COMS BC1016

Introduction to Computational Thinking and Data Science

Lecture 10: Probability and Sampling

Oct 6, 2025

Upcoming Schedule

Today

Date	Topic	Lab	Assignment
10/1	9 - Conditionals and Iteration	Lab 4 - Functions and Visualizations (Due 10/3) Courseworks	HW2 Due
10/6	10 - Probability and Sampling		HW4 - Probability, Simulation, Estimation (Due 10/15) Courseworks
10/8	11 - Models and Empirical Simulations	Lab 5 - Simulations (Due 10/10) Courseworks	HW3 Due
10/13	Programming/Python Review		
10/15	Midterm Review	*No Lab*	HW4 Due
10/20	**Midterm Exam**		
10/22	Special Topics - Bias in AI	*No Lab*	

Lecture Outline

- Control Statements
  - For loops
- Randomness
- Probabilities
- Sampling

Control Statements

Control Statements

Control Statements modify *if* and/or *how many times* a block of code is executed in a program

Control Statements

- Two major types are **`if`** and **`for`**
  - **`if`** statements specify code that should be run conditioned on something being true
    - They can also specify if alternative code should be run otherwise
  - **`for`** loops allow executing code over each element in some sequence of items

`if` statements

- Conditionals begin with an `if` followed by a boolean statement
  - Runs code based on whether a boolean statement evaluates to `True`
- Conditionals can include a combination of `if`, `elif`, and `else` clauses
  - Maximum of one `if` and one `else`

`if` statements

```
if statement_1:
   first_code_block
elif statement_2:
   second_code_block
elif statement_3:
   third_code_block
else:
   fourth_code_block
```

if statements

```
if statement_1:
    first_code_block
elif statement_2:
    second_code_block
elif statement_3:
    third_code_block
else:
    fourth_code_block
```

Runs if statement_1 == True

Runs if statement_1 != True
AND statement_2 == True

statement_1 != True
AND statement_2 != True
AND statement_3 == True

nothing above == True

Iteration

- **Iteration** means to repeat a process or steps
    - For example, coming up with a design, prototyping, testing, and then repeating these steps based on the outcome
- In programming we use this term to refer to executing code repeatedly over every element in a list/array/sequence/collection/…
    - The object being iterated over is referred to as an **iterable**

Iterables

- Formally, an iterable is any Python object capable of returning its members one at a time

- Iterables we've seen in this class include:
  - Arrays
  - Lists
  - String

We'll mostly focus on arrays

```
make_array('a','b','c','d')
```

```
array(['a', 'b', 'c', 'd'],
      dtype='<U1')
```

```
['a','b','c','d']
```

```
['a', 'b', 'c', 'd']
```

```
'abcd'
```

```
'abcd'
```

`for` Statements

- Executing a **`for`** runs code with each element in an iterable

variable name

array of values

```
for item in some_array:
    print(item)
```

code to evaluate in each iteration of the loop

for Example

```python
for i in np.arange(4):
    print('iteration', i)
```

`for` Example

```python
for i in np.arange(4):
    print('iteration', i)
```

```python
np.arange(4)
```

```
array([0, 1, 2, 3])
```

`for` Example

```
for i in np.arange(4):
    print('iteration', i)
```

```
iteration 0
```

```
np.arange(4)
```

```
array([0, 1, 2, 3])
```

i=0

`for` Example

```
for i in np.arange(4):
    print('iteration', i)
```

```
iteration 0
iteration 1
```

```
np.arange(4)
```

```
array([0, 1, 2, 3])
```

i=1

for Example

```
for i in np.arange(4):
    print('iteration', i)
```

```
iteration 0
iteration 1
iteration 2
```

```
np.arange(4)
```

```
array([0, 1, 2, 3])
```

i=2

for Example

```
for i in np.arange(4):
    print('iteration', i)
```

```
iteration 0
iteration 1
iteration 2
iteration 3
```

```
np.arange(4)
```

```
array([0, 1, 2, 3])
```

