Bayesian Statistics Workbook - Chapter 2 Bayes' Rule

Vu Nguyen Quang Duy

Table of contents

1 Chapter Summary

1.1 Conditional vs unconditional probability

Let A and B be two events, The **unconditional probability** of A, measures the probability of observing A, without any knowledge of B. In contrast, the **conditional probability** of A

given B, $P(A|B)$, measures the probability of observing A in light of the information that B occured.

Conditional probabilities are fundamental to Bayesian analyses. In general, comparing the conditional vs unconditional probabilities, $P(A|B)$ vs $P(A)$, reveals the extent to which information about B informs our understanding of A. In some cases, the certainty of an event A might *increase* or *decrease* in light of new data B. In other words:

$$P(A|B) > P(A) \text{ Or } P(A|B) < P(A)$$

The *order* of conditioning is also important. Since they measure two it's typically the case that:

$$P(A|B) \neq P(B|A)$$

1.2 Independent events

Two events A and B are **independent** if and only if the occurrence of B does not tell us anything about the occurrence of A:

$$P(A|B) = P(A)$$

1.3 Probability vs likelihood

When B is known, the **conditional probability function** $P(\cdot|B)$ allows us to compare the probabilities of an unknown event, A and $\overline{A}$, occurring with B:

$$P(A|B) \text{ vs } P(\overline{A}|B)$$

When A is known, the **likelyhood function** $L(\cdot|A) = P(A|\cdot)$ allows us to evaluate the relative compatibility of data A with events B or $\overline{B}$:

$$L(B|A) \text{ vs } L(\overline{B}|A)$$

1.4 Joint and conditional probabilities

For events A and B, the joint probablity of $A \cap B$ is calculated by weighting the conditional probability of A given B by the marginal probability of B:

$$P(A \cap B) = P(A|B)P(B) \tag{1}$$

Thus when A and B are *independent*,

$$P(A \cap B) = P(A)P(B)$$

Dividing both sides of Equation 1 by $P(B)$, and assuming $P(B) \neq 0$, reveals the definition of the conditional probability of A given B:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \tag{2}$$

Thus, to evaluate the chance that A occurs in light of information B we can consider the chance that they occur together, $P(A \cap B)$, relative to the chance that B occurs at all, $P(B)$

1.5 Law of Total Probability (LTP)

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}) \tag{3}$$

1.6 Bayes' Rule for events

For events A and B, the posterior probability of B given A follows by combining Equation 1 with Equation 2 and recognizing that we can evaluate data A through the likelihood function, $L(B|A) = P(A|B)$ and $L(\overline{B}|A) = P(A|\overline{B})$:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B)L(B|A)}{P(A)} \tag{4}$$

where by the Law of Total Probability Equation 3:

$$P(A) = P(B)L(B|A) + P(\overline{B})L(\overline{B}|A) \tag{5}$$

More generally,

$$\text{posterior} = \frac{\text{prior} \cdot \text{likelihood}}{\text{normalizing constant}}$$

3

1.7 Discrete probability model

Let Y be a discrete random variable. The probability model of Y is specified by a **probability mass function (pmf)** $f(y)$. Thi pmf defines the probability of any given outcome y,

$$f(y) = P(Y = y)$$

and has the following properties:

- $0 \leq f(y) \leq 1$ for all y, and

- $\sum_{\text{all } y} f(y) = 1$, i.e., the probabilities of all possible outcomes of y sum to 1.

1.8 Conditional probability model of data Y

Let Y be a discrete random variable and π ve a parameter uop which Y depends. The the conditional probability model of Y given π is specified by conditional pmf $f(y|\pi)$. This pmf specifies the conditional probability of observing y given π,

$$f(y|\pi) = P(Y = y|\pi)$$

and has the following properties:

