Chapter 1: Basic Concepts of Probability

1 Basic Probability Concepts and Definitions

- **Statistical Experiment**: Any procedure that obtains data (observations).
- **Sample Space** (denoted by S): The set of all possible outcomes of a statistical experiment.

  It depends on the problem of interest!
- **Sample Point**: Every outcome (element) in a sample space.
- **Events**: Subset of a sample space.

  Example 1
  Consider an experiment of **tossing a die**.

  - If the problem of interest is "the number shows on the top face", then
    - Sample space: $S = \{1, 2, 3, 4, 5, 6\}$.
    - Sample point: 1 or 2 or 3 or 4 or 5 or 6.
    - Events: (1) An event that an odd number occurs = $\{1, 3, 5\}$; (2) An event that a number greater than 4 occurs = $\{5, 6\}$.
  - If the problem of interest is "whether the number is even or odd", then
    - Sample space: $S = \{\text{even}, \text{odd}\}$.
    - Sample point: "even" or "odd".
    - Events: An event that an odd number occurs = $\{\text{odd}\}$.

REMARK:

- The sample space is itself an event and is called a **sure event**.
- An event that contains no element is the empty set, denoted by $\emptyset$, and is called a **null event**. ■

L–example 1.1

Consider an experiment of **tossing two dice**. Suppose that the problem of interest is "the numbers that show on the top faces".

here is a tuple since the order matters

- If the dices are labelled, S contains 36 elements:
  - Sample space: $S = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6), (2,1),(2,2),\ldots,(6,5),(6,6)\}$.
  - Sample point: $(1,1)$ or $(1,2)$ or ......
  - Events: event $A = \{$the sum of the dice equals 7$\}$,

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

- If the dices are NOT labelled, S contains 21 elements:
  - Sample space:

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

  - Sample point: $\{1,1\}$ or $\{1,2\}$ or ......
  - Events: event $A = \{$the sum of the dice equals 7$\}$,

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

L–example 1.2 Consider a two step experiment:

1. Flip a coin and observe whether the head (H) or the tail (T) is facing up.
2. If H is obtained in step 1, then flip it again; otherwise, roll a die once.

- Sample space: $S = \{(H,H),(H,T),(T,1),(T,2),(T,3),(T,4),(T,5),(T,6)\}$.
- Sample point: (H,H) or (H,T) or ......
- Events: $A =$ no die is thrown,

$$A = \{(H,H),(H,T)\}.$$

L–example 1.3 Consider an experiment of drawing **two** balls from a jar with a blue, a white, and a red ball.

If the problem of interest is the colours of the two drawn balls, then

- Sample space:

$$S = \{(B,W),(B,R),(W,B),(W,R),(R,B),(R,W)\}.$$

- Sample point: (B,W) or (B,R) or ......
- Events: $A = \{\text{a white ball is chosen}\}$,

$$A = \{(W,B),(W,R),(B,W),(R,W)\}$$

2 Event Operations

- Denote by S the sample space; let A and B be two events. Event operations and possible relationships are summarized below.
- Event operations include:

  (1). Union: $A \cup B$; (2). Intersection: $A \cap B$; (3). Complement: A'.

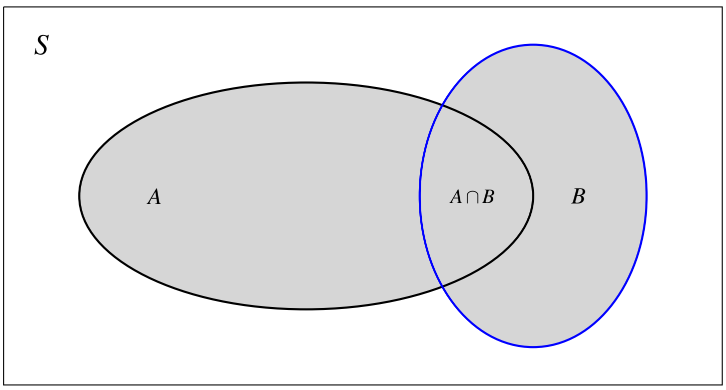
- Possible event relationships:

  (1). Contained: $A \subset B$; (2). Equivalent: $A = B$; (3) Mutually exclusive: $A \cap B = \emptyset$; (4) Independent: $A \perp B$ (postponed).

Union

The **union** of events A and B, denoted by $A \cup B$, is the event containing all elements that belong to A or B or both. That is

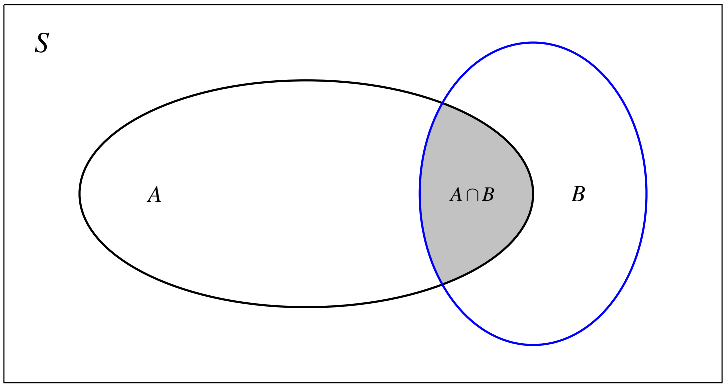
$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$



Intersection

The **intersection** of events A and B, denoted by $A \cap B$ or simply AB, is the event containing elements that belong to both A and B. That is

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



Union and **intersection** can be extended to n events: $A_1, A_2, \ldots, A_n$.

- **Union**:

  $$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \ldots \cup A_n = \{x : x \in A_1 \text{ or } x \in A_2 \text{ or } \ldots \text{ or } x \in A_n\},$$

  composed of elements that belong to one or more of $A_1, \ldots, A_n$.

- **Intersection**:

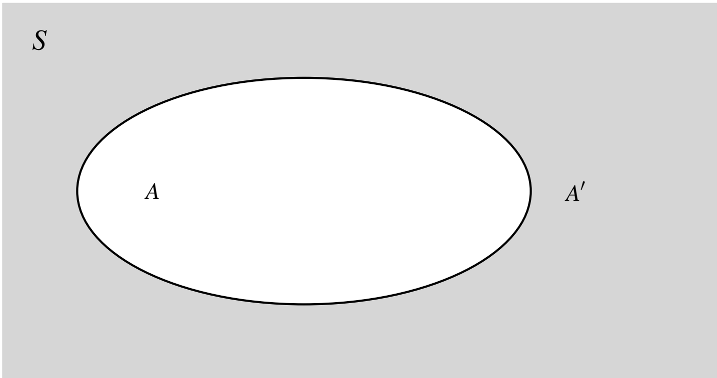
  $$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \ldots \cap A_n = \{x : x \in A_1 \text{ and } x \in A_2 \text{ and } \ldots \text{ and } x \in A_n\},$$

  composed of elements that belong every $A_1, \ldots, A_n$.