i=3

for Example

```python
total = 0
for i in np.arange(4):
    total = total + i
    print(total)
```

```python
np.arange(4)
```

```
array([0, 1, 2, 3])
```

for Example

```
total = 0
for i in np.arange(4):
    total = total + i
    print(total)
```

```
0
```

```
np.arange(4)
```

```
array([0, 1, 2, 3])
```

i=0

for Example

```
total = 0
for i in np.arange(4):
    total = total + i
    print(total)
```

```
0
1
```

```
np.arange(4)
```

```
array([0, 1, 2, 3])
```

i=1

for Example

```python
total = 0
for i in np.arange(4):
    total = total + i
    print(total)
```

```
0
1
3
```

```python
np.arange(4)
```

```
array([0, 1, 2, 3])
```

i=2

for Example

```
total = 0
for i in np.arange(4):
    total = total + i
    print(total)
```

```
0
1
3
6
```

```
np.arange(4)
```

```
array([0, 1, 2, 3])
```

i=3

Simulation

General Process for Simulations

1. Figure out what you want to simulate
   - Example: Outcomes of a coin toss

General Process for Simulations

1. Figure out what you want to simulate
   - Example: Outcomes of a coin toss

2. Write a function whose output is the outcome of a single simulation

General Process for Simulations

1. Figure out what you want to simulate
   - Example: Outcomes of a coin toss
2. Write a function whose output is the outcome of a single simulation
3. Repeat the simulation for some number of iterations
   - Keep track of the results of every iteration in an array

General Process for Simulations

1. Figure out what you want to simulate
   - Example: Outcomes of a coin toss
2. Write a function whose output is the outcome of a single simulation
3. Repeat the simulation for some number of iterations
   - Keep track of the results of every iteration in an array
4. Add results array to a table so you can plot the results

Random Selection

```
import numpy as np
```

To select uniformly at random from array `some_array`

- `np.random.choice(some_array)`

To select `n` number of random elements from array `some_array`

- `np.random.choice(some_array, n)`

Note: Random does not mean arbitrary.

We mean each output has some chance of happening *(probability)*

Appending Arrays

```
import numpy as np
```

Return a copy of `array_1` where `value` is added onto the end

```
np.append(array_1, value)
```

Returns an array with elements of `array_1` followed by elements of `array_2`

```
np.append(array_1, array_2)
```

