Probability
Part 1 - Introduction

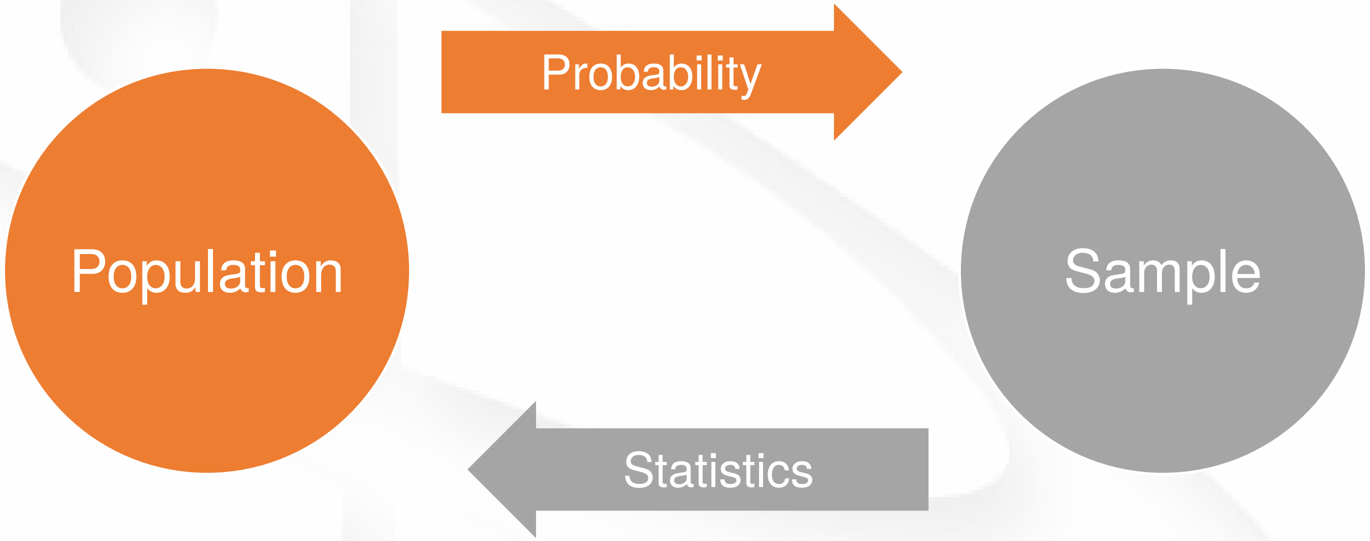
DASC 512

Overview

- Probability and Statistics
- Events
- Sets
- Venn Diagrams
- Rules of Probability

Probability and Statistics

- So far we have used sample data to describe something about the population
  - This is the field of "descriptive statistics"
- Now we'll use assumptions of the population to infer about a sample
  - This is the field of "probability"



Why Probability?

- Questions that we can answer using probability
  - How likely is it that we will destroy a target?
  - What is the chance that the backup system fails if the main system is down?
  - What are the odds of my team winning the big game?
  - Should I attack using my 2D6-damage sword or my 1D12-damage axe?
- Why do we use probability?
  - It allows us to make better decisions
  - Can often exploit known relationships to improve decisions further

Definitions

- Event: A specific set of possible outcomes of an experiment
- Simple Event: Set of a single possible outcome of an experiment
- Compound Event: Set of multiple simple events
- Sample Space: The set of all possible outcomes of an experiment
- p_i: The probability of the ith event in the sample space occurring
- $P(A)$: The probability of any outcome in the event A occurring

Example: Rolling a 6-sided die

	A={6}	A={4,5,6}	A={2,4,6}	A={1,2,3,4,5,6}	A={}
Event					
Simple Event					
Compound Event					
Sample Space					

Example: Rolling a 6-sided die

	A={6}	A={4,5,6}	A={2,4,6}	A={1,2,3,4,5,6}	A={}
Event	X				
Simple Event	X				
Compound Event					
Sample Space					

Example: Rolling a 6-sided die

	A={6}	A={4,5,6}	A={2,4,6}	A={1,2,3,4,5,6}	A={}
Event	X	X			
Simple Event	X				
Compound Event		X			
Sample Space					

Example: Rolling a 6-sided die

	A={6}	A={4,5,6}	A={2,4,6}	A={1,2,3,4,5,6}	A={}
Event	X	X	X		
Simple Event	X				
Compound Event		X	X		
Sample Space					

Example: Rolling a 6-sided die

	A={6}	A={4,5,6}	A={2,4,6}	A={1,2,3,4,5,6}	A={}
Event	X	X	X	X	
Simple Event	X				
Compound Event		X	X	X	
Sample Space				X	

Example: Rolling a 6-sided die

	A={6}	A={4,5,6}	A={2,4,6}	A={1,2,3,4,5,6}	A={}
Event	X	X	X	X	X
Simple Event	X				
Compound Event		X	X	X	
Sample Space				X	

What is Probability

Probability is a map of events from a sample space to the interval [0,1]

- The probability of any event must lie between 0 and 1

$$0 \leq p \leq 1$$

- The total probability of the entire sample space is 1

$$\sum_{i=1}^{|S|} p_i = 1$$

Calculating Probability (Discrete Events)

- Steps for calculating probabilities of events:
  - Define the experiment (type of observation, collection method, etc.)
  - List all simple events in the sample space
  - Assign probabilities to those simple events
  - Determine which simple events are contained in the event of interest
  - Add the simple event probabilities to get the event probability

Dice Example

- Example: Roll two fair 6-sided dice. What is the probability of rolling an 8?

Dice Example

- Example: Roll two fair 6-sided dice. What is the probability of rolling an 8?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Dice Example

- Example: Roll two fair 6-sided dice. What is the probability of rolling an 8?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	**(2,6)**
(3,1)	(3,2)	(3,3)	(3,4)	**(3,5)**	(3,6)
(4,1)	(4,2)	(4,3)	**(4,4)**	(4,5)	(4,6)
(5,1)	(5,2)	**(5,3)**	(5,4)	(5,5)	(5,6)
(6,1)	**(6,2)**	(6,3)	(6,4)	(6,5)	(6,6)

Dice Example

- Example: Roll two fair 6-sided dice. What is the probability of rolling an 8?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	**(2,6)**
(3,1)	(3,2)	(3,3)	(3,4)	**(3,5)**	(3,6)
(4,1)	(4,2)	(4,3)	**(4,4)**	(4,5)	(4,6)
(5,1)	(5,2)	**(5,3)**	(5,4)	(5,5)	(5,6)
(6,1)	**(6,2)**	(6,3)	(6,4)	(6,5)	(6,6)

$$P(8) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{5}{36} \approx 0.14$$

Kids Example

- I’m talking to another dad at the park the other day, and he tells me he has two kids. What is the probability that at least one of his kids is a boy?

Kids Example

- I'm talking to another dad at the park the other day, and he tells me he has two kids. What is the probability that at least one of his kids is a boy?

