04: Conditional Probability and Bayes

Jerry Cain
April 8th, 2024

Lecture Discussion on Ed
Conditional Probability
Roll two, fair 6-sided dice, yielding values $D_1$ and $D_2$.

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

What is $P(E)$?

$|S| = 36$

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

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

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

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

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

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

\[
P(E|F) = \frac{3}{14} \approx 0.21
\]
24 emails are sent, 6 each to 4 users.

- 10 of the 24 emails are spam.
- All possible outcomes are equally likely.

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

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

Let $G =$ user 3 receives 5 spam emails.
What is $P(G|F)$?
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 $P(E)$?

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

Let $F = \text{user 2 receives 6 spam emails.}$

What is $P(E|F)$?

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

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!

Lisa Yan, Chris Piech, Mehran Sahami, and Jerry Cain, CS109, Spring 2024
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
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.

$P(E) = 0.19$  
$P(E) = 0.32$  
$P(E) = 0.20$  
$P(E) = 0.09$  
$P(E) = 0.20$
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?

$$P(E|F)$$

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

$$= \frac{\text{# people who have watched both}}{\text{# people who have watched Amelie}}$$

$$\approx 0.42$$
Let $E$ be the event that a user watches the given movie. Let $F$ be the event that the same user watches Amelie.

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

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

\[ P(EF) \]

Chain rule
(Product rule)

\[ P(E|F) \]

Definition of conditional probability

Law of Total Probability

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

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 $F \cup F^C = S$ \hspace{1cm} \text{Def. of complement}
2. $E = (EF) \cup (EF^C)$ \hspace{1cm} \text{(see diagram)}
3. $P(E) = P(EF) + P(EF^C)$ \hspace{1cm} \text{Additivity axiom}
4. $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$ \hspace{1cm} \text{Chain rule (product rule)}

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

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

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

1. Define events & state goal
   
   Let: $E$: win, $F$: flip heads
   
   Want: $P$(win) = $P(E)$

2. Identify known probabilities

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

3. Solve

   $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$
   
   $P(E) = (1/6)(1/2) + (0)(1/2)$
   
   $= 1/12 \approx 0.083$

Law of Total Probability
Bayes’ Theorem I
Today’s tasks

Law of Total Probability

\[ P(E) \]

\[ P(\overline{E}) \]

\[ P(E|F) \]

\[ P(F|\overline{E}) \]

\[ P(EF) \]

Definition of conditional probability

Chain rule
(Product rule)

Bayes’ Theorem

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

Lisa Yan, Chris Piech, Mehran Sahami, and Jerry Cain, CS109, Spring 2024
Detecting spam email

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

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

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

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

**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!
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
   Let: $E$: "Dear", $F$: spam
   Want: $P(\text{spam} | \text{"Dear"})$
   $= P(F | E)$

2. Identify known probabilities

3. Solve

Bayes’ Theorem

$$P(F | E) = \frac{P(E | F)P(F)}{P(E | F)P(F) + P(E | F^c)P(F^c)}$$
Bayes’ Theorem terminology

- 60% of all email in 2016 is spam. \( P(F) \)
- 20% of spam has the word "Dear" \( P(E|F) \)
- 1% of non-spam (aka ham) has the word "Dear" \( P(E|F^C) \)

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

What is the probability that the email is spam? **Want:** \( P(F|E) \)

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

normalization constant
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

Bayes’

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

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

Mathematically:

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

Real-life application:

Given new evidence \( E \), update belief of fact \( F \)
Prior belief \( P(F) \) → Posterior belief \( P(F|E) \)
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, $F$</th>
<th>Has disease</th>
<th>Evidence, $E$</th>
<th>Test positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>or $F^C$</td>
<td>No disease</td>
<td>or $E^C$</td>
<td>Test negative</td>
</tr>
</tbody>
</table>

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

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

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

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

Lisa Yan, Chris Piech, Mehran Sahami, and Jerry Cain, CS109, Spring 2024
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:
- $E = \text{you test positive}$
- $F = \text{you actually have the disease}$

Want:
- $P(\text{disease | test+}) = P(F|E)$

\[ P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)} \text{ Bayes’ Theorem} \]
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
2. Identify known probabilities
3. Solve

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

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

People who test positive and have Zika

People who test positive but don’t have Zika

People who test positive
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: 
- $E^C = \text{you test negative for Zika with this test.}$

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

\[
P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)} \quad \text{Bayes’ Theorem}
\]

<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></td>
<td>$P(E</td>
<td>F) = 0.98$</td>
</tr>
</tbody>
</table>
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.

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

What is $P(F|E^C)$?
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:  
\[ E = \text{you test positive} \]
\[ F = \text{you actually have the disease} \]

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

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

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

Let the two outcomes be:

\[ F, \text{ disease +} \]
\[ F^C, \text{ disease -} \]

\[ E, \text{ Test +} \]
\[ E^C, \text{ Test -} \]

\[
\begin{array}{c|c|c}
 F & F, \text{ disease +} & F^C, \text{ disease -} \\
\hline
 E, \text{ Test +} & \text{True positive} & \text{False positive} \\
P(E|F) & 0.98 & P(E|F^C) = 0.01 \\
\hline
 E^C, \text{ Test -} & \text{False negative} & \text{True negative} \\
P(E^C|F) & 0.02 & P(E^C|F^C) = 0.99 \\
\end{array}
\]
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!!!

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