Probability

Probability

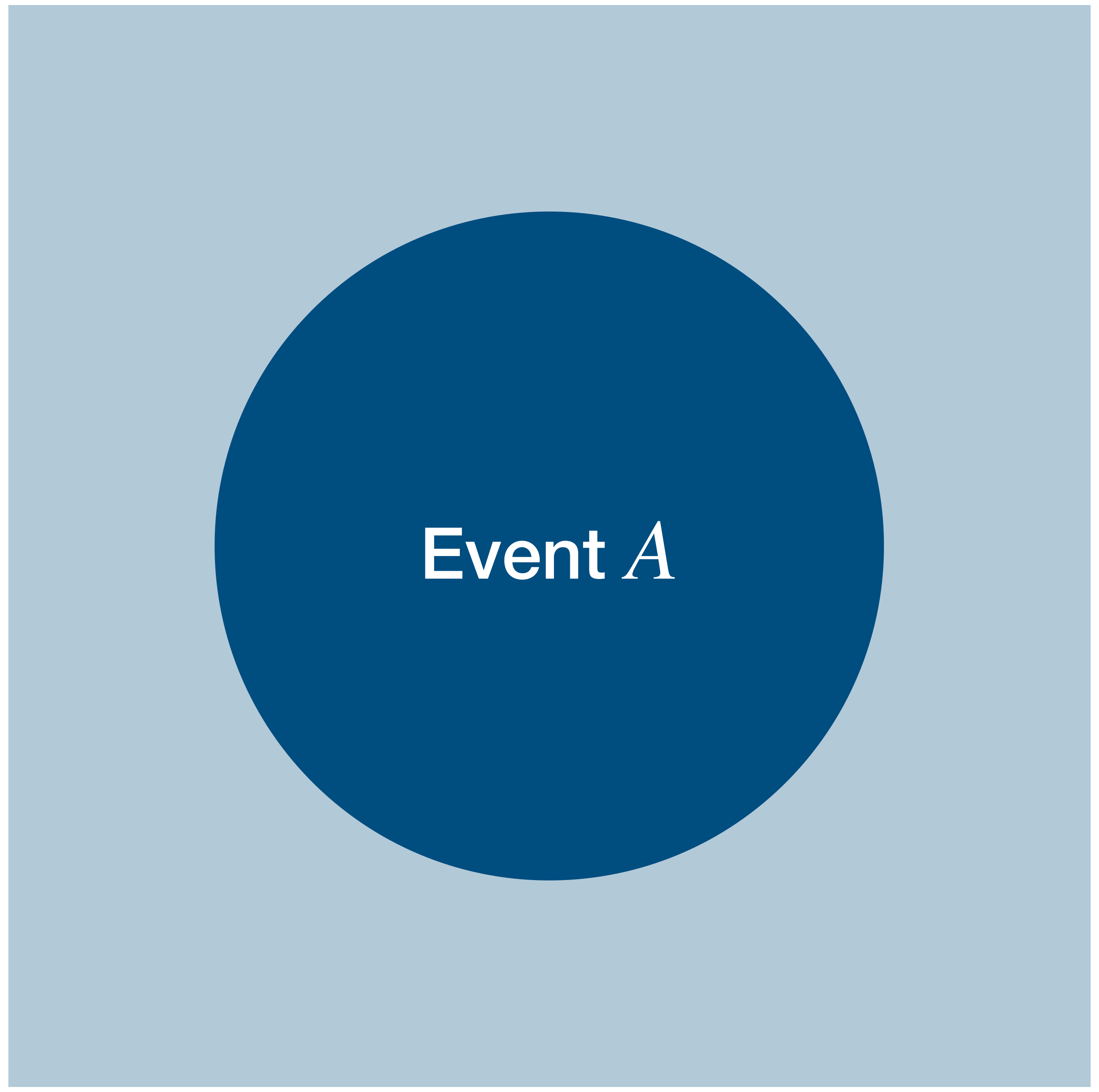
$\text{P}(A) = \text{Probability of event } A \text{ happening}$

- $\text{P}(A) = 0$
    - 0% chance of event A happening
    - Event A is impossible
- $\text{P}(A) = 1$
    - 100% chance of event A happening
    - Event A is certain

Complements

If an event has a chance of happening N, then the chance it *doesn't* happen is 1-N

- e.g., if chance of happening is 70%, chance of not happening is 30%



Equally Likely Outcomes

Assuming all outcomes are equally likely:

$$\mathsf{P}(A) = \frac{\text{number of outcomes that make } A \text{ happen}}{\text{total number of outcomes}}$$

Exercise A

- I have three cards:

  Ace of Hearts, King of Diamonds, and Queen of Spades

- I shuffle them and draw two cards at random without replacement

- What's the chance that I get the Queen followed by the King?

Exercise A

What’s the chance that I get the Queen followed by the King?

- Let A be event “Queen then King”

$$\mathrm{P}(A) = \frac{\text{number of outcomes that make } A \text{ happen}}{\text{total number of outcomes}}$$

Exercise A

What's the chance that I get the Queen followed by the King?

- Let A be event "Queen then King"

$$\mathrm{P}(A) = \frac{\text{number of outcomes that make } A \text{ happen}}{\text{total number of outcomes}}$$

6

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise A

What's the chance that I get the Queen followed by the King?

