04: Conditional Probability and Bayes

Lisa Yan and Jerry Cain
September 21, 2020
<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Conditional Probability + Chain Rule</td>
<td>04a Conditional</td>
</tr>
<tr>
<td>15</td>
<td>Law of Total Probability</td>
<td>04b Total Prob</td>
</tr>
<tr>
<td>22</td>
<td>Bayes’ Theorem I</td>
<td>04c Bayes I</td>
</tr>
<tr>
<td>31</td>
<td>Bayes’ Theorem II</td>
<td>LIVE</td>
</tr>
<tr>
<td>61</td>
<td>Monty Hall Problem</td>
<td>LIVE</td>
</tr>
</tbody>
</table>
Conditional Probability
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 $P(E \text{, given } F \text{ already observed})$?
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: “$P(E$, given $F$ already observed)”

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

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

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

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

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

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

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 $P(G|F)$?

Equally likely outcomes

\[ P(E|F) = \frac{|EF|}{|F|} \]
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!
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.
and Learn
Netflix and Learn

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

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

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

$P(E) = 1/2$ ?

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

$= 10,234,231 / 50,923,123 \approx 0.20$
Let $E$ be the event that a user watches the given movie.

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

$$P(E|F)$$

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

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

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

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

$P(E) = 0.19$ \hspace{1cm} $P(E) = 0.32$ \hspace{1cm} $P(E) = 0.20$ \hspace{1cm} $P(E) = 0.09$ \hspace{1cm} $P(E) = 0.20$

$P(E|F) = 0.14$ \hspace{1cm} $P(E|F) = 0.35$ \hspace{1cm} $P(E|F) = 0.20$ \hspace{1cm} $P(E|F) = 0.72$ \hspace{1cm} $P(E|F) = 0.42$
Law of Total Probability
Today’s tasks

\[ P(E|F) \]

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

\[ P(E) \]

- Law of Total Probability

\[ P(EF) \]
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. $F$ and $F^C$ are disjoint s.t. $F \cup F^C = S$  
   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 by definition are mutually exclusive events
General Law of Total Probability

Thm  For **mutually exclusive 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 $P($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 \( P(\text{winning}) \)?

1. Define events & state goal

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

Want: \( P(\text{win}) = P(E) \)

2. Identify known probabilities

\[
\begin{align*}
P(\text{win} | H) &= P(E | F) = \frac{1}{6} \\
P(H) &= P(F) = \frac{1}{2} \\
P(\text{win} | T) &= P(E | F^C) = 0 \\
P(T) &= P(F^C) = 1 - \frac{1}{2} \\
\end{align*}
\]

3. Solve

\[
P(E) = P(E | F)P(F) + P(E | F^C)P(F^C) \\
= \frac{1}{6} \cdot \frac{1}{2} + 0 \cdot \left(1 - \frac{1}{2}\right) \\
= \frac{1}{12} \approx 0.083
\]
Finding \( P(E) \) from \( P(E|F) \), an understanding

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

   “Probability trees” can help connect your understanding of the experiment with the problem statement.
Bayes’ Theorem
I
Today’s tasks

**Law of Total Probability**

\[ P(E) \]

\[ P(E|F) \]

\[ P(EF) \]

**Chain rule**

(Product rule)

**Definition of conditional probability**

\[ P(F|E) \]

**Bayes’ Theorem**
Thomas Bayes

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

He looked remarkably similar to Charlie Sheen
(but that’s not important right now)
Detecting spam email

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

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

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

\[ P(F|E) = P(\text{Spam email}|“Dear”) \]
(silent drumroll)
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! See board

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! See board
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: $P(\text{spam} | \text{“Dear”})$

$= P(F|E)$
Detecting spam email, an understanding

- 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: $P(\text{spam} | \text{“Dear”})$

$= P(F | E)$

**Note:** You should still know how to use Bayes/ Law of Total Probab., but drawing a probability tree can help you identify which probabilities you have. The branches are determined using the problem setup.
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: \( P(F|E) \)

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

- posterior
- likelihood \( P(E|F) \)
- prior \( P(F) \)
- normalization constant \( P(E) \)