04: Conditional Probability and Bayes

David Varodayan
January 13, 2020
Adapted from slides by Lisa Yan
Key definitions

An experiment in probability:

Sample Space, $S$: The set of all possible outcomes of an experiment.
Event, $E$: Some subset of $S$ ($E \subseteq S$).

We have the power to redesign our experiment, provided we can recreate the set of outcomes!
Card Flipping

In a 52 card deck, cards are flipped one at a time. After the first ace (of any suit) appears, consider the next card. Is $P(\text{next card} = \text{Ace Spades}) < P(\text{next card} = 2 \text{ Clubs})$?
Card Flipping

In a 52 card deck, cards are flipped one at a time. After the first ace (of any suit) appears, consider the next card.

Is $P(\text{next card} = \text{Ace Spades}) < P(\text{next card} = \text{2 Clubs})$?

Sample space $S = 52$ in-order cards (shuffle deck)

Event

$E_{AS}$, next card is Ace Spades
1. Take out Ace of Spades.
2. Shuffle leftover 51 cards.
3. Add Ace Spades after first ace.

$|E_{AS}| = 51! \cdot 1$

$E_{2C}$, next card is 2 Clubs
1. Take out 2 Clubs.
2. Shuffle leftover 51 cards.
3. Add 2 Clubs after first ace.

$|E_{2C}| = 51! \cdot 1$

$P(E_{AS}) = P(E_{2C})$

Equally likely outcomes

Stanford University
Today’s plan

Conditional Probability and Chain Rule

Law of Total Probability

Bayes’ Theorem
Dice, our misunderstood friends

Roll two 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 $P(E, \text{ given } F \text{ already observed})$?

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

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

$P(E, \text{ given } F \text{ observed}) = \frac{1}{6}$
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: “\( P(E, \text{given } F \text{ already observed}) \)”

Sample space \( \rightarrow \) all possible outcomes consistent with \( F \) (i.e. \( S \cap F \))

Event space \( \rightarrow \) all outcomes in \( E \) consistent with \( F \) (i.e. \( E \cap F \))
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:

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

\[
= \frac{|EF|}{|F|}
\]

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

\[
P(E|F) = \frac{3}{14} \approx 0.21
\]
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.

Let $E$ = user 1 receives 3 spam emails.

What is $P(E)$?

$$P(E) = \frac{\binom{10}{3}\binom{4}{3}}{\binom{24}{6}} \approx 0.3245$$

Let $F$ = user 2 receives 6 spam emails.

What is $P(E|F)$?

$$P(E|F) = \frac{\binom{4}{3}\binom{14}{3}}{\binom{18}{6}} \approx 0.0794$$

Let $G$ = user 3 receives 5 spam emails.

What is $P(G|F)$?

$$P(G|F) = 0$$

No way to choose 5 spam emails from 4.

Stanford University
Quick check

You have a flowering plant.
Let \( E = \) Flowers bloom
\( F = \) It gets watered
\( G = \) It gets sun

In English, how do you interpret \( P(E|FG) \)?

The probability that...

A. ...flowers bloom given the probability that it gets water and it gets sun
B. ...flowers bloom given it gets watered given it gets sun
C. ...flowers bloom given (it gets watered and it gets sun)
D. All/none/other
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 $E = \text{a user watches Life is Beautiful}$. What is $P(E)$?

$$P(E) \approx \frac{\# \text{people who have watched movie}}{\# \text{people on Netflix}}$$

$$= \frac{10,234,231}{50,923,123} \approx 0.20$$
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?

$$P(E|F)$$

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

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

$$\approx 0.42 \quad \text{(data)}$$
Let $E$ be the event that a user watches the given movie. Let $F$ be the event that the same user watches Amelie.

\[
P(E) = 0.19 \\
P(E|F) = 0.14
\]

\[
P(E) = 0.32 \\
P(E|F) = 0.35
\]

\[
P(E) = 0.20 \\
P(E|F) = 0.20
\]

\[
P(E) = 0.09 \\
P(E|F) = 0.72
\]

\[
P(E) = 0.20 \\
P(E|F) = 0.42
\]

**Definition of Cond. Probability**

\[
P(E|F) = \frac{P(EF)}{P(F)}
\]
Today’s plan

Conditional Probability and Chain Rule

Law of Total Probability

Bayes’ Theorem
Today’s plan in pictures

\[ P(EF) \]

Chain rule
(Product rule)

Definition of conditional probability

\[ P(E|F) \]

Law of Total Probability

\[ P(E) \]

Bayes’ Theorem

\[ P(F|E) \]
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. $F$ and $F^C$ are disjoint s.t. $F \cup F^C = S$  
   Def. of complement
2. $E = (EF) \cup (EF^C)$  
   (see below)
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)
General Law of Total Probability

Thm For disjoint events $F_1, F_2, ..., F_n$ s.t. $F_1 \cup F_2 \cup \cdots \cup F_n = S$,

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

Let \( E: \text{win} \)
\( F: \text{flip heads} \)

Want \( P(E) \)

So \( P(E) = \frac{1}{6} \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{12} \)

\[
P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)
\]

Law of Total Probability

We know
\[
P(E|F) = P(\text{win | H}) = \frac{1}{6}
\]
\[
P(F) = \frac{1}{2}
\]
\[
P(E|F^c) = P(\text{win | T}) = 0
\]
\[
P(F^c) = P(T) = \frac{1}{2}
\]
Announcements

