

Chapter 2. Probability

Contents

§2.2 Sample Spaces and the Algebra of Sets	2
§2.3 The Probability Function	11

§2.2 Sample Spaces and the Algebra of Sets

► Some terminologies:

- **Experiment:** A repeatable procedure with a set of possible results.
- **Sample Outcome:** (Sample Point) Only one of the possible results of an experiment.
- **Sample Space:** ^{the set of} All the possible outcomes of an experiment. (Usually denoted by S)
- **Event:** Null or one or more outcomes of an experiment.

► **Classical (naive) definition of probability:** $\{ \text{Event} \subseteq S \}$
 Suppose the outcomes of an experiment are all equally likely, and the total number of all possible outcomes is finite.
 (notated w/ $A, B, C, \text{etc.}$)

$$\text{Probability of an event} = \frac{\text{Number of ways it can happen}}{\text{Total number of all possible outcomes}} = \frac{n(A)}{n(S)}$$

The probability of an event is a real number in the interval $[0, 1]$.

Example 1. Experiment: Flipping a fair Coin once.



$$S = \{H, T\}$$

$$A = \text{landing heads} = \{H\}$$

→ one way to achieve $\{H\}$

Event: Landing head.

Number of ways it can happen = 1.

$$P(A) = \frac{n(A)}{n(S)} = \frac{1}{2}$$

Total number of possible outcomes = 2

Probability of 'landing head' = $\frac{1}{2}$.

Example 2. Experiment: Rolling a fair 6-sided die once.



$$S = \{1, 2, 3, 4, 5, 6\}$$

$$A = \{5, 6\}$$

Event: Rolling a number larger than 4 with a die.

Number of ways it can happen = 2.

Total number of possible outcomes = 6

Probability of 'Rolling a number larger than 4' = $\frac{2}{6} = \frac{1}{3}$.

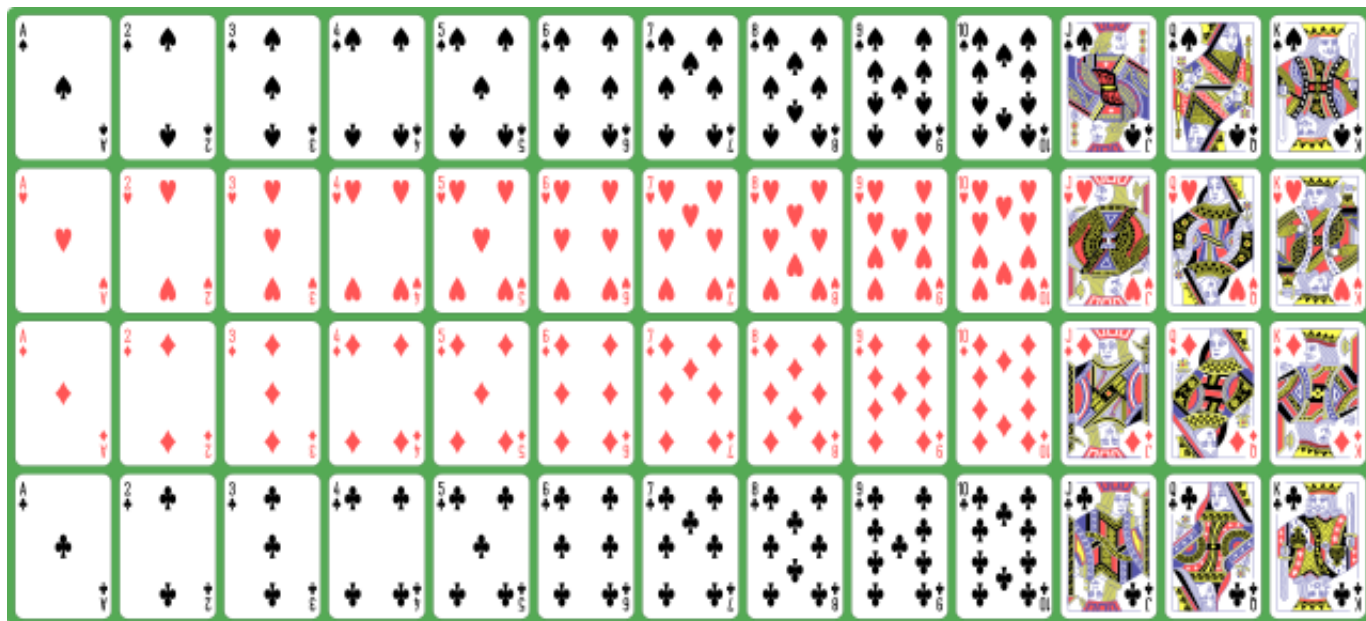
$$P(n > 4) = \frac{|A|}{|S|} = \frac{2}{6} = \frac{1}{3}$$

$$0 \leq P(A) \leq P(S)$$

$$\therefore |0 \leq P(A) \leq 1|$$

Example 3. Randomly draw a card from a standard deck of cards.

A standard deck of playing cards with four suites: Club, Diamond, Spade, and Heart. For each suit, there are 13 values: A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K. There are $4 \times 13 = 52$ cards. (No joker cards.)



