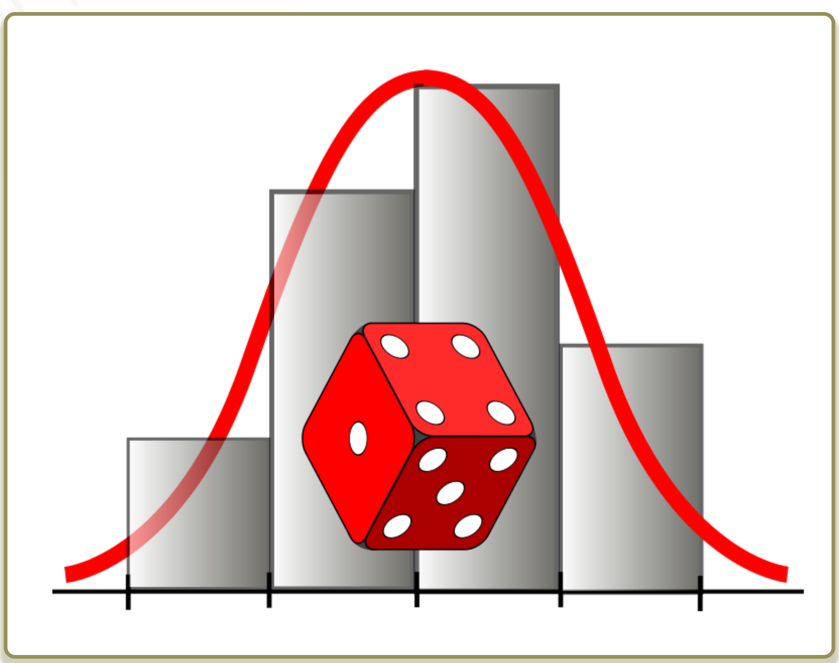
Moulay Ismail University
École Normale Supérieure (ENS–Meknès)
Department of Sciences

Academic Year : *2025–2026*
Program : *Master in Applied Mathematics and Educational Engineering (AMEE)*
Mathematical Sciences and Innovative Educational Systems (MSIES)
Semester : M_1 *(common core)*
Module Coordinator : *Prof. Youssef El Haoui*

Probabilities, Inferential Statistics, and Their Applications



Chapitre 1

Elements of Probability

1) A bit of vocabulary

From a historical point of view, the concept of probability arose from simple examples borrowed from games of chance (the word “hazard/chance” comes from the Arabic *az-zahr* : the die). We introduce this notion through a familiar example : the game of a die.

DEFINITIONS	**EXAMPLES**
A random experiment is an experiment for which the outcome cannot be predicted in advance. We know the set of all possible outcomes, but we do not know which one will occur. This contrasts with a deterministic experiment, whose result is certain.	Throwing a fair six-sided die is a random experiment. The outcome is the number shown on the upper face of the die.
Every random experiment is associated with a set called the sample space (or fundamental set of the experiment), denoted Ω, containing all possible outcomes.	$\Omega = \{1, 2, 3, 4, 5, 6\}$.
A (random) event is any subset of Ω.	The event “obtaining an even number” is $A = \{2, 4, 6\}$.
An event A is said to be realized (or to *occur*) if the outcome of the experiment belongs to A.	If the upper face shows 5, then A does not occur. If it shows 4, then A occurs.
An event containing a single outcome is called an elementary event.	$B = \{1\}$ is one of the six elementary events of Ω.

When carrying out a random experiment, the central question we ask is : **How likely is a given event to occur ?**

2) From frequency to the probability of a random event

The theoretical frequency of an event is defined as the limit of its empirical frequency when the number of repetitions of the same experiment tends to infinity (this is a formulation of the Weak Law of Large Numbers).