- $0 \leq f(y|\pi) \leq 1$ for all y, and

- $\sum_{\text{all } y} f(y|\pi) = 1$

1.9 The Binomal model

Let random variable Y be the *number of successes* in a *fixed number of trials n*. Assume that the trials are *independent* and that the *probability of success* in each trial is π. Then the conditional depdendence of Y on π can be modeled by the Binomal model with **parameters** n and π. In mathematical notation:

$$Y|\pi \sim \text{Bin}(n, \pi)$$

where "$\sim$" can be read as "modeled by". Correspondingly, the Binomal model is specified by **conditional pmf**

$$f(y|\pi) = \binom{n}{y} \pi^2 \text{ for } y \in \{0, 1, 2, ..., n\} \tag{6}$$

1.10 Probability mass functions vs likelihood functions

When π is known, the **conditional pmf** $f(\cdot|\pi)$ allows us to compare the probabilites of different possible values of data Y (e.g., y_1 or y_2) occuring with π:

$$f(y_1|\pi) \text{ vs } f(y_2|\pi)$$

When $Y = y$ is known, the **likelihood function** $L(\cdot|y) = f(y|\cdot)$ allows us to compare the relative likelihood of observing data y under different possible values of π (e.g., π_1 or π_2):

$$L(\pi_1|y) \text{ vs } L(\pi_2|y)$$

Thus, $L(\cdot|y)$ provides the tool we need to evaluate the relative compatibility of data $Y = y$ with various π values.

1.11 Bayes' Rule for variables

For any variables π and Y, let $f(\pi)$ denote the prior pmf of π and $L(\pi|y)$ denote the likelihood function of π given observed data $Y = y$. Then the posterior pmf of π given data $Y = y$ is:

$$f(\pi|y) = \frac{\text{prior} \cdot \text{likelihood}}{\text{normalizing constant}} = \frac{f(\pi)L(\pi|y)}{f(y)} \tag{7}$$

where, by the Law of Total Probability, the overall probability of observing data $Y = y$ accross all possible π is:

$$f(y) = \sum_{\text{all } y} f(\pi)L(\pi|y)$$

1.12 Propotionality

Since $f(y)$ is merely a normalizing constant which does not depend of π, the posterior pmf $f(\pi|y)$ is propotional to the product of $f(\pi)$ and $L(\pi|y)$:

$$f(\pi|y) = \frac{f(\pi)L(\pi|y)}{f(y)} \propto f(\pi)L(\pi|y)$$

That is,

$$\text{posterior} \propto \text{prior} \cdot \text{likelihood}$$

The significance of this proportionality is that all the information we need to build the posterior model is held in the prior and likelihood.

2 Excercises

2.1 Buidling up to Bayes' Rule

Exercise 2.1. *Comparing the prior and posterior*

For each scenario below, you're given a pair of events, A and B. Explain what you believe to be the relationship between the posterior and prior probabilities of B: $P(B|A) > P(B)$ or $P(B|A) < P(B)$

    a) A = you just finished reading Lambda Literary Award-winning author Nicole Dennis-Benn's first novel, and you enjoyed it! B = you will also enjot Benn's newest novel.

    b) A = it's 0 degrees Fahrenheit in Minnesota on a January day. B = it will be 60 degrees tomorrow.

    c) A = the authors only got 3 hours of sleep last night. B = the authors make several typos in their writing today.

    d) A = your friend includes three hashtags in their tweet. B = the tweet gets retweeted.

Solution

    a) **Answer**: $P(B|A) > P(B)$

- The prior probability, $P(B)$: The general probability of enjoying Benn's newest novel before reading any of her previous work.

- The posterior probability, $P(B \mid A)$: The updated probability of enjoying Benn's newest novel, given that her first novel was read and enjoyed.

The event A (enjoying the first novel) is positive evidence that provides a reason to increase belief in event B (enjoying the newest novel). A favorable experience with the author's work makes the updated belief (the posterior) stronger and therefore higher than the initial belief (the prior).

    b) **Answer**: $P(B|A) < P(B)$

- The prior probability, $P(B)$: The general probability that it will be 60 degrees tomorrow.

- The posterior probability, $P(B \mid A)$: The updated probability that it will be 60 degrees tomorrow, given that it was 0 degrees Fahrenheit yesterday.

The event A (a temperature of 0°F yesterday) is negative evidence that provides a reason to decrease the belief in event B (a temperature of 60°F tomorrow). A temperature of 0°F makes it significantly less likely that the temperature will be a relatively mild 60°F the next day. This new information acts as negative evidence, causing a decrease in the belief of event B.

c) **Answer**: $P(B|A) > P(B)$

- The prior probability, $P(B)$: The general probability that the authors will make several typos in their writing today.