If you draw a card randomly, the probability of getting a face (J, or Q, or K) is

$$A = \{J, Q, K \text{ for each suit}\} \quad |A| = 12 \quad |S| = 52 \quad P(\text{Face}) = \frac{12}{52} = \frac{6}{26} = \frac{3}{13}$$

If you draw a card randomly, the probability of getting an Ace is

$$P(\text{Ace}) = \frac{4}{52} = \frac{1}{13}$$

We need to use the **basic set theory** to study probability.

Definition.

- A **set** S is simply a collection of (possibly infinitely many) things.
- If a is an **element** of a set S , we write $a \in S$. If a is **NOT** an element of S , then we write $a \notin S$.
- A **subset** A of S is a set whose elements are in S , denoted as $A \subset S$.

Every set S has at least 2 subsets, itself and the empty set \emptyset .

Each set has exactly 2^n subsets

Extra example

$$n(A) = n$$


↳ A has n elements

Thm A has 2^n subsets

Proof

$$A = \{i_1, i_2, i_3, \dots, i_n\}$$

↑ ↑ ↑
0 or 1 0 or 1 0 or 1 ...

↑
2 possibilities · 2 · 2 ... n times = 2^n 

Countable Sets

1) Finite countable sets Ex. $\{1, 2, 3\}, \{5, 6, 7\}$

2) Infinite countable sets Ex. $\{1, 2, 3, 4, \dots\}$

↑
1-to-1 mapping to \mathbb{N}

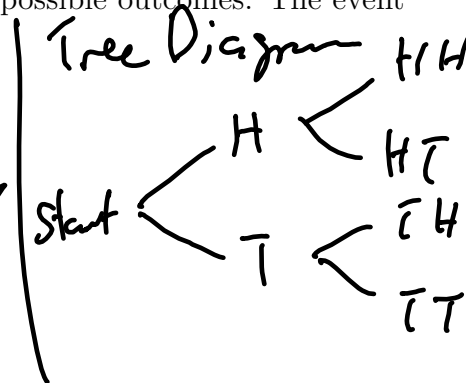
For example, the sample space S of an experiment is the set of all possible outcomes. The event is a subset of S .

Example 4. Experiment: Flipping (tossing) a coin twice.

The sample space $S = \{HH, HT, TH, TT\}$

The event of 'landing head only once' is $A = \{HT, TH\}$

The probability $P(A) = \frac{2}{4} = \frac{1}{2}$



Example 5. Experiment: Flipping (tossing) a coin n times. The size of the sample space is 2^n .

Example 6. Experiment: Rolling a 6-sided die.

The sample space $S = \{1, 2, 3, 4, 5, 6\}$.

The event of 'Rolling a number larger than 4' is $A = \{5, 6\}$ which is a **subset** of S , denoted as $A \subset S$.

The probability of A is $P(A) = 2/6 \approx 0.333$

Example 7. Rolling two 6-sided dice (one red, one blue) once. We can win if we obtain total number larger than 8. What is the probability we can win?

$N = \{1, 2, 3, 4, 5, 6\}$
 $\uparrow 2 \cdot 2 \cdot 2 \dots n \text{ flips}$
 $\rightarrow N \times N = S, |S| = |N|^2$

Cartesian Product, $N \times N = S, |S| = |N|^2$

$$S = \left\{ \begin{array}{cccccc} (1, 1) & (1, 2) & (1, 3) & (1, 4) & (1, 5) & (1, 6) \\ (2, 1) & (2, 2) & (2, 3) & (2, 4) & (2, 5) & (2, 6) \\ (3, 1) & (3, 2) & (3, 3) & (3, 4) & (3, 5) & (3, 6) \\ (4, 1) & (4, 2) & (4, 3) & (4, 4) & (4, 5) & (4, 6) \\ (5, 1) & (5, 2) & (5, 3) & (5, 4) & (5, 5) & (5, 6) \\ (6, 1) & (6, 2) & (6, 3) & (6, 4) & (6, 5) & (6, 6) \end{array} \right\}; A = \left\{ \begin{array}{cccc} & & & (3, 6) \\ & & (4, 5) & (4, 6) \\ & (5, 4) & (5, 5) & (5, 6) \\ (6, 3) & (6, 4) & (6, 5) & (6, 6) \end{array} \right\}$$

So, the probability of event A is $\frac{10}{36} = \frac{5}{18}$

What is the event B that the sum of the two faces showing equal 9? What is the event C that absolute of the difference of the two faces showing equal 3?

$$B = \frac{4}{36} = \frac{2}{18} = \frac{1}{9}$$

$$C = \frac{6}{36} = \frac{1}{6}$$

More generally, if we roll a dice n times, the size of the sample space is 6^n .

All sample spaces in the above two examples are **finite** sets, i.e., there are a finite number of elements in each set. For a finite set S , the number of element in S is called the **cardinality** of S .

In general, a set can contain infinitely many elements.

Example 8. (Countable set)

Experiment: Tossing a coin until we get a head.