- Let A be event "Queen then King"

$$\mathrm{P}(A) = \frac{\overbrace{\text{number of outcomes that make } A \text{ happen}}^{1}}{\underbrace{\text{total number of outcomes}}_{6}}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
$\boxed{\text{Queen}}$	$\boxed{\text{King}}$
Queen	Ace
King	Ace
King	Queen

Exercise A

What's the chance that I get the Queen followed by the King?

- Let A be event "Queen then King"

$$P(A) = \frac{\text{number of outcomes that make } A \text{ happen}}{\text{total number of outcomes}}$$

1

6

$$= \frac{1}{6}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Multiplication Rule

The chance that two events A and B both happen:

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

Exercise A (another way)

What's the chance that I get the Queen followed by the King?

- Let A be event "Queen in the first draw"

- Let B be event "King in the second draw"

P(A and B) $=$ P(A) $\times$ P(B happens given A happens)

Ace

King

Queen

Exercise A (another way)

What's the chance that I get the Queen followed by the King?

- Let A be event "Queen in the first draw"

- Let B be event "King in the second draw"

P(A and B) $=$ P(A) $\times$ P(B happens given A happens)

$$\frac{1}{3}$$

Ace

King

Queen

Exercise A (another way)

What's the chance that I get the Queen followed by the King?

- Let A be event "Queen in the first draw"

- Let B be event "King in the second draw"

P(A and B) $=$ P(A) $\times$ P(B happens given A happens)

$\frac{1}{3}$ $\frac{1}{2}$

Ace

King

Multiplication Rule

The chance that two events A and B both happen:

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

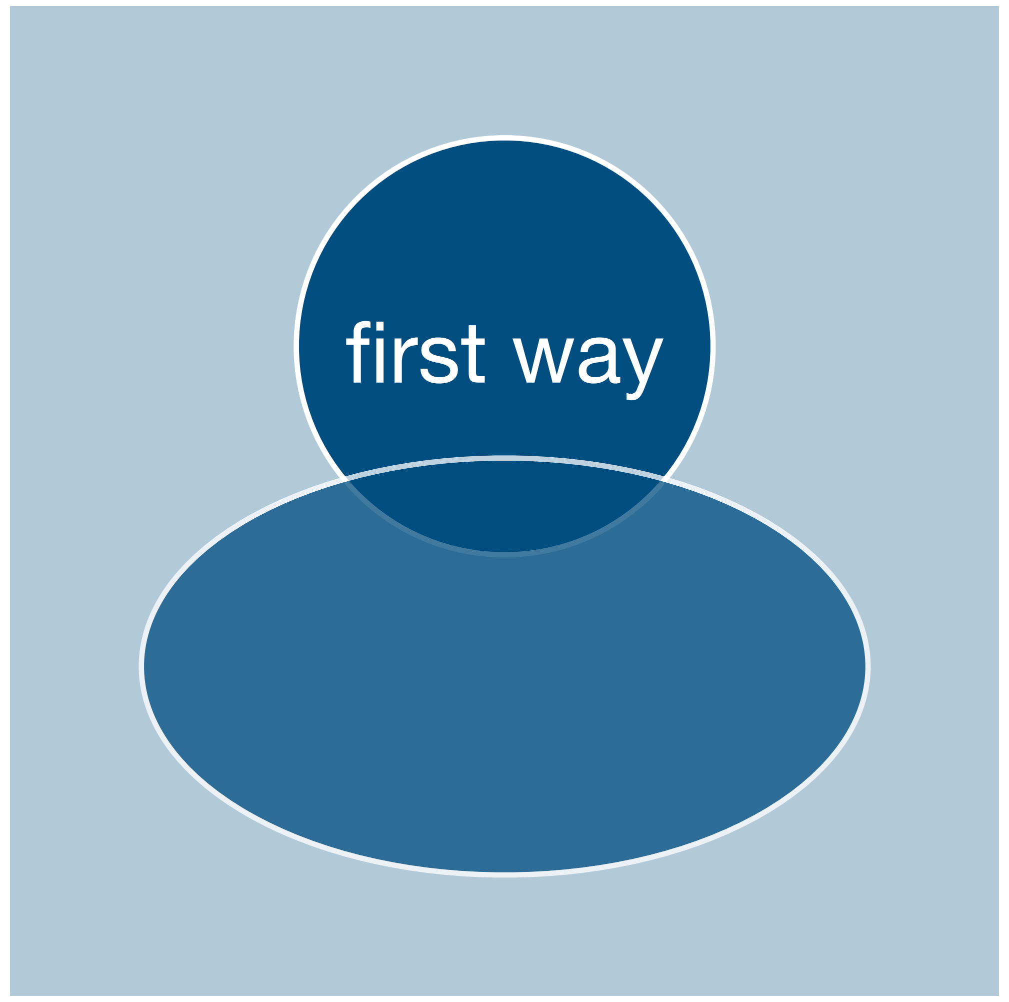
- The answer is less than or equal to each of the two chances being multiplied

