Chapter 2. Probability

He Wang

Contents

§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:** 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:**

Suppose the outcomes of an experiment are all equally likely, and the total number of all possible outcomes is finite.

$$\textbf{Probability of an event} = \frac{\text{Number of ways it can happen}}{\text{Total number of all possible outcomes}}$$

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

Example 1. Experiment: Flipping a fair Coin once.

Event: Landing head.

Number of ways it can happen = 1. Total number of possible outcomes = 2

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

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

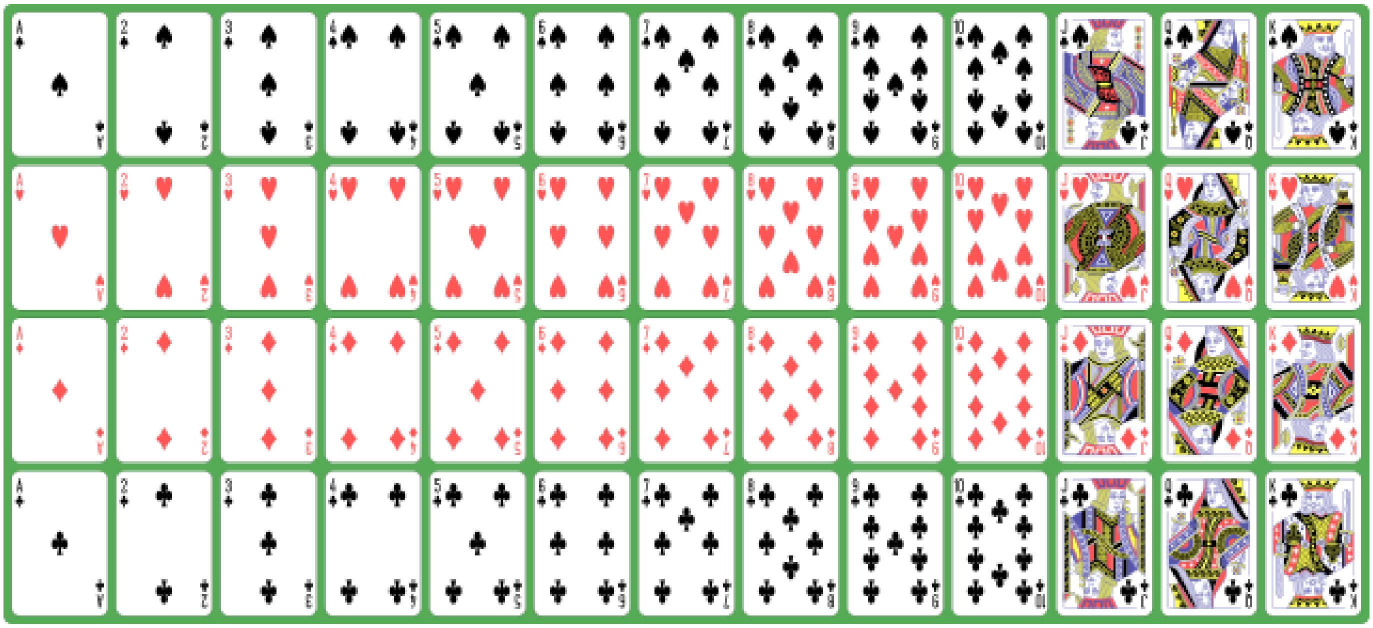
Event: Rolling a number larger than 4 with a die.

Number of ways it can happen = 2. Total number of possible outcomes =6

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

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

$$P(Face) = \frac{4 \times 3}{13 \times 4} = \frac{3}{13}.$$

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

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

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

Definition.

- A **set** S is a *well-defined*, *unordered* collections of *distinct* (possibly infinitely many) elements.
- 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 $A \subset S$.

Non-well-defined example, (Russell's paradox): $S = \{x \mid x \notin x\}$, i.e., set of all sets that are not members of themselves. (The teacher that teaches all who don't teach themselves.)

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

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 $P(A) = 2/4 = 0.5$.

For some calculation purpose, we can also define a sample space as $S = \{0heads, 1heads, 2heads\}$, but each outcome is not equally-likely.

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?

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

So, the probability of event A is $P(A) = \dfrac{10}{36} = \dfrac{5}{18} \approx 27.78\%$

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?

$B = \{(3, 6), (4, 5), (5, 4), (6, 3)\}$.
$C = \{(1, 4), (2, 5), (3, 6), (6, 3), (5, 2), (4, 1)\}$.

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: $S = \{H,\ TH,\ TTH,\ TTTH,\ \ldots\ldots\}$.

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.)

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

Example*: The set of rational numbers $\mathbb{Q}$ is countable infinite.

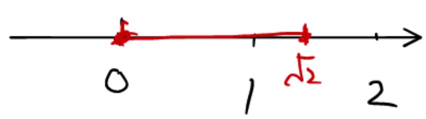
Discrete set means finite or countable set.

Example 9. (Continuous set)

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; $A = [0, 1)$



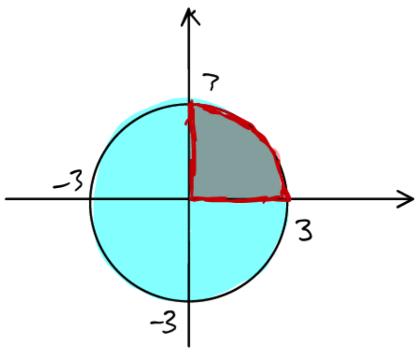
Example 10. (Continuous set)

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

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

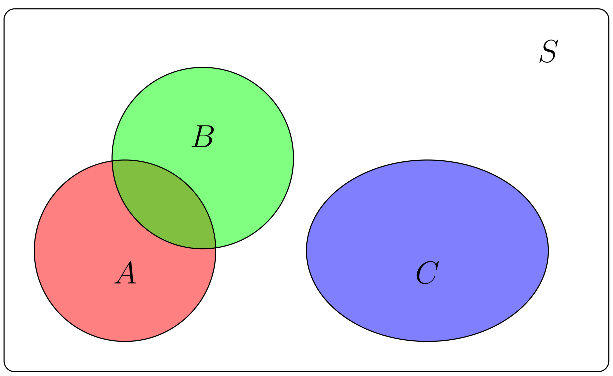
Event A: Get a point (x, y) such that $x \geq 0$ and $y \geq 0$.

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

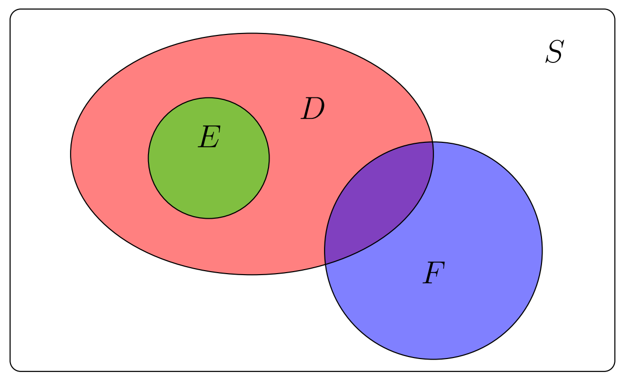


- The **Venn diagram** is a useful visual aid of sets.