Sample Space: $S = \{H, TH, TTH, TTT, \dots\}$.

Event: Getting a head with no more than 3 tosses, $A = \{H, TH, TTH\}$.

infinite countable set

A **countable** infinite set has a one-to-one correspondence to the set of natural numbers \mathbb{N} .

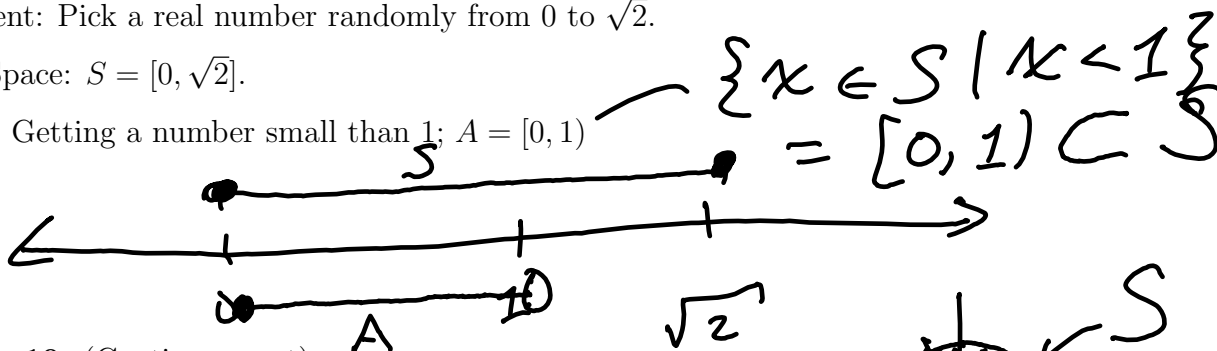
Discrete set means finite or countable set.

Example 9. (Continuous set) *set of values that are measured*

Experiment: Pick a real number randomly from 0 to $\sqrt{2}$.

Sample Space: $S = [0, \sqrt{2}]$.

Event A : Getting a number small than 1; $A = [0, 1)$



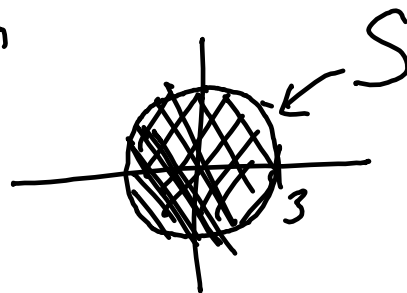
Example 10. (Continuous set)

Experiment: Drop a point in a disc of radius 3.

Sample Space: $S = \{(x, y) \mid x^2 + y^2 \leq 3^2\}$

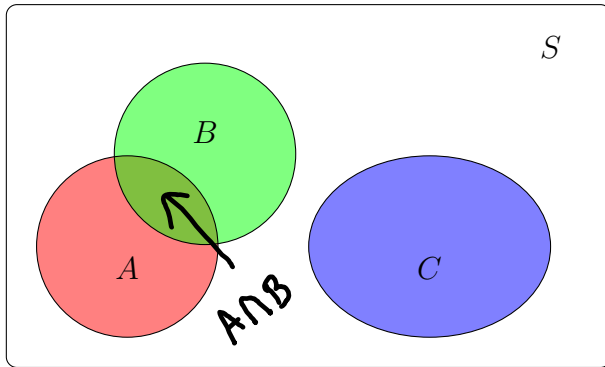
Event A : Get a point (x, y) such that $x \geq 0$ and $y \geq 0$.

$$A = \{(x, y) \in S \mid x \geq 0, y \geq 0\}$$

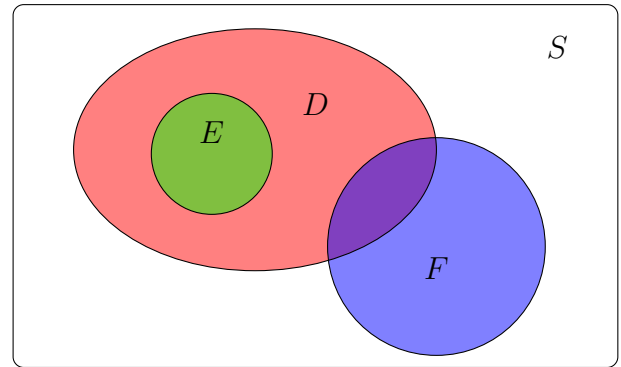


► The **Venn diagram** is a useful visual aid of sets.

The set (or sample space) S is represented as a rectangle and subsets (or events) A, B, C are circles or ellipses.



or



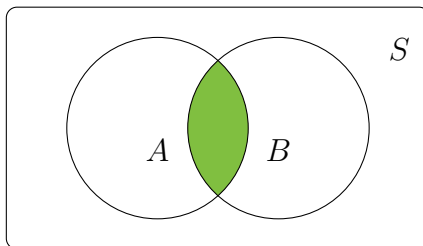
► Basic operations on sets (or events).

1. Intersection

Definition.