- The more conditions you have to satisfy, the less likely you are to satisfy them all

Addition Rule

- If event A can happen in exactly one of two ways, then

$$P(A) \leq P(\text{first way}) + P(\text{second way})$$



https://www.reddit.com/r/mildlyinteresting/comments/1jxn0lw/this_sign_at_the_pool_prohibits_people_with/

Addition Rule

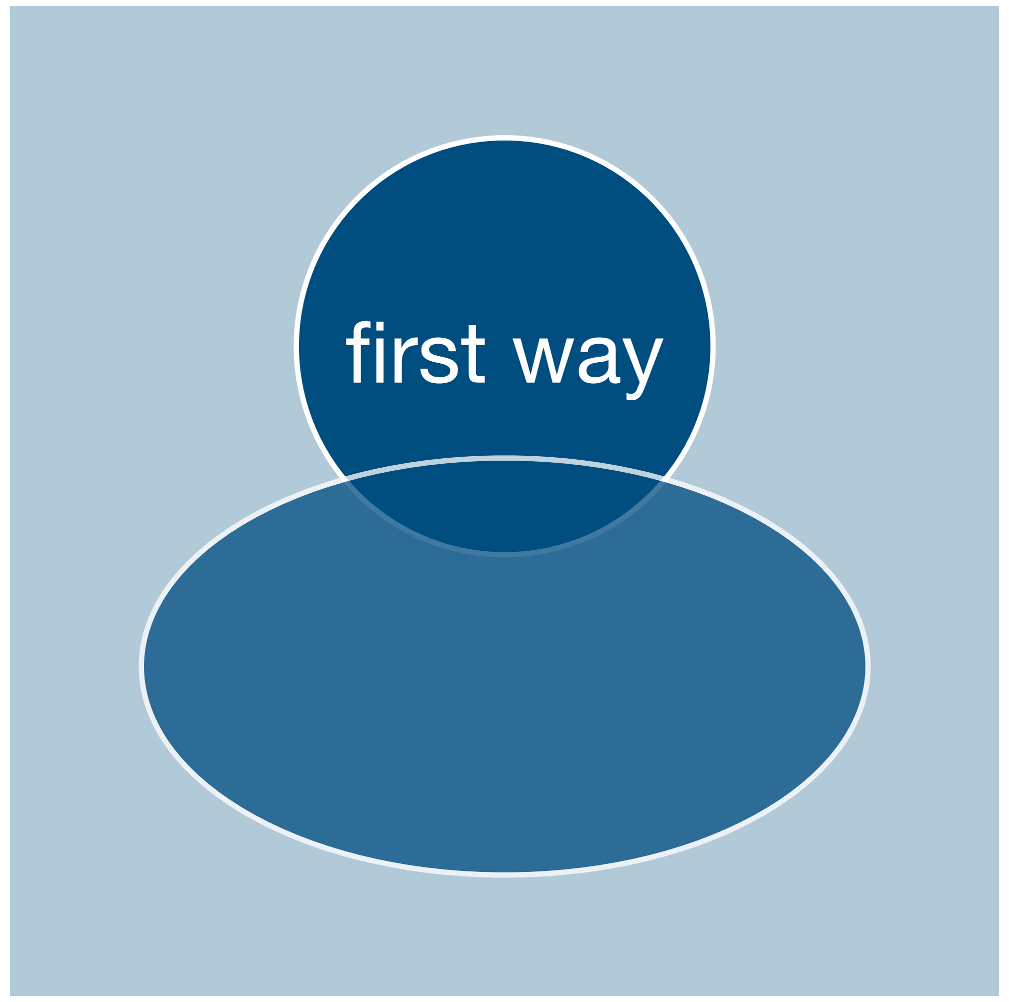
- If event A can happen in exactly one of two ways, then