For example, to estimate the theoretical frequency of obtaining the outcome 3 when throwing a fair die, one may repeat the experiment a large number of times, say 10,000 throws. The theoretical frequency we are looking for, namely, the probability of the elementary event $\{3\}$, will be very close to the empirical frequency observed over these 10,000 trials, and even closer if we increase the number of repetitions to 100,000, and so on. Thus, performing the same experiment many times makes the notion of "chance" much less mysterious.

The formal definition of probability (theoretical frequency) relies on what are known as the **Kolmogorov axioms**.

Three Axioms : Kolmogorov's Axioms

1. **Nonnegativity.** The probability of any event is always nonnegative :

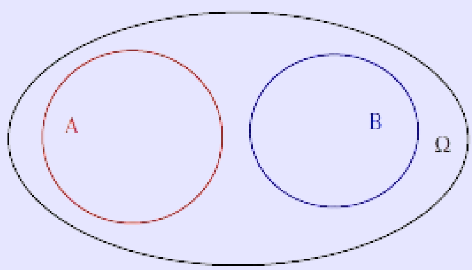
$$p(A) \geqslant 0.$$

2. **Normalization.** Since the sample space Ω represents the certain event, it must have maximal probability :

$$p(\Omega) = 1.$$

3. **(Finite) additivity.** If A and B are *mutually exclusive* (disjoint), then the probability of their union is the sum of their probabilities :

$$p(A \cup B) = p(A) + p(B).$$



I - Probability as a normalized measure

Probability theory, in its modern formulation, is built on the framework of measure theory. A probability is therefore not a new kind of mathematical object : it is simply a *measure* endowed with an additional normalization condition.

Probability as a measure

A probability measure on a measurable space $(\Omega, \mathcal{F})$ is a measure p satisfying

$$p(\Omega) = 1.$$

Here $\mathcal{F}$ denotes a collection of subsets of Ω (the *σ-algebra of events*). In the finite case treated in this chapter, one usually takes $\mathcal{F} = \mathcal{P}(\Omega)$, the set of all subsets of Ω.

In measure theory, a (general) measure μ on $(\Omega, \mathcal{F})$ is a function

$$\mu : \mathcal{F} \to [0, +\infty]$$

satisfying :

1. **Nonnegativity :**
$$\mu(A) \geqslant 0 \qquad \text{for all } A \in \mathcal{F}.$$
2. **Null empty set :**
$$\mu(\emptyset) = 0.$$
3. **Countable additivity :** If $A_1, A_2, \ldots$ are pairwise disjoint,
$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu(A_i).$$

A probability measure p satisfies these *same* three axioms. Thus mathematically :

Probability is a measure

For a general measure μ, the total mass $\mu(\Omega)$ can be :

- any nonnegative finite number, or
- $+\infty$.

A *probability* is simply a measure that assigns mass 1 to the entire sample space :

$$p(\Omega) = 1.$$

The number 1 represents the "unit of certainty" (100% of all possible outcomes). Each event receives a portion of this mass, indicating how likely it is to occur.

Probability theory is therefore a special case of measure theory :

$$\boxed{\text{Probability measure} = \text{measure } p \text{ such that } p(\Omega) = 1.}$$

This perspective is fundamental and will later allow the use of the Lebesgue integral and convergence theorems in probability.

1) Properties of probabilities of a random event

Definition 1.1

If the sample space Ω consists of n elementary events $\{e_i\}$, a *probability measure* on Ω assigns n numbers $p_i \in [0,1]$ (the probabilities of the elementary events) such that

$$\sum_{i=1}^{n} p_i = 1.$$

If an event A is the disjoint union of k elementary events $\{e_i\}$ with $0 < k < n$, then by definition

$$p(A) = p\Big(\cup_{i=1}^{k} \{e_i\} \Big) = \sum_{i=1}^{k} p(e_i) = \sum_{i=1}^{k} p_i,$$

hence $0 \leqslant p(A) \leqslant 1$.

