Introduction to Probability

STA 032: Gateway to data science Lecture 13

Jingwei Xiong

May 1, 2023

Recap

- Introduction to probability
  - Events, sample space
  - Probability rules
    - Complement rule: $P(A) + P(A^c) = 1$
    - Additive rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - Permutation and combinations

Today

- Conditional probability
- Marginal and joint probability
- Independence
- Bayes' Theorem

Conditional Probability

- Often we wish to know the probability an event will occur given that another event has occurred.
- E.g., instead of the **marginal** probability of contracting COVID (regardless of vaccination status), we may wish to know:
  - The probability that someone will contract COVID *given that* they have been vaccinated
  - The probability someone will contract COVID *given that* they have not been vaccinated.
- These are examples of **conditional probability**.
- Our earlier example: A is the event that a person smokes; B is the event that a person identifies as female
  - The conditional probability that someone is a smoker (event A) given that they identify as female (event B) is denoted $P(A|B)$, which we say as "probability of A given B."

Simple Example

- Suppose we have a small population containing 3 female non-smokers, 1 female smoker, 4 non-female non-smokers, and 4 non-female smokers.

	Female	**Non-female**
Smoke	1	4
Does not smoke	3	4

- Each individual's characteristics:

```
   Female    Smoker
1       1         1
2       0         1
3       0         1
4       0         1
5       0         1
6       1         0
7       1         0
8       1         0
9       0         0
10      0         0
11      0         0
12      0         0
```

Conditional probability

- Each individual's characteristics:

```
    Female    Smoker
1        1         1
2        0         1
3        0         1
4        0         1
5        0         1
6        1         0
7        1         0
8        1         0
9        0         0
10       0         0
11       0         0
12       0         0
```

	Female	Non-female
Smoke	1	4
Does not smoke	3	4

- Conditional probability someone is a smoker given that they identify as female = $P(\texttt{smoker}|\texttt{female}) = \frac{1}{3+1}$

- Looking at each individual's characteristics: only look at rows where `female == 1`. Looking at table: only look at `female` column.

- The conditional probability is calculated by changing the denominator to correspond to our smaller population of interest.

Conditional probability: How to do this in R?

- We want to condition on the `female` column, and find the proportion of smokers and non-smokers.
- Recall: **column proportions**

```
outTable <- matrix(c(1, 3, 4, 4), nrow = 2)
rownames(outTable) <- c("Smoker", "Non-smoker")
colnames(outTable) <- c("Female", "Non-female")
prop.table(outTable, margin = 2)
```

```
           Female Non-female
Smoker       0.25        0.5
Non-smoker   0.75        0.5
```

Marginal probability

- Each individual's characteristics:

```
   Female    Smoker
1       1         1
2       0         1
3       0         1
4       0         1
5       0         1
6       1         0
7       1         0
8       1         0
9       0         0
10      0         0
11      0         0
12      0         0
```

- **Marginal probability**: "unconditional" probability; based on a single variable
- One way to think about marginal probability: "throwing away/ignoring the other variable"
- $P(\texttt{female}) = \frac{4}{12}$ (just look at female column)
- $P(\texttt{smoker}) = \frac{5}{12}$ (just look at smoker column)

Marginal probability

- Another way to think about marginal probability: sum over the other variables
  - Might hear people say "marginalize over the other variable"
- Recall: row and column totals in R

```
outTableTotals <- outTable %>%
  cbind(rowTotal = rowSums(outTable))
outTableTotals <- outTableTotals %>%
  rbind(columnTotal = colSums(outTableTotals))
outTableTotals
```

```
            Female Non-female rowTotal
Smoker           1          4        5
Non-smoker       3          4        7
columnTotal      4          8       12
```

- Column totals sum over smoking status; row totals sum over gender identity
- $P(\texttt{female}) = \frac{4}{12}$ (sum over smoking status; look at column totals)
- $P(\texttt{smoker}) = \frac{5}{12}$ (sum over gender identity; look at row totals)

Joint probability

- Joint probability is the probability when we are considering outcomes for **two or more** variables or processes
- E.g.:
  - Flipping a coin *and* rolling a die
  - Identifying as female *and* smoking
- $P(\texttt{smoker} \text{ and } \texttt{female})$, $P(\texttt{smoker}, \texttt{female})$, $P(A \cap B)$

