

# Class 09

## DATA1220-55, Fall 2024

Sarah E. Grabinski

2024-09-18

# Homework 2

- ▶ Instructions (`homework2_instructions.pdf`), a Quarto markdown template (`homework2_template.qmd`), and an example HTML output (`homework2_example.html`) are available for download under Chapter 2 on the Modules page in Canvas.
- ▶ Upload **TWO** (2) documents to Homework 2 on the Assignments page in Canvas by Friday 9/20/2024 by 6:00pm: `homework2_yourlastname.qmd` and `homework2_yourlastname.html`
- ▶ Video walk-through of Homework 2 under Tutorials on the Modules page in Canvas. Make sure you're caught up on the video walk-through of homework 1.

# Late Policy

“This homework is due by 6:00pm on Friday, 9/20/24. No credit will be lost for assignments received by 7:00pm to account for issues with uploading. 10% of the points will be deducted from assignments received by 9:00am on Saturday, 9/21/24.

Assignments turned in after this point are only eligible for 50% credit, so it benefits you to turn in whatever you have completed by the due date.”

# How can I get help with homework?

- ▶ **Read the textbook.** Many of you are asking for additional examples. Luckily, there are tons we didn't go over in the textbook.
- ▶ **Look at the homework early.** Only 1 person has looked at the homework since I posted it. Make sure you leave enough time to get help if you need it.
- ▶ **Ask a question on our Campuswire class feed.** I'm only one person, and I may not be able to give you a prompt answer. However, the 20+ other people in the class might be able to.
- ▶ **Come to office hours.** I will be available after class today Wednesday 9/25/2024 from 2:30pm - 4:00pm. If you cannot make it, reach out to me to try and schedule an appointment.

# Chapter 3 Objectives

- ▶ Define probability, random processes, and the law of large numbers
- ▶ Describe the sample space for disjoint and non-disjoint outcomes
- ▶ Calculate probabilities using the General Addition and Multiplication Rules
- ▶ Create a probability distribution for disjoint outcomes

# Defining Probability

What does the word **probability** mean to you?

# Defining Probability

What does the word **probability** mean to you?

“Highly likely”

# Defining Probability

What does the word **probability** mean to you?

“Highly likely”

“Probably”



# Defining Probability

What does the word **probability** mean to you?

“Highly likely”

“Probably”

“About even”

# Defining Probability

What does the word **probability** mean to you?

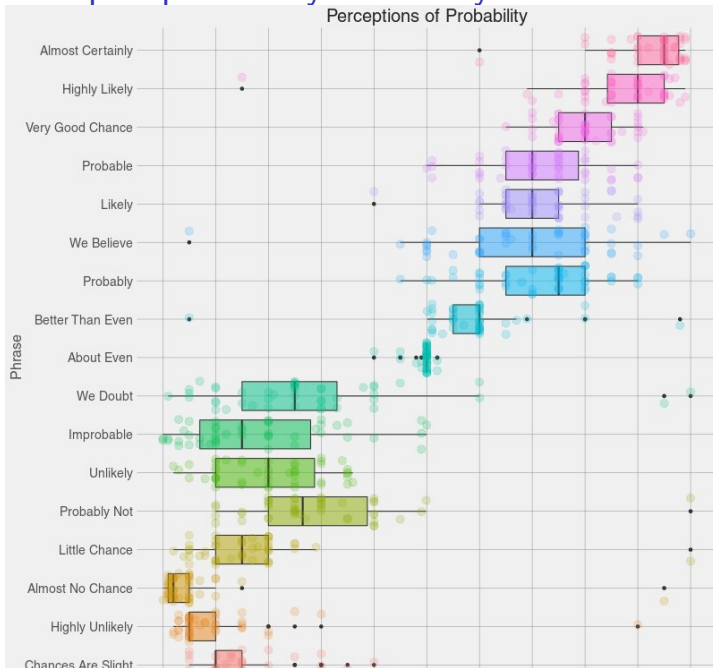
“Highly likely”

“Probably”

“About even”

“Almost no chance”

# People interpret probability differently



# So what is probability?

## **i** Frequentist Definition

The proportion of times that a particular outcome would occur if we observed a random process an infinite number of times.

