Conditional Probability
Reading

  – Chapter 16
Overview

- New information changes a probability
- Definition of conditional probability from regular probability
- Conditional probability traps
  - Sampling bias
  - Transposed conditional
- Law of total probability
  - Probabilistic case-by-case analysis
Suppose I told you that, in general, chances a random person has the flu is about 0.01 (or 1%) (prior probability).

1. Probability of flu: 
   \[ \mathbb{P}[\text{flu}] \approx 0.01 \]

2. Suppose now that you also have a slight fever
   - This is new information. Chances of flu “increase”.

3. Suppose I know that the probability of flu given fever: 
   \[ \mathbb{P}[\text{flu} | \text{fever}] \approx 0.4 \]
   - New information changes the prior probability to the posterior probability.
   - Translate posterior as “After you get the new information.”

4. \( \mathbb{P}[A|B] \) is the (updated) conditional probability of \( A \), given the new information \( B \)

5. Suppose also your roommate has flu (more new information).
   - Flu almost surely!

6. Probability of flu given fever and roommate flu: 
   \[ \mathbb{P}[\text{flu} | \text{fever and roommate flu}] \approx 1 \]
CS, MATH and Dual CS-MATH Majors

• Suppose RPI has 5000 students total, with 1000 in CS (vast underestimate!), 100 in MATH, 80 dual CS-MATH

• If you pick a random student from RPI, what’s the chance the student is in CS
  \[ \Pr[CS] = \frac{1000}{5000} = 0.2 \]
  \[ \Pr[MATH] = \frac{100}{5000} = 0.02 \]
  \[ \Pr[CS \text{ and } MATH] = \frac{80}{5000} = 0.016 \]

• Suppose (after you picked the student), I told you the student is a MATH major
  – New information. What is \( \Pr[CS|MATH] \)?
    • Effectively picking a random student from MATH
    • New probability of CS \( \sim \) striped area \( |CS \cap MATH| \)
    \[ \Pr[CS|MATH] = \frac{\Pr[CS \cap MATH]}{\Pr[MATH]} = \frac{80}{100} = 0.8 \]
  – MATH majors 4 times more likely to be CS majors than a random student.

• Exercise 16.2.
Conditional Probability $\mathbb{P}[A|B]$

- Conditional probability is interpreted as follows:
  \[ \mathbb{P}[A|B] = \text{frequency of outcomes known to be in } B \text{ that are also in } A \]
- Suppose event $B$ contains $n_B$ outcomes when you repeat an experiment $n$ times:
  \[ \mathbb{P}[B] = \frac{n_B}{n} \]
  - Of the $n_B$ outcomes in $B$, the number also in $A$ is $n_{A\cap B}$
  \[ \mathbb{P}[A \cap B] = \frac{n_{A\cap B}}{n} \]
- The frequency of outcomes in $A$ among those outcomes in $B$ is $n_{A\cap B}/n_B$
  \[ \mathbb{P}[A|B] = \frac{n_{A\cap B}}{n_B} = \frac{n_{A\cap B}}{n} \times \frac{n}{n_B} = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \]
- The definition of conditional probability is then:
  \[ \mathbb{P}[A|B] = \frac{n_{A\cap B}}{n_B} = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[A \text{ and } B]}{\mathbb{P}[B]} \]
Chance of Rain Given Clouds

- It is cloudy one in five days, $\mathbb{P}[Clouds] = \frac{1}{5}$
- It rains one in seven days, $\mathbb{P}[Rain] = \frac{1}{7}$
- What are the chances of rain on a cloudy day?
  - Trick question!

\[
\mathbb{P}[Rain|Clouds] = \frac{\mathbb{P}[Rain \cap Clouds]}{\mathbb{P}[Clouds]}
\]

