Discrete Probability

COMPSCI 230 — Discrete Math

April 9, 2015
Outline

1 Conditional Probability Functions

2 Bayesian Inference and the Monty Hall Problem

3 Random Variables and Probability Distributions
Conditional Probability Functions

- Conditional probability: If \( \text{prob}(A) > 0 \),

\[
\text{prob}(B|A) = \frac{\text{prob}(A \cap B)}{\text{prob}(A)}
\]

- For a joint probability function \( P_{XY}(i, j) \)

\[
\begin{array}{cccc}
 & 1 & \ldots & j \\
X |
\hline
1 & & & \\
\vdots & & & \\
i & B & & \\
\vdots & & & \\
m & & & A \cap B
\end{array}
\]

\( \text{prob}(A) = \text{prob}(X=i) \)

\( \text{prob}(B) = \text{prob}(Y=j) \)

\( A = \{(i, 1), (i, 2), \ldots, (i, n)\} = \{X = i\} \) so \( \text{prob}(A) = P_X(i) \), the marginal of \( X \) at \( i \)

\( B = \{(1, j), (2, j), \ldots, (m, j)\} = \{Y = j\} \) so \( \text{prob}(B) = P_Y(j) \), the marginal of \( Y \) at \( j \)

- Define the conditional probability functions

\[
P_{X|Y}(i, j) = \text{prob}(A|B) = \text{prob}(\{X = i\} \mid \{Y = j\}) = \frac{P_{XY}(i, j)}{P_Y(j)}
\]

\[
P_{Y|X}(j, i) = \text{prob}(B|A) = \text{prob}(\{Y = j\} \mid \{X = i\}) = \frac{P_{XY}(i, j)}{P_X(i)}
\]
Example: Two Biased, Dependent Coins

Joint probability function:

|       | coin D       | $P_C$ | $P_{CD}$ | $P_{C|D}(c,d)$ |
|-------|--------------|-------|----------|----------------|
|       |              |       |          |                |
| $P_D$ | $H$          | 0.2   | 0.1      | 0.3            |
|       | $T$          | 0.4   | 0.3      | 0.7            |
| $P_C$ | $H$          | 0.6   | 0.4      | 1              |
|       | $T$          | 0.6   | 0.4      | 1              |

Joint $\rightarrow$ conditional probability functions:

Normalize rows: $P_{D|C}(d, c) = P_{CD}(c, d)/P_C(c)$
or columns: $P_{C|D}(c, d) = P_{CD}(c, d)/P_D(d)$
Conditional $\rightarrow$ Joint, Marginal

- To go back from conditional probability functions to the joint probability function we need the appropriate marginal

$$P_{CD}(c, d) = P_{D|C}(d, c) P_C(c) = P_{C|D}(c, d) P_D(d)$$

- If we have, say, $P_{C|D}(c, d)$ and marginal $P_D(d)$, we can compute the other marginal $P_C(c)$:

  conditional, marginal 1 $\rightarrow$ joint $\rightarrow$ marginal 2

  $$P_{CD}(c, d) = P_{C|D}(c, d) P_D(d)$$

  $$P_C(c) = \sum_{d \in S_D} P_{CD}(c, d)$$

- The summation is called **marginalization** over $d$
- More compactly:

  $$P_C(c) = \sum_{d \in S_D} P_{C|D}(c, d) P_D(d)$$

\[\begin{array}{|c|c|c|}
\hline
\text{coin C} & \multicolumn{2}{|c|}{\text{coin D}} \\
\hline
\text{H} & 1/3 & 1/4 \\
\text{T} & 2/3 & 3/4 \\
\hline
\end{array}\]

\[\begin{array}{|c|c|c|}
\hline
\text{coin D} & \multicolumn{2}{|c|}{\text{coin C}} \\
\hline
\text{H} & 0.2 & 0.1 \\
\text{T} & 0.4 & 0.3 \\
\hline
\end{array}\]