- The posterior probability, $P(B \mid A)$: The updated probability that the authors will make several typos, given they only got 3 hours of sleep last night.

The event A (only 3 hours of sleep) is positive evidence that increases the probability of event B (making typos). Lack of sleep is a well-known factor that impairs cognitive function and attention to detail, making errors like typos more probable. The updated belief is therefore higher than the initial belief.

d) **Answer**: $P(B|A) > P(B)$

- The prior probability, $P(B)$: The general probability that the tweet will be retweeted. This is the baseline likelihood without knowing anything about the tweet's content or format.

- The posterior probability, $P(B \mid A)$: he updated probability that the tweet will be retweeted, given that it includes three hashtags.

The event A (including three hashtags) is positive evidence that increases the probability of event B (the tweet being retweeted). Research on social media engagement shows that tweets with hashtags, especially a moderate number, tend to have wider reach and higher engagement, which includes retweets. Therefore, the updated belief is higher than the initial belief.

Exercise 2.2. *Marginal, conditional, or joint?*

Define the following events for a resident of a fictional town:

- $A =$ drives 10 miles per hour above the speed limit,

- $B =$ gets a speeding ticket,

- $C =$ took statistics at the local college,

- $D =$ has used R,

- $E =$ likes the music of Prince,

- $F =$ is a Minnesotan.

Several facts about these events are listed below. Specify each of these facts using probability notation, paying special attention to whether it's a marginal, conditional, or joint probability.

a) 73% of people that drive 10 miles per hour above the speed limit get a speeding ticket.

b) 20% of residents drive 10 miles per hour above the speed limit.

c) 15% of residents have used R.

d) 91% of statistics students at the local college have used R.

e) 38% of residents are Minnesotans that like the music of Prince.

f) 95% of the Minnesotan residents like the music of Prince.

Solution

a) **Answer**: conditional probability

This fact gives the probability of getting a speeding ticket (B) given that a person is already driving 10 miles per hour above the speed limit (A).

The probability notation for this is: $P(B|A) = 0.73$

b) **Answer**: marginal probability

This facts only describe the proportion of residents that drives 10 miles per hour above the speed limit (A) without any conditions.

The probability notation for this is: $P(A) = 0.20$

c) **Answer**: marginal probability

This facts only describe the proportion of residents that have used R (D) without any conditions.

The probability notation for this is: $P(D) = 0.15$

d) **Answer**: conditional probability

This fact It states the probability of a person having used R (D) given that they took statistics at the local college (C).

The probability notation for this is: $P(D|C) = 0.91$

e) **Answer**: joint probability

This is a joint probability because it refers to the likelihood of two events happening at the same time: being a Minnesotan and liking the music of Prince.

The probability notation for this is: $P(E \cap F) = 0.38$

f) **Answer**: conditional probability

This is a conditional probability. It states the probability of a person liking the music of Prince (E) given that they are a Minnesotan (F).

The probability notation for this is: $P(E|F) = 0.95$

Exercise 2.3. *Binomial practice*

For each variable Y below, determine whether Y is Binomial. If yes, use notation to specify this model and its parameters. If not, explain why the Binomial model is not appropriate for Y.

    a) At a certain hospital, an average of 6 babies are born each hour. Let Y be the number of babies born between 9 a.m. and 10 a.m. tomorrow.

    b) Tulips planted in fall have a 90% chance of blooming in spring. You plant 27 tulips this year. Let Y be the number that bloom.

    c) Each time they try out for the television show *Ru Paul's Drag Race*, Alaska has a 17% probability of succeeding. Let Y be the number of times Alaska has to try out until they're successful.

    d) Y is the amount of time that Henry is late to your lunch date.

    e) Y is the probability that your friends will throw you a surprise birthday party even though you said you hate being the center of attention and just want to go out to eat.

    f) You invite 60 people to your "π day" party, none of whom know each other, and each of whom has an 80% chance of showing up. Let Y be the total number of guests at your party.

Solution

In the Binomal model, we need to specified three parameters:

- Y: number of successes,

- n: fixed number of trials, and each trial must have only two outcomes, typically called "success" and "failure."