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



or



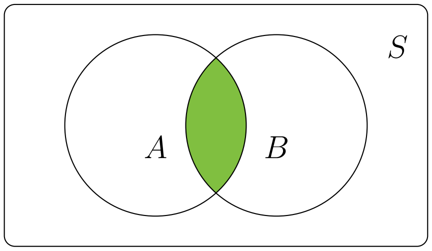
- **Basic operations on sets** (or events).

1. Intersection

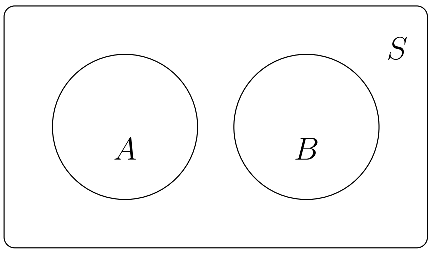
Definition.

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".

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



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



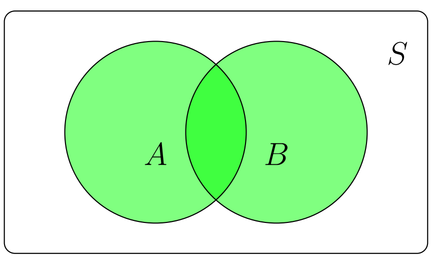
2. Union

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).

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

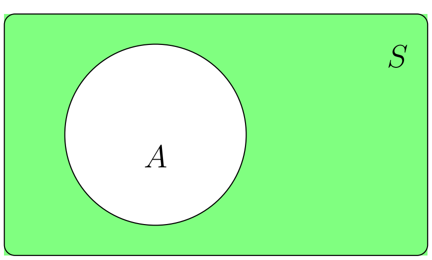
So, $A \cup B$ means that "A or B occurs".



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

$$A \cup A^c = S, \qquad A \cap A^c = \emptyset, \qquad (A^c)^c = A,$$
$$A \cup B = B \cup A, \qquad A \cap B = B \cap A.$$

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 $A \cap B = \{6\}$.

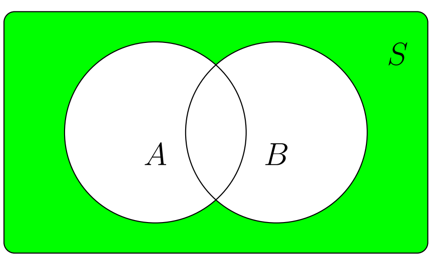
The event of 'Rolling number which is even **or** larger than 4' is the union $A \cup B = \{2, 4, 5, 6\}$.

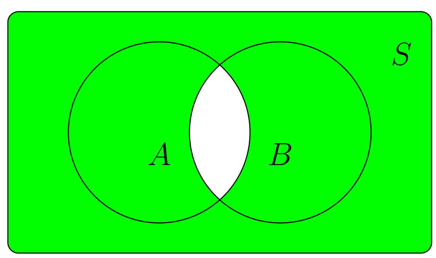
The event of 'Rolling number which is **not** even' is the complement $A^C = \{1, 3, 5\}$.

Theorem. DeMorgan's Law

1. $(A \cup B)^c = A^c \cap B^c$
2. $(A \cap B)^c = A^c \cup B^c$

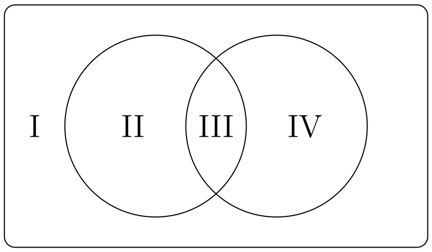
DeMorgan's Law in Venn diagram.

1. 

2. 

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.



$A = \text{II} \cup \text{III}$ and $B = \text{III} \cup \text{IV}$
$A \cap B = \text{III}$
$B^c = \text{I} \cup \text{II}$
$A \cap B^c = \text{II}$
So $(A \cap B) \cap (A \cap B^c) = \text{II} \cup \text{III} = A$

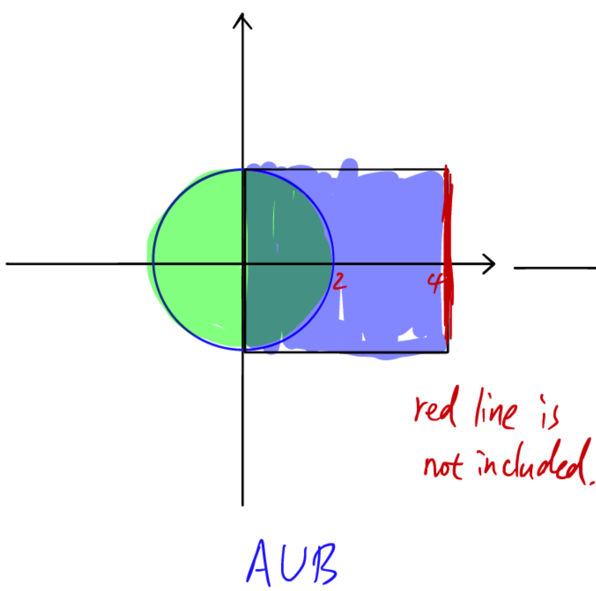
Here, $A \cap B^c$ means that "A occurs but B does not occur".

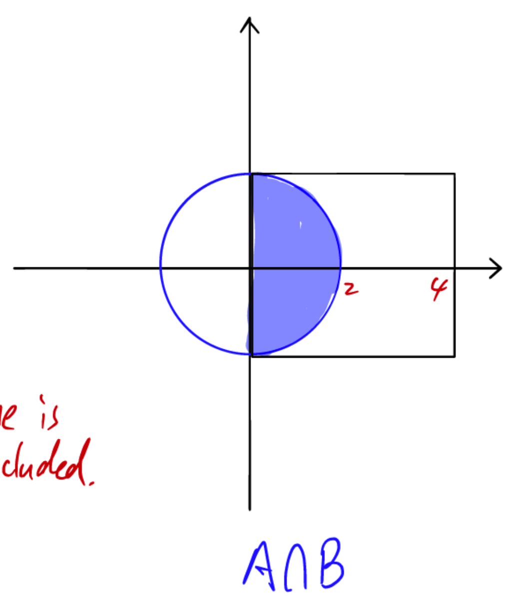
Example 13. Sketch the regions in xy-plane $\mathbb{R}^2$ corresponding $A \cup B$ and $A \cap B$.

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

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

Do not be confused with the Venn diagram.





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 $|S| = 6^4$.

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

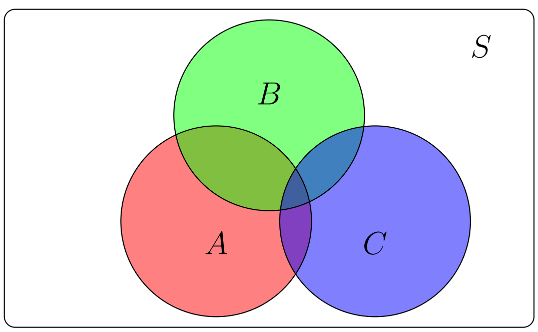
(1) Only B occurs.

(2) exactly one event occurs.

(3) Only A and B occur.

(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.