\[\begin{array}{|c|c|c|}
\hline
\text{coin D} & \multicolumn{2}{|c|}{\text{PC}} \\
\hline
\text{H} & 0.6 \\
\text{T} & 0.4 \\
\hline
\end{array}\]
Bayes’ Theorem for Probability Functions

\[ P_{X|Y}(i, j) = \text{probability of } \{X = i\} \cap \{Y = j\} \text{ given that } Y = j \] (purple intersection area divided by blue area)

\[ P_{Y|X}(j, i) = \text{probability of } \{X = i\} \cap \{Y = j\} \text{ given that } X = i \] (purple intersection area divided by red area)

• Bayes’s theorem for events: If \( \text{prob}(A) > 0 \),

\[ \text{prob}(B|A) = \frac{\text{prob}(A|B)\text{prob}(B)}{\text{prob}(A)} \]

• Bayes’s theorem for probability functions: If \( P_X(i) > 0 \),

\[ P_{Y|X}(j, i) = \frac{P_{X|Y}(i, j)P_Y(j)}{P_X(i)} \]
Revisiting The Monty Hall Problem

- Game show with three closed doors 1, 2, 3
- Car behind one door, goats behind the others
- The host knows where the car is
- You get to choose one door, and you get the prize behind it
- Say you pick door 1
The Monty Hall Problem

- You picked door 1
- Instead of opening door 1 the host opens door 3, which has a goat
  [The host never opens the door with the car]
- Host asks, ”Do you want to pick door 2 instead of 1?”
- **Should you switch to door 2?**
From Intuition to Calculation

- “I get that there is a 2/3 chance that the car is behind door 2”
- “But what is wrong with my other reasoning?”
- “The car is either behind door 1 or behind door 2, so why not 50/50?”
- Answer: Because we know something about door 2 that we did not know before the host opened door 3
- Information skewed probabilities
- This is exactly what Bayes’s theorem is for
- Let us calculate our way to an answer
Formalizing Monty Hall

• Scenario: You have chosen door 1 and the host has not yet opened a door
• Experiment $C$: Generate the position for the car. $\mathcal{S}_C = \{1, 2, 3\}$
• Experiment $H$: Host chooses which door to open. $\mathcal{S}_H = \{1, 2, 3\}$
• Compound experiment $(C, H)$. $\mathcal{S} = \mathcal{S}_C \times \mathcal{S}_H$
• Marginals are easy: Car is placed uniformly at random, so as to give players as little information as possible

$$P_C(1) = P_C(2) = P_C(3) = 1/3$$

• $P_C$ is called the prior probability
What we know about the situation before anything happens
Bayesian Inference and the Monty Hall Problem

What We Know, What We Need

• Scenario: You have chosen door 1 and the host has not yet opened a door
• Host’s rules are best understood through conditional probabilities $P_{H|C}(h, c)$ because we know that he knows where the car is
  • i.e., we know how to compute probabilities conditioned on $C$
• In the end, we are interested in the probability $P_{C|H}(c, 3)$ that the car is behind door $c$ given that the host opens door 3
• **Posterior probability** of $c$ conditioned on $H = 3$
Bayesian Inference and the Monty Hall Problem

Modeling Monty Hall’s Behavior

• Host will never open the door you choose first
• If you chose the door with the car, he opens one of the other two doors uniformly at random
• Otherwise, he opens the only door without the car other than the door you chose

\[
\begin{array}{c|ccc}
C & 1 & 2 & 3 \\
\hline
1 & 1 & 0 & 1/2 \\
2 & 0 & 0 & 1 \\
3 & 0 & 1 & 0 \\
\end{array}
\]

\[
\begin{array}{c|cccc}
H & 1 & 2 & 3 & \\
\hline
1 & 0 & 1/2 & 1/2 & 1 \\
2 & 0 & 0 & 1 & 1 \\
3 & 0 & 1 & 0 & 1 \\
\end{array}
\]