- π: probability of successin each trial and must be the same for every trial.

and it denotes as:

$$f(y|\pi) = \binom{n}{y}\pi^2 \text{ for } y \in \{0, 1, 2, \dots, n\}$$

a) **Answer**: Y is not a binomal variable

because no fixed number of trials (n). The number of "trials" (i.e., moments a baby could be born) is not a fixed, countable number.

b) **Answer**: Y is a binomal variable

$$f(y|0.90) = \binom{27}{y} 0,90^2 \text{ for } y \in \{0, 1, 2, ..., 27\}$$

c) **Answer**: Y is not a binomal variable

Because the number of trials is not predetermined; it could be 1, 2, 3, or any number of attempts until Alaska is successful.

d) **Answer**: Y is not a binomal variable

Because the amount of time Henry is late can be any value within a range (e.g., 5 minutes, 10.5 minutes, 20 minutes, etc.), not just two discrete outcomes. And there is no fixed number of "trials" in the amount of time Henry is late.

e) **Answer**: Y is not a binomal variable

Because the variable is a single event.

f) **Answer**: Y is a binomal variable

$$f(y|0.80) = \binom{60}{y} 0,80^2 \text{ for } y \in \{0, 1, 2, ..., 60\}$$

2.2 Practice Bayes' Rule for events

Exercise 2.4. *Vampires?*

Edward is trying to prove to Bella that vampires exist. Bella thinks there is a 0.05 probability that vampires exist. She also believes that the probability that someone can sparkle like a diamond if vampires exist is 0.7, and the probability that someone can sparkle like a diamond if vampires don't exist is 0.03. Edward then goes into a meadow and shows Bella that he can sparkle like a diamond. Given that Edward sparkled like a diamond, what is the probability that vampires exist?

Solutions

Call V is the event "Vampires exist". We have the prior probability model as follow:

event	V	$\overline{V}$	total
probability	0.05	0.095	1

Call event "Someone can sparkle like a diamond" as S, the conditional probability for S given that vampire exist is $P(S|V)$. It informs us the likelihood of V in light of S, the $L(V|S)$:

event	V	$\overline{V}$	total
probability	0.05	0.095	1
likelihood	0.7	0.03	

Using the Law of Total Probability, we can calculate the normalizing constant $P(S)$:

$$P(S) = P(V)L(V|S) + P(\overline{V})L(\overline{V}|S)$$

Using Bayes' Rule, we can calculate the probablity that vampires exist given that Edward can sparkle like a diamond, $P(V|S)$:

$$P(V|S) = \frac{P(V)L(V|S)}{P(S)}$$

After plug in all the numbers, we have the result of. 0.551. The updated table is:

event	V	$\overline{V}$	total
prior	0.05	0.095	1
likelihood	0.7	0.03	
posterior	0.551	0.449	1

Exercise 2.5. *Sick trees*

A local arboretum contains a variety of tree species, including elms, maples, and others. Unfortunately, 18% of all trees in the arboretum are infected with mold. Among the infected trees, 15% are elms, 80% are maples, and 5% are other species. Among the uninfected trees, 20% are elms, 10% are maples, and 70% are other species. In monitoring the spread of mold, an arboretum employee randomly selects a tree to test.

a) What's the prior probability that the selected tree has mold?

b) The tree happens to be a maple. What's the probability that the employee would have selected a maple?

c) What's the posterior probability that the selected maple tree has mold?

d) Compare the prior and posterior probability of the tree having mold. How did your understanding change in light of the fact that the tree is a maple?

Solution

Before answering these questions, we need to name each event:

- Event E: "A tree is elm"

- Event M: "A tree is maple"

- Event O: "A tree is not elm neither maple"

- Event I: "A tree is infected with mold"

With these notation and clues from the question, we have these data:

- 18% of all trees in the arboretum are infected with mold: $P(I) = 0.18$

- Among the infected trees, 15% are elms: $P(E|I) = 0.15$

- Among the infected trees, 80% are maples: $P(M|I) = 0.80$

- Among the infected trees, 5% are other species: $P(O|I) = 0.05$

- Among the uninfected trees, 20% are elms: $P(E|\bar{I}) = 0.20$

- Among the uninfected trees, 10% are maples: $P(M|\bar{I}) = 0.10$

- Among the uninfected trees, 70% are other species: $P(O|\bar{I}) = 0.70$

a) The prior probability that the selected tree has mold is $P(I) = 0.18$

b) Using the Law of Total Probability, we can calculate he probability that the employee would haveselected a maple, $P(M)$, by:

$$P(M) = P(I)L(I|M) + P(\bar{I})L(\bar{I}|M)$$

The result is 0.226.

c) Using Bayes' Rule, we can calculate the posterior probability that the selected maple tree has mold, $P(I|M)$:

$$P(I|M) = \frac{P(I)L(I|M)}{P(M)}$$

The result is 0.637.

    d) The fact that maples are easy to get infected than other species explain why the posterior is higher than prior probability.

Exercise 2.6. *Restaurant ratings*

The probability that Sandra will like a restaurant is 0.7. Among the restaurants that she likes, 20% have five stars on Yelp, 50% have four stars, and 30% have fewer than four stars. What other information do we need if we want to find the posterior probability that Sandra likes a restaurant given that it has fewer than four stars on Yelp?

Solution

Call L is the event "Sandra will like a restaurant", therefor $P(L) = 0.70$

Call event "The restaurant has five stars on Yelp" E, and "The restaurant has four stars on Yelp" G, and "The restaurant has lower than four stars on Yelp" M, we have $P(E|L) = 0.20$, $P(G|L) = 0.50$, and $P(M|L) = 0.30$.

To find the posterior probability that Sandra likes a restaurant given that it has fewer than four stars on Yelp, $P(L|M)$, we need to know the normalizing constant, $P(M)$, which stands for propotion of every restaurants that lower than four stars on Yelp.

$P(M)$ can easily be calculated using the Law of Total Probablity, which requires:

- $P(L)$: Probability that Sandra will like a restaurant, which is known.

- $L(L|M)$: Likelihood that a restaurant as lower than four stars on Yelp given that Sandra like it, which is known.

- $P(\overline{L})$: Probability that Sandra will like a restaurant, which is known by using $1 - P(L)$.

- $L(\overline{L}|M)$: Likelihood that a restaurant as lower than four stars on Yelp given that Sandra does not like it, which is missing.

In conclusion, we need to know likelihood that a restaurant as lower than four stars on Yelp given that Sandra does not like it in order to find the posterior probability that Sandra likes arestaurant given that it has fewer than four stars on Yelp.

Exercise 2.7. *Dating app*

Matt is on a dating app looking for love. Matt swipes right on 8% of the profiles he views. Of the people that Matt swipes right on, 40% are men, 30% are women, 20% are non-binary, and 10% identify in another way. Of the people that Matt does not swipe right on, 45% are men, 40% are women, 10% are non-binary, and 5% identify in some other way.

    a) What's the probability that a randomly chosen person on this dating app is non-binary?

    b) Given that Matt is looking at the profile of someone who is non-binary, what's the posterior probability that he swipes right?

Solution

Let's call event "Matt swipes right" R, we have $P(R) = 0.08$.

Call event "A person is men" M, "A person is women" W, "A person is non-binary" N, and "A person is identified in another way" O, we have $P(M|R) = 0.40$, $P(W|R) = 0.30$, $P(N|R) = 0.20$, and $P(O|R) = 0.10$. Also $P(M|\overline{R}) = 0.45$, $P(W|\overline{R}) = 0.40$, $P(N|\overline{R}) = 0.10$, and $P(O|\overline{R}) = 0.05$.

    a) Using Law of Total Probability, we can calculate the probability that a randomly chosen person on this dating app is non-binary, $P(N)$ as:

$$P(N) = P(R)L(R|N) + P(\overline{R})L(\overline{R}|N)$$

The result is 0.108.

    b) Using Bayes' Rule, we can calculate the posterior probability that he swipes right Given that Matt is looking at the profile of someone who is non-binary, $P(R|N)$:

$$P(R|N) = \frac{P(R)L(R|N)}{P(N)}$$

The result is 0.148.

