

These lecture slides are designed to accompany:  
**Introductory Statistics, Third Edition**

Other features include:



Chapter summary videos



MP3 audio podcasts



VirtualTutor e-learning



AntiCheat and AutoGrade homework

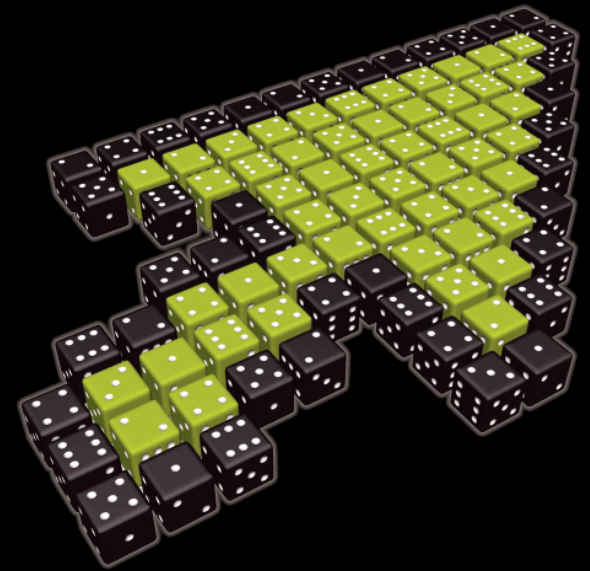


Detailed instructor resources

To find out more, visit: [perdisco.com/stats](http://perdisco.com/stats)

# Chapter 4

## Probability



# What is probability?

---

- In basic terms, probability is a number that we assign to indicate the **likelihood** of an event
- Expressed as:
  - a decimal between 0 and 1
  - a percentage between 0% and 100%
- Two main kinds of probability:
  - **relative frequency probability**
  - **a priori classical probability**

# Patterns in randomness

---

- If a process is ‘random’, we tend to think it is ‘unpredictable’, and without ‘pattern’
- But there is plenty of pattern in randomness!
- Example: Imagine rolling a fair 6-sided die
- This is about as random as you can get
- But you can observe patterns
  - e.g. around 1 in every 6 times you roll, you’ll get a 4
  - that is, around  $1/6$  of the rolls are 4’s



# Relative frequency

---

- In this case, we say that the probability of a 4 turning up is  $1/6$  ( $= 0.16666\dots$ )
- That is, the probability is the proportion of times it is **seen** to occur
- This is the **relative frequency** approach
  
- If a process can be observed over and over again, the **probability** of any outcome of that process is the relative frequency with which it is seen to occur

# Limits to relative frequency

---

- Using relative frequency is practical and empirical
- But this can have disadvantages!
- Theoretically, we would like to make **infinitely** many observations, but this is not possible
  - Even after ‘many’ observations, the probability you assign is only an estimate
- Also, observing a proportion doesn’t tell you anything about **why** the probability is the value it is

# Rules of probability

---

- Two basic rules that probabilities follow:
- **Rule 1:** Probabilities are always between 0 and 1
- **Rule 2:** A probability of 0 means an event is impossible (never occurs), and a probability of 1 means an event is certain (always occurs)
- When we look at the more formal **a priori** definition of probability, we will develop more rules

# Formalizing probability

---

- Relative frequency is not as **formal** as we'd like
- Example: Suppose you roll a die 600 times, and you get a 4 on 98 of those rolls
- Relative frequency would tell you to assign a probability of  $98/600$  (which is not quite  $1/6$ )
- But don't you **know** that the probability is **really** 'meant to be'  $1/6$ ?
- We need a new approach!

# Outcomes

---

- Whenever an observable procedure can occur, **outcomes** are such recordable observations
- Examples
  - Flipping a coin – two outcomes (heads and tails)
  - Rolling a die – six outcomes (1, 2, 3, 4, 5, 6)
- Outcomes can never occur together
  - e.g. you can't have heads **and** tails
- The set of all outcomes covers all possibilities
  - e.g. you flip a coin, you **must** get heads or tails!



