04: Conditional Probability and Bayes

Jerry Cain
April 8th, 2024

Lecture Discussion on Ed
Conditional Probability
Dice, our misunderstood friends

Roll two, fair 6-sided dice, yielding values $D_1$ and $D_2$.

Let $E$ be event: $D_1 + D_2 = 4$.

What is $P(E)?$ 

| $S$ | $|D_1| = 6$ | $|D_2| = 6$ |
|-----|-------------|-------------|
| $|S| = |D_1||D_2| = 36$ |

$E = \{(1,3), (2,2), (3,1)\}$

$P(E) = 3/36 = 1/12$

Let $F$ be event: $D_1 = 2$.

What is $P(E, \text{knowing } F \text{ already observed})?$

$F = \{(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)\}, |F| = 6$

$E = \{(2,2)\}$ when only options are those in $F$.

$P(E, \text{knowing } F \text{ already happened}) = \frac{1}{6}$
Conditional Probability

The conditional probability of \( E \) given \( F \) is the probability that \( E \) occurs given that \( F \) has already occurred. This is known as conditioning on \( F \).

Written as: \( P(E|F) \)

Means: "\( P(E, \text{ knowing } F \text{ already observed}) \)"

Sample space \( \rightarrow \) all possible outcomes in \( F \)

Event \( \rightarrow \) all possible outcomes in \( E \cap F \)
Conditional Probability, equally likely outcomes

The **conditional probability** of $E$ given $F$ is the probability that $E$ occurs given that $F$ has already occurred. This is known as conditioning on $F$.

With **equally likely outcomes**:

$$P(E|F) = \frac{\text{# of outcomes in } E \text{ consistent with } F}{\text{# of outcomes in } S \text{ consistent with } F} = \frac{|E \cap F|}{|S \cap F|} = \frac{|E \cap F|}{|F|}$$

$$P(E|F) = \frac{|EF|}{|F|}$$

$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) = \frac{3}{14} \approx 0.21$$
## Slicing up the spam

24 emails are sent, 6 each to 4 users.
- 10 of the 24 emails are spam.
- All possible outcomes are equally likely.

<table>
<thead>
<tr>
<th>Event</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$</td>
<td>User 1 receives 3 spam emails.</td>
</tr>
<tr>
<td>$F$</td>
<td>User 2 receives 6 spam emails.</td>
</tr>
<tr>
<td>$G$</td>
<td>User 3 receives 5 spam emails.</td>
</tr>
</tbody>
</table>

### What is $P(E)$?

$E = \binom{10}{3} \binom{14}{3}$

### What is $P(E|F)$?

Knowing that $F$ has happened, only 4 spam emails are available to user 1, but all 14 legitimate emails are still available.

### What is $P(E|F)$?

Given that 6 of 10 spam emails have already been directed to user 2, it is impossible for user 3 to receive more than 4 spam.
Slicing up the spam

24 emails are sent, 6 each to 4 users.
- 10 of the 24 emails are spam.
- All possible outcomes are equally likely.

Let $E = \text{user 1 receives 3 spam emails.}$
What is $P(E)$?

$$P(E) = \frac{\binom{10}{3}\binom{14}{6}}{\binom{24}{6}} \approx 0.3245$$

Let $F = \text{user 2 receives 6 spam emails.}$
What is $P(E|F)$?

$$P(E|F) = \frac{\binom{4}{3}\binom{14}{18}}{\binom{18}{6}} \approx 0.0784$$

Let $G = \text{user 3 receives 5 spam emails.}$
What is $P(G|F)$?

$$P(G|F) = \frac{\binom{4}{5}\binom{14}{1}}{\binom{18}{6}} = 0$$

No way to choose 5 spam from 4 remaining spam emails!
Conditional probability in general

General definition of conditional probability:

\[ P(E | F) = \frac{P(EF)}{P(F)} \]

The Chain Rule (aka Product rule):

\[ P(EF) = P(F)P(E | F) \]

These properties hold even when outcomes are not equally likely.
Netflix and Learn

Let $E$ = a user watches Life is Beautiful. What is $P(E)$?

Equally likely outcomes? $S = \{\text{watch, not watch}\}$

$E = \{\text{watch}\}$

$P(E) = 1/2$ ?