Joint probability

- Get joint probabilities from table by dividing by the total count for the entire table

```
outTable / sum(outTable)
```

```
                Female Non-female
Smoker     0.08333333  0.3333333
Non-smoker 0.25000000  0.3333333
```

- This is a **probability distribution**; should sum to 1

- The probabilities we saw in the previous class were joint and marginal probabilities

Recall Our Vaccine Hesitancy Example

Ethnicity	Vaccine Hesitant	Not Hesitant
White British or Irish	1362	7368
Other white background	71	199
Mixed	55	115
Asian or Asian British - Indian	37	143
Asian or Asian British - Pakistani/Bangladeshi	85	115
Asian or Asian British - other	15	95
Black or Black British	136	54
Other Ethnic Group or Not Specified	31	119

Three Probabilities

Define events A=vaccine hesitant and B=Asian or Asian British-Indian. Calculate the following probabilities for a randomly-selected person drawn from the population of 10,000.

- **Marginal probability** of vaccine hesitancy, $P(A) = \frac{1362+71+55+37+85+15+136+31}{10000} = \frac{1792}{10000}$
- **Joint probability** of vaccine hesitancy and Indian ethnicity, $P(A \cap B) = \frac{37}{10000}$
- **Conditional probability** of vaccine hesitancy given a person is of Indian ethnicity, $P(A \mid B) = \frac{37}{37+143}$

Rules for Conditional Probability

- More formally, we define **conditional probability** as $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- Recall: Conditional probability someone is a smoker given that they identify as female = $P(\texttt{smoker}|\texttt{female}) = \frac{1}{3+1}$
  - Calculated by changing the denominator to correspond to our smaller population of interest
  - $= \frac{\#(\texttt{smoker and female})}{\#\texttt{female}}$
- According to new formula: $P(\texttt{smoker}|\texttt{female}) = \frac{P(\texttt{smoker} \cap \texttt{female})}{P(\texttt{female})} = \frac{\#\texttt{smoker and female}/\texttt{total}}{\#\texttt{female}/\texttt{total}}$
- Works because total cancels out

Rules for Conditional Probability

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- Manipulating this formula, we get the **general multiplication rule**: $P(A \cap B) = P(B)P(A|B)$.
- **Sum of conditional probabilities**: Let $A_1, \ldots, A_k$ be disjoint outcomes. Then $P(A_1|B) + \ldots + P(A_k|B) = 1$
  - This is the same idea as the complement rule from last class: $P(A) + P(A^c) = 1$.
- When there are just two events, A and A^c, we have $P(A|B) + P(A^c|B) = 1$, so $P(A|B) = 1 - P(A^c|B)$.

Rules for Conditional Probability

- One more helpful rule is the **law of total probability**:
- $P(B) = P(B \mid A)P(A) + P(B \mid A^c)P(A^c) = P(B \cap A) + P(B \cap A^c)$
  - Translates to the statement that the probability that B occurs is equal to the sum of the probabilities that B occurs with A and that B occurs without A
- Extending this to $A_1, \ldots, A_k$, where $A_1, \ldots, A_k$ are disjoint outcomes: $P(B) = P(B \cap A_1) + \ldots + P(B \cap A_k)$

Independence

- Events A and B are **independent** if knowing the outcome of one provides no useful information about the outcome of the other
- E.g.: Flipping a coin and rolling a die are two independent processes
    - Knowing the coin was heads does not help determine the outcome of a die roll
- Seeing someone with an umbrella and the day being rainy are not independent
    - If we see someone with an umbrella, it is more likely to be a rainy day

Independence

- What is the probability of flipping heads and rolling a 1 on a die?
- Probability distribution

	1	**2**	**3**	**4**	**5**	**6**
Heads	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
Tails	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

- Intuition: 1/2 of the time we get heads, and 1/6 of those times we roll a 1
- Probabilities can therefore be multiplied

Independence

- Multiplication rule for independent processes:
  - If A and B represent events from two different and independent processes, then the probability that both A and B occur can be calculated as the product of their separate probabilities: $P(A \cap B) = P(A) \times P(B)$.
  - This is "if and only if" relationship. $P(A \cap B) = P(A) \times P(B)$, A indepdent of B
  - If there are k events $A_1, \ldots, A_k$ from k independent processes, then the probability they all occur is $P(A_1) \times P(A_2) \times \ldots \times P(A_k)$.