⚠ The concrete meaning of $p(A)$ is : in a random experiment, the closer $p(A)$ is to 1 (resp. to 0), the more (resp. less) likely A is to occur.

Example 1 : **Uniform probability (equiprobability).**

Assume that all elementary outcomes of the finite sample space Ω are **equally likely**.
This means that each outcome $\omega_i \in \Omega$ has the same probability

$$p_i = \frac{1}{n},$$

where $n = \operatorname{card}\Omega$.

Let $A \subset \Omega$ be an event containing k favorable outcomes. Since the probability of each outcome is the same, the probability of A is

$$p(A) = \frac{k}{n} = \frac{\operatorname{card} A}{\operatorname{card}\Omega}.$$

Equivalently,

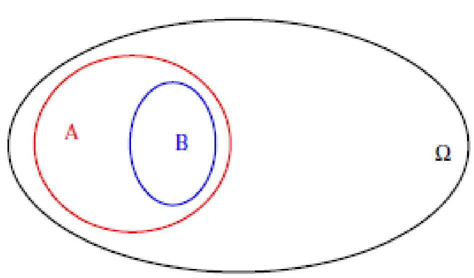
$$p(A) = \frac{\text{number of favorable outcomes}}{\text{number of possible outcomes}}.$$

Property 1.2

If B is a subset of A, then

$$B \subseteq A \;\Rightarrow\; p(B) \leqslant p(A).$$

Indeed, $A = B \cup (A \setminus B)$ with B and $A \setminus B$ disjoint, so $p(A) = p(B) + p(A \setminus B) \geqslant p(B)$.



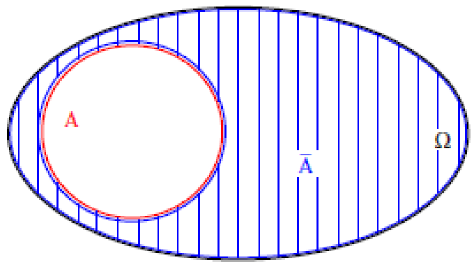
Property 1.3

The event $\emptyset$ is the impossible event ; it never occurs. Its probability is $p(\emptyset) = 0$.

Property 1.4

The *complement* of an event A in Ω is denoted by $\bar{A}$ (or A^c). Its probability is given by

$$p(\bar{A}) = 1 - p(A).$$



Remark 1.5

By Property 1.4 and the non-negativity axiom, we have for every event A :

$$\boxed{0 \leqslant p(A) \leqslant 1}.$$

Indeed,

$$p(A) = 1 - p(\bar{A}) \leqslant 1,$$

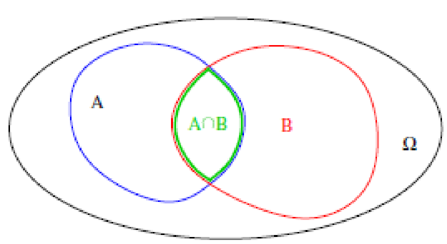
since $p\left(\bar{A}\right) \geqslant 0$.

Property 1.6

For any $A, B \subseteq \Omega$,

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

Proof:
$p(A) = p(A \setminus B) + p(A \cap B)$ and $p(B) = p(B \setminus A) + p(A \cap B)$ (disjoint unions).
Also $p(A \cup B) = p(A \setminus B) + p(B \setminus A) + p(A \cap B)$.
Adding yields $p(A \cup B) = p(A) + p(B) - p(A \cap B)$. ■

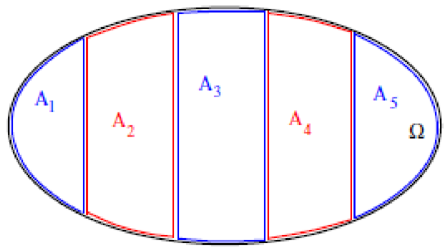


Property 1.7: Addition rule

If $A_1, \ldots, A_k$ form a *partition* of Ω (pairwise disjoint and $\Omega = \bigcup_{j=1}^{k} A_j$), then

$$\sum_{j=1}^{k} p(A_j) = 1.$$

In this case the A_j are sometimes called a *complete system of events*.