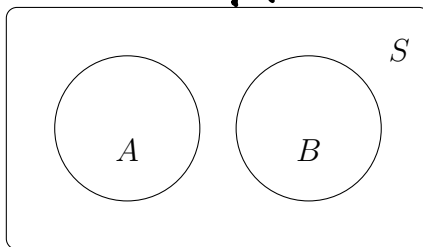
The **intersection** of events A and B , (denoted as $A \cap B$), is the event that whose outcomes belong to both A and B , that is, $A \cap B$ is the event that “both A and B occur”.

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$



m.l.

Events A and B are called mutually exclusive (disjoint) if A and B have no common outcome, i.e., $A \cap B = \emptyset$. **A**



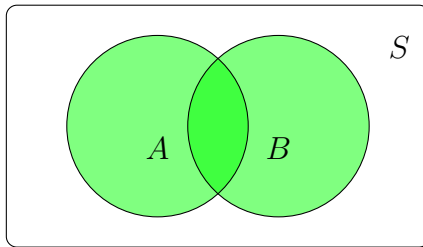
2. Union

Definition.

The **union** of A and B , (denoted as $A \cup B$), is the event whose outcomes belong to either A or B (or both).

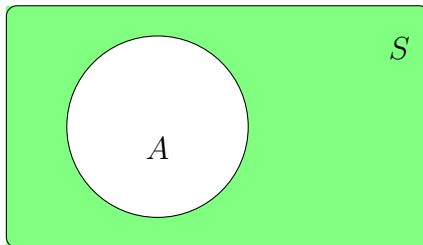
$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$$

So, $A \cup B$ means that “ A or B occurs”.

**3. Complement**

The **complement** of an event A , denoted as A^c (or A^C), is the event whose outcomes in S not belong to A .

$$A^c = \{x \in S \mid x \notin A\}$$



Some quick formulas

$$A \cup A^c = S, \quad A \cap A^c = \emptyset, \quad (A^c)^c = A,$$

$$A \cup B = B \cup A, \quad A \cap B = B \cap A.$$

Example 11. Rolling a 6-sided die once.

The sample space $S = \{1, 2, 3, 4, 5, 6\}$

The event of ‘Rolling a number larger than 4’ is $A = \{5, 6\}$

The event of ‘Rolling even number’ is $B = \{2, 4, 6\}$

The event of ‘Rolling number which is even and larger than 4’ is the intersection $A \cap B = \{6\}$

The event of ‘Rolling number which is even or larger than 4’ is the union $A \cup B = \{2, 4, 5, 6\}$

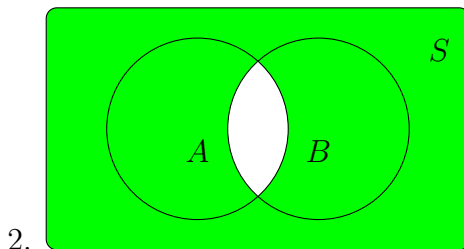
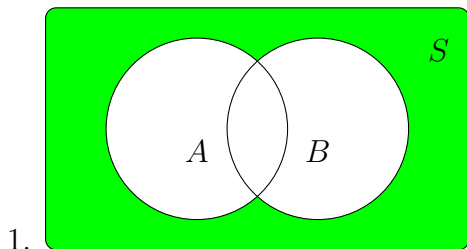
The event of ‘Rolling number which is not even’ is the complement $A^c =$

$$\{1, 3, 5\}$$

Theorem. DeMorgan's Law

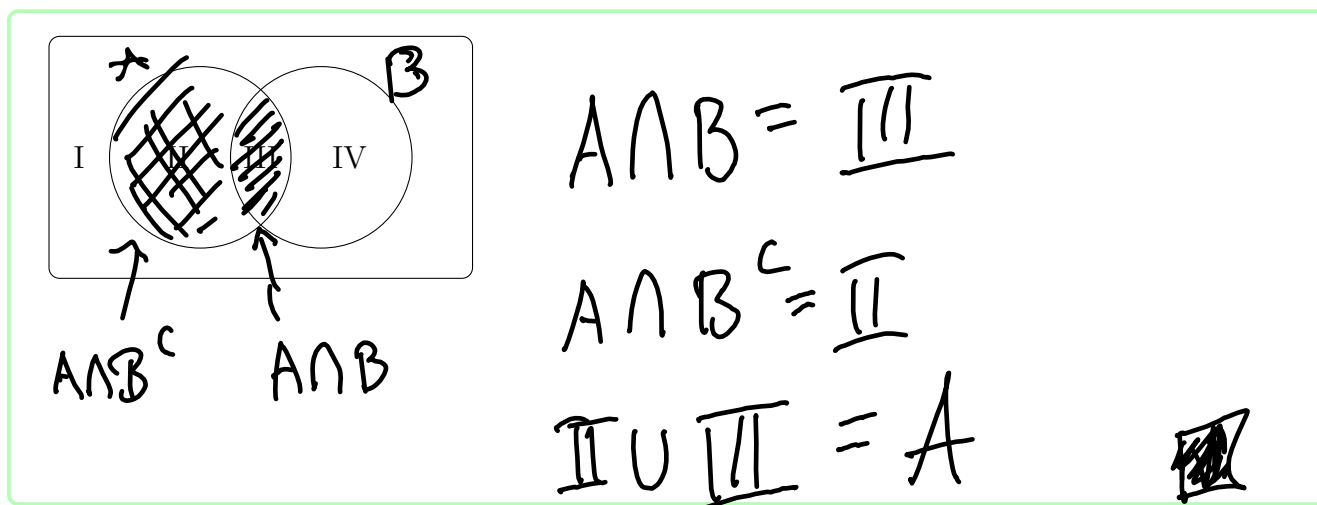
1. $(A \cup B)^c = A^c \cap B^c$ $(\neg(A \vee B) \equiv \neg A \wedge \neg B)$
2. $(A \cap B)^c = A^c \cup B^c$ $(\neg(A \wedge B) \equiv \neg A \vee \neg B)$