- We need to know $\mathbb{P}[Rain \cap Clouds]$
- It is mostly safe to assume that $\{Rainy\ Days\} \subseteq \{Cloudy\ Days\}$
- This means that $\mathbb{P}[Rain \cap Clouds] = \mathbb{P}[Rain]$

\[
\mathbb{P}[Rain|Clouds] = \frac{\mathbb{P}[Rain \cap Clouds]}{\mathbb{P}[Clouds]} = \frac{\frac{1}{7}}{\frac{1}{5}} = \frac{5}{7}
\]

- 5-times more likely to rain on a cloudy day than on a random day.
- Crucial first step: identify the conditional probability. What is the “new information”? 
Dicey Conditions

• Here’s an odd question:

\[ \mathbb{P}[\text{Sum of 2 dice is 10}|\text{Both are Odd}] \]

– Two dice have both rolled odd. What are the chances the sum is 10?

• First, write the definition of conditional probability

\[ \mathbb{P}[\text{Sum is 10}|\text{Both are Odd}] = \frac{\mathbb{P}[\text{(Sum is 10) AND (Both are Odd)}]}{\mathbb{P}[\text{Both are Odd}]} \]

• What is the probability space?

– Let’s get counting!

\[ \mathbb{P}[\text{Sum is 10}] = \frac{3}{36} = \frac{1}{12} \]
\[ \mathbb{P}[\text{Both are Odd}] = \frac{9}{36} = \frac{1}{4} \]
Here’s an odd question:

\[ P[\text{Sum of 2 dice is 10} | \text{Both are Odd}] \]

– Two dice have both rolled odd. What are the chances the sum is 10?

First, write the definition of conditional probability

\[ P[\text{Sum is 10} | \text{Both are Odd}] = \frac{P[(\text{Sum is 10}) \text{ AND (Both are Odd})]}{P[\text{Both are Odd}]} \]

What is the probability space?

– Let’s get counting!

\[ P[\text{Sum is 10}] = \frac{3}{36} = \frac{1}{12} \]

\[ P[\text{Both are Odd}] = \frac{9}{36} = \frac{1}{4} \]

\[ P[(\text{Sum is 10}) \text{ AND (Both are Odd})] = \frac{1}{36} \]

\[ P[\text{Sum is 10} | \text{Both are Odd}] = \frac{1}{36} \div \frac{1}{4} = \frac{1}{9} \]
Computing a Conditional Probability

1. Identify that you need a conditional probability $\mathbb{P}[A|B]$
2. Determine the probability space $(\Omega, \mathbb{P})$ using the outcome-tree method
3. Identify the events $A$ and $B$ appearing in $\mathbb{P}[A|B]$ as subsets of $\Omega$
4. Compute $\mathbb{P}[A \cap B]$ and $\mathbb{P}[B]$
5. Compute $\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$
Monty Prefers Door 3

- Recall the Monty Hall outcome tree
  - Suppose Monty prefers door 3
- What is the best strategy now?
  - You should still switch
  - Winning outcomes are the same!

\[ \mathbb{P}[\text{WinBySwitching}] = \frac{2}{3} \]

- Interestingly, you should feel better if you see Monty opening door 2!
  - Intuitively, why is this true?
  - Monty prefers door 3, so if he opened 2, then 3 very likely contains the prize!

\[ \mathbb{P}[\text{Win} | \text{Monty Opens Door 2}] = \frac{\mathbb{P}[\text{Win AND Monty Opens Door 2}]}{\mathbb{P}[\text{Monty Opens Door 2}]} = \frac{\frac{1}{3}}{\frac{1}{9} + \frac{1}{3}} = \frac{3}{4} \]

- Your chances improved from \( \frac{2}{3} \) to \( \frac{3}{4} \)!
A Pair of Slides

• Suppose I have to make 2 lecture slides in 2 minutes
  • The probability of having a typo on a slide is 0.5
    – What is the probability that both slides have typos?
    – Answer: \( \frac{1}{4} \)

• What is the probability that both slides have typos given the additional information in each of the cases below?