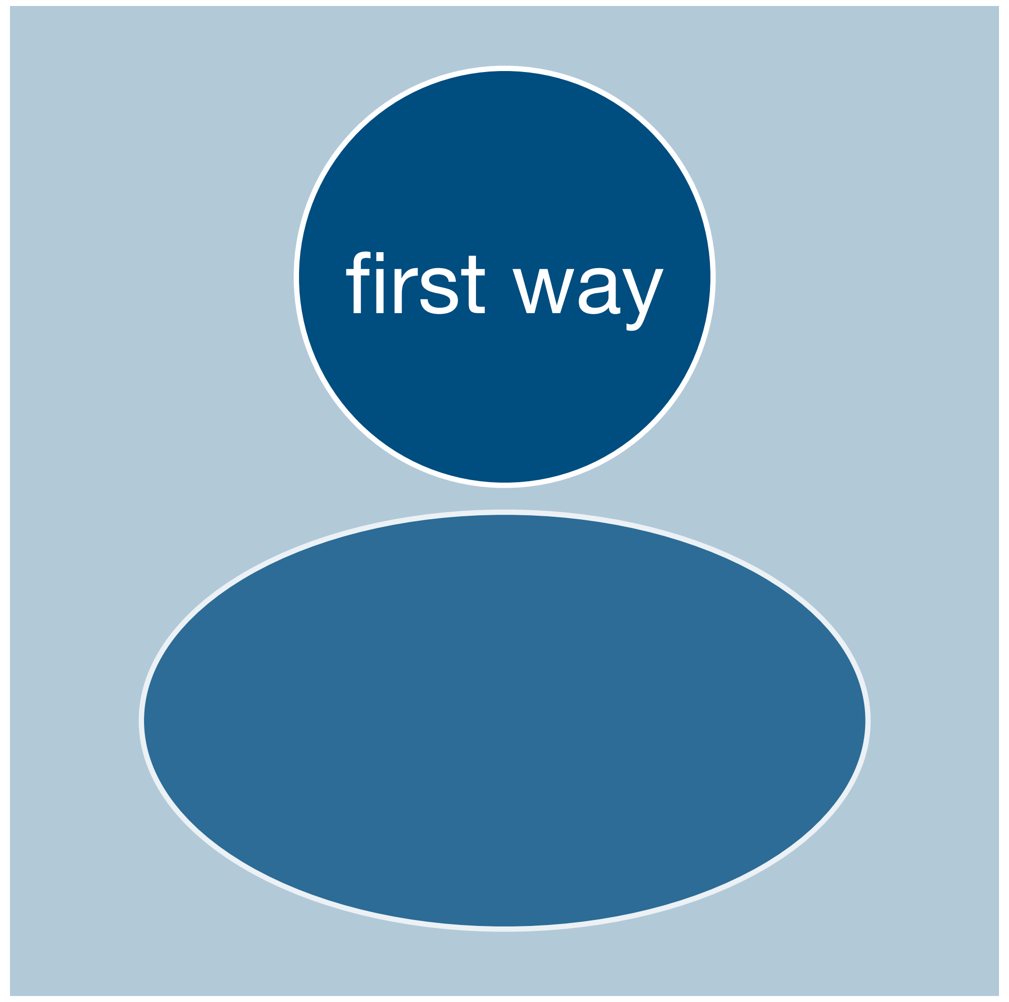
$$\text{P}(A) \leq \text{P}(\text{first way}) + \text{P}(\text{second way})$$