$P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \approx \frac{\# \text{ people who have watched movie}}{\# \text{ people on Netflix}}$

$= \frac{10,234,231}{50,923,123} \approx 0.20$
Netflix and Learn

Let $E$ be the event that a user watches the given movie.

$P(E|F) = \frac{P(EF)}{P(F)}$  
Definition of Cond. Probability

$P(E) = 0.19$  
$P(E) = 0.32$  
$P(E) = 0.20$  
$P(E) = 0.09$  
$P(E) = 0.20$
Netflix and Learn

Let $E$ = a user watches Life is Beautiful.
Let $F$ = a user watches Amelie.

What is the probability that a user watches Life is Beautiful, given they watched Amelie?

$$P(E|F)$$

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\# \text{ people who have watched both}}{\# \text{ people on Netflix}} \cdot \frac{\# \text{ people who have watched Amelie}}{\# \text{ people on Netflix}}$$

$$\approx 0.42$$
Netflix and Learn

Let $E$ be the event that a user watches the given movie. Let $F$ be the event that the same user watches Amelie.

\[
P(E|F) = \frac{P(EF)}{P(F)} \quad \text{Definition of Cond. Probability}
\]

\[
P(E) = 0.19 \\
P(E) = 0.32 \\
P(E) = 0.20 \\
P(E) = 0.09 \\
P(E) = 0.20 \\

P(E|F) = 0.14 \\
P(E|F) = 0.35 \\
P(E|F) = 0.20 \\
P(E|F) = 0.72 \\
P(E|F) = 0.42
\]
Law of Total Probability
Today’s tasks

\[ P(EF) \]

- Law of Total Probability
- Chain rule (Product rule)
- Definition of conditional probability

\[ P(E|F) \]

\[ P(E) \]
Law of Total Probability

**Thm** Let $F$ be an event where $P(F) > 0$. For any event $E$,

$$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$$

**Proof**

1. $F, F^C$ are disjoint such that $F \cup F^C = S$  
   Def. of complement
2. $E = (EF) \cup (EF^C)$  
   (see diagram)
3. $P(E) = P(EF) + P(EF^C)$  
   Additivity axiom
4. $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$  
   Chain rule (product rule)

Note: disjoint sets are, by definition, mutually exclusive events
**General Law of Total Probability**

**Thm**  
For *mutually exclusive events* $F_1, F_2, \ldots, F_n$ such that $F_1 \cup F_2 \cup \ldots \cup F_n = S$,  

$$P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

\[ P(E|F_i) \]

Assume that $n = 5$ in this example, $E \cap F_1 = E \cap F_5 = \emptyset$  

\[ E \cap F_2 \supset E \cap F_3 \supset E \cap F_4 \]
Finding $P(E)$ from $P(E|F)$

- Flip a fair coin.
- If heads: roll a fair 6-sided die.
- Else: roll a fair 3-sided die.

You win if you roll a 6. What is $P($winning$)$?

$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$

Law of Total Probability
Finding $P(E)$ from $P(E|F)$

- Flip a fair coin.
- If heads: roll a fair 6-sided die.
- Else: roll a fair 3-sided die.

You win if you roll a 6. What is $P(\text{winning})$?

1. Define events & state goal

   Let: $E$: win, $F$: flip heads
   Want: $P(\text{win}) = P(E)$

2. Identify known probabilities

   $P(\text{win}|H) = P(E|F) = \frac{1}{6}$
   $P(H) = P(F) = \frac{1}{2}$
   $P(\text{win}|T) = P(E|F^C) = 0$
   $P(T) = P(F^C) = 1 - \frac{1}{2}$

3. Solve

   $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$

   $P(E) = \left(\frac{1}{6}\right)\left(\frac{1}{2}\right) + (0)\left(\frac{1}{2}\right)$

   $= \frac{1}{12} \approx 0.083$
Bayes’ Theorem
Today’s tasks

- Law of Total Probability
- Chain rule (Product rule)
- Definition of conditional probability

Rev. Thomas Bayes (~1701-1761): British mathematician and Presbyterian minister

Bayes’ Theorem

\[ P(E|F) \]

\[ P(E) \]

\[ P(F|E) \]
Detecting spam email

We can easily calculate how many existing spam emails contain "Dear":

\[ P(E|F) = P\left(\text{"Dear"} \mid \text{Spam email}\right) \]

But what is the probability that a mystery email containing "Dear" is spam?

\[ P(F|E) = P\left(\text{Spam email} \mid \text{"Dear"}\right) \]
Bayes’ Theorem

Thm For any events $E$ and $F$ where $P(E) > 0$ and $P(F) > 0,$

\[
P(F|E) = \frac{P(E|F)P(F)}{P(E)}
\]

Proof 2 steps!