DeMorgan's Law in Venn diagram.



Example 12. Prove that $A = (A \cap B) \cup (A \cap B^c)$.

For complicated questions, it is better to label the diagram by disjoint parts I, II, III, IV.



Here, $A \cap B^c$ means that “ A occurs but B does not occur”.

Example 13. Sketch the regions in xy -plane \mathbb{R}^2 corresponding $A \cup B$ and $A \cap B$.

$$A = \{(x, y) \mid x^2 + y^2 \leq 4\}$$

$$B = \{(x, y) \mid 0 \leq x < 4, -2 \leq y \leq 2\}$$

Do not be confused with the Venn diagram.

$$\text{Prove } A = (A \cap B) \cup (A \cap \bar{B})$$

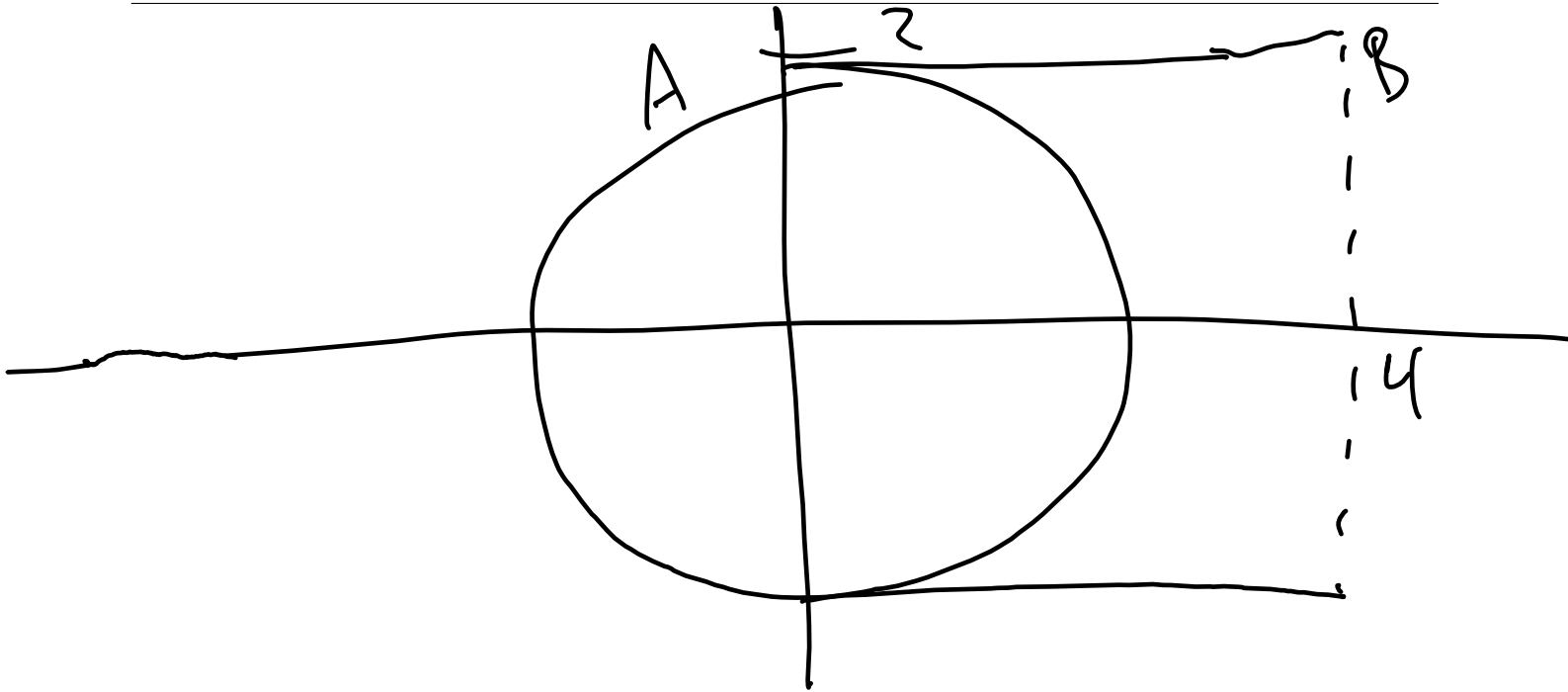
$$= [(A \cap B) \cup A] \cap [(A \cap B) \cup \bar{B}]$$

$$= [\cancel{A \cup A}^A \cap (B \cup A)] \cap [A \cup \bar{B}] \cap \cancel{(B \cup \bar{B})}$$

$$= [B \cup A] \cap [A \cup \bar{B}]$$

$$= [A \cup B] \cap [A \cup \bar{B}]$$

↑ trivially A



Example 14. A dice is tossed 4 times. What outcomes make up the event A that the sum of the four face results showing equal 5? How many outcomes in the sample space?