- ▶ A **random process** is one where you know which outcomes are possible (i.e. the **sample space**) but you don't know which outcome comes next
- ▶ Examples of a **random process**: coin toss, die roll, stock market

# How do you know a process is random?

**POPULAR SCIENCE**SCIENCE • TECHNOLOGY • ENVIRONMENT • DIY • GEAR • MERCH • NEWSLETTER🔍📱📺📺📧

Big Fall Wellness Sale

Our lowest prices of the year!

Now \$394


Shop now


**TECHNOLOGY**

## A brief history of shuffling your songs, from Apple to Adele

Spotify made a change to one of music's most popular features. Here's what that means for how we listen to tunes.

BY [SHIRA FEDER](#) ✓ POSTED ON NOV 30, 2021 3:00 PM EST






**SHIRA FEDER**  
Contributor, Tech

Shira Feder covers tech, science, and health. She holds a master's degree from the Craig Newmark Graduate School of Journalism and has written for The Washington...

2025 Equinox EV LT starting at \$34,995<sup>2</sup>



2024 Equinox EV LT as shown: \$43,295<sup>1</sup>

Learn More


 **CHEVROLET**

Figure 2: Both Apple and Spotify took steps to make their “shuffle” features less random after complaints from users

# A brief history of “Shuffle”

- ▶ January 11, 2005 – Apple releases the iPod Shuffle, a small device capable only of playing music randomly (“true” shuffle)
- ▶ September 7, 2005 – Apple offers “Smart Shuffle” in response to complaints, which controlled how likely songs from the same album or artist would play close together
- ▶ July 2011 – Spotify launches in the United States using the Fisher-Yates Algorithm, which is like picking tickets out of a hat until no more remain
- ▶ February 2014 – Spotify modifies their sampling algorithm to ensure an even distribution across albums/artists

# What went wrong?

- ▶ The human brain is good at finding patterns in noise, even when there are none
- ▶ If an artist is repeated “too soon”, the listener doesn't feel the order is random
- ▶ We perceive a “random” distribution as also being “uniform” and “fair”

# So why didn't we like a “true” random shuffle?

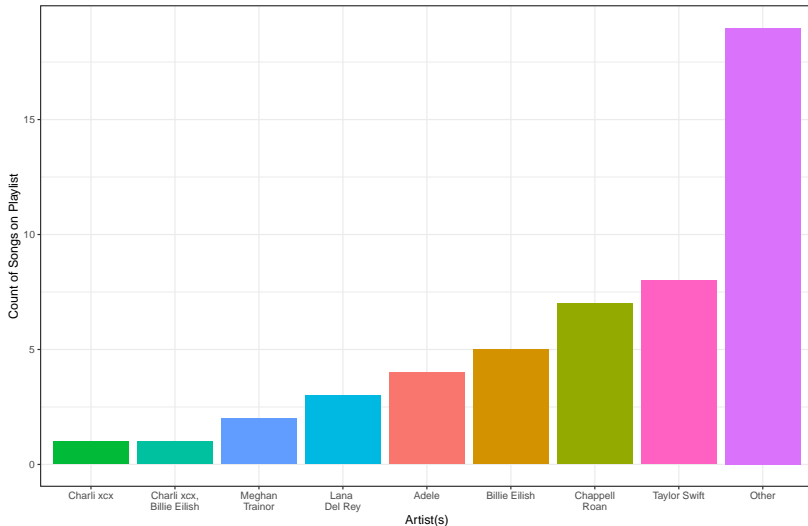
- ▶ Songs not evenly distributed across albums and artists on a playlist
- ▶ Some albums/artists may play more frequently than others simply because they have more songs in the library/on the playlist
- ▶ Each song is equally likely to play next, but not each artist
- ▶ Artists/albums with more songs also more likely to play in a row
- ▶ A true random shuffle might play the same artist multiple times in a row
- ▶ It's unusual but not impossible to roll a 1 on a die 3 times in a row
- ▶ It's also possible for the same song to play twice in a row



# Example: Spotify Playlists

Number of Songs on the 'Taylor Swift Radio' playlist on Spotify by Artist(s)