1. \( P(F|E) = \frac{P(F \cap E)}{P(E)} \)

2. \( \frac{P(F \cap E)}{P(E)} = \frac{P(E|F)P(F)}{P(F)} \)

Expanded form:

\[
P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}
\]

Proof 1 more step!

denominator is just $P(E)$ expanded using LOTP
Detecting spam email

- 60% of all email in 2016 is spam.
- 20% of spam has the word "Dear"
- 1% of non-spam (aka ham) has the word "Dear"

You get an email with the word "Dear" in it.

What is the probability that the email is spam?

1. Define events & state goal
2. Identify known probabilities
3. Solve

Let: \( E \): "Dear", \( F \): spam
Want: \( P(\text{spam} \mid \text{"Dear"}) = P(F \mid E) \)

\[
P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid F^c)P(F^c)}
\]

\[
P(E) = 0.6

P(E \mid F) = 0.2

P(E \mid F^c) = 0.01
\]

\[
P(F \mid E) = \frac{(0.2)(0.6)}{(0.2)(0.6) + (0.01)(0.4)} = 0.967
\]
Bayes’ Theorem terminology

- 60% of all email in 2016 is spam.  
- 20% of spam has the word "Dear"
- 1% of non-spam (aka ham) has the word "Dear"

You get an email with the word "Dear" in it. What is the probability that the email is spam?  

\[ P(F|E) = \frac{P(E|F)P(F)}{P(E)} \]

Want: \( P(F|E) \)
Bayes’ Theorem II
This class going forward

Last week
Equally likely events

\[ P(E \cap F) \quad P(E \cup F) \]
(counting, combinatorics)

Today and for most of this course
Events not always equally likely

\[ P(E = \text{Evidence} \mid F = \text{Fact}) \]
(collected from data)

\[ P(F = \text{Fact} \mid E = \text{Evidence}) \]
(categorize a new datapoint)

Bayes’
Bayes’ Theorem

\[ P(F|E) = \frac{P(E|F)P(F)}{P(E)} \]

Mathematically:

\[ P(E|F) \rightarrow P(F|E) \]

Real-life application:

Given new evidence \( E \), update belief of fact \( F \)
Prior belief \( P(F) \rightarrow P(F|E) \)
Zika, an autoimmune disease


If a test returns positive, what is the likelihood you have the disease?

Ziika Forest, Uganda

Rhesus monkeys

### Taking tests: Confusion matrix

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Fact</th>
<th>Evidence, $E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E, \text{Test } +$</td>
<td>$F$, disease +</td>
<td>True positive $P(E</td>
</tr>
<tr>
<td>$E^C, \text{Test } -$</td>
<td>$F^C$, disease −</td>
<td>False negative $P(E^C</td>
</tr>
</tbody>
</table>

If a test returns positive, what is the likelihood you have the disease?
## Taking tests: Confusion matrix

If a test returns positive, what is the likelihood you have the disease?

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Fact</th>
<th>Evidence, $E$</th>
<th>Test positive</th>
<th>Evidence, $E^C$</th>
<th>Test negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$, Test +</td>
<td>$F$, disease +</td>
<td>True positive</td>
<td>$P(E</td>
<td>F)$</td>
<td>False positive</td>
</tr>
<tr>
<td>$E^C$, Test -</td>
<td>$F^C$, disease -</td>
<td>False negative</td>
<td>$P(E^C</td>
<td>F)$</td>
<td>True negative</td>
</tr>
</tbody>
</table>
Zika Testing

- A test is 98% effective at detecting Zika ("true positive").
- However, the test has a "false positive" rate of 1%.
- 0.5% of the US population has Zika.

What is the likelihood you have Zika if you test positive?

Why would you expect this number?

