Week 3: Conditional Probability and Independence

Gracia Dong

STA 237

Fall 2024

Learning Objectives

By the end of this week, you should …

- Define conditional probability and describe the difference between $P(A|B)$ and $P(B|A)$.
- Distinguish between mutually exclusive (disjoint) and independent events.
- Assess whether or not events are independent.
- Represent an event as several disjoint subsets to compute its probability using the law of total probability.
- Recognize the difference between conditional and unconditional events and use appropriate rules (e.g., multiplication rule, Bayes' theorem) to solve problems.

Review – Counting Techniques

- Multiplication Principle: If there are n outcomes for experiment 1 and m outcomes for experiment 2, then there are $n \times m$ outcomes for the two experiments together
  - Permutations and combinations are applications of the multiplication principle
  - To go from order mattering to order not mattering, divide by the number of possible orders
- Example: There is 1 bottle of each of the following sodas at the store: Coke, Pepsi, and Sprite. You pick 2 sodas at random. How many ways can you pick 2 sodas if the order you pick them matters? What if the order doesn’t matter?

Some notes

- When drawing Venn diagrams to represent events, you must draw the sample space as well
- For R: It is sufficient to just complete the posted R activities for this course, if you do not want to download R/RStudio
- First tutorial is this week, please attend your registered section

Conditional Probability

Chapter 2.1-2.4

Contingency Tables

- Consider a survey that asks:
  - Do you drink coffee? Yes/No
  - Do you drink tea? Yes/No
- The relationship between two the outcomes of two experiments with a finite sample space can be shown using a **contingency table**

	Coffee - No	Coffee - Yes	Total - Tea
Tea - No	44	29	73
Tea - Yes	40	45	85
Total - Coffee	84	74	158

Contingency Tables

- Joint Distribution: This is where we consider both the coffee and tea variables together/jointly (the inside of the table)
- Marginal distribution: We consider only one variable, as if the other one wasn't even there (working only with the outside/margins of the table)

	Coffee - No	Coffee - Yes	Total - Tea
Tea - No	44	29	73
Tea - Yes	40	45	85
Total - Coffee	84	74	158

Joint and Marginal Distributions

- What proportion of people drink tea and don’t drink coffee?

- What proportion of people drink tea?

	Coffee - No	Coffee - Yes	Total - Tea
Tea - No	44	29	73
Tea - Yes	40	45	85
Total - Coffee	84	74	158

Conditional Distributions

- Condition distribution: We condition on one value of one variable (e.g. non- coffee drinkers), and consider what happens with other variable for only these people
  - Out of non-coffee drinkers, how many drink tea?
  - We basically ignore anyone who drinks coffee, as if they were never there
  - The total number of people we are looking at decreases
  - Out of all the non-coffee drinkers (84 people), 40 drink tea

	Coffee - No	Coffee - Yes	Total - Tea
Tea - No	44	29	73
Tea - Yes	40	45	85
Total - Coffee	84	74	158

Conditional Probability

- When we find a conditional probability, we are actually dealing with two events:
  - A = event that we are interested in (e.g. tea drinkers)
  - B = event that we already know has happened and want to condition on (e.g. non-coffee drinkers)
- Since we already know that event B happened, we can ignore anything that is not B
- Venn diagram: If we know we are in the circle that represents B, what's the probability of being in A?
- We do this by conditioning on B
  - This means we remove anything in our collection of outcomes of our experiment that is not included in B.
  - Think of B as our new sample space