	Eldest = Girl	Eldest = Boy
Youngest = Girl	¼	¼
Youngest = Boy	¼	¼

Kids Example

- I'm talking to another dad at the park the other day, and he tells me he has two kids. What is the probability that at least one of his kids is a boy?

	Eldest = Girl	Eldest = Boy
Youngest = Girl	¼	¼
Youngest = Boy	¼	¼

$$P(B) = P(BB) + P(BG) + P(GB) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} = 0.75$$

Sets

- Let *A* and *B* be events.
  - <u>Union</u> of *A* and *B* $(A \cup B)$: Occurs if outcome is in *A*, *B*, or both
  - <u>Intersection</u> of *A* and *B* $(A \cap B)$: Occurs if outcome is in both *A* and *B*
  - <u>Complement</u> of *A* (A^c): Occurs if outcome is not in *A*

Boolean Logic

And / Intercept / $A \cap B$

	B=False	B=True
A=False	False	False
A=True	False	True

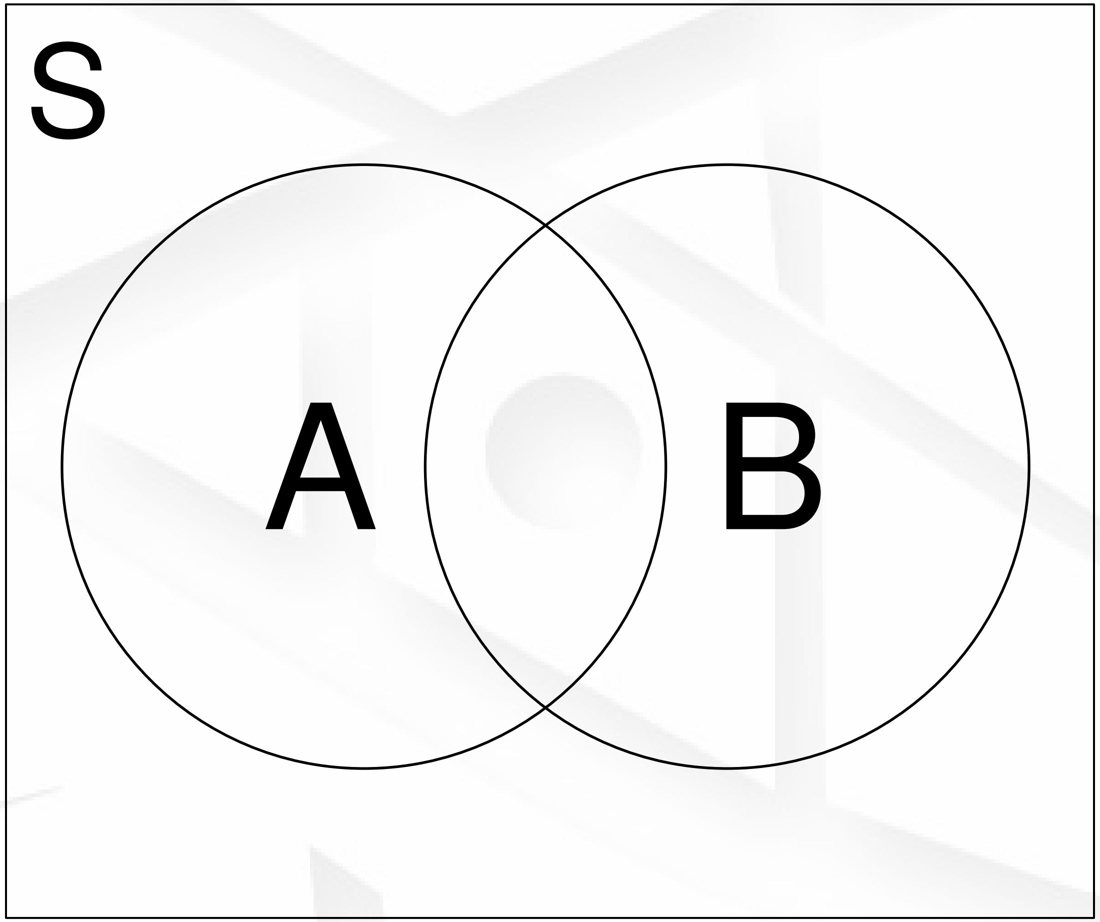
Or / Union / $A \cup B$

	B=False	B=True
A=False	False	True
A=True	True	True

Not / Complement / A^C

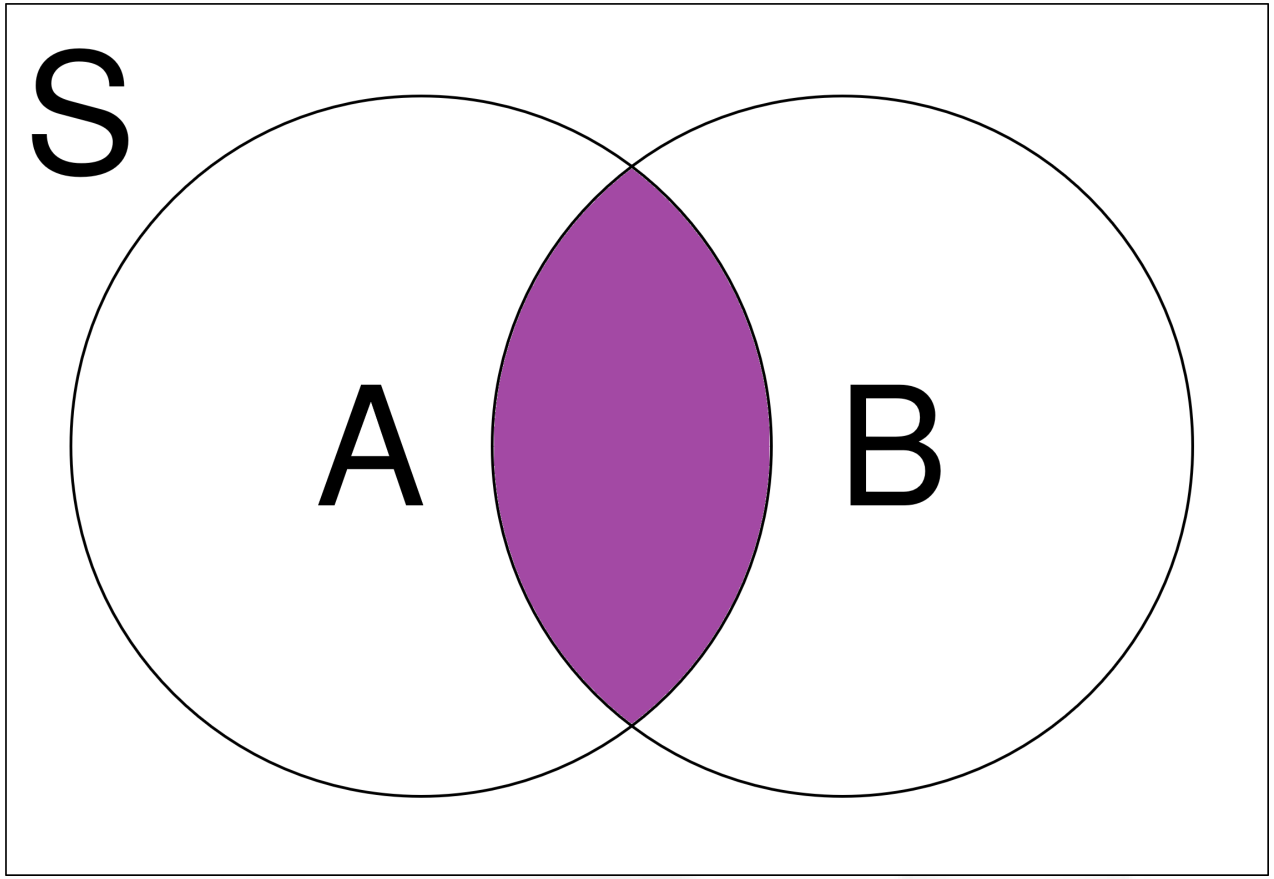
A=False	A^c=True
A=True	A^c=False

Venn Diagrams

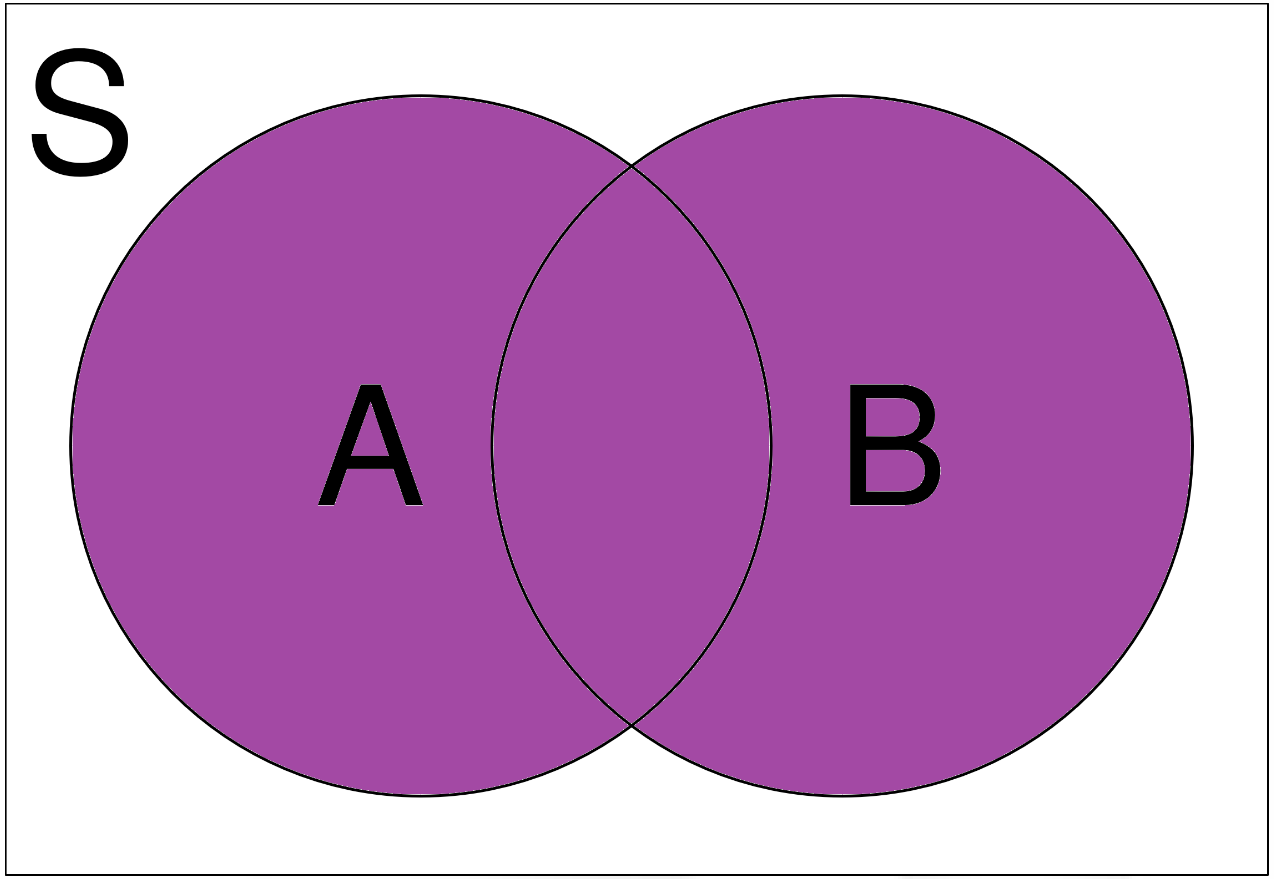


- The area of a section of the Venn diagram represents the probability of that event
- The outermost box represents the sample space S
- $P(S) = 1$