II - Joint probability

The probability that two events A and B both occur is called the joint probability of A and B, denoted $p(A \cap B)$ (read "probability of A and B"). Its computation depends on whether A and B are dependent or independent.

1) Dependent events - conditional probability

If two events are dependent rather than independent, how do we compute the probability that both occur, since the probability of one depends on the other ? We need the notion of conditional probability to quantify the dependence.

Definition 2.1

Let A and B be events, with $p(A) > 0$. The *conditional probability* of B given A is

$$p\big(B\big|_A\big) = \frac{p(A \cap B)}{p(A)}.$$

We read $p\big(B\big|_A\big)$ as "p of B given A".

Remark 2.2

Let $(\Omega, \mathcal{F}, p)$ be a probability space and let $B \in \mathcal{F}$ with $p(B) > 0$. Define

$$p_B : \ \mathcal{F} \to [0,1], \qquad A \longmapsto p_B(A) = p\big(A\big|_B\big) = \frac{p(A \cap B)}{p(B)}.$$

Then p_B is itself a probability measure on $(\Omega, \mathcal{F})$ and satisfies all the usual properties.

Proof:
We check the Kolmogorov axioms for p_B.

(1) Nonnegativity. For any $A \in \mathcal{F}$,

$$p_B(A) = \frac{p(A \cap B)}{p(B)}.$$

Since $p(A \cap B) \geqslant 0$ and $p(B) > 0$, we have $p_B(A) \geqslant 0$.

(2) Normalization. We compute

$$p_B(\Omega) = p\big(\Omega\big|_B\big) = \frac{p(\Omega \cap B)}{p(B)} = \frac{p(B)}{p(B)} = 1.$$

(3) Finite additivity (for two disjoint events). Let A_1 and A_2 be two disjoint events in $\mathcal{F}$. Then $A_1 \cap B$ and $A_2 \cap B$ are also disjoint, and

$$(A_1 \cup A_2) \cap B = (A_1 \cap B) \cup (A_2 \cap B).$$

Therefore, by finite additivity of p,

$$p_B(A_1 \cup A_2) = \frac{p\big((A_1 \cup A_2) \cap B\big)}{p(B)} = \frac{p(A_1 \cap B) + p(A_2 \cap B)}{p(B)} = p_B(A_1) + p_B(A_2).$$

Thus p_B satisfies nonnegativity, normalization, and finite additivity, hence p_B is a probability measure on $(\Omega, \mathcal{F})$. ■

Theorem 2.3: Law of total probability

Let $A_1, \ldots, A_k$ be a complete system of events. Then for any event $B \subseteq \Omega$,

$$p(B) = \sum_{j=1}^{k} p(B \cap A_j) = \sum_{j=1}^{k} p(A_j)\, p\Big(B\big|_{A_j}\Big). \qquad \text{(Total probability formula)}$$

Proof:
If $i \neq j$, then $(A_i \cap B) \cap (A_j \cap B) = (A_i \cap A_j) \cap B = \emptyset$. By distributivity, $B \cap \Omega = B \cap (A_1 \cup \cdots \cup A_k) = (A_1 \cap B) \cup \cdots \cup (A_k \cap B)$. By the additivity property (finite additivity for disjoint unions),

$$p(B) = \sum_{j=1}^{k} p(A_j \cap B).$$

By the definition of conditional probability, $p(A_j \cap B) = p(B\big|_{A_j})\, p(A_j)$, hence