Exercise 2.8. *Flight delays*

For a certain airline, 30% of the flights depart in the morning, 30% depart in the afternoon, and 40% depart in the evening. Frustratingly, 15% of all flights are delayed. Of the delayed flights, 40% are morning flights, 50% are afternoon flights, and 10% are evening flights. Alicia and Mine are taking separate flights to attend a conference.

    a) Mine is on a morning flight. What's the probability that her flight will be delayed?

    b) Alicia's flight is not delayed. What's the probability that she's on a morning flight?

Solution

Let's call event "flight depart in the morning" M, "flight depart in the afternoon" A, "flight depart in the evening" E, we have $P(M) = 0.30$, $P(A) = 0.30$, and $P(E) = 0.10$.

Let's call event "flight is delayed" D, we have $P(D) = 0.15$. Also, $P(M|D) = 0.40$, $P(A|D) = 0.50$, $P(E|D) = 0.10$.

a) To calculate the probability that Mine's flight will be delayed given that her flight is in the morning, $P(D|M)$, we use Bayes' Rule:

$$P(D|M) = \frac{P(D)L(D|M)}{P(M)}$$

The result is 0.2, which mean there are 20% chance her flight will be delayed of she flight in the morning.

b) The probability that Alicia's on a morning flight given that her flight is not delayed is noted as $P(M|\overline{D})$.

Using the Bayes' Rule, we have

$$P(M|\overline{D}) = \frac{P(M)L(M|\overline{D})}{P(\overline{D})}$$

Given that $L(M|\overline{D}) = P(\overline{D}|M) = 1 - P(D|M)$. Which turns the equation to:

$$P(M|\overline{D}) = \frac{P(M)[1 - P(D|M)]}{1 - P(D)}$$

Plug in all the number, the result is 0.2824, which mean there is 28.24% chance her flight is in the morning given that it is not delayed.

Exercise 2.9. *Good mood, bad mood*

Your roommate has two moods, good or bad. In general, they're in a good mood 40% of the time. Yet you've noticed that their moods are related to how many text messages they receive the day before. If they're in a good mood today, there's a 5% chance they had 0 texts, an 84% chance they had between 1 and 45 texts, and an 11% chance they had more than 45 texts yesterday. If they're in a bad mood today, there's a 13% chance they had 0 texts, an 86% chance they had between 1 and 45 texts, and a 1% chance they had more than 45 texts yesterday.

a) Use the provided information to fill in the table.

b) Today's a new day. Without knowing anything about the previous day's text messages, what's the probability that your roommate is in a good mood? What part of the Bayes' Rule equation is this: the prior, likelihood, normalizing constant, or posterior?

c) You surreptitiously took a peek at your roommate's phone (we are attempting to withhold judgment of this dastardly maneuver) and see that your roommate received 50 text messages yesterday. How likely are they to have received this many texts if they're in a good mood today? What part of the Bayes' Rule equation is this?

d) What is the posterior probability that your roommate is in a good mood given that they received 50 text messages yesterday?

Solution

a)

	good mood	bad mood	total
0 texts	0.05	0.13	0.1
1-45 texts	0.84	0.86	0.85
46+ texts	0.11	0.01	0.05
Total	1	1	1

Call the event "roommate is in a good mood" G, we have $P(G) = 0.40$. Call event "they received 0 texts yesterday" A, "they received 1-45 texts yesterday" B, "they received 46+ texts yesterday" C.

To fill the table, we need to calculate $P(A)$, $P(B)$, $P(C)$. Using the Law of Total Probability, we have:

$$P(A) = P(G)L(G|A) + P(\overline{G})L(\overline{G}|A)$$

Apply for $P(B)$, $P(C)$, we can fill the table for `total` column.

b) Without knowing anything about the previous day's text messages, the probability that roommate is in a good mood is $P(G) = 0.40$. This is the prior in the Bayes' Rule equation.

c) The likelihood that they to have received this many texts if they're in a good mood today is $L(G|C) = P(C|G) = 0.11$, which is the likelihood in the Bayes' Rule equation.

d) The posterior probability that roommate is in a good mood given that they received 50 text messages yesterday, $P(G|C)$, is calculated using:

$$P(G|C) = \frac{P(G)L(G|C)}{P(C)}$$

The result is 0.88.