• This is a probabilistic model of Monty Hall’s behavior
**Bayesian Inference and the Monty Hall Problem**

**Before the Host Opens a Door**

- We have the model $P_{H|C}(h, c)$ and $P_C(c)$
- We know how to compute the other marginal $P_H(h)$:

$$P_{CH}(h, c) = P_{H|C}(h, c)P_C(c) \quad P_H(h) = \sum_{c \in S_C} P_{CH}(h, c)$$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{H</td>
<td>C}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1/3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_C$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1/3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1/3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1/3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{CH}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1/3</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1/3</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1/2</th>
<th>1/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_H$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1/2</td>
<td>1/2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
After the Host Opens Door 3

- Condition on $H = 3$: Columns 1, 2 become irrelevant
- The sample space on the host side has shrunk from $\mathbb{S}_H = \{1, 2, 3\}$ with probabilities $P_H = 0, 1/2, 1/2$ to $\mathbb{S}'_H = \{3\}$ with probability 1
- $\mathbb{S}$ has shrunk to $\mathbb{S}' = \mathbb{S}_C \times \mathbb{S}'_H$
- Bayes’s theorem (or definition of conditional probability):

$$P_{C|H}(c, 3) = \frac{P_H|C(3, c) P_C(c)}{P_H(3)} = \frac{P_{CH}(c, 3)}{P_H(3)}$$

- This means: Normalize the only remaining column
- Choose door 2!
Probabilistic Inference

• The probabilities $P_{H|C}(h, c)$ are easy to formulate because we know the host’s behavior well, and we know that he knows $c$

• Together, the prior $P_C$ and the conditional probabilities $P_{H|C}$ are a **probabilistic model** of the situation in the scenario

• We are interested in the **posterior probability** $P_{C|H}(c, 3)$ that the car is behind door $c$ given that the host opens door 3

• Or $P_{C|H}(c, h)$ for more general decision-making

• Bayes’s theorem turns conditional probabilities around

• We **infer** the posterior $P_{C|H}$ from the model $P_{H|C}, P_C$

• **Bayesian inference**: Infer a posterior probability from a probabilistic model, given the data ($H = 3$)
Random Variables: From Outcomes to Numbers

• We often associate numbers to outcomes
• Example: Coin flip, bet on $H$
• Win $w = $10 if head, win $\ell = -$10 if tail
• We need to tie $w$ to $H$ and $\ell$ to $T$
• $H, T$ are in the sample space $\mathcal{S}$
• $w, \ell$ are in $\mathbb{R}$
• Any function $\mathcal{S} \to \mathbb{R}$ is called a random variable
• In the example,

$$W : \{H, T\} \rightarrow \mathbb{R}$$

is defined as

$$W(H) = w \text{ and } W(T) = \ell$$
Random Variables

- Any function $\mathcal{S} \rightarrow \mathbb{R}$ is called a **random variable**
- A random variable is neither random, nor a variable
- It is a deterministic function of the outcome
- If you just observe the values of $W(O)$, they vary randomly because the outcome $O$ does
- Hence the (arguably confusing) name
Probability Distribution

• A probability function defined on a random variable is called a **probability distribution**:

\[ P(X = x) = \sum_{O \in \mathbb{S} : X(O) = x} P(O) \]

• Example 1: Roll of one fair die, \( x = X(O) = (O \mod 2) \)
  - \( X(2) = X(4) = X(6) = 0 \)
  - \( X(1) = X(3) = X(5) = 1 \)
  - \( P(X = 0) = P(X = 1) = 1/2 \)
  (Uniform distribution over \( \{0, 1\} \))

• Example 2: Money won when betting on \( H \) in a coin flip,

\[ W(O) = \begin{cases} w & \text{for } O = H \\ \ell & \text{for } O = T \end{cases} \]

\[ P(W = w) = 0.6 \quad P(W = \ell) = 0.4 \]