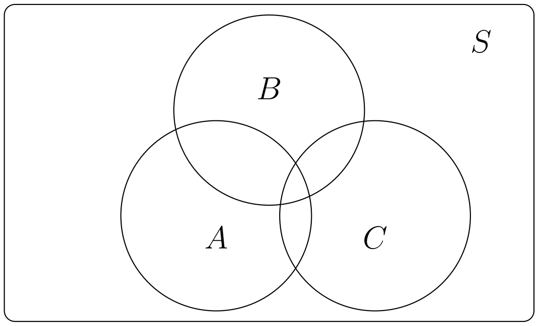


Verify the DeMorgan's law:

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.

For HW 2.2.26, using Venn diagram to check your results. (find the one typo in the solution.)

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


S
B
A
C

§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 $A = P(A) = \dfrac{\text{Cardinality of } A}{\text{Cardinality of } S} = \dfrac{\#(A)}{\#(B)} = \dfrac{|A|}{|B|}$

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?

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

So, the probability of event A is $P(A) = \dfrac{10}{36} = \dfrac{5}{18} \approx 27.78\%$

► 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 a 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,

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

If the sample space S is a finite, a fourth axiom is needed:
- **Axiom 4.** Let $A_1, A_2, A_3, \ldots,$ be events over S.

If any two of them are mutually exclusive, then

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

Remark: In Kolmogorov's definition, conditions of equally likely and finite are NOT needed any more. One can also apply the advanced tools like calculus and real analysis to probability.

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

Example 3. (Countable set)

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

Sample Space: $S = \{H,\ TH,\ TTH,\ TTTH,\ \ldots\ldots\}$.

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.)

Solution: $P(H) = 1/2$, $P(TH) = (1/2)(1/2)$, $P(TTH) = (1/2)^3$.
So, $P(A) = 0.5 + 0.25 + 0.125 = 0.875$.

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

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

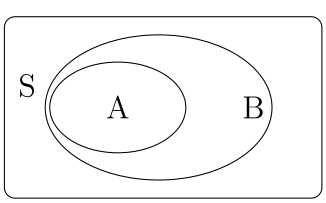
Proof: $S = A \cup A^c$, so $P(S) = P(A \cup A^c)$.
By Axiom 2, $P(S) = 1$.
By Axiom 3, $P(A \cup A^c) = P(A) + P(A^c)$ since $A \cap A^c = \emptyset$.
So, $1 = P(A) + P(A^c)$. Hence $P(A^c) = 1 - P(A)$.

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

Proof: $S^c = \emptyset$. So, $P(\emptyset) = P(S^c) = 1 - P(S) = 1 - 1 = 0$.

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

Proof: $B = A \cup (A \cap A^c)$ where A and $B \cap A^c$ are disjoint.
So, $P(B) = P(A \cup (A \cap A^c)) = P(A) + P(B \cap A^c)$ by Axiom 3.
So, $P(B) \geq P(A)$ since $P(B \cap A^c) \geq 0$ by Axiom 1.

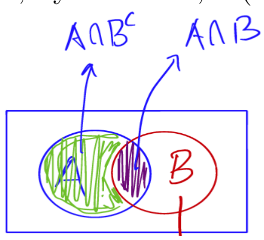


Theorem 4. For every event A, we have $P(A) \leq 1$.

Proof: Since $A \subset S$, by above theorem and axiom 3, we have $P(A) \leq P(S) = 1$.

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

Proof: $A = (A \cap B^c) \cup (A \cap B)$ where $A \cap B^c$ and $A \cap B$ are disjoint.
So, by Axiom 3, $P(A) = P(A \cap B^c) + P(A \cap B)$.



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

Proof:
By above theorem, $P(A) = P(A \cap B^c) + P(A \cap B)$ and $P(B) = P(B \cap A^c) + P(B \cap A)$.
So,

$$\begin{aligned} P(A) + P(B) &= P(A \cap B^c) + P(A \cap B) + P(B \cap A^c) + P(B \cap A) \\ &= P(A \cup B) + P(A \cap B) \end{aligned}$$

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

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?

Solution: Only A occurs = A occurs and B does not occur.
By $P(A) = P(A \cap B) + P(A \cap B^c)$, we have

$$P(A \cap B^c) = P(A) - P(A \cap B) = 0.5 - 0.2 = 0.3.$$

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

Solution:

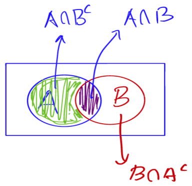
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.6 - 0.2 = 0.9$$

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

Solution: $P(A \cap B) = 0.2$

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

Solution: (Hint: using Venn diagram to rewrite the question.)



$P(A\cap B^c)+P(B\cap A^c) = P(A)-P(A\cap B)+P(B)-P(A\cap B) = 0.5+0.6-2\times 0.2 = 0.7.$

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: the first card is larger
B: the second card is larger
C: the first card is equal to the second card
A, B, C are disjoint(mutually exclusive) and $S = A\cup B\cup C$. So, $1 = P(A)+P(B)+P(C)$.
By symmetry, $P(A) = P(B)$.
First choose a card and count the number, then there are 51 cards left. $P(C) = 3/51$.
So, $P(A) = P(B) = (1-3/51)\frac{1}{2} = 24/51$.

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

A: at most three heads will occur= number of heads ≤ 3.
A^c=number of heads is 4=$\{HHHH\}$
$P(A) = 1 - P(A^c) = 1 - \frac{|A^c|}{|S|} = 1 - \frac{1}{2^4} = \frac{15}{16}.$

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.

Event A: first roll is 1
Event B: the difference (absolute value) of the two numbers is 1.
$P(A\cup B) = P(A)+P(B)-P(A\cap B) = \frac{1}{6}+\frac{5}{18}-\frac{1}{36} = \frac{15}{36}$

§2.4 Conditional Probability

Example 1. Tossing a fair 6-sided die once.

Event $A =$" 2 appears"

Event $B =$ " even number appears"

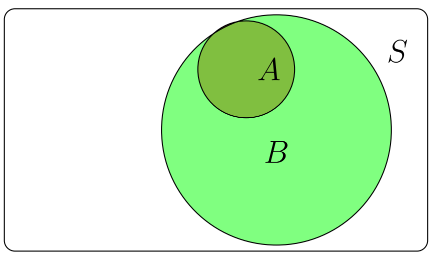
The probability of A is $P(A) = \frac{1}{6}$.

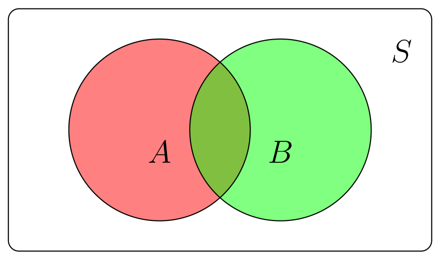
The probability of B is $P(B) = \frac{3}{6} = \frac{1}{2}$.

Suppose the die already tossed. Someone told us that the result is an even number, i.e., B is already occurred. Now, what is the probability that the result is 2?

Probability of event A will occur given that B is already occurred.

Probability of A given B is $\dfrac{1}{3}$.





Definition. Conditional probability

Probability that event A occurs **given** that event B already occurs, denoted by $P(A|B)$ is a **conditional probability**, defined by

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

called **probability of A given B**.