Complement

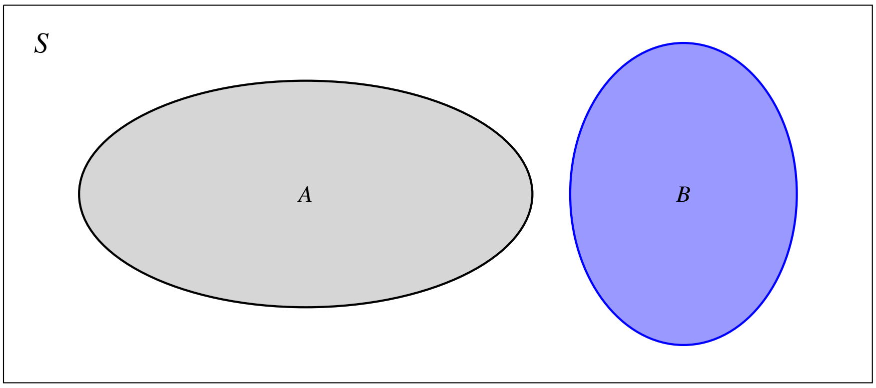
The **complement** of the event A (with respect to S), denoted by A', is the event with elements in S, which are not in A. That is

$$A' = \{x : x \in S \text{ but } x \notin A\}.$$



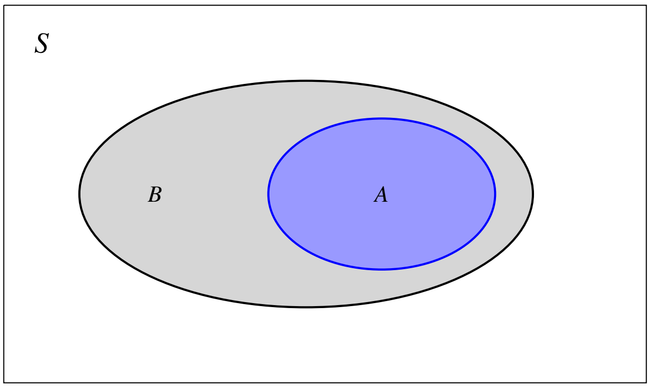
Mutually Exclusive

Events A and B are said to be mutually exclusive or disjoint, if $A \cap B = \emptyset$, i.e., A and B have no element in common.



Contained and Equivalent

- If all elements in A are also elements in B, then we say A is **contained** in B, denoted by $A \subset B$ (or equivalently $B \supset A$).



$A \subset B$:

1) $A \subsetneq B$

2) $A = B$

- If $A \subset B$ and $B \subset A$, then $A = B$, i.e., set A and B are **equivalent**.

EXAMPLE 1

Consider the sample space and events: $S = \{1,2,3,4,5,6\}$, $A = \{1,2,3\}$, $B = \{1,3,5\}$, $C = \{2,4,6\}$. Then

- $A \cup B = \{1,2,3,5\}$; $A \cup C = \{1,2,3,4,6\}$; $B \cup C = S$.
- $A \cap B = \{1,3\}$; $A \cap C = \{2\}$; $B \cap C = \emptyset$.
- $A \cup B \cup C = S$; $A \cap B \cap C = \emptyset$.
- $(A \cap B) \cup C = \{1,3\} \cup \{2,4,6\} = \{1,2,3,4,6\}$.
- $A' = \{4,5,6\}$; $B' = \{2,4,6\} = C$.
- B and C are mutually exclusive, since $B \cap C = \emptyset$; A and B are not mutually exclusive since $A \cap B = \{1,3\} \neq \emptyset$.

Some Basic Properties of Event Operations

(1). $A \cap A' = \emptyset$ (2). $A \cap \emptyset = \emptyset$

(3). $A \cup A' = S$ (4). $(A')' = A$

(5). $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(6). $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(7). $A \cup B = A \cup (B \cap A')$

(8). $A = (A \cap B) \cup (A \cap B')$

De Morgan's Law

For any n events $A_1, A_2, \ldots, A_n$,

(9). $(A_1 \cup A_2 \cup \ldots \cup A_n)' = A_1' \cap A_2' \cap \ldots \cap A_n'$.
A special case: $(A \cup B)' = A' \cap B'$.

(10). $(A_1 \cap A_2 \cap \ldots \cap A_n)' = A_1' \cup A_2' \cup \ldots \cup A_n'$.
A special case: $(A \cap B)' = A' \cup B'$.

EXAMPLE 2

Adopt the setting of Example 1: $S = \{1,2,3,4,5,6\}$, $A = \{1,2,3\}$, $B = \{1,3,5\}$, $C = \{2,4,6\}$. We have

$$A' = \{4,5,6\}, \quad B' = \{2,4,6\}, \quad C' = \{1,3,5\}.$$

- $(A \cup B)' = \{1,2,3,5\}' = \{4,6\}$; $A' \cap B' = \{4,5,6\} \cap \{2,4,6\} = \{4,6\}$. This complies with $(A \cup B)' = A' \cap B'$.
- $(A \cap B)' = \{1,3\}' = \{2,4,5,6\}$; $A' \cup B' = \{4,5,6\} \cap \{2,4,6\} = \{2,4,5,6\}$. This complies with $(A \cap B)' = A' \cup B'$.

- Similarly, we can check $(A \cup B \cup C)' = \emptyset = A' \cap B' \cap C'$; and $(A \cap B \cap C)' = S = A' \cup B' \cup C'$.

L–example 1.4 Revisit L–Example 1. Consider a two step experiment:

1. Flip a coin and observe whether the head (H) or the tail (T) is facing up.
2. If H is obtained in step 1, then flip it again; otherwise, roll a die once.