Exercise 2.10. *LGBTQ students: rural and urban*

A recent study of 415,000 Californian public middle school and high school students found that 8.5% live in rural areas and 91.5% in urban areas. Further, 10% of students in rural areas and 10.5% of students in urban areas identified as Lesbian, Gay, Bisexual, Transgender, or Queer (LGBTQ). Consider one student from the study.

 a) What's the probability they identify as LGBTQ?

 b) If they identify as LGBTQ, what's the probability that they live in a rural area?

 c) If they do not identify as LGBTQ, what's the probability that they live in a rural area?

Solution

Call event U as "The student lives in urban areas", we have $P(U) = 0.915$.

Call event G as "The student identifies themselves as LGBTQ", we have $P(G|U) = 0.105$ and $P(G|\overline{U}) = 0.10$.

 a) The probability a student identifies as LGBTQ is denoted as $P(G)$.

Using the Law of Total Probability:

$$P(G) = P(U)L(U|G) + P(\overline{U})L(\overline{U}|G)$$

The result is 0.1.

 b) If they identify as LGBTQ, the probability that they live in a rural area is denoted as $P(\overline{U}|G)$.

Using Bayes' Rule, we have:

$$P(\overline{U}|G) = \frac{P(\overline{U})L(\overline{U}|G))}{P(G)}$$

The result is 0.08.

c) If they do not identify as LGBTQ, the probability that they live in a rural area is denoted as $P(\overline{U}|\overline{G})$.

Using Bayes' Rule, we have:

$$P(\overline{U}|\overline{G}) = \frac{P(\overline{U})L(\overline{U}|\overline{G}))}{P(\overline{G})}$$

The result is 0.09.

2.3 Practice Bayes' Rule for random variables

Exercise 2.11. *Internship*

Muhammad applies for six equally competitive data science internships. He has the following prior model for his chances of getting into any given internship, π:

π	0.3	0.4	0.5	Total
$f(\pi)$	0.25	0.60	0.15	1

a) Let Y be the number of internship offers that Muhammad gets. Specify the model for the dependence of Y on π and the corresponding pmf, $f(y|\pi)$.

b) Muhammad got some pretty amazing news. He was offered four of the six internships! How likely would this be if $\pi = 0.3$?

c) Construct the posterior model of π in light of Muhammad's internship news.

Solution

a) $f(y|\pi)$ would be a binomal model because:

- It has a fixed number of trials, $n = 6$.

- The outcome of each trial is only "success" or "fail".

- The probability for success each trial is equal across trials, which is π.

The pmf for $f(y|\pi)$ is:

$$f(y|\pi) = \binom{6}{y}\pi^2 \text{ for } y \in \{0, 1, 2, ..., 6\}$$

b) Muhammad was offered four of the six internships, which means $y = 4$.

With $\pi = 0.3$, we have:

$$f(4|0.3) = \binom{6}{4} \times 0.3^2$$

The results is 0.06.

c) The posterior model of in light of Muhammad's internship news is denoted as $f(\pi|y = 4)$.
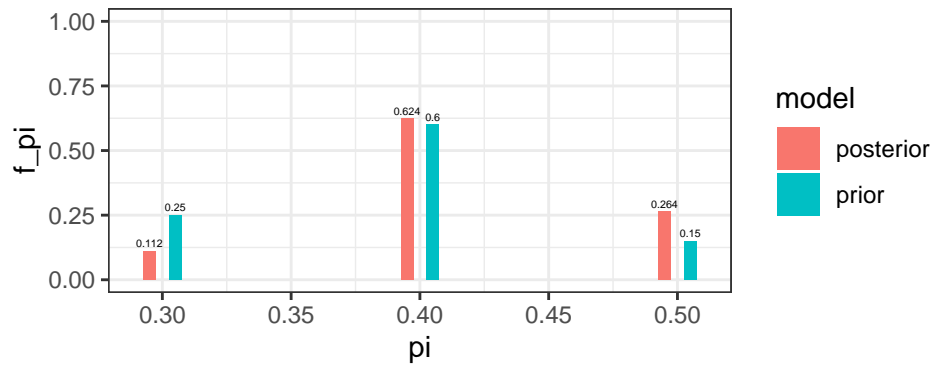
Using Bayes' Rule, we have:

$$f(\pi|y = 4) = \frac{f(\pi)L(\pi|y = 4)}{f(y = 4)}$$

where, by the Law of Total Probability, the overall probability of observing data $Y = y$ accrossall possible π, $f(y)$, is:

$$f(y = 4) = \sum_{\text{all } y} f(\pi)L(\pi|y = 4)$$

With the Muhammad's internship news, we can construct the posterior model of π. The results are:

pi	prior	posterior
0.3	0.25	0.1119214
0.4	0.60	0.6237141
0.5	0.15	0.2643645

Exercise 2.12. *Making mugs*

Miles is learning how to make a mug in his ceramics class. A difficult part of the process is creating or "pulling" the handle. His prior model of π, the probability that one of his handles will actually be good enough for a mug, is below:

π	0.1	0.25	0.4	Total
$f(\pi)$	0.2	0.45	0.35	1

a) Miles has enough clay for 7 handles. Let Y be the number of handles that will be good enough for a mug. Specify the model for the dependence of Y on π and the corresponding pmf, $f(y|\pi)$.

b) Miles pulls 7 handles and only 1 of them is good enough for a mug. What is the posterior pmf of π, $f(\pi|y = 1)$

c) Compare the posterior model to the prior model of π. How would you characterize the differences between them?

d) Miles' instructor Kris had a different prior for his ability to pull a handle (below). Find Kris's posterior $f(\pi|y = 1)$ and compare it to Miles'.

π	0.1	0.25	0.4	Total
$f(\pi)$	0.15	0.15	0.70	1

Solution

a) $f(y|\pi)$ would be a binomal model because:

- It has a fixed number of trials, $n = 7$.

- The outcome of each trial is only "success" or "fail".

- The probability for success each trial is equal across trials, which is π.

The pmf for $f(y|\pi)$ is:

$$f(y|\pi) = \binom{7}{y}\pi^2 \text{ for } y \in \{0, 1, 2, ..., 7\}$$

c) The posterior model of pmf of π given new information is denoted as $f(\pi|y = 1)$.
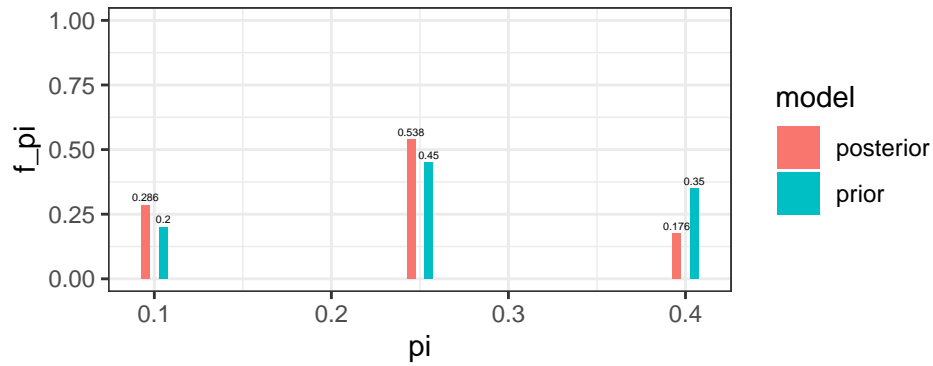
Using Bayes' Rule, we have:

$$f(\pi|y=1) = \frac{f(\pi)L(\pi|y=1)}{f(y=1)}$$

where, by the Law of Total Probability, the overall probability of observing data $Y = y$ accrossall possible π, $f(y)$, is:

$$f(y=1) = \sum_{\text{all } y} f(\pi)L(\pi|y=1)$$

With the Muhammad's internship news, we can construct the posterior model of π. The results are:

pi	prior	posterior
0.10	0.20	0.286
0.25	0.45	0.538
0.40	0.35	0.176

c) The posterior distribution reflects the information from the observation, which suggests that a moderate probability for a good handle is more likely, and the belief in very high or very low probabilities has adjusted accordingly.

d) The results are:

pi	prior	posterior
0.10	0.15	0.288
0.25	0.15	0.241
0.40	0.70	0.471