If $\#S$ is finite, we can calculate conditional probability as

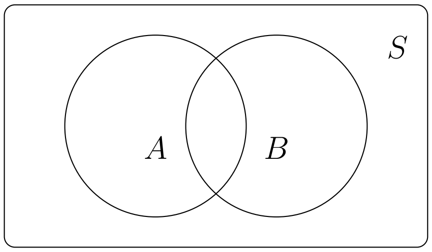
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\#(A \cap B)/\#(S)}{\#(B)/\#(S)} = \frac{\#(A \cap B)}{\#(B)}.$$

Example 2. Suppose $P(A) = 0.45$, $P(B) = 0.6$, and $P(A^c|B) = 0.5$. Find $P(A \cup B)$.

$P(A^c|B) = \dfrac{P(A^c \cap B)}{P(B)}$ implies $0.5 = \dfrac{P(A^c \cap B)}{0.6}$. So, $P(A^c \cap B) = 0.3$. Hence,

$$P(A \cup B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c) = 0.45 + 0.3 = 0.75$$

Example 3. In a town of 5342 residents, there are 3355 voters. The following Venn diagram corresponds to voter information in a mayor election. A is the set of 725 first-time voters, while B is the set of 1588 voters who voted for Bob. There are 260 first-time voters who do NOT vote for Bob.



Question: If a random voter is picked, what is the probability that he/she voted for Bob?

Step 1. Rewrite the question in probability language: $\#(S) = 3355$; $\#(A) = 725$; $\#(B) = 1588$; $\#(A \cap B^c) = 260$.
Hence,

$$\#(A \cap B) = \#(A) - \#(A \cap B^c) = 725 - 260 = 465.$$

Step 2.

$$P(B) = \frac{|B|}{|S|} = \frac{1588}{3355} \approx 0.4733.$$

Question: If a random first-time voter is picked, what is the probability that he/she voted for Bob?

Solution:

$$P(B|A) = \frac{|A \cap B|}{|A|} = \frac{465}{725} \approx 0.6414$$

Example 4. There are 100 marbles distributed as

	Red	White	Blue
Small	5	15	2
Medium	25	5	3
Large	30	10	5

Suppose you choose one marble at random.

(1) Find $P(\text{Blue})$.

From the table, $\#(Small) = 22$, $\#(Large) = 45$, and $\#(Blue) = 10$.
So, $P(Blue) = 10/100 = 0.1$.

(2) Find $P(\text{Blue} \cup \text{small})$.

Let B=Blue, S=Small.

$$P(B \cup S) = P(B) + P(S) - P(B \cap S) = \frac{10}{100} + \frac{22}{100} - \frac{2}{100} = 0.3$$

(3) Find $P(\text{Blue} \mid \text{Large})$.

$$P(B|L) = \frac{P(B \cap L)}{P(L)} = \frac{5/100}{45/100} = 1/9$$

(4) Find $P(\text{Large} \mid \text{Blue})$.

$$P(L|B) = \frac{P(L \cap B)}{P(B)} = \frac{5/100}{10/100} = 1/2 = 0.5$$

Example 5. A family has two children.

(1) Given the first child is a boy, what is the probability that the other child is a boy.

(2) Given that at least one child is a boy, what is the probability that the other child is boy.

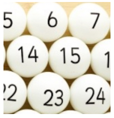
Sample Space $S = \{BB, BG, GB, GG\}$
X: the first child is a boy. $X = \{BB, BG\}$
Y: the second child is a boy. $Y = \{BB, GB\}$
Z: at least one child is a boy. $Z = \{BB, BG, GB\}$
W: both are boys. $W = \{BB\}$
(1). $P(Y|X) = \frac{P(Y \cap X)}{P(X)} = \frac{1/4}{2/4} = 1/2.$
(2). $P(W|Z) = \frac{P(W \cap Z)}{P(Z)} = \frac{1/4}{3/4} = 1/3.$

Remark: In many real world questions, the conditional probability is easy to calculate, because we have more information.

Theorem.

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

Example 6. There are 100 marbles (small glass balls) in a box (or urn):
50 **red** marbles labelled from 1 to 50.
30 **white** marbles labelled from 1 to 30.
20 **blue** marbles labelled from 1 to 20.



(I) Pick **one** marble randomly.

Event A: the ball is white. Event B: the ball is labelled number 6.

Calculate: (1) $P(A)$ (2) $P(A \cap B)$ (3) $P(A \cup B)$

$$P(A) = \frac{30}{100}, \quad P(A \cap B) = \frac{1}{100}, \quad P(A \cup B) = \frac{32}{100}$$

(4) Suppose we have seen that the color is blue, what is the probability that the ball is labelled number 6.

Let C be the event that the ball is blue.

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = 1/20.$$

(II) Now take out **two** without replacement.

(1) If the first one was blue, what is the probability that the second one is red?

R_2: the 2ed one is red. B_1: the 1st one is blue. Because the first one is red, there are 99 balls left with 50 red balls.

$$P(R_2|B_1) = \frac{50}{99}$$

(2) If the first one was red, what is the probability that the second one is red?

R_1: the 1st one is red. Because the first one is red, there are 99 balls left with 49 red balls.

$$P(R_2|R_1) = \frac{49}{99}$$

(3) What is the probability of two red?

$$P(R_1 \cap R_2) = P(R_2|R_1)P(R_1) = \frac{49}{99} \cdot \frac{50}{100} = 49/198$$

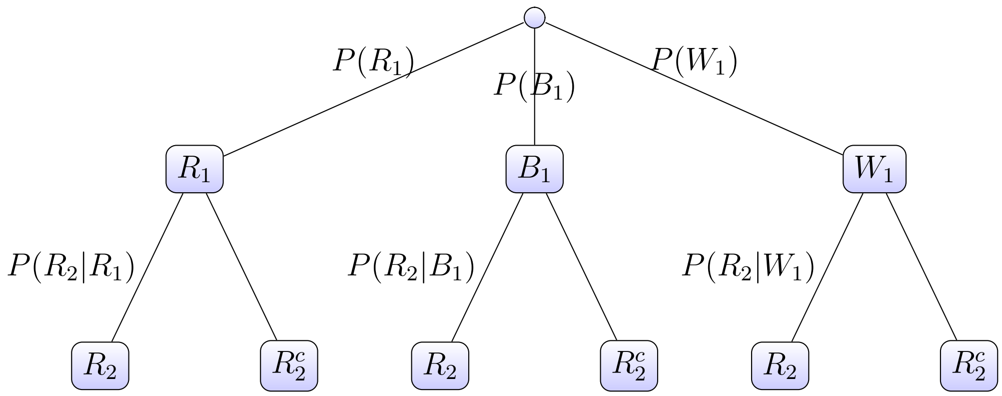
(4) If we don't look at the first one, what is the probability that the second one is red?

R_2: the second one is red.

$$P(R_2) = \frac{50}{100} = 0.5$$

Remark: what we don't know does not matter.