$$p(B) = \sum_{j=1}^{k} p(B\big|_{A_j})\, p(A_j).$$

■

Application exercise :
Suppose there are three different routes to ENS-Meknès : A, B, and C. The probabilities of taking them are 30%, 40%, and 30% respectively. The probability of arriving late if you take A is 10%, if you take B is 5%, and if you take C is 15%. Find the probability of arriving late to ENS-Meknès to study.
Solution : Let R be "Arrive late", A_1 "Take route A", A_2 "Take route B", A_3 "Take route C". We know $p(A_1) = 0.3$, $p(A_2) = 0.4$, $p(A_3) = 0.3$, and $p(R|A_1) = 0.1$, $p(R|A_2) = 0.05$, $p(R|A_3) = 0.15$. By total probability,

$$p(R) = \sum_{j=1}^{3} p(A_j)p(R|A_j) = 0.03 + 0.02 + 0.045 = \boxed{0.095}.$$

Thus the probability of arriving late is 9.5%.

Theorem 2.4: Bayes' formula

Let $A_1, \ldots, A_k$ be a complete system of events and let B be an event with $p(B) > 0$. Then

$$p\big(A_j\big|_B\big) = \frac{p(A_j \cap B)}{p(B)} = \frac{p(A_j)\, p\Big(B\big|_{A_j}\Big)}{\displaystyle\sum_{i=1}^{k} p(A_i)\, p\Big(B\big|_{A_i}\Big)}. \qquad \text{(Bayes' formula)}$$

Application of Bayes' formula :
A medical lab produces a screening test for a disease. Sensitivity is 95% (positives among the truly ill). Specificity is 90% (negatives among the healthy). Prevalence in the population is 1%. If a person tests positive, what is the probability they are actually ill ?
Solution : Let M = "person is ill", T = "test is positive". Given $p(M) = 0.01$, $p(T|M) = 0.95$, $p(T|\bar{M}) = 0.10$. We seek $p(M|T)$:

$$p(M|T) = \frac{p(M)p(T|M)}{p(M)p(T|M) + p(\bar{M})p(T|\bar{M})} = \frac{0.01 \cdot 0.95}{0.01 \cdot 0.95 + 0.99 \cdot 0.10} \approx \boxed{0.087}.$$

So a positive test implies about an 8.7% chance of actually being ill.

Remark 2.5

Sampling *without* replacement is a good illustration of dependent events.

Exercise
An urn contains 5 black balls and 3 white balls. Two balls are drawn without replacement. What is the probability of drawing two white balls ?
Solution : Let B_1 be "white on the first draw", B_2 "white on the second draw". Then $p(B_1 \cap B_2) = p(B_1)\, p(B_2|B_1) = \frac{3}{8} \cdot \frac{2}{7} = \boxed{\frac{3}{28}}$.

2) Independent events

Definition 2.6: Independence of two events

Two events A and B are independent if $p\big(A\big|_B\big) = p(A)$.

Property 2.7

If two events are independent, the probability that both occur equals the product of their probabilities :

$$p(A \cap B) = p(A)\, p(B). \qquad \text{(Product rule for independent events)}$$

Example 2 :

For a fair six-sided die, let A : "the result is even", and B : "the result is a multiple of three". Then $A = \{2, 4, 6\}$, $B = \{3, 6\}$, $A \cap B = \{6\}$ and $p(A) = 3/6$, $p(B) = 2/6$, $p(A \cap B) = 1/6 = p(A)p(B)$; hence A and B are independent.

Example 3 :

Throw a red die and a green die. What is $p(\text{total} = 2)$? We need a 1 on each die. The events "red die = 1" and "green die = 1" are independent, each of probability 1/6, so

$$p(\text{total} = 2) = \boxed{\frac{1}{36}}.$$

Remark 2.8

Sampling *with* replacement is a good illustration of independent events.

Definition 2.9: Mutual independence of n events

For $n \geqslant 2$, events $A_1, \ldots, A_n$ are mutually independent (independent as a family) if

$$p(A_1 \cap \cdots \cap A_n) = \prod_{i=1}^{n} p(A_i).$$

Remark 2.10

vents can be pairwise independent ($p(A_i \cap A_j) = p(A_i)p(A_j)$ for $i \neq j$) without being mutually independent.