$$A = \{(1, 1, 1, 2), (1, 1, 2, 1), (1, 2, 1, 1), (2, 1, 1, 1)\}$$

The size of the sample space is 6^4

Example 15. Three events A , B , and C . Find the following events using union, intersection and complement.

- (1) Only B occurs. $B \cap A^c \cap C^c$
- (2) exactly one event occurs. $(A \cap B^c \cap C^c) \cup (B \cap A^c \cap C^c) \cup (C \cap A^c \cap B^c)$
- (3) Only A and B occur. $A \cap B \cap C^c$

(1) Only B occurs = B occurs **and** A does not occur **and** C does not occur.

The answer is $B \cap A^c \cap C^c$ or $B \cap (A \cup C)^c$

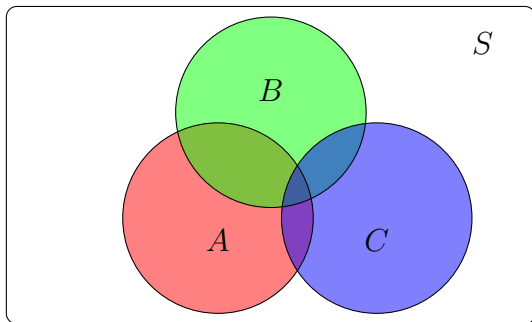
(2) Exactly one event occurs = only A occurs **or** only B occurs **or** only C occurs.

So the answer is $(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C)$

(3) Only A and B occur = A and B occur **and** C does not occur.

So the answer is $A \cap B \cap C^c$

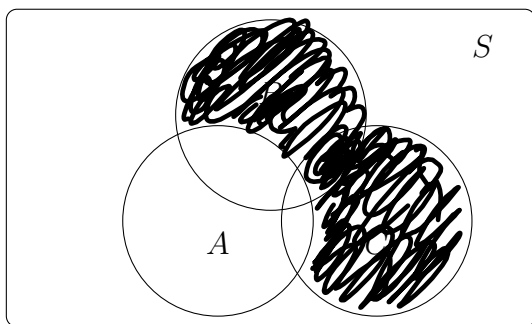
The Venn diagram is very helpful for understanding this kind of questions.



Verify the DeMorgan's law:

1. $(A \cup B \cup C)^c = A^c \cap B^c \cap C^c$ which means none of the three events occurs.
2. $(A \cap B \cap C)^c = A^c \cup B^c \cup C^c$ which means not all three events occur.

Example 16. Find $A^c \cap (B \cup C)$ in the Venn diagram.



§2.3 The Probability Function

Recall that we defined sample space and event of an experiment.

- **Sample Space S :** Set of all the possible outcomes.
- **Event $A \subset S$:** Subset of the sample space.

Recall the classical definition of probability: Suppose the outcomes of an experiment are **all equally likely**, and the sample space is **finite**.

Probability of an event $A = P(A) = \frac{\text{Cardinality of } A}{\text{Cardinality of } S} = \frac{\#(A)}{\#(S)} = \frac{|A|}{|S|}$

Example 1. Rolling two 6-sided dice (one red, one blue) once. Let A be the event that the difference (absolute value) of the two numbers is 1. What is the probability of A ?

$$S = \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\};$$

So, the probability of event A is $P(A) = \frac{10}{36} = \frac{5}{18}$

► In 1930s, Kolmogorov gave a modern axiomatic definition of the probability function P .

Definition. Definition of Probability Function

A **probability function** P assigns a real number to any event of a sample space.

If the sample space S is finite, the probability function satisfies the following axioms.

- **Axiom 1.** $P(A) \geq 0$ for any event A .
- **Axiom 2.** $P(S) = 1$.
- **Axiom 3.** For any two **mutually exclusive** (disjoint) events A and B ,

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

- **Axiom 4.** Let A_1, A_2, A_3, \dots , be events over S .

If any two of them are mutually exclusive, then

(ie. all are disjoint)

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

$A_i \cap A_j = \emptyset$
 $\forall i, j$

$A_1 \cup A_2 \cup A_3 \dots$

Remark: In Kolmogorov's definition, conditions of equally likely and finite are NOT needed any more.

Example 2. Flip a biased coin once, with $P(\text{Head}) = 1/3$ and $P(\text{Tail}) = 2/3$.

Example 3. (Countable set)

Experiment: Tossing a fair coin until we get a head.

Sample Space: $S = \{H, TH, TTH, TTTT, \dots\}$.