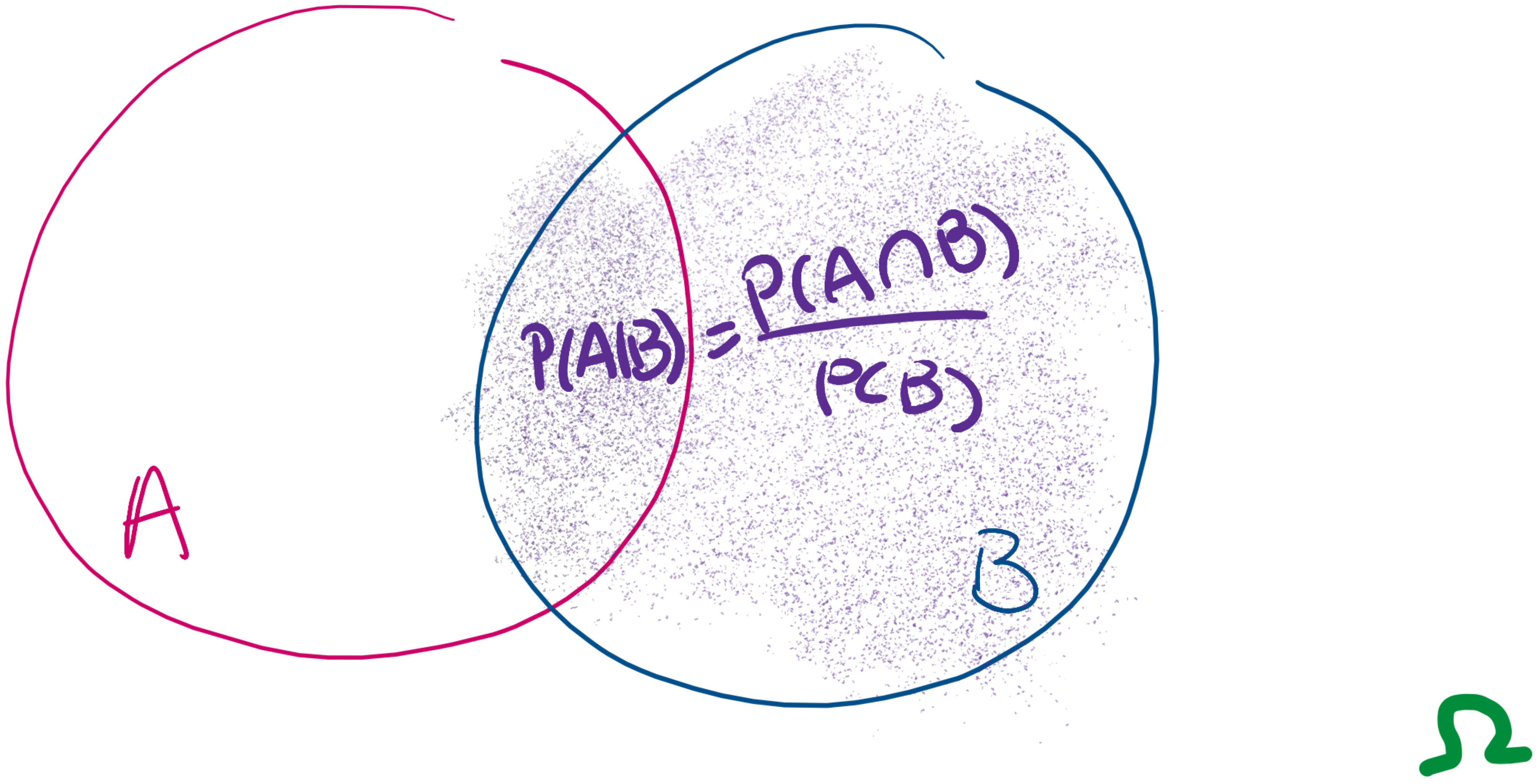
Notation – Conditional Probability

- The general definition of conditional probability has the following form:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{where} \quad P(B) > 0$$

- Since we are conditioning on B, this means that we are only considering how many times A happens out of the number of times B happens
    - i.e. out of only the outcomes in B, how many of them are also outcomes in A?
- First, look at all the outcomes that are in both A and B, and find their probability (out of everything)
- Then we look at just the ones in B, and find that probability (out of everything).

New information changes the sample space



Example: Coin Flipping

- Let's flip a coin 3 times. Then we have 8 possible outcomes:

$$\Omega = \{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\}$$

- Suppose we are interested in the event A = two or more heads
- What is $P(A)$?

