# Introduction

**Probability** is the branch of mathematics concerning numerical descriptions of how likely an event is to occur.

Procedure used in probability theory: logic of randomness / rules.

- Statistical experiment is considered (rolling a dice, tossing a coin).
- Each possible outcome of the experiment is assigned a probability.
- Practically relevant quantities (probabilities of events, expected value, variance) are computed.
- Probabilistic models are developed.

e.g fair coin (ideal)

| H | T |
|---|---|
| $\frac{1}{2}$ | $\frac{1}{2}$ |

data: HTHHT → is the coin fair. (model)

Goals of **statistics**:

- Understand patterns in data
- Testing of hypotheses
- Reliable predictions → confidence level (interval).

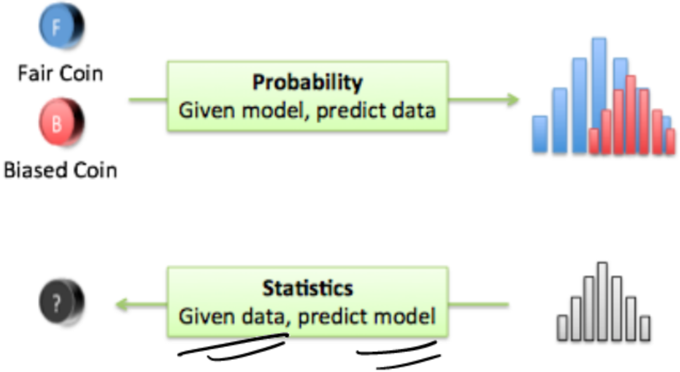
Required:

- Data sampling
- Computation of probabilistic quantities

Statistics = Mathematical Analysis of Data

sample data → inference about population.

# Probability & Statistics



# Methods of Enumeration / Counting

To compute probabilities, we often need to determine the total number of outcomes of an experiment or procedure.

Usually, can find the total number of outcomes by standard counting techniques involving

- multiplication principle
- *n*-tuples
- permutations, *k*-permutations
- combinations
- multisets

| | ordered | unordered. |
|---|---|---|
| repetitions. | | |
| no repetitions. | | |

# Multiplication Principle

$$\text{start} \xrightarrow[n_1]{1} \xrightarrow[n_2]{2} \cdots \xrightarrow[n_k]{k} \text{outcome.}$$

**Theorem 1 (Multiplication Principle)**

*If an experiment consists of $k$ steps and there are $n_i$ possible outcomes for step $i$, then the total number of outcomes is*

$$\prod_{i=1}^{k} n_i = n_1 \cdot n_2 \cdot \dots \cdot n_k.$$

## Example 2

An experiment consists of 3 steps, where possible outcomes are:

step 1: $A$, $B$ ✓ $n_1 = 2$

step 2: $x$, $y$ ✓ $n_2 = 2$

step 3: 1, 2, 3 ✓ $n_3 = 3$.

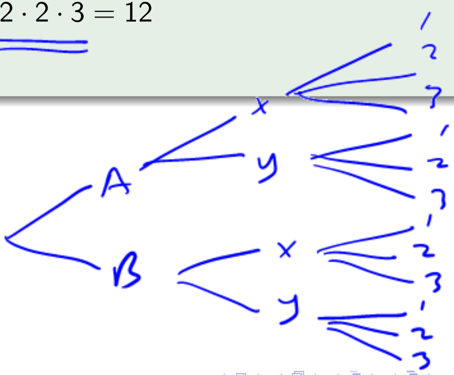
Multiplication principle $\Longrightarrow$ Total number of outcomes is

$$\underline{\underline{2 \cdot 2 \cdot 3}} = 12$$

.

All possible outcomes:

$(A, x, 1)$, $(A, x, 2)$, $(A, x, 3)$,
$(A, y, 1)$, $(A, y, 2)$, $(A, y, 3)$,
$(B, x, 1)$, $(B, x, 2)$, $(B, x, 3)$,
$(B, y, 1)$, $(B, y, 2)$, $(B, y, 3)$.