Then sample space:

$$S = \{(H,H),(H,T),(T,1),(T,2),(T,3),(T,4),(T,5),(T,6)\}$$

- Consider the events:

$$\begin{aligned} A &= \{\text{The die is rolled and the number is no more than } 3\} \\ &= \{(T,1),(T,2),(T,3)\} \\ B &= \{\text{The die is rolled and the number is even}\} \\ &= \{(T,2),(T,4),(T,6)\} \\ C &= \{\text{The die is not rolled}\} = \{(H,H),(H,T)\} \end{aligned}$$

- Then their complements are

$$\begin{aligned} A' &= \{(H,H),(H,T),(T,4),(T,5),(T,6)\} \\ B' &= \{(H,H),(H,T),(T,1),(T,3),(T,5)\} \\ C' &= \{(T,1),(T,2),(T,3),(T,4),(T,5),(T,6)\} \end{aligned}$$

- Some event operations:

$A \cup B = \{(T,1),(T,2),(T,3),(T,4),(T,6)\}$

$B \cup C = \{(T,2),(T,4),(T,6),(H,H),(H,T)\}$

$A \cap B = \{(T,2)\}$

$B \cap C = \emptyset$; so B and C are mutually exclusive

$A \cup B \cup C = \{(H,H),(H,T),(T,1),(T,2),(T,3),(T,4),(T,6)\}$

$A \cap B \cap C = \emptyset \qquad (A \cup B) \cap C = \emptyset$

$A' \cap B' = \{(H,H),(H,T),(T,5)\} = (A \cup B)'$

$A' \cup B' = \{(H,H),(H,T),(T,1),(T,3),(T,4),(T,5),(T,6)\} = (A \cap B)'$.

3 Counting Methods

- In many instances, we need to count the number of ways that some operations can be carried out or that certain situations can happen.
- There are two fundamental principles in counting:

  Multiplication principle

  Addition principle

- They can be applied to obtain some important counting methods: **permutation** and **combination**.

Multiplication Principle

Suppose that r different experiments are to be performed sequentially.

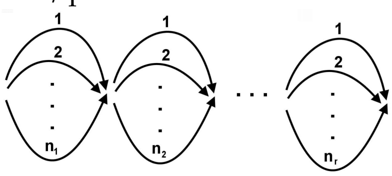
experiment 1 results in n_1 possible outcomes;

for each of the above result, experiment 2 results in n_2 possible outcomes;

$\cdots\ \cdots\ \cdots\ \cdots\ \cdots$

for each of the above result, experiment r results in n_r possible outcomes.

Together there are $n_1 n_2 \cdots n_r$ possible outcomes of the r experiments.



Example 1

How many possible outcomes are there when a die and a coin are thrown together?

Solution:

- experiment 1: throwing a die; it has 6 possible outcomes: $\{1,2,3,4,5,6\}$.
- experiment 2: throwing a coin; for each outcome of experiment 1, it has 2 possible outcomes: $\{H, T\}$

Together there are $6 \times 2 = 12$ possible outcomes.

In fact, the sample space is given by

$$S = \{(x,y) : x = 1,\ldots,6; y = H \text{ or } T\}$$

L–example 1.5 A small community consists of 10 men, each of whom has 3 sons. If one man and one of his sons are to be chosen as "father and son" of the year, how many different choices are possible?

Solution:

- experiment 1: choose the father; it has 10 possible choices.
- experiment 2: choose the son; for each of the father, there are 3 sons to choose from.

Together, there are $10 \times 3 = 30$ choices possible.

L–example 1.6 How many **even three-digit numbers** can be formed from the digits 1,2,5,6, and 9 if each digit can be used **at most once**?

Solution: We can consider the whole task as 3 sequential experiments:

experiment 1: choose the number for ones place; digits 2 and 6 can be used, so there are 2 possibilities.

experiment 2: choose the number for tens place from digits left form experiment 1: 4 possibilities.

experiment 3: choose the number for hundreds place from digits left from experiments 1 and 2: 3 possibilities.

Together, we have $2 \times 4 \times 3 = 24$ possibilities.

L–example 1.7

How many **even three-digit numbers** can be formed from the digits 1,2,5,6, and 9 if **no restriction** on how many times a digit is used?

Solution: Similar to the last L–Example:

experiment 1: choose the number for ones place; digits 2 and 6 can be used, so there are 2 possibilities.

experiment 2: choose the number for tens place from all digits provided: 5 possibilities.

experiment 3: choose the number for hundreds place from all digits provided: 5 possibilities.

Together, we have $2 \times 5 \times 5 = 50$ possibilities.

L–example 1.8

In how many ways can 4 boys and 5 girls sit in a row if the boys and girls must alternate?

Solution: We must have the arrangement: G B G B G B G B G
The number of ways:

$$5(4)(4)(3)(3)(2)(2)(1)(1) = 5!4! = 2880,$$

where $n! = n(n-1)(n-2)\cdots(2)(1)$.

Question: What happens if there are 5 boys and 5 girls?

this will create 2 disjoint events

GBGBGBGBGB or BGBGBGBGBG

this is unlike the previous example where got more girls, so the girls must start the "queue" first

Additional Principle

Suppose that an experiment can be performed by k different procedures.

> Procedure 1 can be carried out in n_1 ways.
>
> Procedure 2 can be carried out in n_2 ways.
>
> $\dots \dots \dots \dots \dots$
>
> Procedure k can be carried out in n_k ways.

Suppose that the "ways" under different procedures are **not overlapped**. Then the total number of ways that we can perform the experiment is

$$n_1 + n_2 + \ldots + n_k.$$

EXAMPLE 2

We can take MRT or bus from home to Orchard road. If there are three bus routes and two MRT routes. How many ways we can go from home to Orchard road?

Solution: Consider that we go from home to Orchard road as an experiment. Two procedures can used to complete the experiment:

> Procedure 1: take MRT: 2 ways.
>
> Procedure 2: take bus: 3 ways.

The ways are not overlapped. So the total number of ways that we can go from home to Orchard road is $2 + 3 = 5$.

L–example 1.9 How many **even three-digit numbers** can be formed from the digits 0, 1, 2, 5, 6, and 9 if each digit can be used **at most once**?

Solution: We consider the whole task as two procedures based on ones place:

> Procedure 1: 0 is used for the ones place. There are $5 \times 4 = 20$ ways to arrange the hundreds and tens place.
>
> Procedure 2: 0 is not used for the ones place. (1) There are two ways (2 or 6) to fill in ones place; (2) as 0 can not be put at the hundreds place, we have 4 digits available for the hundreds place; (3) finally, we have 4 possible choices for the tens place. In summary, we have $2 \times 4 \times 4 = 32$ ways.

With the addition rule, we combine Procedures 1 and 2 to conclude that there are $20 + 32 = 52$ ways.

L–example 1.10 Consider the digits 0,1,2,3,4,5, and 6. If each digit can be used **at most once**, how many 3-digit numbers, which are greater than 420, can be formed?