Example: Coin Flipping

- Let's flip a coin 3 times. Then we have 8 possible outcomes:

  {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

- Suppose we are interested in the event A = two or more heads
- Now let's say we have event B = heads on first toss.
- Our probability of at least two heads (P(A)) will now change if we include the fact that <u>we already know</u> we got a head on the first toss:
  - If we condition on B, then we remove all outcomes that don't have a head on the first toss.
- This lets us calculate $P(A|B)$

Example: Coin Flipping

- The total possible outcomes are:

  {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

- We had event A = two or more heads, and event B = head on first toss
- If we use our formula, we need to find two probabilities:
  - P(A and B) =
  - P(B) =
- So then P(A|B) =

- Note this is the same as when we first removed any outcomes that did not happen in B, and then finding how many were in A, out of what was left.

Example: Coin Flipping

- We could also have looked at this problem with a contingency table:
- We will let the columns of the table be the proportion of possible outcomes that are either in A or not in A
- Then we let the rows of the table be the proportion of possible outcomes that are either in B or not in B.

{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

	A = 2 or more heads	not A = 1 or fewer heads	Row Total
B = head on first toss			
not B = tail on first toss			
Column Total			8/8 = 1

Conditional Probability is a Probability Function

- Recall:
  - $0 \leq \ P(\omega) \leq 1$ for all $\omega \in \Omega$
  - $1 = P(\Omega) \ = \sum_{\omega \in \Omega} P(\omega)$
  - For all events $A \subseteq \Omega$, $\sum_{\omega \in \mathrm{A}} P(\omega) = P(A)$
- Similarly:
  - $0 \leq \ P(b|B) \leq 1$ for all $b \in B$
  - $1 = P(B|B) \ = \sum_{\mathrm{b} \in B} P(b|B)$
  - For all events $A \subseteq \mathrm{B}$, $\sum_{\mathrm{b} \in \mathrm{A}} P(b|B) = P(A|\mathrm{B})$

General Formula (General Multiplication Rule) for finding $P(A \cap B)$

- $P(A \cap B) \ = \ P(A|B)P(B)$
- Can be extended to more events
- $P(A_1 \cap A_2 \cap A_3) \ = \ P(A_3|A_1 \cap A_2)P(A_2|A_1)P(A_1)$
- $P(A_1 \cap \cdots \cap A_k) =$
  $P(A_k|A_1 \cap \cdots \cap A_{k-1})P(A_{k-1}|A_1 \cap \cdots \cap A_{k-2}) \ldots P(A_2|A_1)P(A_1)$
  - Proof: By induction (not covered in this class)

