Reasoning Under Uncertainty: Conditional Probability

CPSC 322 – Uncertainty 2

Textbook §6.1
Lecture Overview

1. Recap
2. Probability Distributions
3. Conditional Probability
4. Bayes’ Theorem
A random variable is a variable that is randomly assigned one of a number of different values.

The domain of a variable $X$, written $\text{dom}(X)$, is the set of values $X$ can take.

A possible world specifies an assignment of one value to each random variable.

$w \models \phi$ means the proposition $\phi$ is true in world $w$.

Let $\Omega$ 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 $\phi$ is defined by:

$$P(\phi) = \sum_{w \models \phi} \mu(w).$$
<table>
<thead>
<tr>
<th></th>
<th>Recap</th>
<th>Probability Distributions</th>
<th>Conditional Probability</th>
<th>Bayes’ Theorem</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Recap</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Probability Distributions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Conditional Probability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Bayes’ Theorem</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Consider the case where possible worlds are simply assignments to one random variable.

**Definition (probability distribution)**

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

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

- When $\text{dom}(X)$ is infinite we need a *probability density* function.
Joint Distribution

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 $P(X_1 = x_1 \land \ldots \land X_n = x_n)$.
- The sum of entries across the whole table is 1.
Consider the following example, describing what a given day might be like in Vancouver.

- We have two random variables:
  - weather, with domain \{Sunny, Cloudy\};
  - temperature, with domain \{Hot, Mild, Cold\}.

- Then we have the joint distribution \( P(\text{weather}, \text{temperature}) \) given as follows:

<table>
<thead>
<tr>
<th>weather</th>
<th>temperature</th>
<th>Hot</th>
<th>Mild</th>
<th>Cold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>0.10</td>
<td>0.20</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Cloudy</td>
<td>0.05</td>
<td>0.35</td>
<td>0.20</td>
<td></td>
</tr>
</tbody>
</table>
Marginalization

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.
Marginalization Example

<table>
<thead>
<tr>
<th>weather</th>
<th>Sunny</th>
<th>Mild</th>
<th>Cold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hot</td>
<td>0.10</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>Cloudy</td>
<td>0.05</td>
<td>0.35</td>
<td>0.20</td>
</tr>
</tbody>
</table>

If we marginalize out *weather*, we get

\[ P(temperature) = \begin{bmatrix} Hot & Mild & Cold \\ 0.15 & 0.55 & 0.30 \end{bmatrix} \]

If we marginalize out *temperature*, we get

\[ P(weather) = \begin{bmatrix} Sunny & Cloudy \\ 0.40 & 0.60 \end{bmatrix} \]
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 $h$. 
Evidence $e$ rules out possible worlds incompatible with $e$.

We can represent this using a new measure, $\mu_e$, 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

$$P(h|e) = \sum_{\omega \models h} \mu_e(\omega) = \frac{P(h \land e)}{P(e)}$$
### Conditional Probability Example

<table>
<thead>
<tr>
<th>weather</th>
<th>Sunny</th>
<th>Cloudy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>0.10</td>
<td>0.05</td>
</tr>
<tr>
<td>Cloudy</td>
<td>0.35</td>
<td>0.20</td>
</tr>
</tbody>
</table>

If we condition on $\text{weather} = \text{Sunny}$, we get

$$P(\text{temperature}|\text{Weather} = \text{Sunny}) = \begin{bmatrix} 0.25 & 0.50 & 0.25 \end{bmatrix}$$

Conditioning on $\text{temperature}$, we get $P(\text{weather}|\text{temperature})$:

<table>
<thead>
<tr>
<th>weather</th>
<th>Sunny</th>
<th>Cloudy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>0.67</td>
<td>0.33</td>
</tr>
<tr>
<td>Cloudy</td>
<td>0.33</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Note that each column now sums to one.
Chain Rule

**Definition (Chain Rule)**

\[
P(f_1 \land f_2 \land \ldots \land f_n) = P(f_n | f_1 \land \ldots \land f_{n-1}) \times P(f_1 \land \ldots \land f_{n-1})
\]

\[
= P(f_n | f_1 \land \ldots \land f_{n-1}) \times P(f_{n-1} | f_1 \land \ldots \land f_{n-2}) \times P(f_1 \land \ldots \land f_{n-2})
\]

\[
= P(f_n | f_1 \land \ldots \land f_{n-1}) \times P(f_{n-1} | f_1 \land \ldots \land f_{n-2}) \times \ldots \times P(f_3 | f_1 \land f_2) \times P(f_2 | f_1) \times P(f_1)
\]

\[
= \prod_{i=1}^{n} P(f_i | f_1 \land \ldots \land f_{i-1})
\]

E.g., \(P(\text{weather, temperature}) = P(\text{weather} | \text{temperature}) \cdot P(\text{temperature}).\)
### Lecture Overview

<table>
<thead>
<tr>
<th>1. Recap</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Probability Distributions</td>
</tr>
<tr>
<td>3. Conditional Probability</td>
</tr>
<tr>
<td>4. Bayes’ Theorem</td>
</tr>
</tbody>
</table>
The chain rule and commutativity of conjunction (\(h \land e\) is equivalent to \(e \land h\)) gives us:

\[
P(h \land e) = P(h|e) \times P(e) = P(e|h) \times P(h).
\]

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:

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

...and you want to do evidential reasoning:

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