## Example 3

Bob has 5 coats, 6 shirts, and 3 pairs of shoes in his closet. In how many ways can he choose one coat, one shirt, and one pair of shoes to get dressed?

To get dressed, Bob needs to perform all 3 steps: choose a coat, choose a short and choose a pair of shoes. By the Multiplication Principle, the total number of ways is

$$5 \cdot 6 \cdot 3 = 90.$$

$\square$

5 6 3

## Example 4

Suppose a license plate number consists of 2 letters in front of 4 digits, excluding 0000. For example, a license plate number can be *AZ*0001, *BB*0102, *YC*9120 etc. How many possible licence plate numbers are there? ~~Are these enough for all cars in Singapore?~~

*Give it a try!*

$\boxed{\text{Allow } 0000}$

$\square\ \square\ \square\ \square\ \square\ \square$

26 ways 26 ways. 0, 1, 2, ..., 9 10 ways 10 ways 10 ways 10 ways.

Total number (allow 0000) $= 26 \times 26 \times 10 \times 10 \times 10 \times 10$

$= 26^2 \cdot 10^4$

Total number. $(\underline{\underline{XX}}\ 0000)$ $= 26 \times 26 \times 1 \times 1 \times 1 \times 1$

$= 26^2$ ✓

Answer $= 26^2 \times 10^4 - \underline{\underline{26^2}}$ ✓

# $n$-Tuples

Often we need to select/draw a sample of objects from a given set. The **order** in which the objects are selected/drawn may or may not be important.

If $k$ objects are selected from a set of $n$ objects, and if the order of selection is noted (i.e the order matters), then the selected set of $k$ objects is called an **ordered sample** of size $k$.

It is convenient to represent an ordered sample of size $k$ by an $k$-tuple:

$$(x_1, x_2, \ldots, x_k) \quad \text{or just} \quad x_1 x_2 \cdots x_k$$

(repetitions).

- **Sampling with replacement** occurs when an object is selected and then replaced before the next object is selected.

- **Sampling without replacement** occurs when an object is not replaced after it has been selected.

(no repetitions).

## Theorem 5

*The total number of ordered sample of size $k$ from a set of $n$ different objects, with replacement, is*

$$n^k.$$

Reason: Represent the ordered sample by $x_1 x_2 \cdots x_k$. There are $n$ choices for each $x_i$. By the Multiplication Principle, the total number of such samples is

$$\underbrace{n \cdot n \cdots n}_{k \text{ times}} = n^k.$$

n choices. n choices.

n choices.

. . .

k boxes.

## Example 6

How many ways are there to assign 3 courses to 5 professors (given that each professor can be assigned up to 3 courses)?

Professors : A, B, C, D, E.

Courses : $T_1, T_2, T_3$

assignment : $(\bigcirc, \bigcirc, \bigcirc)$

# ways $= 5 \times 5 \times 5$
$= 5^3$.

e.g. (A, A, B)
(A, A, A)

$$\underline{A} \quad B \quad C \quad \underline{D} \quad E$$

$$(\underline{T_1}, T_2, T_3, \underline{T_1}, T_2) \qquad \times$$

options: A, B, C, D.

## Example 7

How many ways are there to answer a multiple-choice test with 8 questions, where each question has 4 options and exactly one option has to be chosen?

*Give it a try!*

outcome: (A, A, B, B, C, C, D, A)

$$\text{total \# ways} = \underbrace{4 \times 4 \times \cdots \times 4}_{8 \text{ times}} = 4^8$$

# Permutations and $k$-Permutations

A **permutation** of a set with $n$ different objects is an ordered sample of size $n$ without replacement.