Law of Total Probability

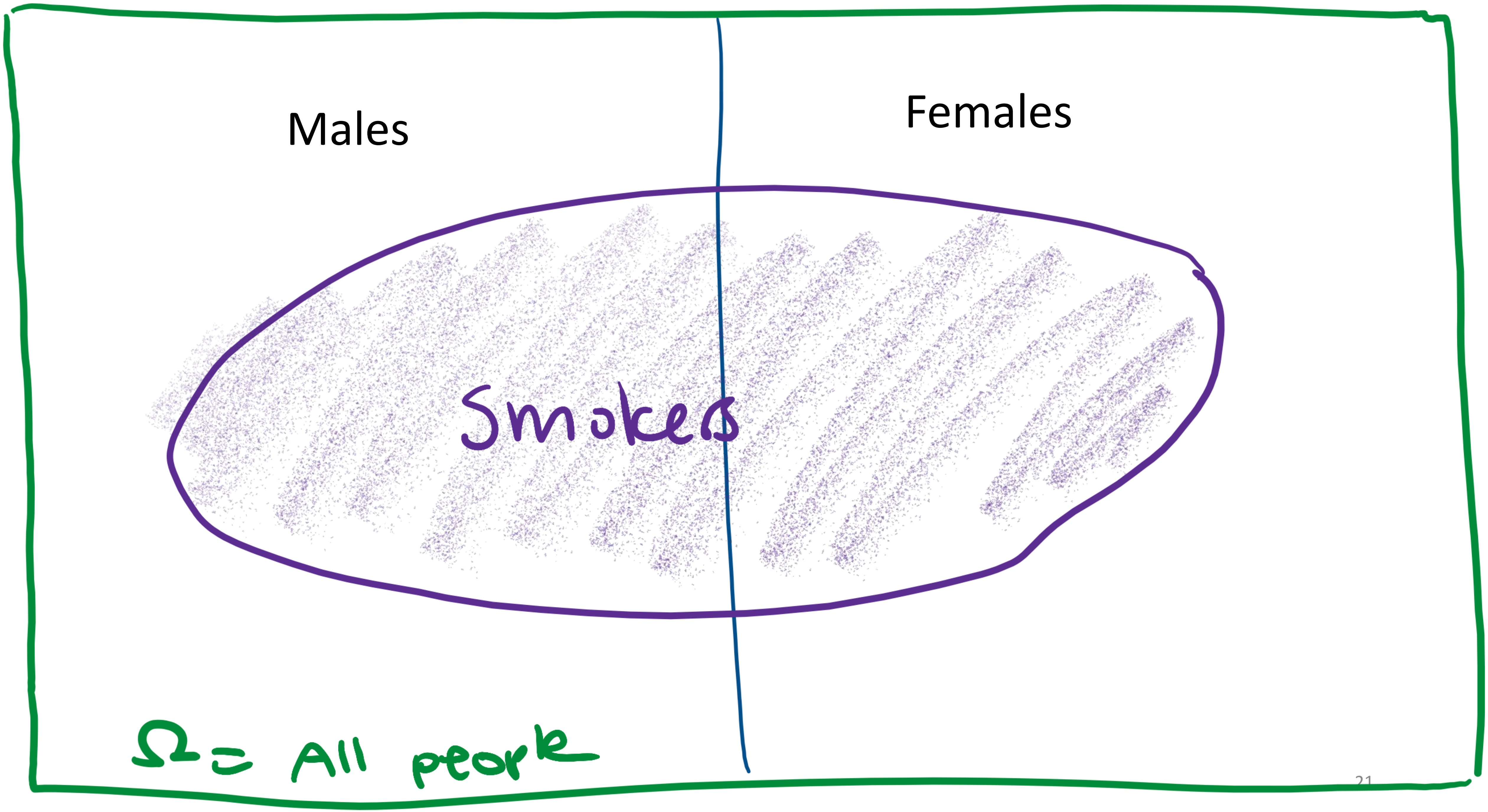
- Recall: Partition rule:
  - $P(A) = P(A \cap B) + P(A \cap B^c)$
  - $P(A) = P(A \cap B_1) + P(A \cap B_2)$ + …. $P(A \cap B_k)$ Where the B's make up the entire sample space and are mutually exclusive.
- We now combine the partition rule and the general multiplication rule
  - 2 partitions: $P(A) \ = \ P(A|B)P(B) \ + \ P(A|B^c)P(B^c)$
  - k partitions: $P(A) \ = \ P(A|B_1)P(B_1) \ + \cdots \ + P(A|B_k)P(B_k)$ Where the B's make up the entire sample space and are mutually exclusive.

Example: Smoking

- We know 60% of the population smokes, and 50% of the population is male. Given that someone is male, there is a 40% chance that they are a smoker. What is the probability of a female being a smoker?

- Step 1: Set up notation for events and list all the probabilities we know:

- Step 2: Use the law of total probability


Males
Females
Smokers
Ω = All people

Bayes' Rule and Inverting a Conditional Probability

Chapter 2.5

Bayes’ Rule

- Also: Bayes’ Formula
- This is a handy rule that will let us use conditional probabilities that we already know and “reverse” them to find a probability of interest
- A direct extension of the definition of conditional probability

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

- The difference is that instead of being given a value for P(B), you need to find it from other probabilities

Bayes' Rule

- Recall: $P(B) = P(B \cap A) + P(B \cap A^c) = P(B|A)P(A) + P(B|A^c)P(A^c)$ by law of total probability
- Recall: $P(A \cap B) = P(B|A)P(A)$ by general multiplication rule
- Expanding the numerator and denominator:

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

Tree Diagrams

- Tree diagrams are a tool you can use when you deal with conditional probabilities.
  - Multiply as you go left to right (as you go across a branch)
  - Add as you go up and down (within the same column)
- Tree diagrams let us see how many combinations of events there are
  - Recall: Multiplication Principle
- Tree diagrams are used in upper year finance courses as well, for example when pricing stock options.