• At least one slide contains a typo
  – Answer: \( \frac{1}{3} \)

  \[
  \mathbb{P}[\text{2 typos}|\text{at least 1 typo}] = \frac{\mathbb{P}[\text{2 typos AND at least 1 typo}]}{\mathbb{P}[\text{at least 1 typo}]} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}
  \]

• The second slide has a typo
  – Answer: \( \frac{1}{2} \)

  \[
  \mathbb{P}[\text{2 typos}|\text{second slide has a typo}] = \frac{\mathbb{P}[\text{2 typos AND second slide has a typo}]}{\mathbb{P}[\text{second slide has a typo}]} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}
  \]

• I made one of the slides during office hours and made a typo
  – Answer: \( \frac{1}{2} \)

  \[
  \mathbb{P}[\text{2 typos}|\text{1 typo during OH}] = \frac{\mathbb{P}[\text{2 typos AND 1 typo during OH}]}{\mathbb{P}[\text{1 typo during OH}]} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}
  \]

• I wanted to illustrate my typo-making process, so I showed one slide with a typo
  – Answer: \( \frac{1}{3} \)

  \[
  \mathbb{P}[\text{2 typos}|\text{at least 1 typo}] = \frac{\mathbb{P}[\text{2 typos AND at least 1 typo}]}{\mathbb{P}[\text{at least 1 typo}]} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}
  \]

• Same question in each case, but with slightly different additional information.
Conditional Probability Traps

• These four probabilities are all different
  \[ \mathbb{P}[A] \quad \mathbb{P}[A|B] \quad \mathbb{P}[B|A] \quad \mathbb{P}[A \cap B] \]

• Be careful which one you use!!

• **Sampling bias:** Using \( \mathbb{P}[A] \) instead of \( \mathbb{P}[A|B] \)
  
  – Pollsters have a very tricky job when conducting surveys of public opinion
  – How do they contact people (landline, cell phone, social media, snail mail?)
  – Suppose you ask people “What is the probability that AI goes rogue?” and the only responses you get are through mail
  – What value are you surveying?
    \[ \mathbb{P}[\text{rogue AI|responder uses mail}] \]
    – This number is likely very different from \( \mathbb{P}[\text{rogue AI}] \)

• Transposed conditional: using \( \mathbb{P}[B|A] \) instead of \( \mathbb{P}[A|B] \)
  
  – See book for the famous Lombard study
The Covid Test and Transposed Conditionals

- At-home antigen covid tests have different accuracy depending on the case
  - If you have covid, the test will make a mistake ~20% of the time
  - If you don’t have covid, the test will make a mistake ~1% of the time
  - Source: https://www.cochrane.org/CD013705/INFECTN_how-accurate-are-rapid-antigen-tests-diagnosing-covid-19

- Suppose you tested positive. What are the chances you actually have covid?
  - If you don’t have covid, the test is unlikely to be wrong, so you are likely sick

- It is wrong to look at $\mathbb{P}[positive \mid not \, Covid]$. We already have this probability!
  - We need $\mathbb{P}[not \, Covid \mid positive]$
The Covid Test and Transposed Conditionals

• It is wrong to look at \( \mathbb{P}[\text{positive} | \text{not Covid}] \). We already have this probability!
  – We need \( \mathbb{P}[\text{not Covid} | \text{positive}] \)

\[
\mathbb{P}[\text{not Covid} | \text{positive}] = \frac{\mathbb{P}[\text{not Covid AND YES}]}{\mathbb{P}[\text{YES}]}
= \frac{0.9 \times 0.01}{0.1 \times 0.8 + 0.9 \times 0.01}
\approx 10\%
\]

• The less accurate test says YES but the chances are 90% that you have covid
  – Two possibilities:
    – You don’t have covid (likely) and test made a mistake (very rare)
    – You have covid (rare) and test was correct (likely). Wins!
Total Probability: Case by Case Probability

• Two types of outcomes in any event $A$:
  – The outcomes in $B$ (green)
  – The outcomes not in $B$ (red)

\[ \Pr[A] = \Pr[A \cap B] + \Pr[A \cap \bar{B}] \quad (*) \]