Example: The set $\{a, b, c\}$ has a total of 6 permutations:

$$abc, acb, bac, bca, cab, cba.$$

**Theorem 8 (Number of permutation)**

*A set with $n$ different objects has exactly $n!$ permutations.*

*Here, $n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1$ (n-factorial).*

fill in the boxes from left to right without replacement

$\square \; \square \cdot \square \cdot - \quad \square$

n choices, n−1 choices, n−2 choices, 1 choices.

A $k$-**permutation** of a set with $n$ objects is an ordered sample of size $k$, where $k \leq n$, without replacement.

$= 4 \times 3$ .

Example: The set $\{6, 7, 8, 9\}$ has exactly 12 2-permutations

$$67, 68, 69, 76, 78, 79, 86, 87, 89, 96, 97, 98.$$

4 choices 3 choices.

Note: A $n$-permutation of a set with $n$ objects is just a permutation.

## Theorem 9 (Number of $k$-permutations)

*A set with $n$ different objects has exactly*

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

*$k$-permutations.*

Reason: Represent a $k$-permutation by a $k$-tuple $x_1 x_2 \cdots x_k$. We will select $x_1$ followed by $x_2$ etc. Without replacement, the number of choices for $x_1$ is $n$, the number of choices for $x_2$ is $n-1$, ...., the number of choices for $x_k$ is $n-k+1$. By the Multiplication principle, the total number is

$(n-(k-1))$

$$n \cdot (n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}.$$

$\boxed{1}$ $\boxed{2}$ $\boxed{k}$

# Combinations

Sometimes, the order of selection is **not** important. Such a selection is called a **combination**. That is, we are interested in the number of (unordered) subsets of size $k$ from a set with $n$ different objects.

## Example 10

There are exactly 10 2-element combinations of the set $\{1, 2, 3, 4, 5\}$:

$$12, 13, 14, 15, 23, 24, 25, 34, 35, 45.$$

12

$\{1, 2\}$

$P(n,k)$.

k boxes

## Theorem 11 (Number of combinations)

*The number of k-element combinations of a set with n different objects is*

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

"n choose k"

k! ways to arrange the boxes

## Example 12

How many ways are there to select a team of 3 players from total of 10 players?

$$\binom{10}{3} = \frac{10!}{3!7!} = 120.$$

# Multisets

A **multiset** is a collection of elements where elements can occur repeatedly. If an element $a$ occurs $t$ times in a multiset, then $t$ is called the multiplicity of $a$.

Example: $\{a, a, a, b, c, c\}$ is a multiset where $a$ has multiplicity 3, $b$ has multiplicity 1, and $c$ has multiplicity 2.

A permutation of a multiset is an ordering of its elements.

Example: $\{1, 1, 2, 2\}$ has exactly 6 permutations:

$$1122, 1212, 1221, 2112, 2121, 2211.$$

$1_1, 1_2, 2_1, 2_2 \longrightarrow 4!$ permutations.

$2!$ $\quad$ $2!$

$$\frac{4!}{2!\,2!} = \frac{24}{4} = 6.$$

## Theorem 13 (Number of permutations of a multiset)

*Suppose $S$ is a multiset whose elements have multiplicities $n_1, \ldots, n_k$ where $n_1 + \cdots + n_k = n$. The number of permutations of $S$ is*

$$\binom{n}{n_1, \ldots, n_k} = \frac{n!}{n_1! \cdots n_k!}.$$

$$\frac{n!}{n_1!\, n_2! \ldots n_k!}$$

## Example 14

In how many ways can the letter of the word "abracadabra" be permuted?

The letter this word form the multiset $\{a, a, a, a, a, b, b, c, d, r, r\}$ with $n = 11$ and multiplicities

$$n_1 = 5, \; n_2 = 2, \; n_3 = 1, \; n_4 = 1, \; n_5 = 2.$$

Hence the total number of permutations is

$$\binom{11}{5, 2, 1, 1, 2} = \frac{11!}{5!2!1!1!2!} = 83160.$$

□

## Example 15

A coin is flipped 10 times and the sequence of heads and tails is observed. Find the number of possible 10-tuples that result in four heads and six tails.

