Full Joint Probability Distributions

<table>
<thead>
<tr>
<th>Coin</th>
<th>Card</th>
<th>Candy</th>
<th>P(Coin, Card, Candy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tails</td>
<td>black</td>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>tails</td>
<td>black</td>
<td>2</td>
<td>0.06</td>
</tr>
<tr>
<td>tails</td>
<td>black</td>
<td>3</td>
<td>0.09</td>
</tr>
<tr>
<td>tails</td>
<td>red</td>
<td>1</td>
<td>0.02</td>
</tr>
<tr>
<td>tails</td>
<td>red</td>
<td>2</td>
<td>0.06</td>
</tr>
<tr>
<td>tails</td>
<td>red</td>
<td>3</td>
<td>0.12</td>
</tr>
<tr>
<td>heads</td>
<td>black</td>
<td>1</td>
<td>0.075</td>
</tr>
<tr>
<td>heads</td>
<td>black</td>
<td>2</td>
<td>0.03</td>
</tr>
<tr>
<td>heads</td>
<td>black</td>
<td>3</td>
<td>0.045</td>
</tr>
<tr>
<td>heads</td>
<td>red</td>
<td>1</td>
<td>0.035</td>
</tr>
<tr>
<td>heads</td>
<td>red</td>
<td>2</td>
<td>0.10</td>
</tr>
<tr>
<td>heads</td>
<td>red</td>
<td>3</td>
<td>0.21</td>
</tr>
</tbody>
</table>

This cell means $P(\text{Coin=heads, Card=red, Candy}=3) = 0.21$

The probabilities in the last column sum to 1

Marginalization

The general marginalization rule for any sets of variables $Y$ and $Z$:

$$ P(Y) = \sum_{Z} P(Y, Z) $$

or

$$ P(Y) = \sum_{Z} P(Y | z) P(z) $$

Conditional Probabilities

We can also compute conditional probabilities from the joint. Recall:

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

Bayes’ Rule

The product rule can be written in two ways:

$$ P(A, B) = P(A | B)P(B) $$

$$ P(A, B) = P(B | A)P(A) $$

You can combine the equations above to get:

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

Inference

We will write the query as $P(X | e)$

$$ P(X | e) = \alpha P(X, e) = \alpha \sum_{Y} P(X, e, y) $$

Summation is over all possible combinations of values of the unobserved variables $Y$
Bayes’ Rule
More generally, the following is known as Bayes’ Rule:

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

Note that these are distributions

Sometimes, you can treat \( P(B) \) as a normalization constant \( \alpha \)

\[ P(A | B) = \alpha P(B | A)P(A) \]

When is Bayes Rule Useful?
Sometimes it’s easier to get \( P(X|Y) \) than \( P(Y|X) \).

Information is typically available in the form \( P(\text{effect | cause}) \) rather than \( P(\text{cause | effect}) \)

For example, \( P(\text{symptom | disease}) \) is easy to measure empirically but obtaining \( P(\text{disease | symptom}) \) is harder

Bayes Rule Example
Meningitis causes stiff necks with probability 0.5. The prior probability of having meningitis is 0.00002. The prior probability of having a stiff neck is 0.05. What is the probability of having meningitis given that you have a stiff neck?

Let \( m \) = patient has meningitis
Let \( s \) = patient has stiff neck
\( P(s | m) = 0.5 \)
\( P(m) = 0.00002 \)
\( P(s) = 0.05 \)
\( P(m | s) = \frac{P(s | m)P(m)}{P(s)} = \frac{(0.5)(0.00002)}{0.05} = 0.0002 \)

Bayes Rule Example
In machine learning, we use Bayes rule in the following way:

\[ P(h | D) = \frac{P(D | h)P(h)}{P(D)} \]

H = hypothesis
D = data

Likelihood of the data
Prior probability
Posterior probability
Bayes Rule With More Than One Piece of Evidence

Suppose you now have 2 evidence variables Card=red and Candy=1 (note that Coin is uninstantiated below)

\[ P(\text{Coin} \mid \text{Card}=\text{red}, \text{Candy}=1) = \alpha P(\text{Card}=\text{red}, \text{Candy}=1 \mid \text{Coin}) P(\text{Coin}) \]

In order to calculate \( P(\text{Card}=\text{red}, \text{Candy}=1 \mid \text{Coin}) \), you need a table of 6 probability values. With N Boolean evidence variables, you need \( 2^N \) probability values.