Tree Diagram for conditional probability and intersections:



$P(R_2 \cap R_1)$
$= P(R_2|R_1)P(R_1)$

$P(R_2 \cap B_1)$
$= P(R_2|B_1)P(B_1)$

$P(R_2 \cap W_1)$
$= P(R_2|W_1)P(W_1)$

More generally, by induction, we have

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

or

$$P(A \cap B \cap C \cap D) = P(A)P(B|A)P(C|A \cap B)P(D|A \cap B \cap C)$$

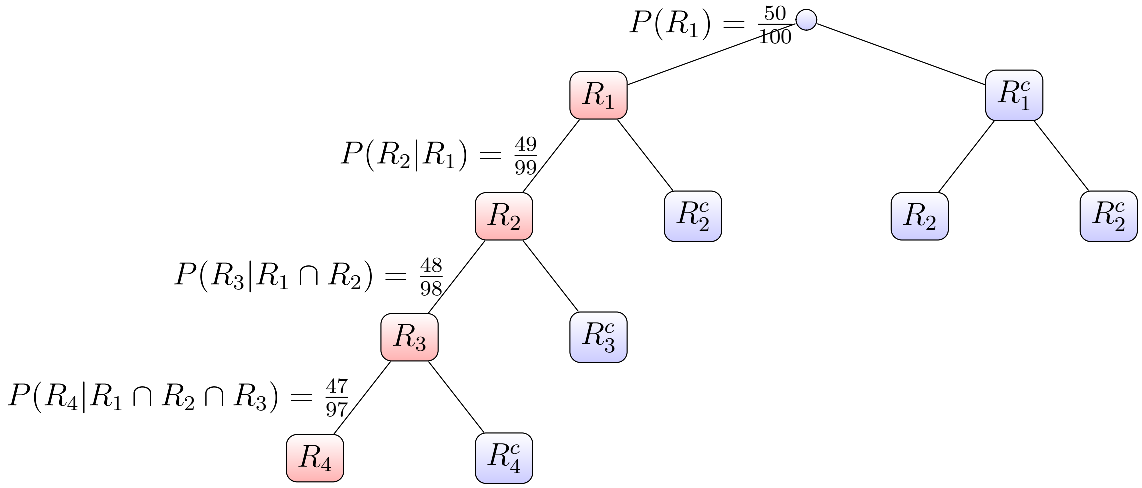
or ...

(III) Now take out **four** without replacement.

(1) What is the probability all four are red?

R_i : the i-th ball is red.

$$P(\text{all four are red}) = P(R_1 \cap R_2 \cap R_3 \cap R_4) = \frac{50}{100} \cdot \frac{49}{99} \cdot \frac{48}{98} \cdot \frac{47}{97}.$$



(2) What is the probability that at least one is blue?

D: at least one is blue. D^c : no blue.

$$P(D) = 1 - P(D^c) = 1 - \frac{80}{100} \cdot \frac{79}{99} \cdot \frac{78}{98} \cdot \frac{77}{97}$$

Recall our Theorem $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$.

Example 7. (A model for **contagious diseases**)

An urn contains 3 red chips and 6 white chips. A chip is drawn at random. If it is red, itself and an additional red chip is put back in the urn. If it is white, the chip is simply returned to the urn. Next a second chip is drawn.

(1) What is the probability that both chips are red?

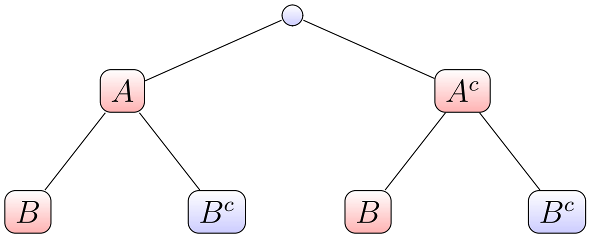
A: first is red. B: second is red.

$$P(A \cap B) = P(B|A)P(A) = \frac{4}{10} \cdot \frac{3}{9} = \frac{2}{15}$$

(2) What is the probability that the second chip is red?

A: first is red. B: second is red.

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap A^c) \\ &= P(B|A)P(A) + P(B|A^c)P(A^c) \\ &= \frac{4}{10} \cdot \frac{3}{9} + \frac{3}{9} \cdot \frac{6}{9} \\ &= \frac{16}{45} \end{aligned}$$



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

▶ **Law of Total Probability (Unconditional Probability)**

Theorem. Law of Total Probability

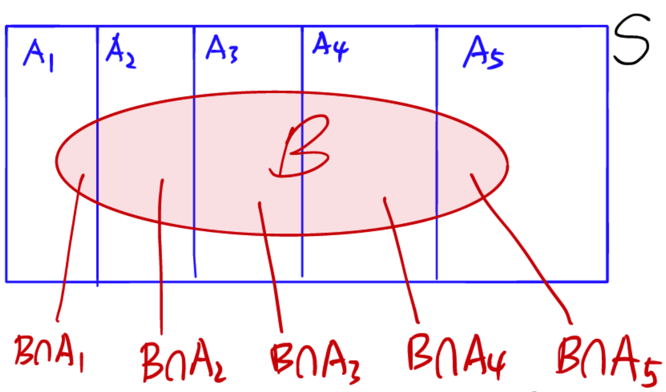
Let $A_1, A_2, \ldots, A_n$ be a sequence of events such that $S = \bigcup_{i=1}^n A_i$ and $A_i \cap A_j = \emptyset$. Then, for any event B,

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

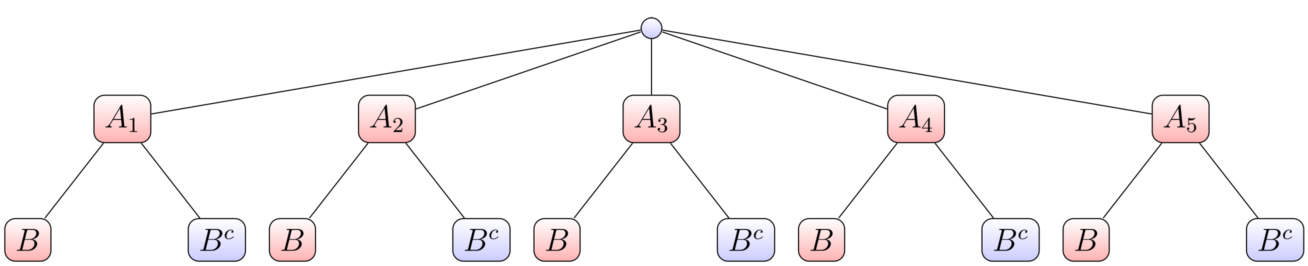
Proof of Law of Total Probability for $n = 5$,

$$\begin{aligned} P(B) &= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) + P(B \cap A_4) + P(B \cap A_5) \\ &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) + P(B|A_4)P(A_4) + P(B|A_5)P(A_5) \end{aligned}$$

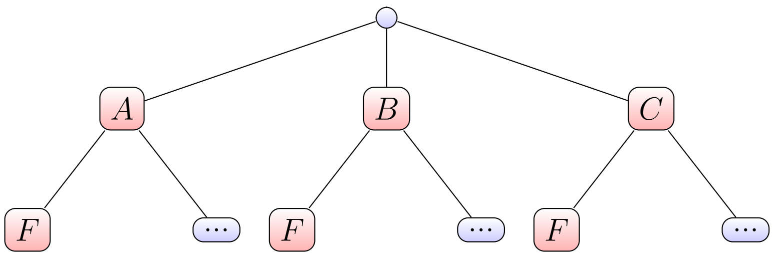
Venn Diagram explanation for Law of Total Probability when $n = 5$:



Tree Diagram explanation for law of total probability:



Example 8. A computer manufacturer uses chips from three suppliers. Based on past performance, it is known that the chips from supplier A will fail with probability 0.01, the chips from supplier B will fail with probability 0.02, and the chips from supplier C will fail with probability 0.05. The manufacturer buys 50% of chips from supplier A, 40% from supplier B and 10% from supplier C. What is the probability that a chip chosen randomly from her mixture will fail?