Example 4 :

Throw two fair dice and consider
A_1 : "the first die is even",
A_2 : "the second die is even",
A_3 : "the sum of the two throws is even".
They are pairwise independent but not mutually independent. Indeed,

$$\text{card}(\Omega) = 36, \quad p(A_1) = p(A_2) = p(A_3) = \tfrac{1}{2}, \quad p(A_i \cap A_j) = \tfrac{1}{4} \ (i \neq j),$$

but $p(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq \frac{1}{8} = p(A_1)p(A_2)p(A_3)$.

III - How to tackle a probability exercise ?

In many situations, problem solving is helped by the following approach.

1 List the elementary events or describe the sample space Ω.

2 Determine the probability measure on Ω.

- If the probability is uniform, then $p(A) = \dfrac{\text{card}\, A}{\text{card}\, \Omega}$.
- Otherwise, determine each elementary probability, remembering that their sum is 1.

3 Identify precisely the event(s) whose probability is sought.

4 Use the appropriate formula : Are we computing the probability of

- an elementary event ?
- a complementary event ?
- compatible/incompatible events (total probability) ?
- dependent/independent events (product rule) ?

Exercise

1. Two fair dice are thrown. What is the probability of getting a total of 7 ?
2. Now the dice are loaded : even numbers are twice as likely as odd numbers. What is $p(\text{sum} \neq 8)$?

Solution.

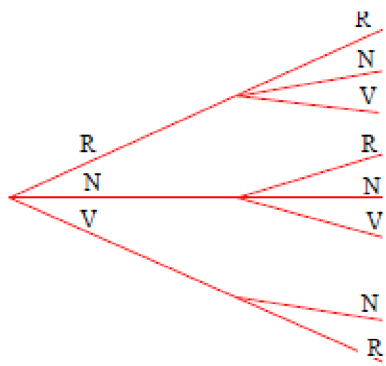
1.
   - $\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$, so $\text{card}(\Omega) = 36$.
   - The measure is uniform.
   - The event $(\text{sum} = 7)$ has 6 favorable pairs : $(1,6), (6,1), (2,5), (5,2), (3,4), (4,3)$.
   - Hence $p(\text{sum} = 7) = \dfrac{6}{36} = \boxed{\dfrac{1}{6}}$.
2.
   - Same Ω.
   - Let p_p be the probability of an even face and p_i that of an odd face ; the statement gives $p_p = 2p_i$. Since the six outcomes exhaust the space : $3p_p + 3p_i = 1 \Rightarrow 9p_i = 1$, so $p_i = \frac{1}{9}$, $p_p = \frac{2}{9}$.
   - Compute $p(\text{sum} = 8)$ first : elementary outcomes are $(2,6), (6,2), (3,5), (5,3), (4,4)$.
   - Using independence across dice, $p\{(2,6)\} = p\{2\}p\{6\} = \frac{2}{9} \cdot \frac{2}{9} = \frac{4}{81}$; likewise $p\{(6,2)\} = p\{(4,4)\} = \frac{4}{81}$ and $p\{(5,3)\} = p\{(3,5)\} = \frac{1}{81}$.
   - Thus $p(\text{sum} = 8) = \frac{14}{81}$ and $p(\text{sum} \neq 8) = 1 - \frac{14}{81} = \boxed{\dfrac{67}{81}}$.

IV - Counting techniques

1) Trees

Example 5 :

Consider an urn with two red, two black, and one green ball. Draw two balls without replacement. This is a two-stage experiment; its possibilities are represented by a (horizontal) tree. We get three main branches and three secondary branches at each stage, except when a green ball is drawn first. The number of terminal branches equals the number of outcomes in Ω.