Solution: consider three procedures:

> Procedure 1: the hundreds place is 4 and the tens place is 2: $(1)(1)(4) = 4$ ways.

Procedure 2: the hundreds place is 4 and the tens place is 3,5,or 6: $(1)(3)(5) = 15$ ways.

Procedure 3: the hundreds place is 5 or 6: $(2)(6)(5) = 60$ ways

In total, $4 + 15 + 60 = 79$ ways.

Permutation

- A **permutation** is a selection and arrangement of r objects out of n.

  In this case, **order** is taken into consideration.

- The number of ways to choose and arrange r objects out of n ($r \leq n$) is denoted by P_r^n, which has the value:

$$P_r^n = \frac{n!}{(n-r)!} = n(n-1)(n-2)\dots(n-(r-1)).$$

ob1	ob2	ob3	$\cdots$	ob$r$
n ways	$(n-1)$ ways	$(n-2)$ ways	$\cdots$	$(n-(r-1))$ ways

- A special case: when $r = n$, $P_n^n = n!$. Practically, it is the number of ways to arrange n objects in order.

  EXAMPLE 3
  Find the number of all possible four-letter code words in which all letters are different.

  Solution: $n = 26$, $r = 4$. So the number of all possible four-letter code words is
  $$P_4^{26} = (26)(25)(24)(23) = 358800.$$

L–example 1.11

(a) How many ways can 6 persons line up to get on a bus?

(b) If certain 3 persons insisting on following each other, how many ways can these 6 persons line up?

(c) If 2 persons refuse to follow each other, how many ways of lining up are possible?

Solution:

(a) $n = r = 6$, so $P_6^6 = 6! = 720$ ways.

(b) Let a, b, c, d, e, f be the names of 6 persons.

- Without loss of generality, assume that a, b, c insist on following each other.
- Group them into one group, denoted by $A = \{a, b, c\}$, which can now be viewed as one single person.
- We need to line up four persons, i.e., A, d, e, f, in a row. So $P_4^4 = 4! = 24$ ways.

On the other hand, for each permutation above, such as (A, d, f, e), a, b, c within A can be ordered differently. The number of ways of ordering them within A is $P_3^3 = 3! = 6$.

Therefore, applying the multiplication rule, the number of ways to line them up is $24 \times 6 = 144$.

(c) With the same principle as Part (b), we can first count the number of ways of lining up if they two persons are following each other:

$$P_5^5 \times P_2^2 = 5! \times 2! = 240.$$

From Part (a), the total number of ways for lining up 6 persons is 720. Therefore, we have $720 - 240 = 480$ ways of lining up 6 persons such that two given persons are not following each other.

Combination

- A combination is a selection of r objects out of n, **without regard to the order**.
- The number of combinations of choosing r objects out of n, denoted by C_r^n or $\binom{n}{r}$, is given by

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

- Note that this formula immediately implies $\binom{n}{r} = \binom{n}{n-r}$.
- The derivation is as follows:
  - Based on permutation, **the number of ways to choose and arrange** r **objects out** n is P_r^n.
  - On the other hand, this permutation task can be achieved by conducting two experiments sequentially:

  Exp. 1 get a combination, i.e., select r objects out n without regard to the order; there are $\binom{n}{r}$ ways.

Exp. 2 for each combination, get a permutation on its r objects; there are P_r^r ways.

Therefore, by multiplication rule, **the number of ways to choose and arrange r objects out n** is $\binom{n}{r} \times P_r^r$.

– As a consequent, $\binom{n}{r} \times P_r^r = P_r^n$, and we obtain

$$\binom{n}{r} = \frac{P_r^n}{P_r^r} = \frac{n!/(n-r)!}{r!} = \frac{n!}{r!(n-r)!}$$

EXAMPLE 4

From 4 women and 3 men, find the number of committees of 3 that can be formed with 2 women and 1 man.

Solution:

- The number of ways to select 2 women from 4 is $\binom{4}{2} = 6$;
- The number of ways to select 1 man from 3 is $\binom{3}{1} = 3$;
- By the multiplication rule, the number of committees formed with 2 women and 1 man is $\binom{4}{2} \times \binom{3}{1} = 6 \times 3 = 18$.

L–example 1.12 From a group of 4 men and 5 women, how many committees of size 3 are possible

(a) with no restriction?

(b) with 2 men and 1 woman if a certain man must be on the committee?

(c) with 2 men and 1 woman if 2 of the men are feuding and refuse to serve on the committee together?

Solution:

(a) The number of committee is $\binom{9}{3} = 9!/(3!6!) = 84$.

(b) Since a certain man must be on the committee, we only need to choose one man from the remaining 3 men and 1 woman from 5 woman: $\binom{1}{1}\binom{3}{1}\binom{5}{1} = 15$.

(c) The number of committees such that both "particular" men serve is $\binom{2}{2}\binom{5}{1} = 5$. As this includes all the "undesirable" cases, the number of "desirable" cases are $\binom{4}{2}\binom{5}{1} - 5 = 30 - 5 = 25$.

L–example 1.13

Shortly after being put into service, some buses manufactured by a certain company have developed cracks on the underside of the main frame.
Suppose a particular bus company has 20 of these buses, and the cracks have actually appeared in 8 of them.

(a) How many ways are there to select a sample of 5 buses from the 20 for a thorough inspection?

(b) In how many ways can a sample of 5 buses contain exactly 4 buses with visible cracks?

Solution:

(a) It is the number of ways for selecting 5 buses out of 20: $\binom{20}{5} = 20!/(5!15!) = 15504$.

(b) Now 4 buses need to be selected from the 8 buses with visible cracks: $\binom{8}{4} = 70$; then 1 bus need to be selected from the remaining 12 buses: $\binom{12}{1} = 12$. Applying the multiplication rule, the number of ways that satisfy the condition is $70 \times 12 = 840$.

4 Probability

- Intuitively, "probability" is understood as the chance or how likely a certain "event" may occur.
- More specifically, let A be an event in an experiment. We typically associate a number, called "probability", to quantify how likely the event A occurs. We denote "$P(A)$".
- But..., how could we obtain such a number?
- Even more, as a fundamental concept, it has been extended from the intuition to more rigorous, abstract, and advanced mathematical theory and has wide applications in scientific disciplines.