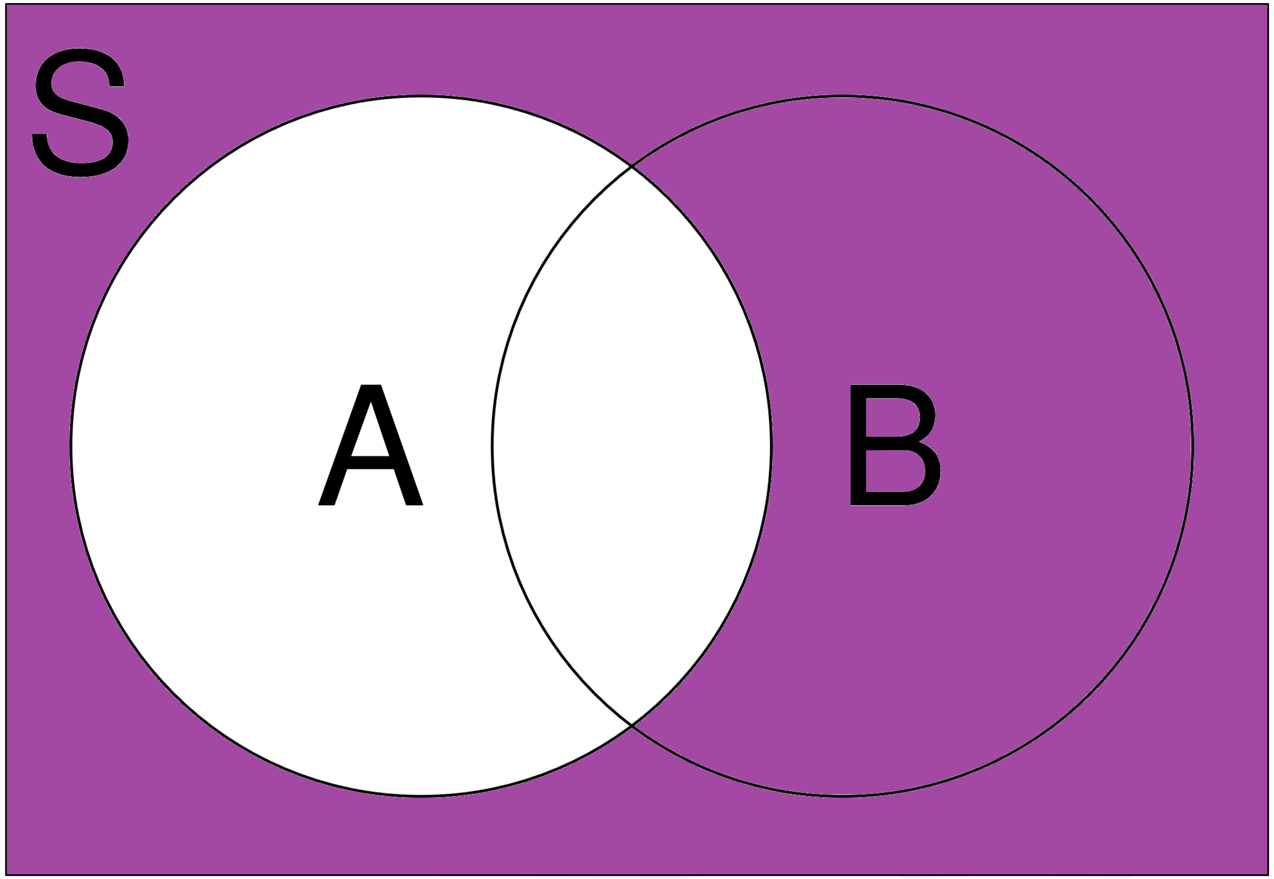
A and B – $A \cap B$ – Intersection



A or B – $A \cup B$ – Union



Not A – A^c – Complement



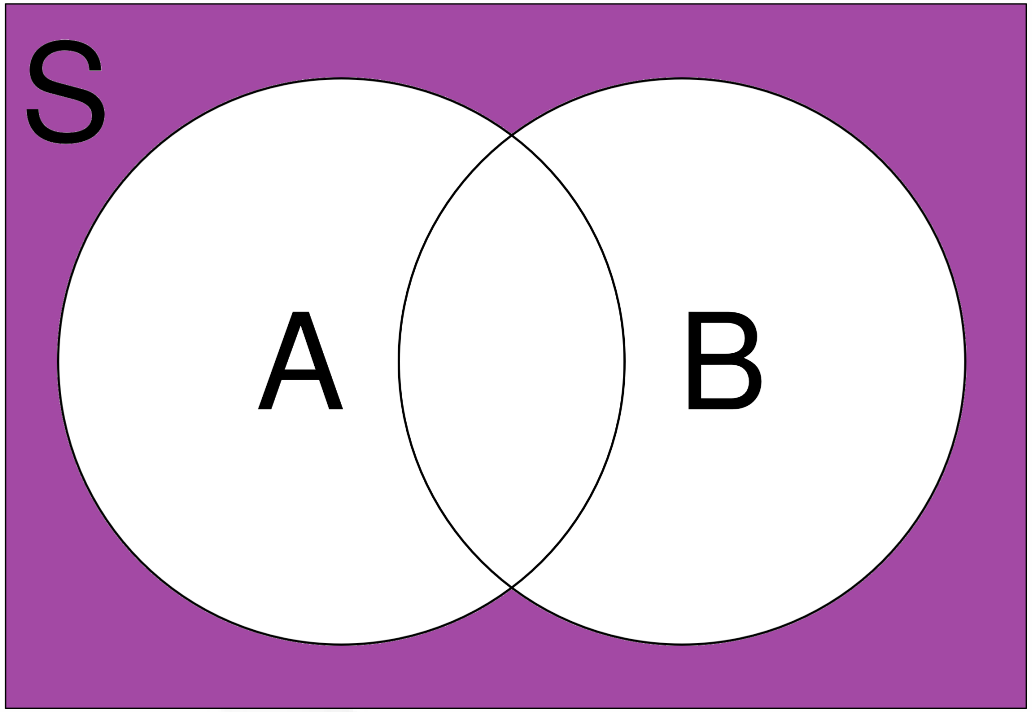
Combining Set Operators

- Try writing the English phrase and drawing Venn diagrams of the following:
  - $(A \cup B)^C$
  - $(A \cap B)^C$
  - $A \cap (B^C)$