Total artists = 26



# What if shuffle was truly random?

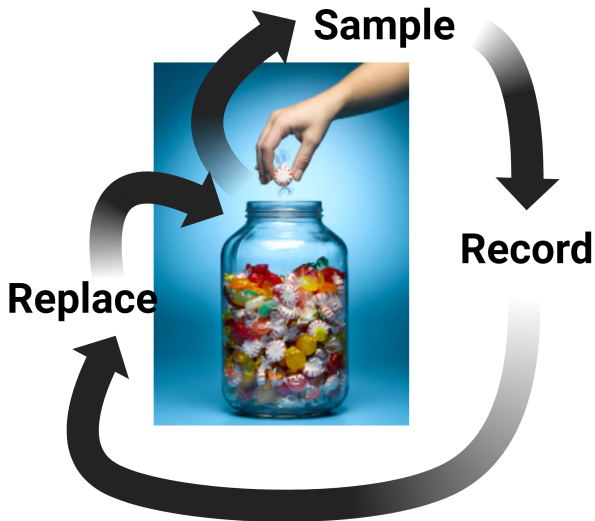
Each time the song changes, every song on the playlist is eligible to be played next

- ▶ Does not matter if the song was just played
- ▶ Does not matter who the artist is

We call this ***sampling with replacement***.

- ▶ Like drawing a playing card, looking at it, then putting it back in the deck before the next draw
- ▶ Repetition of outcomes is possible

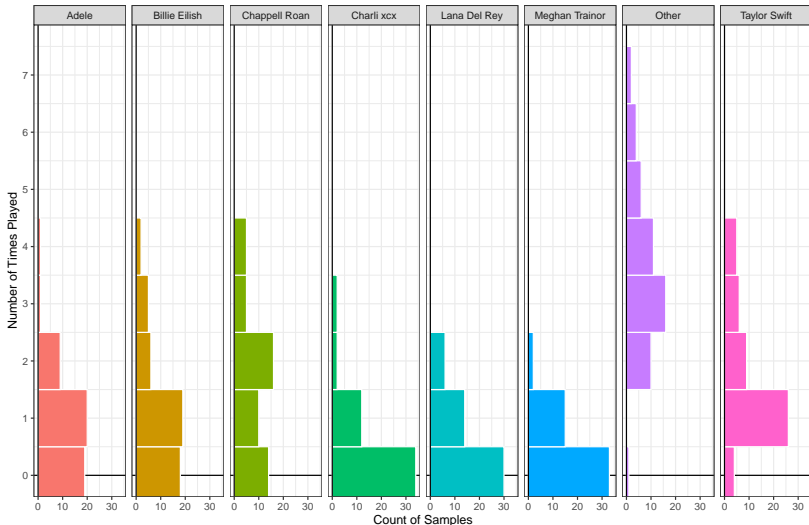
# Sampling With Replacement



# How often were artists repeated during Spotify's original shuffle?

Number of Times Artist is Played in 10 Songs

n = 50



# What is the probability of hearing a song by Chappell Roan?

- ▶ There is our “observed” probability that the next song is by Chappell Roan
- ▶ Sample proportion ( $\hat{p}_n$ )
- ▶ There is some “true” real-world probability that the next song is by Chappell Roan
- ▶ Population proportion ( $p$ )

# Defining the sample space

The **sample space**  $s$  or  $S$  is the total collection of possible outcomes for a **random process**.

- ▶ Die rolls: 1, 2, 3, 4, 5, 6
- ▶ Coin flips: heads, tails
- ▶ Stock market: up, down, no change

Here, the **sample space** could be all the songs on the playlist ( $n = 50$ ) or all the artists who perform them ( $n = 26$ ).

# Defining the Sample Space

Flipping a Coin



SAMPLE SPACE

{Head, Tail}  
Uniform

Rolling a Six Sided Dice



SAMPLE SPACE

{1, 2, 3, 4, 5, 6}  
Uniform

Spinning a 4 color spinner



SAMPLE SPACE

{Red, Yellow, Green, Blue}  
Uniform

Rolling a Weighted Dice



SAMPLE SPACE

{4, 5, 6}  
Not Uniform

Picking a flavor of ice cream



SAMPLE SPACE

{Chocolate, Vanilla, Strawberry}  
Uniform

Determining the gender of baby



SAMPLE SPACE

{Boy, Girl}  
Uniform

Picking from a bag of marbles



SAMPLE SPACE

{Blue, Red}  
Not Uniform