Event: Getting a head with no more than 3 tosses, $A = \{H, TH, TTH\}$.

What is the probability of A ? (infinite S , not equally-likely.)

$$P(S) = P(H) + P(TH) + \dots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1$$

$$\left\{ \begin{array}{l} P(A) = P(H) + P(TH) + P(TTH) \\ = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8} \end{array} \right.$$

Handwritten notes: Recall $\sum_{k=1}^{\infty} ar^{k-1} = \frac{a}{1-r}$ if $|r| < 1$. Start $\frac{1}{2}$ H $\frac{1}{4}$ TH $\frac{1}{8}$ TTH \dots

► Some properties can be derived easily from Kolmogorov's axioms. They are extremely important in solving problems.

Theorem 1. $P(A^c) = 1 - P(A)$.

$$S = A \cup A^c \text{ where } A \cap A^c = \emptyset$$

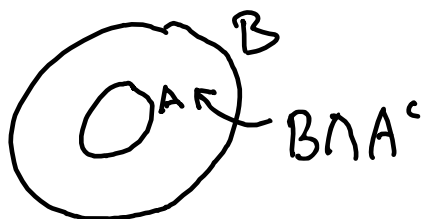
$$P(S) = P(A \cup A^c) = P(A) + P(A^c) = 1 \Rightarrow P(A^c) = 1 - P(A)$$

Theorem 2. $P(\emptyset) = 0$.

$$P(\emptyset) = P(S^c) = 1 - P(S) = 1 - 1 = 0$$

Theorem 3. If $A \subset B$ then $P(A) \leq P(B)$.

Proof: $B = A \cup (B \cap A^c)$ where A and $B \cap A^c$ are disjoint.



$$P(B) = P(A) + P(B \cap A^c)$$

$$P(B) - P(A) = P(B \cap A^c) \geq 0$$

$$\Rightarrow P(B) \geq P(A) \Rightarrow P(A) \leq P(B)$$

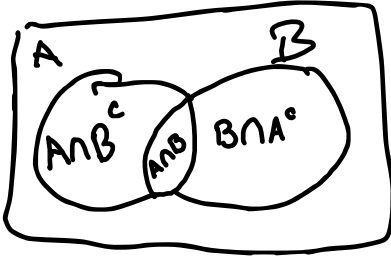
Theorem 4. For every event A , we have $P(A) \leq 1$.

$A \subset S$

Proof: By Thm 3, $P(A) \leq P(S) = 1$

Theorem 5. $P(A) = P(A \cap B^c) + P(A \cap B)$.

Proof:



$$A = (A \cap B^c) \cup (A \cap B)$$

$$P(A) = P(A \cap B^c) + P(A \cap B)$$

Theorem 6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Inclusion - Exclusion Principle $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof:

$$A = (A \cap B^c) \cup (A \cap B) \Rightarrow P(A) = P(A \cap B^c) + P(A \cap B)$$

$$B = (A \cap B) \cup (A^c \cap B) \Rightarrow P(B) = P(A \cap B) + P(A^c \cap B)$$

$$A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B) \Rightarrow P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B)$$

Example 4. Let A and B be two events on S . Suppose $P(A) = 0.5$, $P(B) = 0.6$ and $P((A \cap B)^c) = 0.8$. Answer the following questions:

1. What is the probability that **only** A occurs?

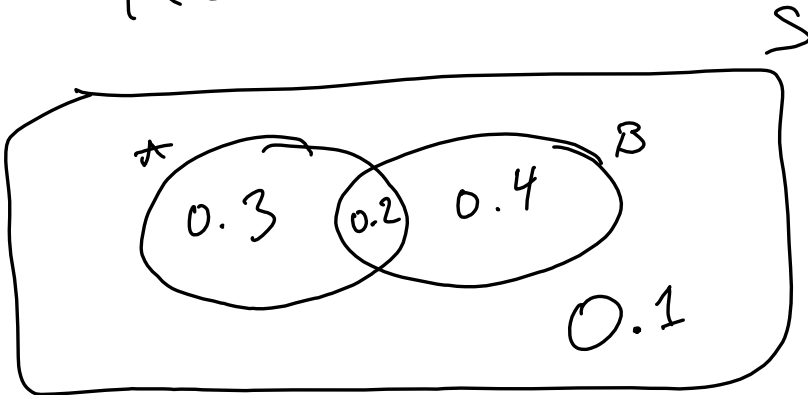
2. What is the probability that A or B occurs?

3. What is the probability that both A and B occur?

$$P(A) = 0.5$$

$$P(B) = 0.6$$

$$P((A \cap B)^c) = 0.8 \Rightarrow P(A \cap B) = 0.2$$



$$\begin{aligned} P(\text{only } A) &= P(A) - P(A \cap B) \\ &= 0.5 - 0.2 = 0.3 \end{aligned}$$

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.5 + 0.6 - 0.2 \\ &= 1.1 - 0.2 = 0.9 \end{aligned}$$

$$\begin{aligned} P(A \cap B) &= 0.2 \quad \therefore A \cap B = (A \cap B)^c \\ \Rightarrow P(A \cap B) &= 1 - P((A \cap B)^c) \\ &= 1 - 0.8 = 0.2 \end{aligned}$$

4. What is the probability that A or B occurs, but not both occurs?

$$P(A \cap B^c) + P(A^c \cap B) = 0.7$$

5. What is the probability that neither A nor B occurs?

$$P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 0.9 = 0.1$$

Example 5. Draw 2 cards from a standard deck. What is the probability that the first card is larger than the second card.