- Pause the video and work this on your own

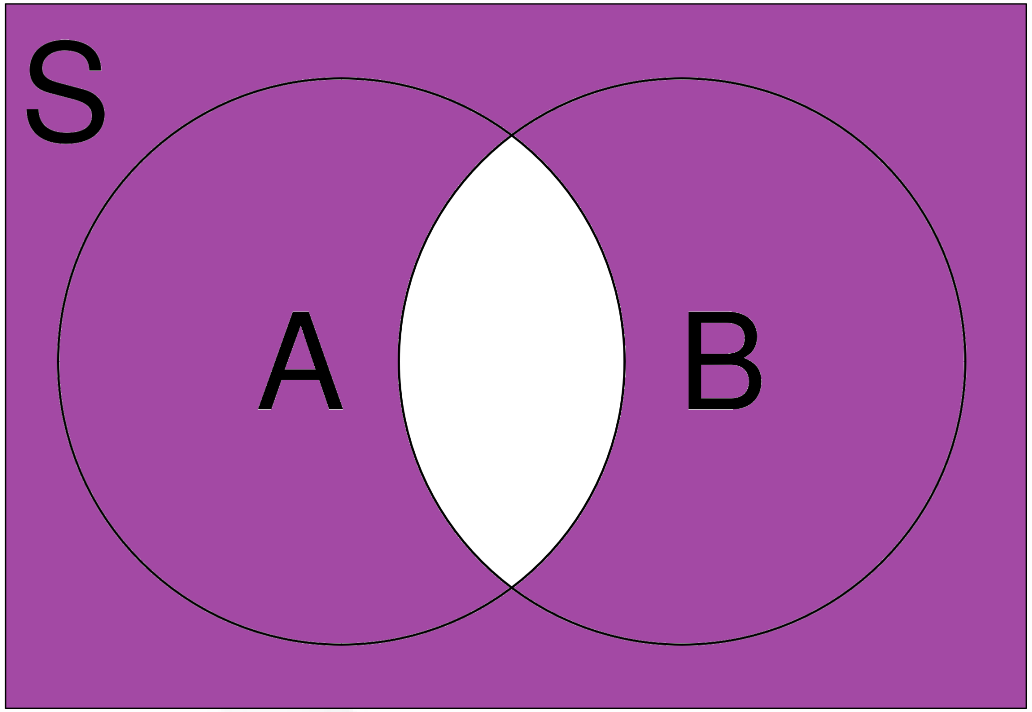
$(A \cup B)^C$

- Neither A nor B
- This is equivalent to $A^C \cap B^C$



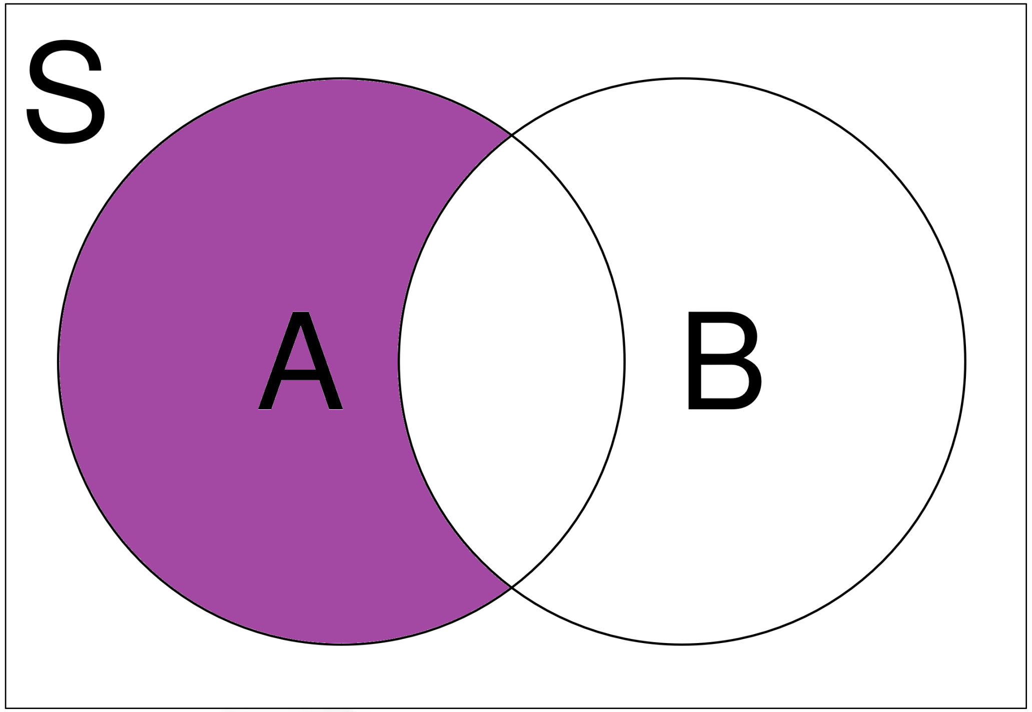
$(A \cap B)^C$

- Not both A and B
- This is equivalent to $A^C \cup B^C$



$$A \cap (B^C)$$

- A but not B



More Definitions

- Marginal Probability: Probability as a function of one event:

$$P(A = A_i) = P(A_i) = p_i$$

- Joint Probability: Joint probability is a function of two events:

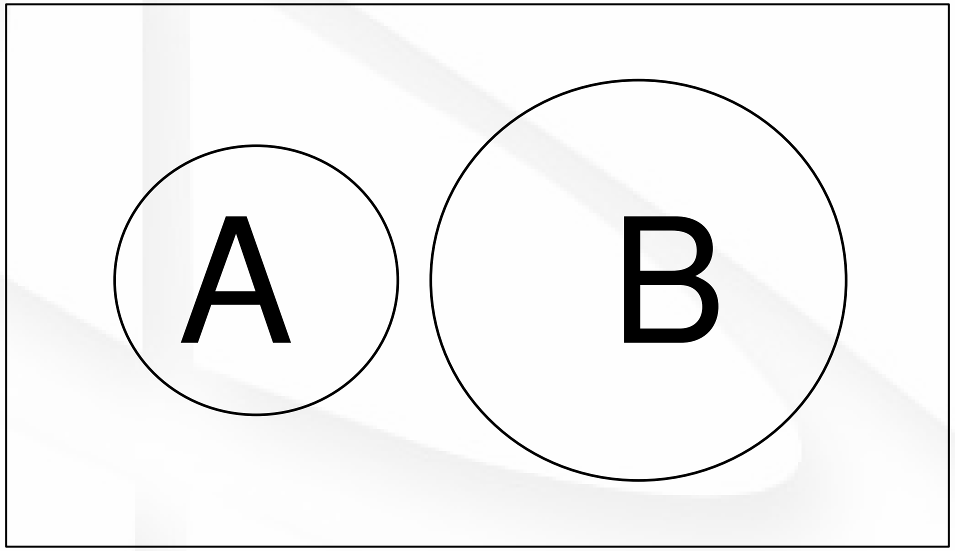
$$P\left((A = A_i) \cap \left(B = B_j\right)\right) = P\left(A_i \cap B_j\right) = p_{i,j}$$

More Definitions

- Disjoint (Mutually Exclusive): Events that have no overlap are disjoint. Mathematically, let A be a set of n disjoint events A_i, $i = 1, \ldots, n$.

$$A_i \cap A_j = \{\ \}, i \neq j$$

- Example: A and B below are disjoint events.

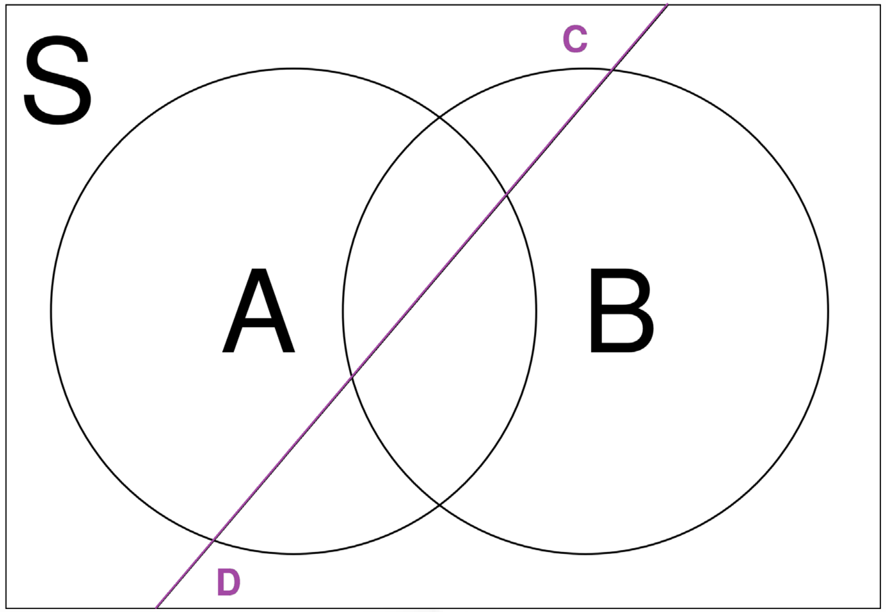


More Definitions

- Collectively Exhaustive: Events that completely cover the sample space are collectively exhaustive.
  Mathematically, let A be a set of n collectively exhaustive events A_i.