# Events

---

- Set of all outcomes is called the **sample space,  $S$** 
  - e.g. sample space for a die roll  $S = \{1, 2, 3, 4, 5, 6\}$
- An **event** is any outcome or combination of outcomes in a sample space
- Example: Roll a die, you can get an even number
- If we call this event  $A$ , it is made up of three outcomes written like this:

$$A = \{2, 4, 6\}$$

# Complement

---

- For an event,  $A$ , the **complement** of  $A$ , denoted  $A^c$ , is the event that  $A$  does **not** occur
- In other words,  $A^c$  is the set of all outcomes in the sample space that are **not** in  $A$
- Example
  - roll a die, consider event that we get 2 or 3,  $A = \{2, 3\}$
  - $A^c$  is the event that we **don't** get a 2 or a 3
  - that is,  $A^c = \{1, 4, 5, 6\}$

# Union and intersection

---

- For two events A and B, the **union** of A with B is the event that **at least one of** the two events occurs
- We refer to the union as 'A or B'
- The **intersection** of A and B is the event that **both** of the events occur
- We refer to the intersection as 'A and B'
- Example: if  $A = \{2, 3\}$  and  $B = \{2, 4, 6\}$  then  
 $A \text{ or } B = \{2, 3, 4, 6\}$        $A \text{ and } B = \{2\}$

# Properties of events

---

- Two events are **mutually exclusive** if it is impossible that they occur simultaneously
- That is, if their intersection has no outcomes
- A set of events is **collectively exhaustive** if at least one of the events must occur
- That is, if their union contains all of the outcomes in the sample space
- Note: Outcomes are always mutually exclusive, and the set of all of them is collectively exhaustive!

# A priori classical probability

---

- The probability of an event is defined in terms of the number of outcomes in that event
- An assumption: all outcomes are **equally likely**
- Then, in a sample space of  $n$  outcomes, each outcome is assigned a probability of  $1/n$
- And the **probability** of an event  $A$  is defined as:

$$P(A) = \frac{\text{number of outcomes in } A}{n}$$

# Example of a priori probability

---

- If you roll a fair six-side die, you **assume** that all six outcomes are equally likely
- So you assign a probability of  $1/6$  to each outcome
- What is the probability of getting an even number?
- There are **3** outcomes in this event  $A = \{2, 4, 6\}$
- So the probability is

$$P(A) = \frac{3}{6} = \frac{1}{2}$$

# Rules of probability

---

- We can now expand our probability rules
- **Rule 1:** The probability that some outcome in the sample space will occur is 1
- **Rule 2:** The probability that no outcome in the sample space will occur is 0
- **Rule 3:** All probabilities are between 0 and 1
- **Rule 4:**  $P(A \text{ or } B) = P(A) + P(B)$ , provided A and B are mutually exclusive
- **Rule 5:**  $P(A^c) = 1 - P(A)$

# Calculating probabilities

---

- To calculate, we typically use **a priori** definition
- So to answer: What is the probability of an event?
- We need to ask:
  - How many different ways can the event occur?
  - How many different outcomes are in the sample space?
- Therefore, **counting** is very important

# Contingency table

---

- Tables can be used to help us enumerate events
- Example: 1000 people asked about gender and employment status

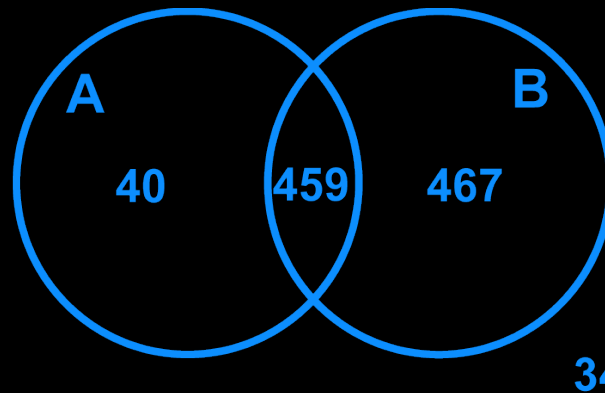
| Employed | Gender |        |
|----------|--------|--------|
|          | Male   | Female |
| Yes      | 459    | 467    |
| No       | 40     | 34     |

- From this you can tell, for example:
  - 459 are male and employed
  - 499 ( $= 459 + 40$ ) are male
  - 926 ( $= 459 + 467$ ) are employed

# Venn diagram

---

- Suppose  $A$  = event that a person chosen is male
- And  $B$  = event that person is employed
- Then this is shown in a Venn diagram like this:



- Area covered by both circles is intersection A and B
- That is, 459 people are male **and** employed

# Example of calculating a probability

---

| Employed | Gender |        |
|----------|--------|--------|
|          | Male   | Female |
| Yes      | 459    | 467    |
| No       | 40     | 34     |

- We can use this table to calculate probabilities
- Example: Probability that a randomly chosen person from the 1,000 is male **and** employed,  $P(A \text{ and } B)$ ?
- Well, 459 out of 1,000 possible outcomes lead to this event
- So  $P(A \text{ and } B) = 459/1000 = 0.459$

# Another example

---

| Employed | Gender |        |
|----------|--------|--------|
|          | Male   | Female |
| Yes      | 459    | 467    |
| No       | 40     | 34     |

- What about the probability that a person chosen is male,  $P(A)$ ?
- Well,  $459 + 40 = 499$  are male
- So  $P(A) = 499/1000 = 0.499$

# The general addition rule

---

| Employed | Gender |        |
|----------|--------|--------|
|          | Male   | Female |
| Yes      | 459    | 467    |
| No       | 40     | 34     |

- What about the probability that a person is male **or** employed,  $P(A \text{ or } B)$ ?
- Is it equal to  $P(A) + P(B)$ ? No!
- Adding all males (499) and all employed people (926), means 459 people (male **and** employed) get counted twice!