2) Fundamental Principle of Counting

Fundamental Principle of Counting (Multiplication Axiom)

If one experiment can be performed in m ways and a second in n ways, then doing the first followed by the second can be performed in $m \times n$ ways.

This principle is not limited to two experiments; it applies to any finite number.

Example 6 :

You have 3 shirts (red, blue, green) and 4 pants (black, beige, blue, gray). In how many ways can you choose a shirt followed by pants to make an outfit? Answer : $3 \times 4 = 12$ outfits.

3) Arrangements and permutations

Consider a set of n distinct objects. Choose p of them and order them.

Definition 4.1

An ordered selection of p distinct objects from n is called an *arrangement* (with $p \leqslant n$).

How many ? View the p positions as fixed ; the number of choices is n for the first, $n-1$ for the second, ..., $n-p+1$ for the p-th.

Proposition 4.2

$$A_n^p = n(n-1)\cdots(n-p+1) = \frac{n!}{(n-p)!}.$$

⚠ $n!$ ("n factorial") is the product of all integers up to n : $n! = n(n-1)\cdots 3\times 2\times 1$, with the convention $0! = 1$.

Example 7 :

Arrangements of two letters from $\{a, b, c, d\}$: $A_4^2 = 4 \cdot 3 = \boxed{12}$,
namely $(a, b), (a, c), (a, d), (b, a), (b, c), (b, d), (c, a), (c, b), (c, d), (da), (d, b), (d, c)$.

Example 8 :

If $\mathrm{card}(A) = s$ and $\mathrm{card}(B) = t$, the number of injective functions $A \to B$ is A_t^s.
If $t < s$, no injection exists.

Definition 4.3

A *permutation* of n elements is an ordered arrangement of all n elements.

Proposition 4.4

The number of permutations of n elements is $A_n^n = n!$.

4) Combinations

Definition 4.5

A selection of p distinct objects from n *without regard to order* is called a *combination* of p from n.

In the previous example with $\{a,b,c,d\}$, the combination $\{a,b\}$ is the same as $\{b,a\}$, whereas the arrangements (a,b) and (b,a) are different.

How many ? The number of combinations of p from n is denoted by the English notation

$$\binom{n}{p} \quad \text{(read : “}n\text{ choose }p\text{”).}$$

Since each unordered selection corresponds to $p!$ ordered arrangements :

Proposition 4.6

$$\binom{n}{p} = \frac{A_n^p}{p!} = \frac{n(n-1)\cdots(n-p+1)}{p!} = \frac{n!}{p!\,(n-p)!}.$$

Example 9 : Combinations of two letters from $\{a,b,c,d\}$:

$$\binom{4}{2} = \frac{4!}{2!\,2!} = \boxed{6},$$

namely $\{a,b\}, \{a,c\}, \{a,d\}, \{b,c\}, \{b,d\}, \{c,d\}$.

Example 10 : At a party of ten people, how many handshakes occur if everyone shakes hands with everyone else ?

$$\boxed{\tbinom{10}{2} = 45 \text{ handshakes.}}$$

5) Case where elements need not be distinct

How many ordered selections of p elements from n when repetition is allowed ? Each position has n choices, so by the multiplication axiom there are n^p choices (here p may exceed n).

Example 11 : If $E = \{a,b,c,d,e\}$, there are $5^3 = 125$ three-letter words.

6) Summary

	Conditions	Number of possible selections	A common example :
possible with $p \geq n$	p elements not necessarily distinct, ordered	p-lists of elements of E : n^p	successive draws *with* replacement of p objects from n.
$p < n$	p elements all distinct and ordered	arrangements of p elements of E : A_n^p	successive draws *without* replacement.
$p = n$	n elements distinct and ordered	permutations of the n elements : $n!$	anagrams of a word with distinct letters.
$p < n$	p elements distinct and unordered	combinations of p from n : $\binom{n}{p}$	simultaneous draw of p objects from n.