$$\bigcup_{i=1}^{n} A_i = S$$

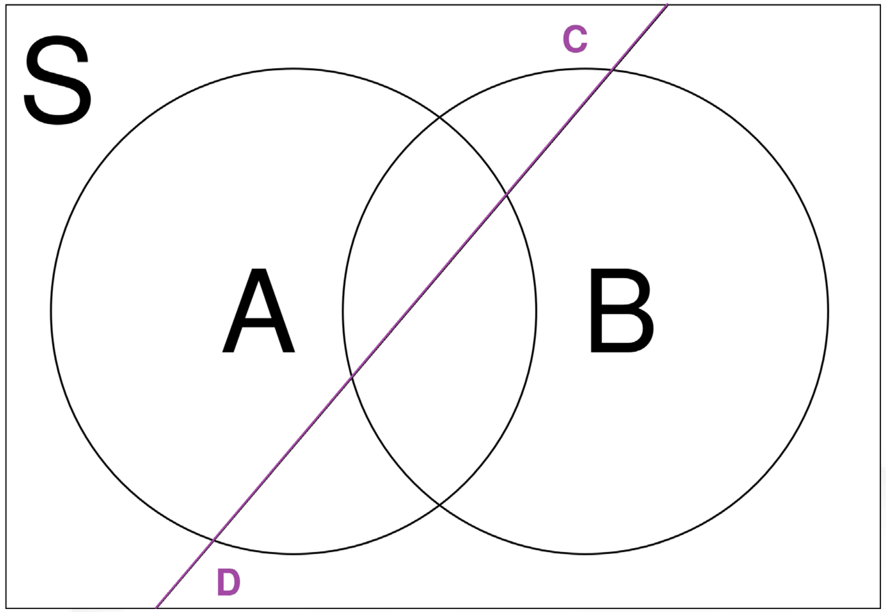
- Example: $E = \{A, B, C, D\}$ is a collectively exhaustive set of events.



More Definitions

- Partition: A set of events that is both disjoint and collectively exhaustive is said to form a partition of the sample space.

- Example: C and D form a partition of the sample space S.

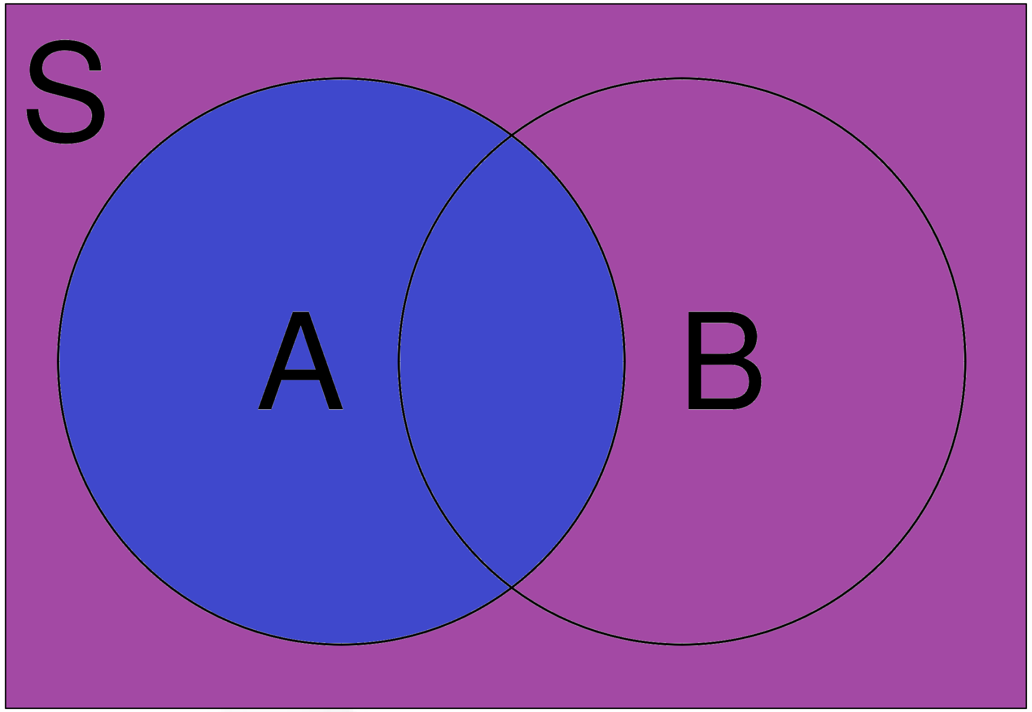


Probability Rules
Complement

Let A and B be events in the sample space S

$$P(A) + P(A^C) = 1$$

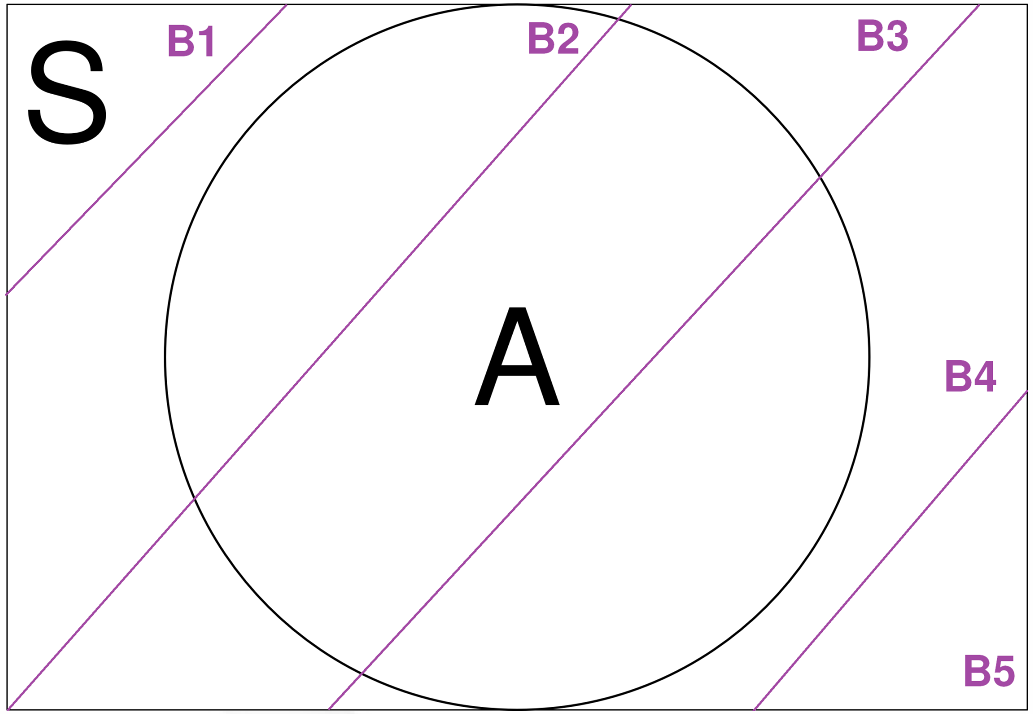
Any event and its complement form a partition.



