Chapter 1
Conditional Probability and Independence
Bayes Formula and Independence
Outline

Tree Diagrams

Terminology for Conditional Probabilities

Weighing the Odds
Bayes Formula

- Let $A$ be the event that an individual tests positive for some disease and
- let $C$ be the event that the person has the disease.
- We can perform clinical trials to estimate the probability that a randomly chosen individual tests positive given that they have the disease,

$$P\{\text{tests positive|has the disease}\} = P(A|C),$$

by taking individuals with the disease and applying the test.
- We would like to use the test as a method of diagnosis of the disease. In other words, we would like to give the test and assert the chance that the person has the disease. That is, we want to know the probability with the reverse conditioning

$$P\{\text{has the disease|tests positive}\} = P(C|A).$$
Bayes Formula

The Public Health Department gives us the following information.

- The test yields a **positive** result 90% of the time when the disease is **present**.
- The test yields a **positive** result 1% of the time when the disease is **not present**.
- One person in 1,000 has the disease.

Let’s first think about this intuitively to find the probability

\[ P(C|A) = P\{\text{has the disease}|\text{tests positive}\}. \]

- In a city with a population of **1 million** people, on average,
  
  1,000 have the disease and 999,000 do not

- Of the **1,000** that have the disease, on average,
  
  900 test positive and 100 test negative

- Of the **999,000** that do not have the disease, on average,
  
  \[ 999,000 \times 0.01 = 9990 \] test positive and \( 989,010 \) test negative.
Bayes Formula

Consequently, among those that test positive, the odds of having the disease is

\[(\text{have the disease}):\text{(does not have the disease)}\]

900:9990

and converting odds to probability we see that

\[P\{\text{have the disease}|\text{test is positive}\} = \frac{900}{900 + 9990} = 0.0826\]
Tree Diagrams

1,000,000 people

1,000 have the disease

P(C) = 0.001

P(A|C)P(C) = 0.0009
900 test positive

P(A^c|C)P(C) = 0.0001
100 test negative

999,000 do not have the disease

P(C^c) = 0.999

P(A|C^c)P(C^c) = 0.00999
9,990 test positive

P(A^c|C^c)P(C^c) = 0.98901
989,010 test negative
Bayes Formula

Consequently, among those that test positive, the *odds* of having the disease is

\[
\frac{\text{#(have the disease)}}{\text{#(does not have the disease)}} = \frac{900}{9990}
\]

and converting odds to probability we see that

\[
P(C|A) = \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|C^c)P(C^c)} = \frac{900}{900 + 9990} = 0.0826
\]

\[
P(C|A) = \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|C^c)P(C^c)} = \frac{0.0009}{0.0009 + 0.0999} = 0.0826
\]
Bayes Formula

Let’s check the formula

$$P(C|A) = \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|C^c)P(C^c)}$$

(Bayes Formula)

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

(Law of Total Probability)

$$= \frac{P(A \cap C)}{P(A)}$$

(Multiplication Principle)

which is just the definition of conditional probability.

Generally, we call $P(C)$ the prior probability of $C$. With $A$ given, we call $P(C|A)$ the posterior probability of $A$. The Bayes factor

$$\frac{P(A|C)}{P(A)}.$$

is their ratio. Thus, $A$ and $C$ are independent if and only if the Bayes factor is 1.
Bayes Formula

Example. Both autism $A$ and epilepsy $C$ exists at approximately 1% in human populations. In this case,

$$P(C|A) = P(A|C) \frac{P(C)}{P(A)} \approx P(A|C)$$

Clinical evidence shows that this common value is about 30%. The Bayes factor is

$$\frac{P(A|C)}{P(A)} \approx \frac{0.3}{0.01} = 30.$$ 

Thus, the knowledge of one disease increases the chance of the other by a factor of 30.
Bayes Formula

Two traits are called genetically linked if the appearance of one increases the probability of the other. In addition,

\[ P\{\text{individual has allele for trait 1}|\text{individual has allele for trait 2}\} > P\{\text{individual has allele for trait 1}\}. \]

implies \[ P\{\text{individual has allele for trait 2}|\text{individual has allele for trait 1}\} > P\{\text{individual has allele for trait 2}\}. \]

Exercise. For events \( A \) and \( C \),

\[ P(A|C) > P(A) \quad \text{if and only if} \quad P(C|A) > P(C) \]

In this case, we say that \( A \) and \( C \) are positively associated.

