Reasoning Under Uncertainty: Conditional Probability

CPSC 322 Lecture 24

March 19, 2007
Textbook §9.1 – §9.3

Lecture Overview

1. Recap
2. Conditional Probability
3. Bayes' Theorem

Probability

- Probability is formal measure of uncertainty. There are two camps:
- Frequentists: believe that probability represents something *objective*, and compute probabilities by counting the frequencies of different events
- Bayesians: believe that probability represents something *subjective*, and understand probabilities as degrees of belief.
  - They compute probabilities by starting with prior beliefs, and then updating beliefs when they get new data.
  - Example: Your degree of belief that a bird can fly is your measure of belief in the flying ability of an individual based only on the knowledge that the individual is a bird.
  - Other agents may have different probabilities, as they may have had different experiences with birds or different knowledge about this particular bird.
  - An agent's belief in a bird's flying ability is affected by what the agent knows about that bird.

Possible World Semantics

- 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:

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

Axioms of Probability: finite case

- Four axioms define what follows from a set of probabilities:
  - Axiom 1 $P(f) = P(g)$ if $f \leftrightarrow g$ is a tautology. That is, logically equivalent formulae have the same probability.
  - Axiom 2 $0 \leq P(f)$ for any formula f.
  - Axiom 3 $P(\tau) = 1$ if τ is a tautology.
  - Axiom 4 $P(f \vee g) = P(f) + P(g)$ if $\neg(f \wedge g)$ is a tautology.
- You can think of these axioms as constraints on which functions P we can treat as probabilities.
- These axioms are sound and complete with respect to the semantics.
  - if you obey these axioms, there will exist some μ which is consistent with your P
  - there exists some P which obeys these axioms for any given μ

Probability Distributions

Definition (probability distribution)

A probability distribution P on a random variable X is a function $dom(X) \to [0, 1]$ 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 $X_1 = x_1, \ldots, X_n = x_n$, corresponds to $P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$.
  - 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., $P(X, Y) = \sum_{z \in dom(Z)} P(X, Y, Z = z)$.
  - This corresponds to summing out a dimension in the table.
  - The new table still sums to 1.

Lecture Overview

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

$$P(h|e) = \sum_{\omega \models h} \mu_e(w) = \frac{P(h \wedge e)}{P(e)}$$

Chain Rule

Definition (Chain Rule)

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

Lecture Overview

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:

- $P(symptom \mid disease)$
- $P($*light is off* $\mid$ *status of switches and switch positions*$)$
- $P(alarm \mid fire)$
- $P($*image looks like*  $\mid$ *a tree is in front of a car*$)$

...and you want to do evidential reasoning:

- $P(disease \mid symptom)$
- $P($*status of switches* $\mid$ *light is off and switch positions*$)$
- $P(fire \mid alarm)$.
- $P($*a tree is in front of a car* $\mid$ *image looks like* $)$