Independence

- Recall this probability distribution:

	1	**2**	**3**	**4**	**5**	**6**
Heads	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
Tails	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

- $P(\texttt{roll 1}|\texttt{heads}) = \frac{P(\texttt{1 and heads})}{P(\texttt{heads})} = \frac{1/12}{1/2} = 1/6$
- $P(\texttt{roll 1}) = 1/6$
- Events A and B are **independent** if and only if $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$.

Revisiting Independence

- Independence can be check using conditional probability.
- Intuitively, A independent of B means the condition of B won't affect A.
- Events A and B are **independent** if and only if $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$.
    - This comes from the multiplication rule $P(A \cap B) = P(A) \times P(B)$ and the definition of conditional probability, $P(A|B) = \frac{P(A \cap B)}{P(B)}$
    - $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$
    - $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$

Checking Independence

Are vaccine hesitancy and Indian ethnicity independent in our population?

- **Marginal probability** of vaccine hesitancy, $P(A) = \frac{1362+71+55+37+85+15+136+31}{10000} = \frac{1792}{10000} = .1792$
- **Conditional probability** of vaccine hesitancy given a person is of Indian ethnicity, $P(A \mid B) = \frac{37}{37+143} = .206$
- $P(A) \neq P(A \mid B)$, so they are not independent.

Independent vs Disjoint Events

- For **independent events** A and B, $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$, so knowing one event occurred tells us *nothing* about the chances the other event will occur.

- For two **disjoint** or **mutually exclusive** events, knowing that one event has occurred tells us that the other event definitely has not occurred, i.e., $P(A \cap B) = 0$.

- Disjoint events are therefore *not* independent!

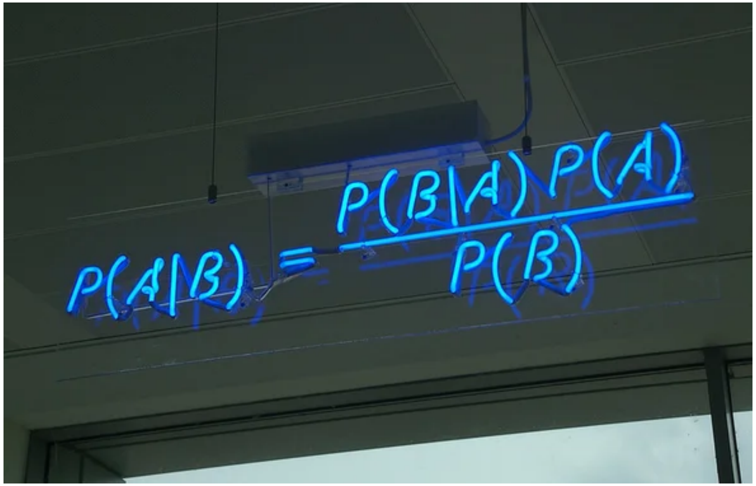
Bayes' Theorem

- Often we know $P(B|A)$ when we really want $P(A|B)$
- For example, imagine a hypothetical scenario of a 40-year-old woman with a positive screening mammogram. Let the event A be having cancer and B be the positive mammogram screening result.
- We want to know $P(A|B)$, which is the probability of having cancer given a positive screening result.
- $P(B|A)$ depends on how good the screening tool is (called **sensitivity**, the probability of a positive result given that a person has cancer)
- We will need to know some properties of the screening test, and $P(A)$, the prevalance of breast cancer among 40-year-old women.
- Using Bayes' Theorem is sometimes described as "updating our beliefs": without any information on the woman's test result, the probability of cancer is just $P(A)$; with the test result we can calculate $P(A|B)$

Bayes' Theorem

Bayes' theorem says that $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)}$.