Interpretation of Probability by Relative Frequency

- Suppose that we repeat an experiment E for n times.
- Let n_A be the number of times that the event A occurs.
- Then $f_A = n_A/n$ is called the "relative frequency" of event A in the n repetition of E.
- Clearly, f_A may not equal to $P(A)$ exactly. But when n grows large, we may expect that f_A may "approach" it; in a sense $f_A \approx P(A)$. Or more mathematically,
$$f_A \to P(A), \quad \text{as } n \to \infty.$$
- Therefore f_A "mimics" $P(A)$, and it has the following properties:

  (1) $0 \le f_A \le 1$;

  (2) $f_A = 1$ if A occurs in every repetition.

  (3) If A and B are mutually exclusive events, $f_{A\cup B} = f_A + f_B$.
- Extending this, we define **probability on a sample space** mathematically.

Axioms of Probability

Probability, denoted by $P(\cdot)$, is a **function** on the collection of events of the sample space S, satisfying:

(i) For any event A,
$$0 \le P(A) \le 1.$$

(ii) For the sample space,
$$P(S) = 1.$$

(iii) For any two mutually exclusive events A and B, i.e., $A \cap B = \emptyset$,
$$P(A \cup B) = P(A) + P(B).$$

EXAMPLE 1
Let H denote the event of getting head when tossing a coin. Find $P(H)$, if

(a) the coin is fair;

(b) the coin is biased and a head is twice as likely to appear as a tail.

Solution:

(a) The sample space is $S = \{H, T\}$.

- "Fair" means $P(H) = P(T)$.
- The events $\{H\}$ and $\{T\}$ are mutually exclusive.
- Based on Axioms 2 and 3, we have

$$1 = P(S) = P(\{H\} \cup \{T\}) = P(H) + P(T) = 2P(H),$$

which implies $P(H) = 1/2$.

(b) "A head is twice likely to appear as a tail" means $P(H) = 2P(T)$; therefore

$$1 = P(S) = P(\{H\} \cup \{T\}) = P(H) + P(T) = 3P(T),$$

which leads to $P(T) = 1/3$ and $P(H) = 2/3$.

L–example 1.14 A fair die is tossed. Let

$$A = \{\text{an even number turns up}\}$$

$$B = \{\text{1 or 3 turns up}\}$$

$$C = \{\text{a number divisible by 3 is obtained}\}$$

Find $P(A), P(B), P(C), P(A \cup B)$, and $P(A \cup C)$.
Solution: $A = \{2, 4, 6\}$, $B = \{1, 3\}$, $C = \{3, 6\}$. We have

- $P(A) = 3/6 = 1/2$; $P(B) = P(C) = 1/3$.
- Since $A \cap B = \emptyset$, based on Axiom 3, we have $P(A \cup B) = P(A) + P(B) = 5/6$.
- But $A \cap C = \{6\} \neq \emptyset$; Axiom 3 is not applicable. Instead, $A \cup C = \{2, 3, 4, 6\}$ leads to $P(A \cup C) = 4/6 = 2/3$.

Basic Properties of Probability

Using the axioms, we can derive the following propositions.

Proposition 2
The probability of the empty set is $P(\emptyset) = 0$.

Proof Since $\emptyset \cap \emptyset = \emptyset$ and $\emptyset = \emptyset \cup \emptyset$, applying Axiom 3 leads to

$$P(\emptyset) = P(\emptyset \cup \emptyset) = P(\emptyset) + P(\emptyset) = 2P(\emptyset),$$

which implies $P(\emptyset) = 0$.

PROPOSITION 3
If $A_1, A_2, \ldots, A_n$ are mutually exclusive events ($A_i \cap A_j = \emptyset$ for any $i \neq j$), then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n).$$

Proof This proposition is immediately observed based on the induction.

PROPOSITION 4
For any event A, we have

$$P(A') = 1 - P(A).$$

Proof Since $S = A \cup A'$ and $A \cap A' = \emptyset$, based on Axioms 2 and 3, we have

$$1 = P(S) = P(A \cup A') = P(A) + P(A').$$

The result follows.

PROPOSITION 5
For any two events A and B,

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

Proof Based on property (8) of event operations, i.e., $A = (A \cap B) \cup (A \cap B')$, and $(A \cap B) \cap (A \cap B') = \emptyset$, we have

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

PROPOSITION 6
For any two events A and B,

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

Proof Based on property (7) of event operations, i.e., $A \cup B = B \cup (A \cap B')$, and $B \cap (A \cap B') = \emptyset$; and proposition 5 that $P(A \cap B') = P(A) - P(A \cap B)$, we have

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

PROPOSITION 7
If $A \subset B$, then $P(A) \leq P(B)$.

Proof Since $A \subset B$, we have $A \cup B = B$; based on property (7) of the event operations, i.e, $A \cup B = A \cup (B \cap A')$; and based on $A \cap (B \cap A') = A \cap B \cap A' = \emptyset$, we have

$$P(B) = P(A \cup B) = P(A \cup (B \cap A')) = P(A) + P(B \cap A') \geq P(A).$$

EXAMPLE 8

- A retail establishment accepts either the American Express or the VISA credit card.
- A total of 24% of it customer carry an American Express card, 61% carry a VISA card, and 11% carry both.
- What is the probability that a customer carries a credit card that the establishment will accept?

Solution:

- Let $A = \{$the customer carries an American Express Card$\}$; $V = \{$the customer carries an VISA Card$\}$.
- Then $P(A) = 0.24$; $P(V) = 0.61$; and $P(A \cap V) = 0.11$.
- The question is asking $P(A \cup V)$, and

$$P(A \cup V) = P(A) + P(V) - P(A \cap V) = 0.24 + 0.61 - 0.11 = 0.74.$$

L–example 1.15 [Hall Pageant]

Audrey is taking part in her hall's pageant. The probability that she will **win the crown** is 0.14; the probability that she will **win Miss Photogenic** is 0.3; the probability that she will **win both** is 0.11.

(a) What is the probability that she wins at least one of two?

(b) What is the probability that she wins only one of two?

Solution: Let $A = \{$win the crown$\}$ and $B = \{$win Miss Photogenic$\}$.

(a) The probability that she wins at least one of the two titles

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.14 + 0.3 - 0.11 = 0.33.$$

(b) The event that she wins the crown but not the Miss Photogenic is $A \cap B'$; based on Proposition 5 $P(A \cap B') = P(A) - P(A \cap B) = 0.14 - 0.11 = 0.03$. Similarly the event that she wins the Miss Photogenic but not the crown is $B \cap A'$, and $P(B \cap A') = 0.3 - 0.11 = 0.19$. We have

$$P((A \cap B') \cup (B \cap A')) = 0.03 + 0.19 = 0.22,$$

since $(A \cap B') \cap (B \cap A') = \emptyset$.