Example: Gender of children

- There is a 50% of having a female firstborn child
- If we have a female firstborn, then the probability the second child is female is 1/3
- If we have a male firstborn, then the probability the second child is female is 2/5
- Set up notation and list all the probabilities we have:

- Note: Whenever you see the word “if” or “given” in a word problem, then that’s a sign you should consider using conditional probabilities!

Example: Gender of Children – Tree Diagram

Example: Gender of children

- What is the probability of a female first child given that we had a female second child? $P(F1|F2)$

Example: Gender of children

- The denominator is $P(F2)$
- Going back to our tree diagram, we have 2 tree branches that end at the second child being female
  - We need to add the probability of both to get the total probability of the second child being female
  - We have the following information:
    - Female second child if we already had female first, $P(F2|F1)$
    - Female second child if we had a male child first, $P(F2|M1)$
    - Female first child, $P(F1)$
    - Male first child, $P(M1)$

Example: Gender of Children

- Let's put everything together. What is $P(F1|F2)$?

Independence and Dependence

Chapter 2.6

Independence of Events

- The general idea of independence between 2 events is when knowing one event already happened, it does not influence whether or not the other event happens.
- As an example, suppose my experiment is to roll a dice and flip a coin.
  - Event A = roll a 3
  - Event B = land on heads
- Does knowing that I flipped a head with my coin give me any information (that I didn't already have) about the chance that I will roll a 3? i.e. are A and B independent?
  - The key to determining independence is this idea of providing information that didn't already exist.

Independence: Definition

We can show independence in 2 ways:

1. Conditional probability: if knowing that B happens doesn't add information about A happening, then

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

   - Similarly, if knowing A happens doesn't give me information about the chances of B happening, then

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

2. Multiplication Rule for independent events: We can say that events A and B are independent when

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

Independence: Definition

- The two definitions of independence are equivalent.
  - Recall that a conditional probability can be written as

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

  - If A and B are independent, then by the multiplication rule,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \times P(B)}{P(B)} = P(A)$$

  - By using the multiplication rule, we get back out first definition of independence

Example: Roll dice/toss coin

- Again, A = {roll a 3} and B = {flip a head}.
- Let's check P(A) and P(A|B) and see if they are equal.
- Let's also check to see if P(A) x P(B) = P(A and B)
- Let's take this back to contingency tables:

S = {(H,1), (H,2), (H,3), (H,4), (H,5), (H,6), (T,1), (T,2), (T,3), (T,4), (T,5), (T,6)}

	A = roll a 3	not A = did not roll a 3	
B = flip a head			
not B = flip a tail			
			12/12

Mutually Exclusive Events

- **Mutually exclusive and independent are not the same thing**
- The key difference between independent events and disjoint/mutually exclusive events is <u>what is going on in the overlap between the events</u>, i.e. P(A and B):
  - **Independent events**: there are common outcomes in the overlap, but they can be split nicely into P(A)P(B)
  - **Mutually exclusive events**: there are no common outcomes in the overlap, i.e. the overlap is empty.

Can disjoint (mutually exclusive) events ever be independent?

- What are the definitions of mutually exclusive and independent?

Example: Blood Types

Canadian Blood services says that about 46% of the Canadian population have Type O blood, 42% have Type A blood, 9% have Type B blood, and the rest, which is 3% have Type AB.

Let’s first discuss some assumptions we need to make about:

- Mutually exclusive events
- Independent events

We always have to make assumptions whenever we work with data or probability models

If you examine one person, are the events that the person is Type A and that the person is Type B mutually exclusive or independent or neither? Explain.

- Event A: The person has blood type A
- Event B: The person has blood type B

If you examine two people, are the events that the first is Type A and the second is Type B mutually exclusive or independent or neither? Explain.