- The last equality is by the law of total probability



Credit: Matt Buck, Flickr, CC BY-SA 2.0

Bayes' Theorem

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)}$$

- A = cancer, B = positive screening result.
- $P(B|A)$ is the probability of a positive screening result, given that a person has cancer. This is called the **sensitivity** of the test. Say it is .85.
- We also need to know $P(A)$, the prevalance of breast cancer among 40-year-old women. Say this is .01.
- The last ingredient is $P(B|A^c)$, which is the probability of a positive screening result given that a person does not have cancer. Say it is .1.

Bayes' Theorem in Action

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

- $P(B|A) = .85$
- $P(A) = .01$
- $P(B|A^c) = .1$
- Then $P(A \mid B) = \frac{.85 * .01}{.85 * .01 + .1 * (1 - .01)} = 0.079$

Behind the scenes: Hypothetical 10,000

Consider a hypothetical population of 10,000 40-year-old women.

We have

1. The prevalance of breast cancer among 40-year-old women, $P(A) = .01$.
2. The sensitivity of a screening mammogram for diagnosing cancer, $P(B|A) = .85$.
3. The probability of a positive screening result given that a person does not have cancer, $P(B|A^c) = .1$.

	Cancer (A)	**No Cancer**	**Total**
Mammo + (B)			
Mammo -			
Total			10000

Behind the scenes: Hypothetical 10,000

We have

1. **The prevalance of breast cancer among 40-year-old women,** $P(A) = .01$**.**
2. The sensitivity of a screening mammogram for diagnosing cancer, $P(B|A) = .85$.
3. The probability of a positive screening result given that a person does not have cancer, $P(B|A^c) = .1$.

Item 1 says the prevalence in this group is 1%, so then we expect to have $10000 \times 0.01 = 100$ cases and $10000 \times 0.99 = 9900$ cancer-free women.

	Cancer (A)	**No Cancer**	**Total**
Mammo + (B)			
Mammo -			
Total	100	9900	10000

Behind the scenes: Hypothetical 10,000

We have

1. The prevalance of breast cancer among 40-year-old women, $P(A) = .01$.
2. **The sensitivity of a screening mammogram for diagnosing cancer,** $P(B|A) = .85$**.**
3. The probability of a positive screening result given that a person does not have cancer, $P(B|A^c) = .1$.

In the group of 100 women with cancer, the mammogram should pick up $100 \times 0.85 = 85$ of them, and miss the remaining $100 - 85 = 15$.

	Cancer (A)	No Cancer	Total
Mammo + (B)	85		
Mammo -	15		
Total	100	9900	10000

Behind the scenes: Hypothetical 10,000

We have

1. The prevalance of breast cancer among 40-year-old women, $P(A) = .01$.
2. The sensitivity of a screening mammogram for diagnosing cancer, $P(B|A) = .85$.
3. **The probability of a positive screening result given that a person does not have cancer,** $P(B|A^c) = .1$**.**

In the group of 9900 women without cancer, the mammogram should correctly identify $9900 * 0.90 = 8910$ of them as being cancer-free, and it will mistakenly identify $9900 - 8910 = 990$ as having cancer.

	Cancer (A)	**No Cancer**	**Total**
Mammo + (B)	85	990	
Mammo -	15	8910	
Total	100	9900	10000

Behind the scenes: Hypothetical 10,000

- Now we complete the table by filling in the row totals.

- At this point, it's easy to calculate the conditional probability of cancer given a positive mammogram as $\frac{85}{1075} = 0.079$.

- This entire computation is equivalent to doing $P(A \mid B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)}$

	Cancer (A)	**No Cancer**	**Total**
Mammo + (B)	85	990	1075
Mammo -	15	8910	8925
Total	100	9900	10000

Summary

- Conditional probability
  - General multiplication rule: $P(A \cap B) = P(B)P(A|B)$
  - Sum of conditional probabilities: $P(A_1|B) + \ldots + P(A_k|B) = 1$
  - Law of total probability: $P(B) = P(B \cap A_1) + \ldots + P(B \cap A_k) = P(B \mid A_1)P(A_1) + \ldots + P(B \mid A_k)P(A_k)$
- Marginal and joint probability
- Revisiting independence
  - $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$
- Bayes' Theorem
  - $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$