Finite Sample Space with Equally Likely Outcomes

- Consider a sample space $S = \{a_1, a_2, \ldots, a_k\}$.
- Assume that all outcomes in the sample space are **equally likely** to occur, i.e.,
$$P(a_1) = P(a_2) = \ldots = P(a_k).$$
- Then for any event $A \subset S$,
$$P(A) = \frac{\text{Number of sample points in } A}{\text{Number of sample points in } S}.$$

EXAMPLE 9
- A box contains 50 bolts and 150 nuts.
- Half of the bolts and half of the nuts are rusted.
- If one item is chosen at random, what is the probability that it is rusted or is a bolt?

Solution:
- Let $A = \{\text{the item is rusted}\}$, $B = \{\text{the item is a bolt}\}$, $S = \{\text{all the items}\}$.
- Since the item is selected at random, the elements in S are equally likely. S contains 200 elements. A contain $25 + 75 = 100$ elements, B contains 50, and $A \cap B$ contains 25.
- $P(A) = 100/200$, $P(B) = 50/200$, $P(A \cap B) = 25/200$;
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 5/8.$$

L–example 1.16 [Birthday Problem]

- Here's a useful party trick: walk into a room or bar with at least 50 people.
- Boldly claim that you sense two people sharing the same birthday. Act awesome afterwards.

- How often are you right?

Solution: We can cast this as a probability question:

> There are n people in a room, what is the probability that there are at least two people with the same birthday?

Some assumptions:

- Each day is **equally likely** to be a birthday of everyone.
- there is no leap year.

We can then have the following:

- The sample space is

  $$S = \{\text{all possible combinations of birthdays of } n \text{ people}\}.$$

  It is formed of equally likely sample points.

- Let

  $$A = \{\text{at least two people share the same birthday}\},$$

  then

  $$A' = \{\text{all people have different birthdays}\}.$$

- We count the number of sample points in S and A':
  - We call n people as $p_1, p_2, \cdots, p_n$, and consider the number of choices of birthdays for them one by one.
  - If all people have different birthdays,
    * for p_1, it has 365 possibilities as his/her birthday;
    * for p_2, $365 - 1$;
    * $\cdots\cdots\cdots$;
    * for p_n, $365 - (n-1)$.

    As a consequence, $\#(A') = 365(364)\cdots[365-(n-1)]$.
  - Similarly $\#S = 365^n$.

- Therefore

  $$P(A') = \frac{\#(A')}{\#S} = \frac{365(364)\cdots[365-(n-1)]}{365^n},$$

  and hence

  $$P(A) = 1 - P(A') = 1 - \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\cdots\left(1 - \frac{n-1}{365}\right)$$

Let $q_n = P(A')$ when there are n people, and

$$p_n = P(A) = 1 - q_n.$$

The values of p_n and q_n for selected values of n are tabulated:

n	q_n	p_n
1	1	0
2	0.99726	0.00274
3	0.99180	0.00820
10	0.88305	0.11695
15	0.74710	0.25290
20	0.58856	0.41144
21	0.55631	0.44369
22	0.52430	0.47570
23	0.49270	0.50730
30	0.29368	0.70632
40	0.10877	0.89123
50	0.029626	0.979374
100	3.0725×10^{-7}	≈ 1
253	6.9854×10^{-53}	≈ 1

- For 50 people, 98% of the time you can find at least two people with the same birthday.
- The probability of having at least two people sharing the same birthday exceeds 1/2 once you have 23 people.
- When there are 100 people, almost for sure you can find two people sharing the same birthday!

L–example 1.17 [Inverse Birthday Problem]

How large does a group of (randomly selected) people have to be such that the probability that someone shares his or her birthday **with you** is larger that 0.5?

Solution: The probability that n persons all have different birthdays from you is $\left(\frac{364}{365}\right)^n$.

So we need n such that $1 - (364/365)^n \geq 0.5$. Solving it, we obtain

$$n \geq \frac{\log(0.5)}{\log(364/365)} = 252.7.$$

We need at least 253 people (excluding youself).

REMARK (BIRTHDAY PROBLEMS):

Why there is a big difference in the answers between the two birthday problems?

- The inverse birthday problem requires the sharing of a **particular** day as the common birthday;
- The birthday problem allows that **any** day is the shared birthday. ■

5 Conditional Probability

- Sometimes, we need to compute the probability of some events when some **partial information** is available.
- Specifically, we might need to compute the probability of an event B, given that we have the information "an event A has occurred".
- Mathematically, we denote

$$P(B|A)$$

as the **conditional probability** of the event B, given that event A has occurred.

Definition 1 (Conditional Probability)
For any two events A and B with $P(A) > 0$, the ***conditional probability*** *of B given that A has occurred is defined by*

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

Example 2
A fair die is rolled twice.

(a) What is the probability that the sum of the 2 rolls is even?

(b) Given that the first roll is a 5, what is the (conditional) probability that the sum of the 2 rolls is even?

Solution:
We define the following events:

$$\begin{aligned} B &= \{\text{the sum of the 2 rolls is even}\} \\ A &= \{\text{the first roll is a 5}\} \end{aligned}$$

(a) The sample space is given by

		2nd roll					
		1	2	3	4	5	6
1st roll	1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

It is easy to see that $P(B) = 18/36$.

(b) Since we know that A has already happened, we can just look at the fifth row:

		2nd roll					
		1	2	3	4	5	6
1st roll	5	$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$	$(5,6)$

We are interested to look instances among this row that gives an even sum. So $P(B|A) = 3/6$.

Alternatively, we can use the formula:

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{\frac{3}{36}}{\frac{6}{36}}.$$

REMARK (REDUCED SAMPLE SPACE):

- $P(B|A)$ can also be read as:

  "the conditional probability that B occurs given that A has occurred."
  ■

- Since we know that A has occurred, regard A as our new, or **reduced sample space**.

- The conditional probability that the event B given A will equal the probability of $A \cap B$ relative to the probability of A.

L–example 1.18

- Suppose two fair dice are rolled.

- Given that the first die is less than 3, what is the probability that the sum of the 2 dice is more than 7?

<u>Solution</u>: Define the events:

$$\begin{aligned} B &= \{\text{the sum of the 2 dice is more than 7}\} \\ A &= \{\text{the first die is less than 3}\} \end{aligned}$$

Consider the reduced sample space, i.e., event A, with the following 12 equally likely sample points:

$$\{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(2,6)\}$$

The required probability is $P(B|A) = 1/12$ since there is only one point $(2,6)$ in the reduced sample space that gives a sum more than 7.