1. Define events & state goal

Let: \( E \) = you test positive
\( F \) = you actually have the disease

Want: \( P(\text{disease} \mid \text{test+}) \)
\( = P(F \mid E) \)

\[
P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid F^C)P(F^C)} \quad \text{Bayes' Theorem}
\]
Zika Testing

- A test is 98% effective at detecting Zika ("true positive").
- However, the test has a "false positive" rate of 1%.
- 0.5% of the US population has Zika.

What is the likelihood you have Zika if you test positive?

Why would you expect this number?

1. Define events & state goal
   Let: $E = \text{you test positive}$
   $F = \text{you actually have the disease}$

2. Identify known probabilities
   - $P(E) = 0.005$
   - $P(F|E) = 0.98$
   - $P(E|F) = 0.01$

3. Solve
   $$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)}$$
   \[= \frac{(0.98)(0.005)}{(0.98)(0.005)+(0.01)(0.995)}\]
   $\approx 0.030$

Bayes’ Theorem
Bayes’ Theorem intuition

**Original question:**
What is the likelihood you have Zika if you test positive for the disease?
Bayes’ Theorem intuition

**Original question:**
What is the likelihood you have Zika if you test positive for the disease?

**Interpretation:**
Of the people who test positive, how many actually have Zika?
Bayes’ Theorem intuition

Original question:
What is the likelihood you have Zika if you test positive for the disease?

Interpretation:
Of the people who test positive, how many actually have Zika?

The space of facts, conditioned on a positive test result.
Update your beliefs with Bayes’ Theorem

\[ E = \text{you test positive for Zika} \]
\[ F = \text{you have the disease} \]

I have a 0.5% chance of having Zika.

Take test, results positive

With these test results, I now have a 33% chance of having Zika!!!
Why it’s still good to get tested

- A test is 98% effective at detecting Zika (“true positive”).
- However, the test has a “false positive” rate of 1%.
- 0.5% of the US population has Zika.

Let:

- $E = \text{you test positive}$
- $F = \text{you actually have the disease}$

Let: $E^C = \text{you test negative for Zika with this test.}$

<table>
<thead>
<tr>
<th>$F$, disease +</th>
<th>$F^C$, disease -</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$, Test +</td>
<td>True positive</td>
</tr>
<tr>
<td></td>
<td>$P(E</td>
</tr>
</tbody>
</table>

What is $P(F|E^C)$?

Bayes’ Theorem

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$
Why it’s still good to get tested

A test is 98% effective at detecting Zika (“true positive”).
However, the test has a “false positive” rate of 1%.
0.5% of the US population has Zika.

Let: \( E \) = you test positive
\( F \) = you actually have the disease

Let: \( E^C \) = you test negative for Zika with this test.

What is \( P(F|E^C) \)?

\[
P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}
\]

Bayes’ Theorem
Why it’s still good to get tested

- A test is 98% effective at detecting Zika ("true positive").
- However, the test has a "false positive" rate of 1%.
- 0.5% of the US population has Zika.

Let:
- $E$ = you test positive
- $F$ = you actually have the disease

Let:
- $E^C$ = you test negative for Zika with this test.

What is $P(F|E^C)$?

\[
P(F|E^C) = \frac{P(E^C|F)P(F)}{P(E^C|F)P(F) + P(E^C|F^C)P(F^C)}
\]

<table>
<thead>
<tr>
<th></th>
<th>$F$, disease +</th>
<th>$F^C$, disease -</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$, Test +</td>
<td>True positive</td>
<td>False positive</td>
</tr>
<tr>
<td>$P(E</td>
<td>F)$</td>
<td>0.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$E^C$, Test -</th>
<th>False negative</th>
<th>True negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(E^C</td>
<td>F)$</td>
<td>0.02</td>
</tr>
</tbody>
</table>

\[
P(F|E^C) = \frac{(0.02)(0.005)}{(0.02)(0.005) + (0.99)(0.995)} = 0.001
\]
Why it’s still good to get tested

\[ E = \text{you test positive for Zika} \]
\[ F = \text{you actually have the disease} \]
\[ E^C = \text{you test negative for Zika} \]

I have a 0.5\% chance of having Zika disease.

\[ P(F) \]

Take test, results positive

With these test results, I now have a 33\% chance of having Zika!!!

\[ P(F|E) \]

Take test, results negative

With these test results, I now have a 0.01\% chance of having Zika disease!!!

\[ P(F|E^C) \]