Section sign-ups
Results: later today
Late signups/changes: later today

Problem set 1 autograder issues
Read problem carefully: see pinned Piazza post
Syntax issue: \texttt{np.random.randint(1, 101)}
Today’s plan

Conditional Probability and Chain Rule

Law of Total Probability

Bayes’ Theorem
Today’s tasks

\[ P(EF) \]

- Chain rule (Product rule)
- Definition of conditional probability

\[ P(E|F) \]

- Law of Total Probability
- Bayes’ Theorem

\[ P(E) \]

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

Spam volume as percentage of total email traffic worldwide

We can easily calculate how many spam emails contain “Dear”:

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

But what is the probability that an email containing “Dear” is spam?

\[ P(F|E) = P(\text{Spam email} | \text{“Dear”}) \]
Bayes’ Theorem

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

\[
\begin{align*}
P(EF) &= P(E|F)P(F) \\
- P(FE) &= P(F|E)P(E) \\
\hline
P(F|E) &= \frac{P(E|F)P(F)}{P(E)}
\end{align*}
\]
Bayes’ Theorem (expanded form)

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

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)}$$

**Proof**

Substitute Law of Total Probability into Bayes’ Theorem
Detecting spam email

- 60% of all email in 2019 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?

Let E: “Dear”
F: spam

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

\[
P(F) = 0.6 \quad P(F^c) = 0.4
\]

\[
P(E|F) = 0.2 \quad P(E|F^c) = 0.01
\]

\[
\text{Want } P(F|E) \approx 0.967
\]
Bayes’ Theorem terminology

- 60% of all email in 2019 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?  

Want: \( P(F|E) \)  posterior

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

\( P(F) \)  prior
\( P(E|F) \)  likelihood
\( P(E|F^C) \)
Zika, an autoimmune disease

A disease spread through mosquito bites. Usually no symptoms; worst case paralysis. During pregnancy: may cause birth defects

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

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

- **Fact**: $F$ or $F^C$
  - Has disease: $F$, disease +
  - No disease: $F^C$, disease -

- **Evidence**: $E$ or $E^C$
  - Test positive: $E$, Test +
  - Test negative: $E^C$, Test -

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

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

- **False negative** $P(E^C|F)$
- **True negative** $P(E^C|F^C)$
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?

$$E: \text{test positive}$$
$$F: \text{actually positive}$$

Want $$P(F|E)$$

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

$$P(E|F) = 0.98$$
$$P(F|E^C) = 0.01$$
$$P(F) = 0.005$$
$$P(F^C) = 0.995$$

$$P(F|E) = \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.01)(0.995)}$$

$$\approx 0.330$$
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?

Interpret

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

The space of facts, conditioned on a positive test result

Stanford University
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?

Say we have 1000 people:

- 5 have Zika and test positive
- 985 do not have Zika and test negative.
- 10 do not have Zika and test positive.

\[ \approx 0.333 \]
Update your beliefs with Bayes’ Theorem

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

I have a 0.5% chance of having Zika.

\[ P(F) \]

Take test, results positive

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

\[ P(F|E) \]
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
- \( E^C \) = you test negative for Zika with this test.

### Bayes’ Theorem

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

<table>
<thead>
<tr>
<th>( F ), disease +</th>
<th>( F^C ), disease –</th>
</tr>
</thead>
</table>
| \( E \), Test +     | True positive  
| \( P(E|F) = 0.98 \) | False positive \n| \( P(E|F^C) = 0.01 \) |

What is \( P(F|E^C) \)? ≠ 1 - \( P(F|E) \)

\[\text{unrelated}\]
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 
- $E^C$ = you test negative for Zika with this test.

<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>$P(E</td>
<td>F') = 0.98$</td>
<td></td>
</tr>
<tr>
<td>$E^C$, Test –</td>
<td>False negative</td>
<td>True negative</td>
</tr>
<tr>
<td>$P(E^C</td>
<td>F) = 0.02$</td>
<td></td>
</tr>
</tbody>
</table>

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

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

Bayes’ Theorem

- $P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$ 
- $P(E|F)$ 
- $P(E|F^C)$ 
- $P(F)$ 

Stanford University
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 \textbf{0.5\%} chance of having Zika disease.

\[ P(F) \]

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

\[ P(F|E) \]

Take test, results \textbf{positive}

Take test, results \textbf{negative}

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

\[ P(F|E^C) \]
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
Not equally likely events

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

\[ P(F = \text{Fact} \mid E = \text{Evidence}) \]
(categorize a new datapoint)
Monty Hall, 1000 envelope version

Start with 1000 envelopes (of which 1 is the prize).

1. You choose 1 envelope.

\[
\begin{align*}
\frac{1}{1000} &= P(\text{envelope is prize}) \\
\frac{999}{1000} &= P(\text{other 999 envelopes have prize})
\end{align*}
\]

2. I open 998 of remaining 999 (showing they are empty).

\[
\begin{align*}
\frac{999}{1000} &= P(\text{998 empty envelopes had prize}) \\
&+ P(\text{last other envelope has prize}) \\
&= P(\text{last other envelope has prize})
\end{align*}
\]

3. Should you switch?

\[
\begin{align*}
P(\text{you win without switching}) &= \frac{1}{\text{original # envelopes}} \\
P(\text{you win with switching}) &= \frac{1}{\text{original # envelopes} - 1}
\end{align*}
\]