Multiplication Rule

Rearranging the definition of the conditional probability, we have

$$\begin{aligned} & P(A \cap B) &=& \; P(A)P(B|A), \quad \text{if } P(A) \neq 0 \\ \text{or} \quad & P(A \cap B) &=& \; P(B)P(A|B), \quad \text{if } P(B) \neq 0. \end{aligned}$$

This together with the definition of the conditional probability, we have the inverse probability formula:

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

EXAMPLE 3

Deal 2 cards from a regular playing deck without replacement. What is the probability that both cards are aces?

Solution:

$$\begin{aligned} P(\text{both aces}) &= P(\text{1st card is ace and 2nd card is ace}) \\ &= P(\text{1st card ace}) \cdot P(\text{2nd card ace}|\text{1st card ace}) \\ &= \frac{4}{52} \cdot \frac{3}{51} = \frac{1}{221}. \end{aligned}$$

L–example 1.19

Suppose that among 12 shirts, 3 are white. Two shirts are chosen randomly one by one **without replacement**

(a) What is the probability that **both shirts** that being picked are white?

(b) What is the probability that there is **only one** white shirt being picked?

Solution: Define the events:

$$\begin{aligned} A_1 &= \{\text{the first shirt is white}\} \\ A_2 &= \{\text{the second shirt is white}\}. \end{aligned}$$

(a) We have $P(A_1) = 3/12$; given that the first shirt is white, then there are 2 white shirts among the remaining 11 shirts, therefore

$$P(A_2|A_1) = 2/11.$$

$P(A_1 \cap A_2) = P(A_1)P(A_2|A_1) = (3/12)(2/11) = 1/22.$

(b) we need to compute the probability for $(A_1 \cap A_2') \cup (A_1' \cap A_2)$. We note that $(A_1 \cap A_2') \cap (A_1' \cap A_2) = \emptyset$.

$$P((A_1 \cap A_2') \cup (A_1' \cap A_2)) = P(A_1 \cap A_2') + P(A_1' \cap A_2).$$

On the other hand, with similar argument as Part (a), we have

$$\begin{aligned} P(A_1 \cap A_2') &= P(A_1)P(A_2'|A_1) = (3/12) \cdot (9/11) \\ P(A_1' \cap A_2) &= P(A_1')P(A_2|A_1') = (9/12) \cdot (3/11). \end{aligned}$$

As a consequence

$$P((A_1 \cap A_2') \cup (A_1' \cap A_2)) = (3/12) \cdot (9/11) + (9/12) \cdot (3/11) = 9/22.$$

6 Independence

Independence is one of the most important concepts in probability.

Definition 1 (Independence)
Two events A and B are ***independent*** *if and only if*

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

We denote it as $A \perp B$.

If A and B are not independent, they are **dependent**, denoted by $A \not\perp B$.

Remark:

- If $P(A) \neq 0$, $A \perp B$ if and only if $P(B|A) = P(B)$.
- This is observed from the definition of conditional probability:

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

- Intuitively, this is stating: A and B independent if the knowledge of A does not change the probability of B.
- Likewise, if $P(B) \neq 0$, $A \perp B$ if and only if $P(A|B) = P(A)$.

Example 2
Suppose we roll 2 fair dice.

to check dependence must use the mathematical definition

(a) Let

$$\begin{aligned} A_6 &= \{\text{the sum of two dice is } 6\} \\ B &= \{\text{the first die equals } 4\}. \end{aligned}$$

Hence, $P(A_6) = 5/36$, $P(B) = 6/36 = 1/6$ and $P(A_6 \cap B) = 1/36$. Since

$$P(A_6 \cap B) \neq P(A_6)P(B),$$

we say that A_6 and B are **dependent**.

(b) Let $A_7 = \{\text{the sum of two dice is } 7\}$. Then $P(A_7 \cap B) = 1/36$, $P(A_7) = 1/6$ and $P(B) = 1/6$. Hence

$$P(A_7 \cap B) = P(A_7)P(B),$$

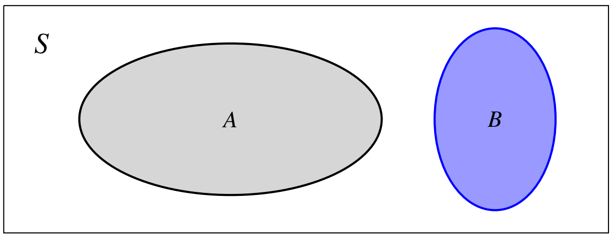
and we say that A_7 and B are **independent**.

REMARK (INDEPENDENT VS MUTUALLY EXCLUSIVE):

Independent and **mutually exclusive** are totally different concepts:

$$\begin{aligned} A, B \text{ independent } &\Leftrightarrow P(A \cap B) = P(A)P(B) \\ A, B \text{ mutually exclusive } &\Leftrightarrow A \cap B = \emptyset \end{aligned}$$

"Mutually exclusive" can be illustrated by the Venn diagram, but "independent" can not.



■

L–example 1.20 [Some properties of independence]

Decide whether the following statements are TRUE or FALSE.

(a) Suppose $P(A) > 0$, $P(B) > 0$, then if $A \perp B$, then A and B are not mutually exclusive.

(b) Suppose $P(A) > 0$, $P(B) > 0$, then if A and B are mutually exclusive, then $A \not\perp B$.

(c) S and $\emptyset$ are independent of any event.

(d) If $A \perp B$, then $A \perp B'$, $A' \perp B$, and $A' \perp B'$.

Solution:

(a) TRUE; based on independence $P(A \cap B) = P(A)P(B) > 0$.

(b) TRUE; based on mutually exclusive $P(A \cap B) = 0 \neq P(A)P(B)$.

(c) TRUE; for any event A; $P(A \cap S) = P(A) = P(A)P(S)$; $P(A \cap \emptyset) = P(\emptyset) = 0 = P(A)P(\emptyset)$.