# The general addition rule (cont'd)

---

- So when calculating  $P(A \text{ or } B)$  in general, you must subtract  $P(A \text{ and } B)$  to get answer:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

- This is the **general addition rule**

# Conditional probability

---

- The probability of an event can change if:
  - we are given some new condition
  - we are told that some other event has occurred



- Example:
  - What is the probability that a person has children?
  - What if you were told that the person was married?

# Conditional probability defined

---

- The **conditional probability** of A, given that another event B has occurred is:

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

- We refer to  $P(A|B)$  as the 'probability of A, given B'
- It can also be thought of as the following ratio:

$$P(A | B) = \frac{\text{number of ways 'A and B' can occur}}{\text{number of ways B can occur}}$$

# Decision tree

---

- Can be used to help calculate conditional probability

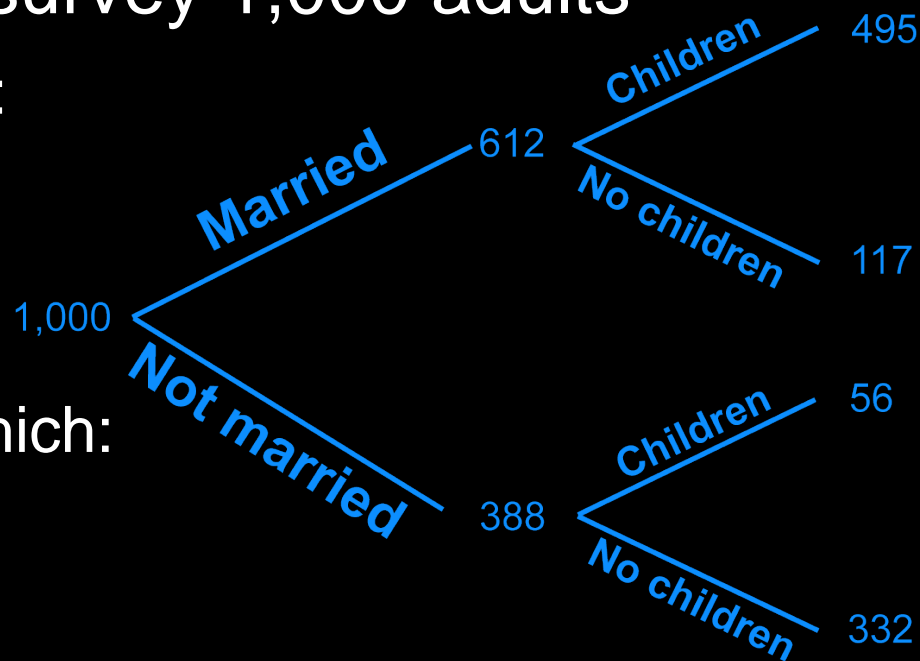
- Example: Suppose you survey 1,000 adults

- 612 are married, of which:

- 495 have children
- 117 do not have children

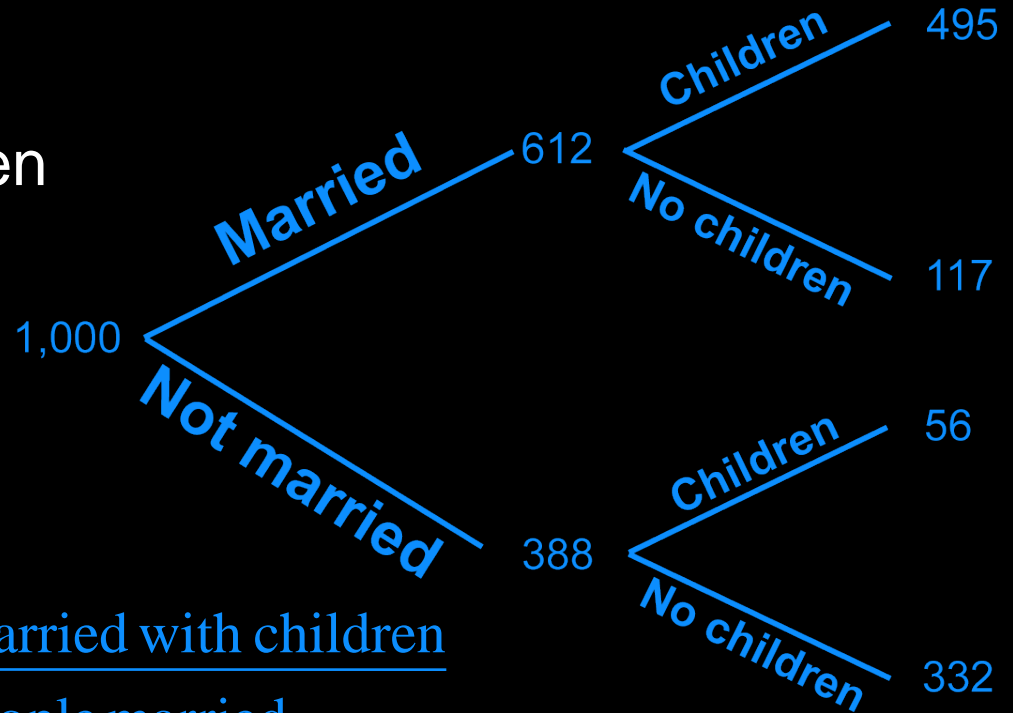
- 388 are not married, of which:

- 56 have children
- 332 do not have children



# Using the decision tree

- Denote:
  - A = person has children
  - B = person is married
- Then  $P(A|B)$  is:



$$\begin{aligned}P(A|B) &= \frac{\text{number of people married with children}}{\text{number of people married}} \\ &= \frac{495}{612} \\ &= 0.8088 \dots\end{aligned}$$

# The general multiplication rule

---

- Recall the conditional probability of A given B:

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

- This formula can be re-arranged to give:

$$P(A \text{ and } B) = P(A | B) \times P(B)$$

- This is the **general multiplication rule**

# The general multiplication rule (cont'd)

---

- Example: Suppose 60% of statistics student receive tutoring.
- Of the students that get tuition, 80% get a credit or better.
- What proportion get tuition **and** get a credit or better?
- Let  $A$  = gets credit+,  $B$  = gets tuition
- Then  $P(B) = 0.6$  and  $P(A|B) = 0.8$
- So  $P(A \text{ and } B) = P(A|B) \times P(B) = 0.8 \times 0.6 = 0.48$

# Independence

---

- Sometimes, the probability of A doesn't change, regardless of whether or not B occurred
- That is:  $P(A | B) = P(A)$
- When this occurs, we say A and B are **independent**
- Example: Roll two dice. The outcome on one die is independent of what happens to the other
- For independent events, the general multiplication rule is simplified:

$$P(A \text{ and } B) = P(A) \times P(B)$$

# Bayes' Theorem

---

- We might want to reverse a conditional probability
- We might know  $P(B|A)$ , but want to know  $P(A|B)$
- Are they the same? No!
- There are various versions of Bayes' Theorem to help calculate  $P(A|B)$  from  $P(B|A)$
- Simplified Bayes' Theorem:

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

# Another version of Bayes' Theorem

---

- We often don't have enough information to use the simplified version!
- That is, we don't (directly) know all three probabilities  $P(A)$ ,  $P(B)$ ,  $P(B|A)$
- The most common version of Bayes' Theorem is:

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

# Example of using Bayes' Theorem

---

- Suppose you play tennis against your friend
- You win 56% of the time, lose 44% of the time
- Of the games you won, 90% of the time you trained before the game
- When you lose, 20% of the time you trained before the game
- You are about to play your friend today, and you've just had a training session.
- What is the probability that you will win?

# Example of using Bayes' Theorem (cont'd)

---

- Let  $A$  = you win,  $B$  = you train before the game
- You want to know  $P(A|B)$
- You know:
  - $P(A) = 0.56$ ,  $P(A^c) = 0.44$
  - $P(B|A) = 0.9$ ,  $P(B|A^c) = 0.2$

▪ So:

$$\begin{aligned}P(A | B) &= \frac{P(B | A) \times P(A)}{P(B | A) \times P(A) + P(B | A^c) \times P(A^c)} \\ &= \frac{0.9 \times 0.56}{0.9 \times 0.56 + 0.2 \times 0.44} \\ &= 0.8514\dots\end{aligned}$$

# Full version of Bayes' Theorem

---

- There is actually a more complex version of the theorem
- Suppose B is any event, and  $\{A_1, A_2, \dots, A_n\}$  are mutually exclusive, collectively exhaustive events
- Then for any  $A_i$

$$P(A_i | B) = \frac{P(B | A_i) \times P(A_i)}{P(B | A_1) \times P(A_1) + \dots + P(B | A_n) \times P(A_n)}$$

- That's as complex as it gets!