Examples of such a tuple are $HHHHTTTTTT$, $THTHTHTHTT$. These are permutations of the multiset $\{H, H, H, H, T, T, T, T, T, T\}$ with multiplicities of 4 and 6. The total number of such tuples is

$$\binom{10}{4, 6} = \frac{10!}{4!6!} = 210.$$

$\square$

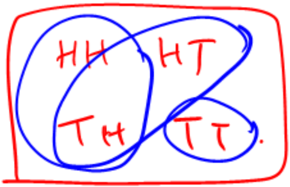
# Probability

# Sample space

Mathematically, a statistical experiment is represented by its possible outcomes.

The **sample space** of a statistical experiment is the set of all its possible outcomes. The sample space usually is denoted by $\Omega$.

toss a coin twice




$\Omega$

## Example 16

- Roll a dice one time:

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

$2^3$

- Toss a coin 3 times:

$$\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}.$$

- Measure systolic blood pressure of one patient

$$\Omega = \mathbb{R}^+ (\text{set of positive real numbers}).$$

- Count number of customers in a queue

$$\Omega = \mathbb{Z}^+ (\text{set of nonnegative integers}).$$

An **event** is a subset of the sample space $\Omega$.

**Example 17**

Toss a coin 3 times. Find the event $A$ of outcomes with at least two heads.

The event is

$$A = \{HHH, HHT, HTH, THH\}.$$

## Example 18

Toss a coin 2 times. Find all the events of this experiment.

The events are:

$$\emptyset \text{ (empty set)}$$

$$\{HH\}, \{HT\}, \{TH\}, \{TT\}$$

$$\{HH, HT\}, \{HH, TH\}, \{HH, TT\}, \{HT, TH\}, \{HT, TT\}, \{TH, TT\}$$

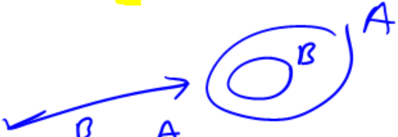
$$\{HH, HT, TH\}, \{HH, HT, TT\}, \{HH, TH, TT\}, \{HT, TH, TT\}$$

$$\{HH, HT, TH, TT\} = \Omega.$$

□

Consider a sample space $\Omega$ and let $A$ and $B$ be events.

- $|A|$: number of elements of $A$
- $\emptyset$ empty set
- $A \subseteq B$: $A$ is a subset of $B$




- $A \cup B$: **union** of $A$ and $B$


- $A \cap B$: **intersection** of $A$ and $B$


- $A$ and $B$ are **disjoint** (mutually exclusive) if $A \cap B = \emptyset$


- $\overline{A}$: complemntary event of $A$, i.e. $\overline{A} = \Omega \setminus A$.

De Morgan's Laws:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

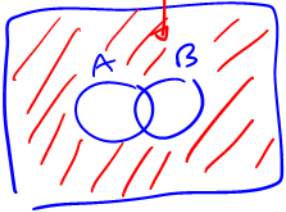
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

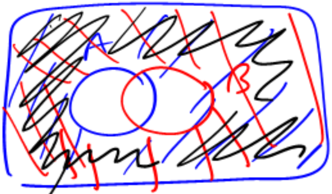
Distributive laws (for events $A$, $B$, $C$):

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

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

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$



# Definition of probability

Consider a statistical experiment with sample space $\Omega$. $\mathbb{P}$ is a **probability measure** on $\Omega$ if it satisfies the following conditions

(P1) $\mathbb{P}(A)$ is real number with $0 \leq \mathbb{P}(A) \leq 1$ for each event $A$ of $\Omega$.

(P2) $\mathbb{P}(\Omega) = 1$.