- Event A: Person 1 is type A
- Event B: Person 2 is type B
- No information is given on if the patients are related – let's assume they are not for the purposes of this question.

a) We randomly pick 4 people. What’s the probability all 4 individuals are type O

b) We randomly pick 4 people. What's the probability that no one is type AB

c) We randomly pick 4 people. What's the probability at least 1 person is Type B

What if we have more than 2 events?

- Mutual independence: For a general collections of events, independence means that for every finite subcollection $A_i, \ldots, A_k$, $P(A_1, \ldots, A_k) = P(A_1) \times \ldots \times P(A_k)$
- Pairwise independence: A collection of events is pairwise independent if $P(A_i A_j) = P(A_i)P(A_j)$ for all pairs of events
- A collection of mutually independent events is pairwise independent, BUT, events that are pairwise independent are not necessarily mutually independent

Example: Two Coin Flips (Ex 2.21)

- Flip two coins. Let A be the event that the first coin comes up heads; B the event that the second comes up heads; and C the event that both coins come up the same, either heads or tails.
- P(A) = P(B) = P(C) = 1/2.
- Are all 3 events mutually independent? $P(A \cap B \cap C) =$ $P(\text{First coin is heads, second coin is heads, both coin are the same}) =$ $P(HH) = 1/4$
  - However, P(A)P(B)P(C) = 1/8, so they are not mutually independent!
- But, they are pairwise independent!
  - $P(A \cap B) = P(A \cap C) = P(B \cap C) = P(HH) = \frac{1}{4} = P(A)P(B) = P(A)P(C)$ $= P(B)P(C)$

Example: Thrombosis

A genetic test is used to determine if people have a predisposition for thrombosis (i.e. blood clotting that blocks blood flow). It is believed that 3% of people actually have this predisposition. The genetic test has probability 99% of giving positive result when person actually has it. It also has probability 98% of giving a negative result when the person does not have the predisposition. What is the probability that someone who tests positive for the predisposition actually has it?

Example: Thrombosis

- Set up notation

- Calculate conditional probability

Summary

- Conditional probability formula: $P(A|B) \; = \; P(A \cap B)/P(B)$
- General multiplication rule: $P(A \cap B) \; = \; P(A|B)P(B) \; = \; P(B|A)P(A)$
- Law of total probability: $P(A) \; = \; P(A|B)P(B) \; + \; P(A|B^c)P(B^c)$
- Bayes' formula: $P(B|A) \; = \; \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$
- Independent events: Events A and B are independent if $P(A|B) \; = \; P(A)$. Equivalently, $P(A \cap B) \; = \; P(A)P(B)$
- Mutual independence: $P(A_1, \ldots, A_k) = \; P(A_1) \times \ldots \times P(A_k)$, for every finite subcollection $A_i, \ldots, A_k$
- Pairwise independence: $P(A_iA_j) \; = \; P(A_i)P(A_j)$ for all pairs of events

Useful Tools and Tips

- Tree diagrams
  - For ordered events
- Contingency tables
  - When there are two random experiments with finite sample spaces
- Conditional probabilities as a tool
  - Similar to partition rule, it's often easier to find $P(A|B_i)$ than $P(A)$
- Don't memorize all the formulas! They're all applications the same couple of rules!
- Always write down the probabilities you are given, and what you are trying to find. What can you use to connect them?

Next Week

- We will cover start talking about discrete random variables and various probability distributions (Chapters 3-5)

Your tasks this week

- Weekly Reflection: Due Sunday at 11:59pm
- Textbook Readings: Sections 2.1-2.6 (pp 45-80)
- Recommended Textbook Practice Problems: 2.2, 2.6, 2.8, 2.9, 2.11, 2.14, 2.17, 2.26 (Hint: Conditioning on tails restarts the process), 2.28, 2.36, 2.37 (pp 83-89)
- Do R Activity 1 if you haven't already:
  - https://rconnect.utstat.utoronto.ca/content/b38a0e6b-221b-4103-ab76-512bdce9d260/
- Tutorials start this week! Materials for Tutorial 1 are here:
  - https://q.utoronto.ca/courses/354355/files/33250251