Answer.

\[ P(C|A) = \frac{P(A|C)}{P(A)} P(C) > P(C). \]
Bayes Formula

**Exercise.** If $A$ and $C$ are positively associated, then $A$ and $C^c$ are negatively associated.

**Answer.**

$$P(C^c|A) < P(C^c) \text{ if and only if } P(C|A) > P(C)$$

Bayes formula can be generalized to the case of a partition

$\{C_1, C_2, \ldots, C_n\}$

of $\Omega$ chosen so that all the $P(C_i) > 0$.

Then, for any event $A$ and any $j$

$$P(C_j|A) = \frac{P(A|C_j)P(C_j)}{\sum_{i=1}^{n} P(A|C_i)P(C_i)}.$$
Terminology for Conditional Probabilities

<table>
<thead>
<tr>
<th>true positive (TPR)</th>
<th>false positive (FPR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(A</td>
<td>C)$</td>
</tr>
<tr>
<td>false negative (FNR)</td>
<td>true negative (TNR)</td>
</tr>
<tr>
<td>$P(A^c</td>
<td>C)$</td>
</tr>
</tbody>
</table>

Table: $A = \{\text{test positive}\}$ and $C = \{\text{has the disease}\}$. Notice that the columns sum to one.

<table>
<thead>
<tr>
<th>has the disease</th>
<th>does not have the disease</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>test positive</td>
<td>900 = TPR $\times$ #disease</td>
<td>9,990 = FPR $\times$ #no disease</td>
</tr>
<tr>
<td>test negative</td>
<td>100 = FNR $\times$ #disease</td>
<td>989,010 = TNR $\times$ #no disease</td>
</tr>
<tr>
<td>total</td>
<td>1,000</td>
<td>990,000</td>
</tr>
</tbody>
</table>

Having filled in the columns, Bayes formula has us look at odds along the rows.
Weighing the Odds

The answer may be surprising. Only 8% of those who test positive actually have the disease. This example underscores the fact that good predictions based on intuition can be hard to make.

To determine the probability $P(C|A)$, we must weigh the odds of two terms, each of them itself a product.

- $P(A|C)P(C)$, a big number (the true positive probability) times a small number (the probability of having the disease) versus
- $P(A|C^c)P(C^c)$, a small number (the false positive probability) times a large number (the probability of being disease free).
Bayes Formula

**Exercise.** Fill in the table of values for $P(C|A) = P\{\text{has the disease|tests positive}\}$.

<table>
<thead>
<tr>
<th>percent with disease</th>
<th>0.1%</th>
<th>0.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>false positive</td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>true positive</td>
<td>90%</td>
<td>0.0826</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td></td>
</tr>
</tbody>
</table>

Which has a bigger impact - a change in the false positive rate or the true positive rate? Give an intuitive explanation for your answer?
Bayes Formula

Exercise. Fill in the table of values for $P(C|A) = P\{\text{has the disease} | \text{tests positive}\}$.

<table>
<thead>
<tr>
<th>percent with disease</th>
<th>0.1%</th>
<th>0.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>false positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>5%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>true positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>0.0826</td>
<td>0.3114</td>
</tr>
<tr>
<td>95%</td>
<td>0.0868</td>
<td>0.3231</td>
</tr>
<tr>
<td>positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0177</td>
<td>0.0829</td>
<td>0.0872</td>
</tr>
<tr>
<td>0.0187</td>
<td>0.0177</td>
<td>0.0872</td>
</tr>
<tr>
<td>0.3114</td>
<td>0.3114</td>
<td>0.3231</td>
</tr>
<tr>
<td>0.0829</td>
<td>0.0829</td>
<td>0.0872</td>
</tr>
<tr>
<td>0.0868</td>
<td>0.0868</td>
<td>0.0872</td>
</tr>
<tr>
<td>0.0177</td>
<td>0.0177</td>
<td>0.0872</td>
</tr>
<tr>
<td>0.3231</td>
<td>0.3231</td>
<td>0.0872</td>
</tr>
<tr>
<td>0.0829</td>
<td>0.0829</td>
<td>0.0872</td>
</tr>
<tr>
<td>0.0868</td>
<td>0.0868</td>
<td>0.0872</td>
</tr>
</tbody>
</table>

Which has a bigger impact - a change in the false positive rate or the true positive rate? Give an intuitive explanation for your answer?
Bayes Formula

A box has a 2-headed coin and a fair coin. It is flipped \( n \) times, yielding heads each time. What is the probability that the 2-headed coin is chosen? To solve this, note that

\[
P\{\text{2-headed coin}\} = \frac{1}{2}, \quad P\{\text{fair coin}\} = \frac{1}{2}.
\]

and

\[
P\{n \text{ heads}|\text{2-headed coin}\} = 1, \quad P\{n \text{ heads}|\text{fair coin}\} = 2^{-n}.
\]