(P3) If $A_1, A_2, \ldots$ are events that are pairwise disjoint i.e. $A_i \cap A_j = \emptyset$ for all $i \neq j$. Then

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_k) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_k)$$

for each positive $k$, and

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots$$

for an infinite, but countable, number of events.

## Theorem 19 (Finite sample space with fair outcomes)

*Let $\Omega$ be a finite sample space, and $A$ be an event. Assume that all possible outcomes have the same probability, The probability of the event $A$ is given by*

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

$\Omega$

$\boxed{\underline{H} \quad \underline{T}}$

$\mathbb{P}(H) = \frac{1}{2}$

## Example 20

A dice is rolled 2 times. What is the probability that the total is at least 11? Assume that all outcomes in the sample space have the same probability $\frac{1}{36}$.

$$\Omega = \left\{ (x_1, x_2) : \begin{array}{l} x_1 \in \{1,2,3,4,5,6\} \\ x_2 \in \{1,2,3,4,5,6\} \end{array} \right\}$$

$$|\Omega| = 6 \times 6 = 36.$$

$$A = \{ (x_1, x_2) : x_1 + x_2 \geq 11 \}$$

$$= \{ (5,6), (6,5), (6,6) \}$$

$$|A| = 3. \Rightarrow \mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{3}{36} = \frac{1}{12}.$$

## Example 21

Consider a standard poker deck of 52 cards. Find the probability of drawing a poker hand of 5 cards which form a **three of a kind**? Assume all outcomes in the sample space have the same probability.

Note that $|\Omega| = \binom{52}{5}$ since each draw of 5 cards forms a 5-element combination of a set of 52 cards.

$A = \{$ three of a kind $\}$.

(unordered)

e.g

$\underbrace{1\heartsuit \quad 1\clubsuit \quad 1\diamondsuit}_{\text{same rank}} \quad \underbrace{K\diamondsuit \quad 9\heartsuit}$

rank must be different among themselves & from the rank of 3 kind.

step 1 $\to$ step 2 $\to$ step 3 $\to$ step 4.

choose rank for 3 of a kind.

choose 3 out of the 4 cards in step 1.

choose 2 ranks out of 12 ranks remaining.

choose 1 card out of 4 cards of each rank in step 3.

$$\binom{13}{1} = 13 \qquad \binom{4}{3} \qquad \binom{12}{2} \qquad \binom{4}{1}\binom{4}{1}$$

Total number of outcomes in $A$ $= |A| = \binom{13}{1}\binom{4}{3}\binom{12}{2}\binom{4}{1}\binom{4}{1}$

$$= 54{,}912.$$

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{54{,}912}{\binom{52}{5}} \approx 0.02$$

| | ordered. | unordered |
| --- | --- | --- |
| replace | $n^k$ | ? |
| no replacement. | permutations.<br>k-permutations. | $\binom{n}{k}$. |

$\cancel{\binom{n+k-1}{k}}$

$\underbrace{\square\square\square\ldots\square}_{k}$ out $n$ objects.

Wooclap Question:

For combinations we divide by k to account for k ways to permute k objects. But then for multisets we divide the ways to permute all the objects by the products of the possible permutations of (each object), can I clarify on the difference between these 2 cases

Choose $\underline{\underline{3 \text{ objects}}}$ from $\{1, 2, 3, 4, 5\}$.

$\downarrow$

Combinations: outcome

$$\# \text{ways} = \binom{5}{3}$$

$\{1,2,4\} \longleftrightarrow 1\ 1\ 0\ 1\ 0$ ✓

$\{1,3,5\} \longleftrightarrow 1\ 0\ 1\ 0\ 1$ ✓

$\{2,4,3\} \longleftrightarrow 0\ 1\ 1\ 1\ 0$ ✓

$\{\underline{\underline{0,0}}, \underline{\underline{1,1,1}}\}$.

# permutations of this multiset

$$= \frac{5!}{2!\,3!}$$