Probability Rules
Law of Total Probability

Let A be an event in the sample space S and let $B = \{B_1, \ldots, B_n\}$ form a partition of S.

$$P(A) = \sum_{n} P(A \cap B_n)$$

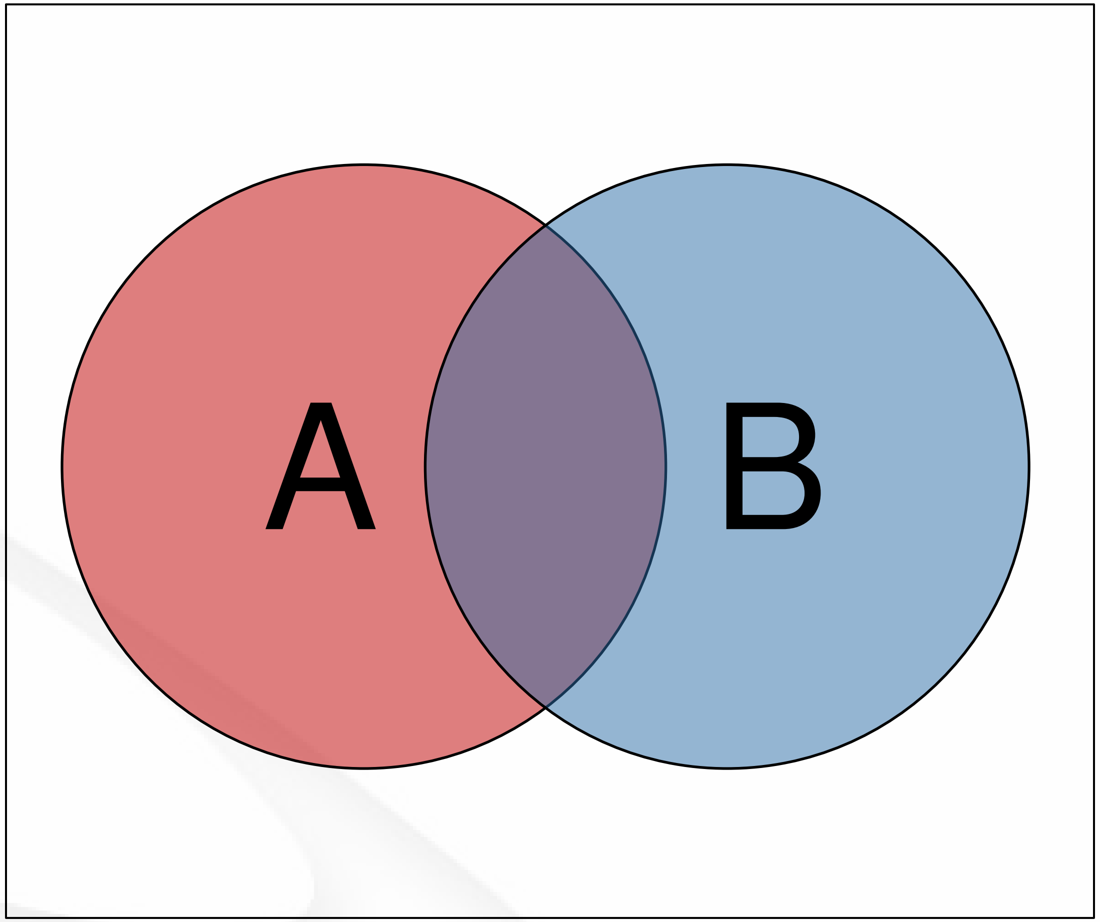


Probability Rules

Additive Rule

Let A and B be events in the sample space S

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

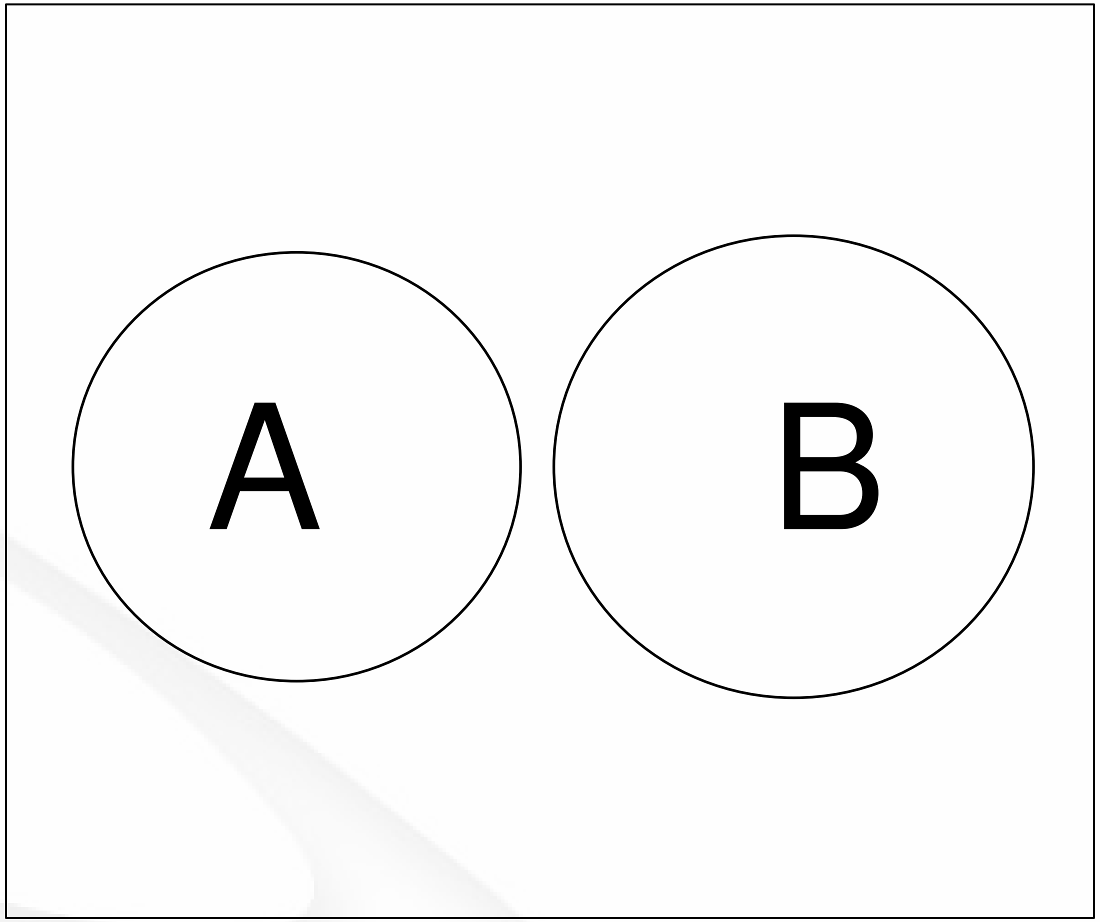


Probability Rules
Disjoint Events

Let A and B be events in sample space S

If, and only if, A and B are disjoint,

$$P(A \cap B) = 0$$

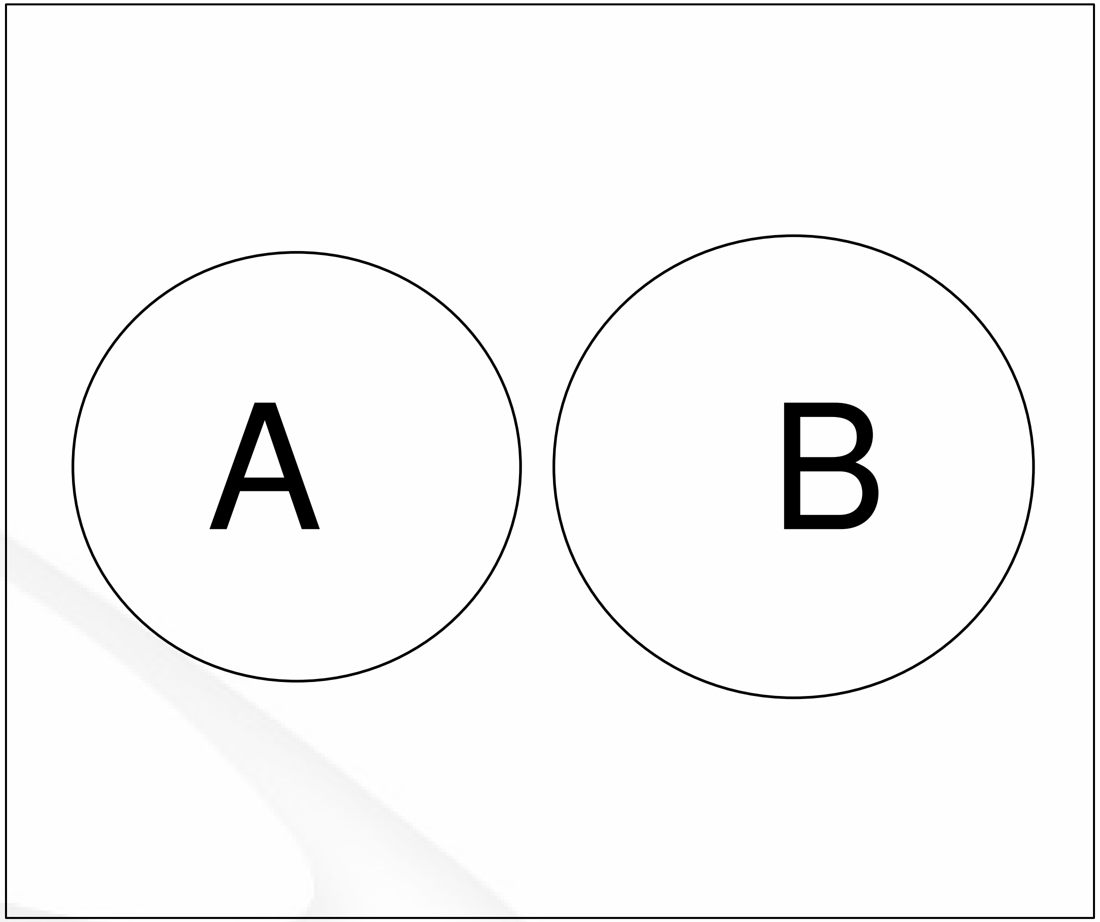


Probability Rules
Additive Rule – Special Case for Disjoint Events

Let A and B be events in sample space S

If, and only if, A and B are disjoint,

$$P(A \cup B) = P(A) + P(B)$$



Example

- A sample of Google Play Store apps in 2019 had the following attributes:

	Everyone	Ages 10+	Teen 13+ (A)	Mature 17+	Total
Free	8019	380	1156	479	**10034**
Paid (B)	695	33	52	20	**800**
Total	**8714**	**413**	**1208**	**499**	**10834**

- Let us define the following events, if we randomly select an app.
  - *A*: The app is rated Teen 13+.
  - *B*: The app is not free.
- Convert this contingency table into a relative frequency table.

https://www.kaggle.com/lava18/google-play-store-apps

Example

- A sample of Google Play Store apps in 2019 had the following attributes:

	Everyone	Ages 10+	Teen 13+ (A)	Mature 17+	Total
Free	0.7402	0.0351	0.1067	0.0442	**0.9262**
Paid (B)	0.0641	0.0030	0.0048	0.0018	**0.0738**
Total	**0.8043**	**0.0381**	**0.1115**	**0.0461**	**1**

- Calculate the following:
  - P(A)
  - P(B)

https://www.kaggle.com/lava18/google-play-store-apps

Example

- A sample of Google Play Store apps in 2019 had the following attributes:

	Everyone	Ages 10+	Teen 13+ (A)	Mature 17+	Total
Free	0.7402	0.0351	0.1067	0.0442	0.9262
Paid (B)	0.0641	0.0030	0.0048	0.0018	0.0738
Total	0.8043	0.0381	0.1115	0.0461	1

- Calculate the following:
  - $P(A) = 0.1115$
