conjunction ($h \wedge e$ is equivalent to $e \wedge h$) gives us:

$$\begin{aligned}P(h \wedge e) &= P(h|e) \times P(e) \\ &= P(e|h) \times P(h).\end{aligned}$$

If $P(e) \neq 0$, you can divide the right hand sides by $P(e)$, giving us Bayes' theorem.

Definition (Bayes' theorem)

$$P(h|e) = \frac{P(e|h) \times P(h)}{P(e)}.$$

Why is Bayes' theorem interesting?

Often you have causal knowledge:

- $P(\textit{symptom} \mid \textit{disease})$
- $P(\textit{light is off} \mid \textit{status of switches and switch positions})$
- $P(\textit{alarm} \mid \textit{fire})$
- $P(\textit{image looks like } \img alt="stick figure" data-bbox="380 445 415 495" \mid \textit{a tree is in front of a car})$

...and you want to do evidential reasoning:

- $P(\textit{disease} \mid \textit{symptom})$
- $P(\textit{status of switches} \mid \textit{light is off and switch positions})$
- $P(\textit{fire} \mid \textit{alarm})$.
- $P(\textit{a tree is in front of a car} \mid \textit{image looks like } \img alt="stick figure" data-bbox="715 805 755 855" \textit{)})$