By the law of total probability, \( P\{n \text{ heads}\} \)

\[
= P\{n \text{ heads}|\text{2-headed coin}\} P\{\text{2-headed coin}\}
+ P\{n \text{ heads}|\text{fair coin}\} P\{\text{fair coin}\}
= 1 \cdot \frac{1}{2} + 2^{-n} \cdot \frac{1}{2} = \frac{2^n + 1}{2^{n+1}}.
\]

Next, we use Bayes formula. \( P\{\text{2-headed coin}|n \text{ heads}\} \)

\[
= \frac{P\{n \text{ heads}|\text{2-headed coin}\} P\{\text{2-headed coin}\}}{P\{n \text{ heads}\}} = \frac{1 \cdot (1/2)}{(2^n + 1)/2^{n+1}} = \frac{2^n}{2^n + 1} < 1.
\]
Bayes Formula

This simple and seemingly silly example is mathematically identical to a question in the vertical transmission of a genetic disease.

- A female knows that she has a history of a allele on her X chromosome for a recessive genetic condition.
- She does not have the condition. So, she knows that she cannot be homozygous for the recessive allele. Consequently, she wants to know her chance of being
  - a carrier (heterozygous for a recessive allele) or
  - not a carrier (homozygous for the common genetic type).
- The female is a mother with $n$ sons, none of which show the recessive allele on their single X chromosome and so do not have the condition.
Bayes Formula

A box has a 2-headed coin and a fair coin. It is flipped \(n\) times, yielding heads each time. What is the probability that the 2-headed coin is chosen? To solve this, note that

\[ P\{\text{2-headed coin}\} = \frac{1}{2}, \quad P\{\text{fair coin}\} = \frac{1}{2}. \]

and

\[ P\{n \text{ heads}|2\text{-headed coin}\} = 1, \quad P\{n \text{ heads}|\text{fair coin}\} = 2^{-n}. \]

By the law of total probability, \(P\{n\ \text{heads}\}\)

\[
= P\{n \ \text{heads}|2\text{-headed coin}\} P\{2\text{-headed coin}\} \\
+ P\{n \ \text{heads}|\text{fair coin}\} P\{\text{fair coin}\} \\
= 1 \cdot \frac{1}{2} + 2^{-n} \cdot \frac{1}{2} = \frac{2^n + 1}{2^{n+1}}. 
\]

Next, we use Bayes formula.. \(P\{2\text{-headed coin}|n\ \text{heads}\}\)

\[
= \frac{P\{n \ \text{heads}|2\text{-headed coin}\} P\{2\text{-headed coin}\}}{P\{n \ \text{heads}\}} = \frac{1 \cdot (1/2)}{(2^n + 1)/2^{n+1}} = \frac{2^n}{2^n + 1} < 1. 
\]
Tree Diagrams

 Terminology for Conditional Probabilities

 Bayes Formula

What is the probability that the female is not a carrier? Let’s look at the computation above again. Based on her pedigree, the female estimates that

\[ P\{\text{mother not a carrier}\} = p, \quad P\{\text{mother a carrier}\} = 1 - p. \]

and

\[ P\{n \text{ sons condition free}|\text{mother not a carrier}\} = 1, \quad P\{n \text{ sons condition free}|\text{mother a carrier}\} = 2^{-n}. \]

By the law of total probability,

\[
P\{n \text{ sons condition free}\} = P\{n \text{ sons condition free}|\text{mother not a carrier}\}P\{\text{mother is not a carrier}\} \\
+ P\{n \text{ sons condition free}|\text{mother a carrier}\}P\{\text{mother a carrier}\} \\
= 1 \cdot p + 2^{-n} \cdot (1 - p).
\]

Next, we use Bayes formula.

\[
P\{\text{mother not a carrier}|n \text{ sons condition free}\} = \frac{P\{n \text{ sons condition free}|\text{mother not a carrier}\}P\{\text{mother not a carrier}\}}{P\{n \text{ sons condition free}\}} = \frac{p}{p + 2^{-n}(1 - p)}. \]
Bayes Formula

Figure: Probability of mother being carrier free, given \( n \) sons are disease free for \( n = 1 \) (black), 2 (orange), 3 (red), 4 (magenta), and 5 (blue), The vertical dashed line at \( p = 1/2 \) is the case for the boxes, one with a fair coin and one with a two-headed coin.