(d) TRUE; we derive one only. Note that $A = (A \cap B) \cup (A \cap B')$, we have

$$\begin{aligned} P(A \cap B') &= P(A) - P(A \cap B) = P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) = P(A)P(B'). \end{aligned}$$

L–example 1.21

- The probability that Tom will be alive in 20 years is 0.7.
- The probability that Jack will be alive in 20 years is 0.9.
- What is the probability that neither will be alive in 20 years?
- Define

$$\begin{aligned} A &= \{\text{Tom would be alive in 20 years.}\} \\ B &= \{\text{Jack would be alive in 20 years.}\} \end{aligned}$$

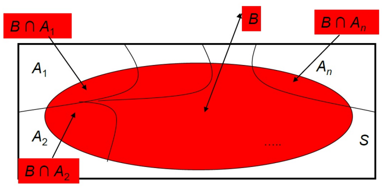
- A and B are independent, as whether one is alive would not affect the other.
- A' and B' are also independent.
- The desired probability is given by

$$P(A' \cap B') = P(A')P(B') = (1 - 0.7)(1 - 0.9) = 0.03.$$

7 The Law of Total Probability

Definition 1 (Partition)
If $A_1, A_2, \ldots, A_n$ are mutually exclusive events and $\cup_{i=1}^{n} A_i = S$, we call $A_1, A_2, \ldots, A_n$ a partition of S.

in this case here it becomes a partition for B as well



$B = (A_1 \cap B) \cup (A_2 \cap B) \ldots$
$\ldots \cup (A_n \cap B)$

$\emptyset = (A_i \cap B) \cup (A_j \cap B)$
$i \neq j$

Theorem 2 (The law of total probability)

If $A_1, A_2, \ldots, A_n$ is a partition of S, then for any event B, we have

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

A special case: For any events A and B, we have

$$P(B) = P(A)P(B|A) + P(A')P(B|A').$$

Example 3 (Frying Fish)

- At a nasi lemak stall, the chef and his assistant take turns to fry fish.
- The chef burns his fish with probability 0.1, his assistant burns his fish with probability 0.23.
- If the chef is frying fish 80% of the time, what is the probability that the fish you order is burnt?

Solution:

- Let

$$\begin{aligned} B &= \{\text{the fish is burnt}\} \\ C &= \{\text{the fish is fried by the chef}\}; \end{aligned}$$

  we then need to compute $P(B)$.
- Use the law of total probability

$$P(B) = P(C)P(B|C) + P(C')P(B|C') = 0.8 \times 0.1 + 0.2 \times 0.23.$$

L–example 1.22 [The Monty Hall Problem]

- Suppose you are on a game show, and you are given the choice of three doors: behind one door is a car; behind the others, goats.
- You pick a door, say No. 1, and the host Monty, who knows what is behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?"
- Is it to your advantage to switch your choice?

Solution: Let's formulate it as a probability question. Denote the events:

$$\begin{aligned} W &= \{\text{Win the car}\} \\ A &= \{\text{car is behind the door of the initial pick}\} \end{aligned}$$

Then, our interest is $P(W)$, i.e., the probability of wining the car.

Applying the law of total probability, we have

$$\begin{aligned} P(W) &= P(A)P(W|A) + P(A')P(W|A') \\ &= \frac{1}{3}P(W|A) + \frac{2}{3}P(W|A'). \end{aligned}$$

- If a "stick" strategy is used, i.e., not to switch your choice,
  - $P(W|A) = 1$; that is, you are sure to win the car if the initial choice is the car door;
  - $P(W|A') = 0$; that is, you are sure to lose the car if the initial choice is not the car door.

  As a consequence $P(W) = (1/3) \cdot 1 + (2/3) \cdot 0 = 1/3$.
- If a "switch" strategy is used, i.e., switch to another door when asked, then $P(W|A) = 0$ and $P(W|A') = 1$. We then have $P(W) = (1/3) \cdot 0 + (2/3) \cdot 1 = 2/3$.

Conclusion: "Switch" double the chance of wining the car!

REMARK (MONTY HALL):

Still confused? Watch the following videos:

http://www.youtube.com/watch?v=mhlc7peGlGg

http://www.youtube.com/watch?v=P9WFKmLK0dc ■

THEOREM 1 (BAYES' THEOREM)
Let $A_1, A_2, \ldots, A_n$ be a partition of S, then for any event B and $k = 1, 2, \ldots, n$,

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

- A special case of Bayes' theorem: take $n = 2$. A and A' become a partition of S. We have

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')}.$$

- This formula is practically meaningful. For example:
  - A = disease status of a person;
  - B = symptom observed;
  - $P(A)$: the probability of a disease in general;
  - $P(B|A)$: if diseased, probability of observing symptom;
  - $P(A|B)$: if symptom observed, probability of diseased.
- Bayes' theorem can be derived based on the conditional probability, multiplication rule, and the law of the total probability.
- In particular,

$$\begin{aligned} P(A_k|B) &= \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^n P(B \cap A_i)} \\ &= \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^n P(A_i)P(B|A_i)}. \end{aligned}$$

EXAMPLE 2
- Historically we observe some newly constructed house to collapse.
- The chance that the design is faulty is 1%.
- If the design is faulty, the chance that the house is to collapse is 75%, otherwise, the chance is 0.01%.
- If we observed that a newly constructed house collapsed, what is the probability that the design is faulty?

<u>Solution</u>:

- Let

$$\begin{aligned} B &= \{\text{The design is faulty}\}, \\ A &= \{\text{The house collapses}\}. \end{aligned}$$

- We have $P(B) = 0.01$, $P(A|B) = 0.75$, and $P(A|B') = 0.0001$.
- The question is asking to compute $P(B|A)$.
- We compute it based on Bayes' theorem. The denominator can be computed based on the law of total probability:

$$\begin{aligned} P(A) &= P(B)P(A|B) + P(B')P(A|B') \\ &= (0.01)(0.75) + (0.99)(0.00001) = 0.007599. \end{aligned}$$

- The numerator is

$$P(A|B)P(B) = 0.75(0.01) = 0.0075.$$

- As a consequence $P(B|A) = P(A|B)P(B)/P(A) = 0.9870$.

L–example 1.23

- An insurance company believes that people can be divided into two classes: **accident prone** and **not accident prone**.
- Historically, they observe that the probability that an accident-prone person will have an accident within a fixed 1-year period is 0.04; otherwise, the probability is 0.02.
- Assume that 30% of the population is accident prone.

(a) What is the probability that a new policyholder will have an accident within a year of purchasing a policy?

(b) If a new policyholder has an accident within a year of purchasing a policy, what is the probability that he or she is accident prone?

Solution:

- Define the events:

$$\begin{aligned} B &= \{\text{a new policy holder has an accident within a year}\} \\ A &= \{\text{a new policy holder is accident prone}\}. \end{aligned}$$

- Based on the question, $P(A) = 0.3$; $P(B|A) = 0.04$; $P(B|A') = 0.02$.

(a) $P(B) = P(A)P(B|A) + P(A')P(B|A') = 0.3(0.04) + 0.7(0.02) = 0.26$.

(b) $P(A|B) = P(A)P(B|A)/P(B) = 0.3(0.04)/0.26 = 6/13$.