  - $P(B) = 0.0738$

https://www.kaggle.com/lava18/google-play-store-apps

Example

- A sample of Google Play Store apps in 2019 had the following attributes:

	Everyone	Ages 10+	Teen 13+ (A)	Mature 17+	Total
Free	0.7402	0.0351	0.1067	0.0442	**0.9262**
Paid (B)	0.0641	0.0030	0.0048	0.0018	**0.0738**
Total	**0.8043**	**0.0381**	**0.1115**	**0.0461**	**1**

- What is the probability that the app is both free and rated Teen 13+?

https://www.kaggle.com/lava18/google-play-store-apps

Example

- A sample of Google Play Store apps in 2019 had the following attributes:

	Everyone	Ages 10+	Teen 13+ (A)	Mature 17+	Total
Free	0.7402	0.0351	$\boxed{0.1067}$	0.0442	0.9262
Paid (B)	0.0641	0.0030	0.0048	0.0018	0.0738
Total	0.8043	0.0381	0.1115	0.0461	1

- What is the probability that the app is both free and rated Teen 13+?

- $P(A \cap B^c) = 0.1067$

https://www.kaggle.com/lava18/google-play-store-apps

Example

- A sample of Google Play Store apps in 2019 had the following attributes:

	Everyone	Ages 10+	Teen 13+ (A)	Mature 17+	Total
Free	0.7402	0.0351	0.1067	0.0442	**0.9262**
Paid (B)	0.0641	0.0030	0.0048	0.0018	**0.0738**
Total	**0.8043**	**0.0381**	**0.1115**	**0.0461**	**1**

- What is the probability that the app is either paid or rated Teen 13+?

https://www.kaggle.com/lava18/google-play-store-apps

Example

- A sample of Google Play Store apps in 2019 had the following attributes:

	Everyone	Ages 10+	Teen 13+ (A)	Mature 17+	Total
Free	0.7402	0.0351	0.1067	0.0442	**0.9262**
Paid (B)	0.0641	0.0030	$\boxed{0.0048}$	0.0018	$\boxed{\mathbf{0.0738}}$
Total	**0.8043**	**0.0381**	$\boxed{\mathbf{0.1115}}$	**0.0461**	**1**

- What is the probability that the app is either paid or rated Teen 13+?

- $$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.1115 + 0.0738 - 0.0048$$
$$= 0.1805$$

https://www.kaggle.com/lava18/google-play-store-apps

Recap

- Probability and Statistics
- Events
- Sets
- Venn Diagrams
- Rules of Probability