Solution:
A: Chips from supplier A; $P(A) = 0.5$
B: Chips from supplier B; $P(B) = 0.4$
C: Chips from supplier C; $P(C) = 0.1$
F: Chips fail.
So,

$$P(F|A) = 0.01, \qquad P(F|B) = 0.02, \qquad P(F|C) = 0.05$$

Hence,

$$\begin{aligned} P(F) &= P(F|A)P(A) + P(F|B)P(B) + P(F|C)P(C) \\ &= (0.01)0.5 + (0.02)0.4 + (0.05)0.1 \\ &= 0.018 \end{aligned}$$

Example 9. Three friends are playing poker. Each one draw a card without looking the face of the card. What is the probability that the third card is an Ace?

If we already know the first two cards are club Ace and heart 4, what is the probability that the third card is an Ace?

The answer for the first question is $P(Ace) = 4/52 = 1/13$.
The answer for the second question is $P(Ace|data) = 3/50$.

If you don't trust the first answer, you can have a detailed calculation.
A_1: the first is Ace.
A_2: The second is Ace.
A_3: The third is Ace.
Find $P(A_3)$ using tree diagram for the law of total probability.

▶ **Bayes' Theorem**

Theorem. Bayes' Theorem

Let $A_1, A_2, \ldots, A_n$ be a sequence of events such that $S = \bigcup_{i=1}^n A_i$ and $A_i \cap A_j = \emptyset$. Then, for any event B,

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

for any $j = 1, \cdots, n$.

Proof is easy by the law of total probability:

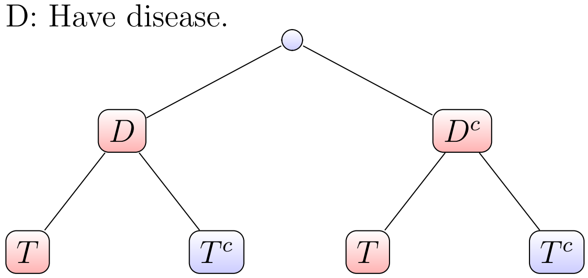
$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

Example 10. Suppose the probability that a person has a disease is 0.02. There is a test that is 95% accurate when given to someone infected, (which says P(test positive | have disease)=0.95) and 96% accurate when given to someone not infected, P(test negative | do not have disease)=0.96)

Find: (1) Randomly choose a person, what is the probability that the test is positive. Find P(test positive).

(2) Randomly choose a person, if the test is positive, what is the probability that this person has disease. Find P(have disease | test positive).

T: Test positive.
D: Have disease.



From the question, $P(D) = 0.02$, $P(T|D) = 0.95$, $P(T^c|D^c) = 0.96$.
So, $P(T|D^c) = 0.04$ and $P(D^c) = 0.98$.
(1) By law of total probability,

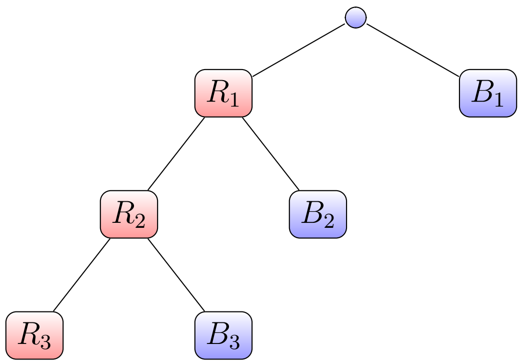
$$P(T) = P(T|D)P(D) + P(T|D^c)P(D^c) = (0.95)0.02 + (0.04)0.98 = 0.0582$$

(2) By Bayes' theorem,

$$P(D|T)\frac{P(T|D)P(D)}{P(T)} = \frac{(0.95)0.02}{0.0582} \approx 0.3265$$

Example 11. A bowl has 6 marbles, 4 red and 2 blue. Take out marbles until either you have one blue or 3 marbles.

(1) Find $P(\text{blue})$.



So,

$$P(\text{blue}) = P(B_1) + P(B_2) + P(B_3) = \frac{2}{6} + \frac{4}{6}\cdot\frac{2}{5} + \frac{4}{6}\cdot\frac{3}{5}\cdot\frac{2}{4} = 0.8$$

(2) Find $P(\text{have 2 marbles} \mid \text{blue})$

$$P(\text{have 2 marbles}|blue) = \frac{P(\text{have 2 marbles} \cap blue)}{P(blue)} = \frac{(4/6)(2/5)}{(4/5)} = 1/3$$

Example 12. Go back to the computer manufacturer problem. Suppose a customer mailed back a failed product. What is the probability that the item came from supplier A?

$$\begin{aligned} P(A|F) &= \frac{P(A\cap F)}{P(F)} \\ &= \frac{P(F|A)P(A)}{P(F|A)P(A) + P(F|B)P(B) + P(F|C)P(C)} \\ &= \frac{0.01(0.5)}{0.018} \\ &\approx 27.78\% \end{aligned}$$

Example 13. *Prize behind door problem (Monty Hall problem.)

In a game show, there are three doors with is a big prize behind only one door. You choose one of them, then one of the left is opened and there is no prize behind the opened door. You have a chance to switch your choice. Will you switch?

C: 1st choice is correct. $P(C) = 1/3$
S: win after switch.
T: win without switch.

$$P(S) = P(S|C)P(C) + P(S|C^c)P(C^c) = 0(\frac{1}{3}) + 1(\frac{2}{3}) = \frac{2}{3}$$

$$P(T) = P(T|C)P(C) + P(T|C^c)P(C^c) = 1(\frac{1}{3}) + 0(\frac{2}{3}) = \frac{1}{3}$$

Conclusion: you should switch.

Example 14. * I take a test with 60 multiple choice questions, where each question has 5 possible answers. I knows some of the answers, and guesses the others. Given that I answers a question correctly, the probability that I knows the answer is 0.9. What is the probability that I knows the answer to a question?

A: I know the answer.
B: I answers the question correctly.
So, from the question $P(A|B) = 0.9$. We want to find $P(A)$.
We also know the following information: $P(A|B^c) = 0$; and $P(B|A) = 1$ because I know the answer; and $P(B|A^c) = 1/5$ because I guess the answer.
By law of total probability, $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$. So $P(A) = 0.9P(B)$.
By law of total probability, $P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$. So, $P(B) = P(A) + 0.2(1 - P(A))$.
Solve the linear equations, we have $P(A) = 9/14 \approx 64.3\%$.

§2.5 Independence

Example 1. Roulette wheel

A Roulette (a wheel gamble) has 18 red, 18 black, and 2 green. If you spin the wheel 2 times, what is the probability of getting 2 red.

R_1 : the first result is red.
R_2 : the second result is red.
$P(R_1 \cap R_2) = P(R_2 \cap R_1) = P(R_1)P(R_2|R_1) = (18/38)^2$.