Independence

We say that variables X and Y are independent if any of the following hold: (note that they are all equivalent)

\[ P(X \mid Y) = P(X) \quad \text{or} \quad P(Y \mid X) = P(Y) \quad \text{or} \quad P(X, Y) = P(X)P(Y) \]

Why is independence useful?

\[ P(\text{Card}, \text{Candy}) = P(\text{Card})P(\text{Candy}) \]

This table has 2 values

This table has 3 values

• You now need to store 5 values to calculate \( P(\text{Coin}, \text{Card}, \text{Candy}) \)
• Without independence, we needed 6

Conditional Independence

Suppose I tell you that to select a piece of Candy, I first flip a Coin. If heads, I select a Card from one (stacked) deck; if tails, I select from a different (stacked) deck. The color of the card determines the bag I select the Candy from, and each bag has a different mix of the types of Candy.

Are Coin and Candy independent?

Conditional Independence

Suppose I tell you that to select a piece of Candy, I first flip a Coin. If heads, I select a Card from one deck; if tails, I select from a different deck. The color of the card determines the bag I select the Candy from, and each bag has a different mix of the types of Candy.

Are Coin and Candy independent?  No.

But given Card, they are independent!

\[ P(\text{Coin} = \text{heads}, \text{Candy} = 3 \mid \text{Card}) = P(\text{Coin} = \text{heads} \mid \text{Card}) \times P(\text{Candy} = 3 \mid \text{Card}) \]

Conditional Independence

General form:

\[ P(A, B \mid C) = P(A \mid C)P(B \mid C) \]

Or equivalently:

\[ P(A \mid B, C) = P(A \mid C) \quad \text{and} \quad P(B \mid A, C) = P(B \mid C) \]

How to think about conditional independence:

In \( P(A \mid B, C) = P(A \mid C) \): if knowing \( C \) tells me everything about \( A \), I don’t gain anything by knowing \( B \).
Conditional Independence

\[ P(Coin, Card, Candy) = P(Candy | Coin, Card) P(Coin, Card) = P(Candy | Card) P(Card | Coin) P(Coin) \]

11 independent values in table (have to sum to 1)

4 independent values in table
2 independent values in table
1 independent value in table

4 independent values in table
2 independent values in table
1 independent value in table

Conditional independence permits probabilistic systems to scale up!

### Candy Example

| Coin | P(Coin) | Card | P(Card | Coin) | Card | Candy | P(Candy | Card) |
|------|---------|------|-------------|------|-------|-------------|
| tails | 0.5     | black | 0.6         | black | 1     | 0.5         |
| heads | 0.5     | red   | 0.4         | black | 2     | 0.2         |
|       |         |       |             | red   | 1     | 0.1         |
|       |         |       |             |       |       |             |
| tails | 0.5     | black | 0.3         | black | 3     | 0.3         |
| heads | 0.5     | red   | 0.7         | red   | 2     | 0.3         |
|       |         |       |             |       |       |             |
| heads | 0.5     | red   | 0.3         | red   | 3     | 0.3         |

\[ P(Coin = heads, Card = red, Candy = 3) = P(Coin = heads) \times P(Card = red | Coin = heads) \times P(Candy = 3 | Card = red) = 0.5 \times 0.7 \times 0.6 = 0.21 \]

### Practice

| Coin  | P(Coin) | Coin | Card | P(Card | Coin) | Card | Candy | P(Candy | Card) |
|-------|---------|------|------|---------|------|-------|---------|
| tails | 0.5     | tails| black| 0.6     | black| 1     | 0.5     |
| heads | 0.5     | tails| red  | 0.4     | black| 2     | 0.2     |
|       |         | heads| black| 0.3     | black| 3     | 0.3     |
|       |         | heads| red  | 0.7     | red  | 1     | 0.1     |
|       |         |       |       |         | red  | 2     | 0.3     |
|       |         |       |       |         | red  | 3     | 0.3     |

Compute \( P(Coin = tails | Card = red) \)

### What You Should Know

- How to do inference in joint probability distributions
- How to use Bayes Rule
- Why independence and conditional independence is useful