$A = \text{first is larger}$ $S = A \cup B \cup C$ $\left. \begin{array}{l} A = \text{first is larger} \\ B = \text{second card larger} \\ C = \text{equal} \end{array} \right\} \begin{array}{l} P(S) = P(A) + P(B) + P(C) \\ A = \{(3D, 2H), \dots\} \\ C = \{(10, 10)\} \end{array}$

$B = \text{second card larger}$ $B = \{\text{invert all in } A\}$

$C = \text{equal}$

see next page

Example 6. A fair coin is tossed four times. What is the probability that at most three heads will occur?

$n(S) = 2^4$ $\underline{H} \quad \underline{H} \quad \underline{H} \quad \underline{T}$

$A = \text{3 or less H in 4 tosses}$ $P(A) = 1 - P(A^c) = 1 - \frac{1}{16} = \frac{15}{16}$

Example 7. Rolling two 6-sided dice (one red, one blue) once. Find the probability that the first roll is 1, or the absolute value of the difference is 1.

$P(\text{first 1} \cup \text{abs diff 1}) = P(\text{first 1}) + P(\text{abs diff 1}) - P(\text{first 1} \cap \text{abs diff 1})$

$n(\text{first 1} \cap \text{abs diff 1}) = 1 \quad \{(1, 2)\}$

$n(\text{first 1}) = 6, \quad n(S) = 36 \quad = \frac{6}{36} + \frac{10}{36} - \frac{1}{36}$

$n(\text{abs diff 1}) = 10$

14

$= \frac{15}{36} = \frac{5}{12}$

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

Ex. 5 cont

$$A = \{ (3\Diamond, 2\heartsuit), \dots \}$$

$$B = \{ (2\heartsuit, 3\Diamond), \dots \}$$

$$n(A) = n(B) \text{ by symmetry}$$

$$\Downarrow$$
$$P(A) = P(B)$$

$$1 = 2P(A) + P(C)$$

$$C = \left\{ \left(\overset{1st}{\boxed{\text{any}}}, \overset{2nd}{\boxed{3 \text{ opts}}} \right) \right\}$$

$$P(C) = \frac{52}{52} \cdot \frac{3}{51} = \frac{3}{51}$$

$$1 = 2P(A) + \frac{3}{51}$$

$$\frac{48}{51} = 2P(A) \Rightarrow \boxed{P(A) = \frac{24}{51}}$$

► Counting and Probability*

For a **finite** sample space with all **equally likely** outcomes, it is still very important to use the classical definition of probability to compute classical examples.

A very basic principle of counting is the **multiplication rule**:

If operation A can be performed in m different ways and operation B in n different ways, the **sequence (operation A , operation B)** can be performed in $m \cdot n$ different ways.

Example 8. When we buy a cup of smoothie, we can choose Large, Medium, or Small for the cup, then choose Banana, Chocolate, Strawberry, Vanilla for the flavor.

How many ways we can buy a cup of smoothie?

$$3 \text{ size} \times 4 \text{ flavor} = 12 \text{ smoothies}$$

Definition.

The number of ways to *arrange* k objects of a set of n distinct elements (**permutations**), repetitions not allowed, is denoted by the symbol ${}_nP_k$, or P_k^n , or $P(n, k)$,

$$P_k^n = n(n-1)(n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!} \quad \begin{array}{l} \text{order} \\ \text{matter} \end{array}$$

Example 9. What is the probability that we get the word NBA if we arrange the letters A, B, N randomly?

$$\begin{aligned} |S| &= {}_3P_3 = \frac{3!}{0!} = 6 \\ |A| &= 1 \Rightarrow \{NBA\} \end{aligned} \quad \left\{ \begin{array}{l} P(A) = \frac{|S|}{|A|} = \frac{1}{6} \end{array} \right.$$

Example 10. What is the probability that at least two students in our class (³⁰~~70~~ students) share the same birthday?

Example - 6 people (A, B, C, D, E, F)

1) # of ways to choose pres, VP, + secretary

$${}^6P_3 = \frac{6!}{(6-3)!} = \frac{6!}{3!} = 6 \cdot 5 \cdot 4 = 120$$

2) Choose committee of size 3

$$\begin{aligned} \binom{6}{3} &= \frac{6!}{3!(6-3)!} = \frac{6!}{3! \cdot 3!} = \frac{6 \cdot 5 \cdot 4}{3!} \\ &= \frac{120}{6} = 20 \end{aligned}$$

Hypergeometric probability (Ex. 3.2)

$$\frac{{}^{13}C_3 \cdot {}^{39}C_2}{{}^{52}C_5}$$

add to 52