Some other roulette does not have the 2 green numbers.(e.g., homework 2.5.12)

Definition.

The sets A and B are called **independent** if

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

Recall our Theorem $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$.

If If A and B are not empty set, we have the following equivalent definition.

Theorem.

The sets A and B are independent iff

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

From the theorem, independent means the the probability of A does not depending on the result of B, vice versa.

Example 2. Draw a card from a standard poker deck.

Event A: the card is a King.

Event B: the card is a Diamond.

Are the sets A and B independent?

$$P(A) = \frac{4}{52} = 1/13, \text{ and } P(B) = \frac{13}{52} = 1/4$$

$$P(A \cap B) = P(\text{the card is the Diamond King}) = 1/52$$

So, $P(A \cap B) = P(A)P(B)$. Hence, A and B are independent.

Remark: It is important not to confuse "mutually exclusive" and "independence". In the above example, A and B are not disjoint.

Consider Event C: the card is a Jack. Then A and C are disjoint but not independent.)

Example 3. Let A and B be two independent events on S, and $P(A) = 0.3$ and $P(B) = 0.8$. Find $P(A \cup B)$.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B) = 0.3 + 0.8 - 0.3(0.8) = 0.86$$

Example 4. Roll two fair 6-sides dice.

Consider the sets: A={first roll =3}, B={sum=8}, C={sum=7}, D={first roll =1},

(1) Are the sets A and B independent? Are they disjoint?

$P(A) = \frac{1}{6}$ and $P(B) = \frac{5}{36}$

$P(A \cap B) = \frac{1}{36}$ so $P(A)P(B) \neq P(A \cap B)$.

Hence, A and B are neither independent nor disjoint.

(2) Are the sets A and C independent? Are they disjoint?

$P(C) = \frac{6}{36} = \frac{1}{6}$

$P(A \cap C) = \frac{1}{36}$. So, $P(A \cap C) = P(A)P(C)$.

Hence, A and C are independent, but not disjoint.

(3) Are the sets B and D independent? Are they disjoint?

$P(D) = \frac{1}{6}$

$P(B \cap D) = 0$. Hence, B and D are not independent. B and D are disjoint.

More than Two Sets:

Definition.

The sets A, B, and C are called **independent** if:

(1) $P(A \cap B \cap C) = P(A)P(B)P(C)$, and

(2) $P(A \cap B) = P(A)P(B)$,

$P(A \cap C) = P(A)P(C)$,

$P(B \cap C) = P(B)P(C)$.

This means that knowing one event or two events does not affect the other events.

Those 4 equations do not depend each other. See the following example.

Example 5. Homework 11 and 12.

Most of the time, we know independent from the real world questions. (For example, roll a coin or dice n times.)

Then we can use one side of the property: If A_1, A_2, ..., A_n are independent, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2)\cdots P(A_n)$$

Example 6. Consider a string of 20 Christmas tree lights connected in series. Suppose the probability that a light bulb fails is 2%. What is the probability that the string fails?

F: the sting fails.
F_i: the i-th fails. So, $P(F_i) = 0.02$ and $P(F_i^c) = 0.98$.

$$\begin{aligned} P(F) &= 1 - P(\text{the string work}) \\ &= 1 - P(F_1^c \cap \cdots \cap F_{20}^c) \\ &= 1 - \big(P(F_1^c)\cdots P(F_{20}^c)\big) \\ &= 1 - (0.98)^{20} \\ &\approx 0.3324 \end{aligned}$$

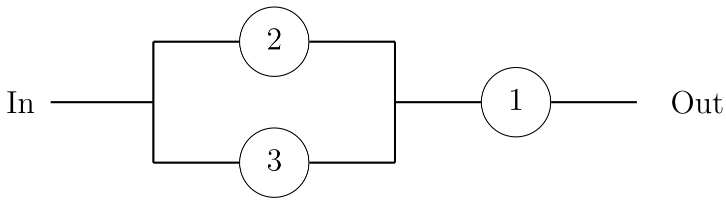
Why $P(F) = P(F_1) + P(F_2) + \cdots + P(F_{20})$ is wrong?
$F = F_1 \cup F_2 \cup \cdots \cup F_{20}$ but F_i are not disjoint.

Example 7. Players A and B toss a fair coin in order. The first player to throw a head wins and ends the game. What are their respective chances of winning?

$$\begin{aligned} P(\text{A throws a head}) &= P\big(A_H \cup (A_T \cap B_T \cap A_H) \cup (A_T \cap B_T \cap A_T \cap B_T \cap A_H) \cup \cdots\big) \\ &= P(A_H) + P(A_T \cap B_T \cap A_H) + P(A_T \cap B_T \cap A_T \cap B_T \cap A_H) + \cdots \\ &= \frac{1}{2} + \frac{1}{2}(1/4) + \frac{1}{2}(1/4)^2 + \cdots \\ &= \frac{1}{2}\left(\frac{1}{1-1/4}\right) \\ &= 2/3 \end{aligned}$$

$P(\text{B wins}) = 1 - 2/3 = 1/3.$

Example 8. (Circuit Problem.) Consider the following 3 bulbs circuit. Suppose each bulb fail independently with probability p. What is the probability of the circuit fail?(String, or circuit fail means no electricity can pass.)



F_i: component i fail. So, $P(F_i) = p$ for i=1,2,3.
C: the circuit fail.

$$\begin{aligned} P(C) &= P\big((F_2 \cap F_3) \cup F_1\big) \\ &= P(F_2 \cap F_3) + P(F_1) - P(F_2 \cap F_3 \cap F_1) \\ &= P(F_2)P(F_3) + P(F_1) - P(F_2)P(F_3)P(F_1) \\ &= p^2 + p - p^3 \end{aligned}$$

Example 9. Roll a unfair (biased) coin 9 times. (Or, roll 9 coin once.) Suppose the probability of getting Head is $P(H) = p = 1/3$.

Find: (1). $P(\text{all Heads})$

By independent, $P(\text{all heads}) = p^9 = (1/3)^9$.

(2). $P(\text{no Head})$

$$P(\text{no head}) = (1-p)^9 = (2/3)^9$$

(3). $P(\text{Exactly one Head})$

A_i: **only** the i-th is head.

$P(\text{Exactly one Head}) = P(A_1) + \cdots P(A_9) = 9P(A_1) = 9p(1-p)^8 = 9\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^8 = \frac{2^8}{3^7}$

(4). $P(\text{Exactly three Heads})$

Choose 3 positions from the 9 places. There are $\binom{9}{3}$ ways.
For each choice, the probability is $p^3(1-p)^6$.
So,

$$P(\text{Exactly three Heads}) = \binom{9}{3} p^3(1-p)^6$$

You can change the example to "Roll a fair die 9 times" and change Head to "6".

Example 10. (Bayesian Inference) I tell you that I can toss coin such that it always comes up Heads. You are 95% certain that I am lying. I tossed a coin 5 times in front of you and comes up Head every time. How certain are you now that I am lying?