• (Similar to sum rule from counting)
• (Also known as marginalization)

• From the definition of conditional probability:

\[ \Pr[A \cap B] = \Pr[A \text{ AND } B] = \Pr[A \mid B] \times \Pr[B] \]
\[ \Pr[A \cap \bar{B}] = \Pr[A \text{ AND } \bar{B}] = \Pr[A \mid \bar{B}] \times \Pr[\bar{B}] \]

• Plugging these in $(*)$, we get a FUNDAMENTAL result for case-by-case analysis

• **LAW OF TOTAL PROBABILITY:**

\[ \Pr[A] = \Pr[A \mid B] \times \Pr[B] + \Pr[A \mid \bar{B}] \times \Pr[\bar{B}] \]

• (Weight conditional probabilities for each case by probabilities of each case and add.)
Three Coins: Two Are Fair, One is 2-Headed

• Pick a random coin and flip. What is the probability of H?
• First, let’s use the outcome-tree method as before

<table>
<thead>
<tr>
<th>Coin</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>biased</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>T</td>
</tr>
<tr>
<td>fair</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>T</td>
</tr>
</tbody>
</table>

• What is the probability of H?

\[
\mathbb{P}[H] = \frac{1}{3} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3}
\]
Three Coins: Two Are Fair, One is 2-Headed, cont’d

- Pick a random coin and flip. What is the probability of H?
- Now, let’s use the law of total probability:
  - Case 1. \( B \): You picked one of the fair coins
  - Case 2. \( \bar{B} \): You picked the two-headed coin

\[
P[H] = P[H|B] \times P[B] + P[H|\bar{B}] \times P[\bar{B}]
\]

\[
= \frac{1}{2} \times \frac{2}{3} + 1 \times \frac{1}{3} = \frac{2}{3}
\]

- Note that we don’t have to draw the (potentially exponentially growing) outcome tree anymore

- **Exercise.** A box has 10 coins: 6 fair and 4 biased (probability of heads \( \frac{2}{3} \)). What is \( P[2 \text{ heads}] \) in each case?
  - Pick a single random coin and flip it 3 times.
  - Flip 3 times. For each flip, pick a random coin, flip it and then put the coin back.
Fair Toss from Biased Coin (*unknown* probability $p$ of heads)?

• Suppose I would like to use a biased coin in order to obtain a fair outcome
  – The challenge is that I don’t even know the probability of H or T for this coin
  – How do I do it?
  – One option would be to toss $10^{1000}$ times and estimate $p$
  – A more pragmatic method is to notice that $\mathbb{P}[\text{‘ht’}] = \mathbb{P}[\text{‘th’}] = p(1 − p)$
    • (Lower case ‘h’ and ‘t’ denote the outcomes of a toss.)
  – This suggests that an H is as likely as a T.
  – I can now design the following algorithm:

1. Make two tosses of the biased coin
2. If you get ‘ht’ output H; ‘th’ output T; otherwise RESTART.
Fair Toss from Biased Coin (*unknown* probability $p$ of heads)?, cont’d

- Let’s use the law of total probability to estimate the value of $H$ output by our algorithm

$$
P[H] = P[H|\text{RESTART}] \cdot P[\text{RESTART}] + P[H|'ht'] \cdot P['ht'] + P[H|'th'] \cdot P['th']$$

$$= P[H] \cdot (p^2 + (1-p)^2) + 1 \cdot p(1-p) + 0 \cdot p(1-p)$$

$$= P[H] \cdot (p^2 + (1-p)^2) + p(1-p)$$

- Solve for $P[H]$:

$$P[H] = \frac{p(1-p)}{1 - (p^2 + (1-p)^2)} = \frac{p(1-p)}{2p - 2p^2} = \frac{p(1-p)}{2p(1-p)} = \frac{1}{2}$$

- (You can also solve this problem using an infinite outcome tree and computing an infinite sum.)