- If the two ways are independent (i.e., no overlap), then

$$\text{P}(A) = \text{P}(\text{first way}) + \text{P}(\text{second way})$$

In this class, we'll mostly deal with independent events





Exercise B

- I have three cards:

  Ace of Hearts, King of Diamonds, and Queen of Spades

- I shuffle them and draw two cards at random without replacement

- What's the chance that one is a Queen *and* one is a King?

Exercise B: Equal Probabilities

What's the chance that one is the Queen and one is the King?

- Let A be drawing a Queen and King

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Equal Probabilities

What's the chance that one is the Queen and one is the King?

- Let A be drawing a Queen and King

$$P(A) = \frac{\text{number of outcomes that make } A \text{ happen}}{\text{total number of outcomes}}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Equal Probabilities

What's the chance that one is the Queen and one is the King?

- Let A be drawing a Queen and King

$$P(A) = \frac{\text{number of outcomes that make } A \text{ happen}}{\text{total number of outcomes}}$$

$$= \frac{2}{6}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Equal Probabilities

What's the chance that one is the Queen and one is the King?

- Let A be drawing a Queen and King

$$\mathrm{P}(A) = \frac{\text{number of outcomes that make } A \text{ happen}}{\text{total number of outcomes}}$$

$$= \frac{2}{6}$$

$$= \frac{1}{3}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Addition Rule

What's the chance that one is the Queen and one is the King?

- Let A be drawing a Queen and King

$$\mathrm{P}(A) = \mathrm{P}(\text{first way}) + \mathrm{P}(\text{second way})$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Addition Rule

What's the chance that one is the Queen and one is the King?

- Let A be drawing a Queen and King

$$P(A) = P(\text{first way}) + P(\text{second way})$$
$$= P(\text{Queen then King}) + P(\text{King then Queen})$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Addition Rule

What's the chance that one is the Queen and one is the King?

- Let A be drawing a Queen and King

$$\begin{aligned} \text{P}(A) &= \text{P}(\text{first way}) + \text{P}(\text{second way}) \\ &= \text{P}(\text{Queen then King}) + \text{P}(\text{King then Queen}) \\ &= \frac{1}{6} + \frac{1}{6} \end{aligned}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Addition Rule

What's the chance that one is the Queen and one is the King?

- Let A be drawing a Queen and King

$$\begin{aligned} \mathrm{P}(A) &= \mathrm{P}(\text{first way}) + \mathrm{P}(\text{second way}) \\ &= \mathrm{P}(\text{Queen then King}) + \mathrm{P}(\text{King then Queen}) \\ &= \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{3} \end{aligned}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
$\boxed{\text{Queen}}$	$\boxed{\text{King}}$
Queen	Ace
King	Ace
$\boxed{\text{King}}$	$\boxed{\text{Queen}}$

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

What are the following probabilities?

1. P(neither Ruby nor Gertrude are in the sample)

2. P(both Ruby and Gertrude are in the sample)

Pictured: Ruby and Gertrude

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

1. P(neither Ruby nor Gertrude are in the sample)

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

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

1. P(neither Ruby nor Gertrude are in the sample)

First pick:

100 cats

98 are not Ruby or Gertrude

$\hookrightarrow P(A) = \frac{98}{100}$

P(A and B) $=$ P(A) $\times$ P(B happens given A happens)

A: Gertrude and Ruby are not chosen in the first pick

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

1. P(neither Ruby nor Gertrude are in the sample)

First pick:

100 cats

98 are not Ruby or Gertrude

$\hookrightarrow P(A) = \frac{98}{100}$

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

A: Gertrude and Ruby are not chosen in the first pick

B: Gertrude and Ruby are not chosen in the second pick

Second pick:

99 cats after A (removed 1)

97 are not Ruby or Gertrude

$\hookrightarrow P(B \text{ given } A) = \frac{97}{99}$

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

1. P(neither Ruby nor Gertrude are in the sample)

First pick:
100 cats
98 are not Ruby or Gertrude
$\hookrightarrow P(A) = \frac{98}{100}$

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

$$= \frac{98}{100} \times \frac{97}{99}$$

$$= 0.96$$

Second pick:
99 cats after A (removed 1)
97 are not Ruby or Gertrude
$\hookrightarrow P(B \text{ given } A) = \frac{97}{99}$

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

2. P(both Ruby and Gertrude are in the sample)

$$\mathrm{P}(A) = \mathrm{P}(\text{first way}) + \mathrm{P}(\text{second way})$$

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

2. P(both Ruby and Gertrude are in the sample)

$$\begin{aligned} \mathrm{P}(A) &= \mathrm{P}(\text{first way}) + \mathrm{P}(\text{second way}) \\ &= \mathrm{P}(\underbrace{\text{Ruby}}_{\frac{1}{100}}\ \underbrace{\text{then Gertrude}}_{\frac{1}{99}}) + \mathrm{P}(\text{Gertrude then Ruby}) \end{aligned}$$

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

2. P(both Ruby and Gertrude are in the sample)

$$\text{P}(A) = \text{P}(\text{first way}) + \text{P}(\text{second way})$$
$$= \text{P}(\underbrace{\text{Ruby}}_{\frac{1}{100}}\ \underbrace{\text{then Gertrude}}_{\frac{1}{99}}) + \text{P}(\underbrace{\text{Gertrude}}_{1/100}\ \underbrace{\text{then Ruby}}_{\frac{1}{99}})$$

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

2. P(both Ruby and Gertrude are in the sample)

$$
\begin{aligned}
\mathsf{P}(A) &= \mathsf{P}(\text{first way}) + \mathsf{P}(\text{second way}) \\
&= \mathsf{P}(\text{Ruby then Gertrude}) + \mathsf{P}(\text{Gertrude then Ruby}) \\
&= \frac{1}{100} \times \frac{1}{99} + \frac{1}{100} \times \frac{1}{99} \\
&= 0.0002
\end{aligned}
$$

Sampling

Sample

- A subset of your population you choose to utilize in your analysis
- Picking samples is a fundamental part of Data Science
    - Did you sample enough / collect enough data?
    - Is the data representative?

Deterministic vs Random Samples

- Deterministic Sample:
  - Sampling scheme doesn’t involve chance, results are always the same
  - Example: `cat_tbl.where('Coloring', 'tuxedo')`
- Random Sample:
  - Elements are chosen probabilistically
    - Selection probabilities for each element are known *before the sample is drawn*
    - Not all individuals or groups have to have equal chance of being selected
  - Example: `np.random.choice(np.arange(10))`

Convenience Sampling

Random sampling requires knowing the probability of selection *ahead of time*

- Not fully controlling what is picked (e.g., “choose the first 10 people to walk by”) doesn’t necessarily make it a random sample

If you can’t figure out ahead of time

- what’s the population

- what’s the chance of selection, for each group in the population

then it is a **sample of convenience** and not a random sample!

Next Time

Date	Topic	Lab	Assignment
10/1	9 - Conditionals and Iteration	Lab 4 - Functions and Visualizations (Due 10/3) Courseworks	HW2 Due
10/6	10 - Probability and Sampling		HW4 - Probability, Simulation, Estimation (Due 10/15) Courseworks
10/8	11 - Models and Empirical Simulations	Lab 5 - Simulations (Due 10/10) Courseworks	HW3 Due
10/13	Programming/Python Review		
10/15	Midterm Review	*No Lab*	HW4 Due
10/20	**Midterm Exam**		
10/22	Special Topics - Bias in AI	*No Lab*	

Wed