H_1: I am lying (the coin if fair).
$H_2 = H_1^c$: I can always toss Heads.
D=Data:= { I tossed 5 times and got 5 heads }
Prior probabilities: $P(H_1) = 0.95$ and $P(H_2) = 0.05$
Posterior probabilities: (after experiments) $p_1 = P(H_1|D)$ updated probability for H_1 given data D. By Bayes's Theorem,

$$P(H_1|D) = \frac{P(D|H_1)P(H_1)}{P(D|H_1)P(H_1) + P(D|H_2)P(H_2)} == \frac{(0.5)^5 0.95}{(0.5)^5 0.95 + 1(0.05)} = 0.37$$

Summary: based on the data D, your new degree of certainty that I am lying is 37%.
We used independence to calculate $P(D|H_1) = P(HHHHH) = P(H)^5 = 0.5^5$.

Further Reading materials about Bayesian Inference:

1. Naive Bayes spam filtering https://en.wikipedia.org/wiki/Naive_Bayes_spam_filtering
2. Bayesian poisoning https://en.wikipedia.org/wiki/Bayesian_poisoning

§2.6 Combinatorics

▶ 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**:

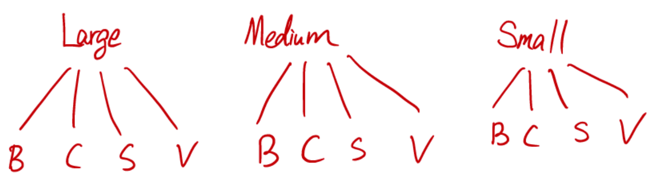
Theorem. Multiplication rule

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 1. 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?

Solution: $3 \times 4 = 12$.



Example 2. Suppose that two cards are drawn-in order- from a standard 52-card poker deck. In how many ways can the first card be a heart and the second card be a King?

If the first card is Heart-King, then there are 1×3 ways.
If the first card is Heart-not-King, then there are 12×4 ways. The total is then $1 \times 3 + 12 \times 4 = 51$ ways.

More generally,

Theorem. Multiplication rule of more operations

If each operation A_i can be performed in n_i different ways, the **ordered sequence** $A_1, A_2, \cdots, A_k$) can be performed in $n_1 n_2 \cdots n_k$ different ways.

Example 3. Suppose you choose a password of length 6. You choose 4 numbers first, then 1 small letter and then 1 capital letter. How many different ways to choose the password.

$10 \cdot 10 \cdot 10 \cdot 10 \cdot 26 \cdot 26 = 6,760,000$

▶ **Permutations (all distinct)**

Theorem.

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 $P(n,k)$,

$$P_k^n = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

In particular, the number of ways to *arrange* n distinct objects is $P_n^n = n!$.

Example 4. What is the probability that we get the word NBA if we arrange the letters A, B, N randomly?

Solution: All possible ways to arrange A, B, N :

$$ABN, ANB, BAN, BNA, NAB, NBA$$

In fact, there are $P_3^3 = 3\times 2\times 1 = 6$ ways. So, the probability $P(A) = 1/6$.

Example 5. What is the probability that we get the word NBA if we arrange 3 of letters randomly?

The number of possible ways to arrange 3 letter from 26 letters are $P_3^{26} = 26\cdot 25\cdot 24 = 15600$. So the probability is $1/15600$

Example 6. What is the probability that at least two students in a class (50 students) share the same birthday?

Solution:

$$\begin{aligned} P(A) &= 1 - P(\text{all have different birthdays}) \\ &= 1 - \frac{P_{50}^{365}}{365^{50}} \\ &\approx 97\% \end{aligned}$$

▶ **Permutations (Not all distinct)**

Theorem.

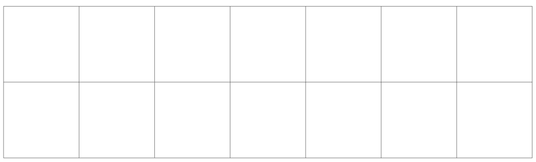
The number of ways to arrange n objects, n_1 of the same kind, n_2 of the same kind, ..., n_k of the same kind, $(n = n_1 + n_2 + \cdots + n_k)$, is given by

$$\frac{n!}{n_1!\cdot n_2!\cdots n_k!}$$

Example 7. How many different ways to permute letters in word ARRANGE.

$$\frac{7!}{2!2!} = 1260$$

Example 8. A point is currently at $O = (0, 0)$ and moving to $A = (7, 2)$. Each move must be a positive 1 unit along x or y. How many different routes can it take.?



It can only x or y positive direction. So one of the result will looks like $xxyxxxyxx$ or $xyxxxxxxy$ with 2 y and 7 x. So the number of routes will be

$$\frac{9!}{7!2!} = 36$$

Example 9. What is the coefficient of $a^2b^3c^5$ in the expansion of $(a + b + c)^{10}$?

The coefficient is $\frac{10!}{2!3!5!} = 2520$.

▶ Combinations

Theorem.

The number of ways to *choose* a subset of k objects from n distinct objects (**combinations**), denoted by C_k^n or $\binom{n}{k}$, or $C(n, k)$,

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

Example 10. Five cards are drawn from a standard deck. What is the probability of getting a "royal flush"(10, J, Q, K, A of the same suit)?

$$\frac{4}{\binom{52}{5}} = \frac{4}{2598960} \approx 0.00000154$$

Example 11. Five cards are drawn from a standard deck. What is the probability of getting a "straight flush"(5 cards in order of the same suit, e.g., A,2,3,4,5 from club or 10, J, Q, K, A from heart)?

$$\frac{40}{\binom{52}{5}} = \frac{40}{2598960} \approx 0.0000154$$

Example 12. Five cards are drawn from a standard deck. What is the probability that there are exactly 3 diamonds?

Solution: In a standard deck, there are 13 diamonds and 39 non-diamonds. Let A be the event that exactly 3 diamonds in 5 random cards.

$$P(A) = \frac{\binom{13}{3}\binom{39}{2}}{\binom{52}{5}} = \frac{286 \times 741}{2598960} \approx 8.15\%$$

Example 13. Recall the binomial formulas:

$$\begin{aligned} (x+y)^2 &= x^2 + 2xy + y^2 \\ (x+y)^3 &= x^3 + 3x^2y + 3xy^2 + y^3 \\ (x+y)^4 &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4 \\ &\vdots \end{aligned}$$

More generally,

$$(x+y)^n = x^n + \binom{n}{n-1}x^{n-1}y + \cdots + \binom{n}{k}x^k y^{n-k} + \cdots + \binom{n}{1}xy^{n-1} + y^n$$

In particular, when $x = y = 1$, we get an interesting equality

$$2^n = \binom{n}{n} + \binom{n}{n-1} + \cdots + \binom{n}{k} + \cdots + \binom{n}{1} + \binom{n}{0}$$

Example 14. Suppose you need to finish reading a assignment with 20 pages in 5 week days. You intend to read the first x_1 pages Monday, the next x_2 pages Tuesday, and the next x_3 pages Wednesday, the next x_4 pages on Thursday and the final x_5 pages on Friday, such that $x_1 + x_2 + x_3 + x_4 + x_5 = 20$, and each $x_i \geq 1$. how many ways can you complete the assignment? That is, how many different sets of values can be chosen for x_1, x_2, x_3, x_4, x_5?

Number the spaces between the twenty pages from 1 to 19. Choosing any four of these spaces partitions the reading assignment into five non-zero numbers.
So the number of ways is $\binom{19}{4} = 3876$