Chapitre 3

Classical Probability Distributions

I - Discrete Probability Distributions

1) The Binomial Distribution

1- a) Bernoulli Trials

Let us consider a random experiment represented by a sample space Ω.
We choose an event A and call its occurrence a "success," with probability $p(A)$, denoted p.
Its complement, called a "failure," has probability $q = 1 - p$.

Definition 1.1: Bernoulli Distribution

A random variable X follows a Bernoulli distribution with parameter p, denoted $X \hookrightarrow \mathcal{B}(p)$, if :

$$X(\Omega) = \{0, 1\}$$

and

$$p(X = x) = \begin{cases} p & \text{if } x = 1, \\ 1 - p & \text{if } x = 0. \end{cases}$$

The event $(X = 1)$ represents success, and the event $(X = 0)$ represents failure.

Example 1 : In a coin-toss experiment, tossing a fair coin gives heads with probability $\frac{1}{2}$ and tails with probability $\frac{1}{2}$. If the coin is biased, then heads appears with probability $p \neq \frac{1}{2}$ and tails with probability $q = 1 - p \neq \frac{1}{2}$.
By assigning heads the value 1 and tails the value 0, we obtain a Bernoulli distribution.

Proposition 1.2: Expectation and Variance of the Bernoulli Distribution

$$\boxed{E(X) = p} \quad \boxed{V(X) = p(1-p)}$$

1- b) Binomial Distribution

Parameters : positive integer n and $p \in [0, 1]$.

- **Experiment :** n independent repetitions of a trial with success probability p.
- **Sample space :** the set of all sequences of H and T of length n.
- **Random variable** X **:** number of successes (Heads) observed in the n trials.
- **Model of :** number of successes in a fixed number of independent Bernoulli trials.

For example, when $n = 3$, the tree of possible outcomes is :

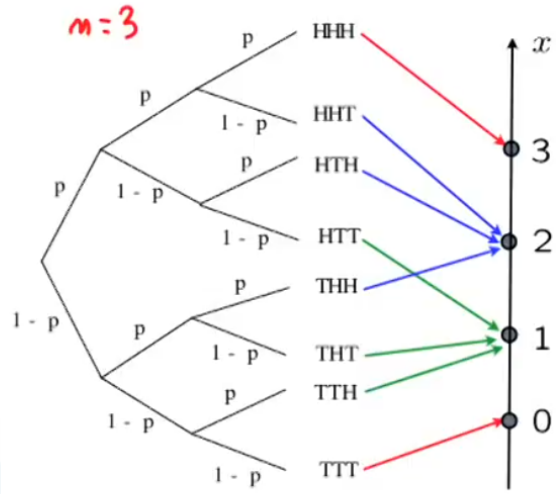


Figure 3.1 – Three independent tosses of a coin with $p(H) = p$.

Since X counts the number of Heads, for $n = 3$ the possible values are $0, 1, 2, 3$. For instance,

$$p_X(2) = p(HHT) + p(HTH) + p(THH) = 3\,p^2(1-p) = \binom{3}{2} p^2(1-p).$$

Probability mass function of a binomial random variable

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

Definition 1.3: Binomial Distribution

A random variable X follows a binomial distribution, written

$$X \hookrightarrow \mathcal{B}(n, p),$$

if

- an experiment is repeated independently n times, each trial resulting in success with probability p,
- X counts the total number of successes among the n trials,

$$X(\Omega) = \mathbb{N}.$$

Example 2 : We toss a coin n times. For each i with $1 \leqslant i \leqslant n$, define the Bernoulli variable

$$X_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ toss gives Heads,} \\ 0 & \text{otherwise.} \end{cases}$$

Let

$$Y = X_1 + X_2 + \cdots + X_n.$$

Then Y counts the number of Heads observed in the n tosses, and hence

$$Y \hookrightarrow \mathcal{B}(n, p).$$

Theorem 1.4: Relationship Between the Binomial and Bernoulli Distributions

If

$$Y = X_1 + X_2 + \cdots + X_n,$$

where $X_1, \ldots, X_n$ are independent Bernoulli random variables with parameter p, then

$$Y \hookrightarrow \mathcal{B}(n, p).$$

To determine the probabilities of the elementary events of a random variable following a binomial law, it is necessary to calculate the number of ways to obtain k successes in n trials.
This means finding the number of (unordered) combinations of k objects among n, with $k \leqslant n$. These combinations are unordered, since only the total number of successes matters, not their order of occurrence.
It is known that the number of ways to obtain k successes and $n-k$ failures is given by the binomial coefficient C_n^k. Multiplying this coefficient by the corresponding probabilities of success and failure yields the binomial law.
We therefore have :

Proposition 1.5

The elementary probabilities of a random variable X following a binomial distribution $\mathcal{B}(n, p)$ are given, for every number of successes $k = 0, 1, \ldots, n$, by :

$$\boxed{p(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.}$$

Proposition 1.6: Expectation and Variance of the Binomial Distribution

$$\boxed{E(X) = np} \quad \boxed{V(X) = np(1-p)}$$

Good exercise : prove this proposition !

Property 1.7: Sum

Let X_1 and X_2 be two *independent* random variables following binomial distributions $\mathcal{B}(n_1, p)$ and $\mathcal{B}(n_2, p)$ respectively. Then the sum $X_1 + X_2$ follows a binomial distribution $\mathcal{B}(n_1 + n_2, p)$.

This property is easy to interpret : if X_1 represents the number of successes during n_1 identical and independent trials, and X_2 the number of successes during n_2 independent trials (also independent of the first n_1), with the same probability of success, then $X_1 + X_2$ represents the total number of successes in $n_1 + n_2$ identical and independent trials.

1- c) Geometric Distribution

Parameter : probability of success p, with $0 < p \leqslant 1$.

- **Experiment :** infinitely many independent tosses of a coin with $p(\text{Heads}) = p$.
- **Sample space :** the set of infinite sequences of H and T.
- **Random variable X :** number of trials until the first success (first Heads).
- **Model of :** waiting times ; number of trials required to obtain the first success.

For example, if the sequence begins as

$$\text{TTTTHHT}\ldots$$

then the first Heads occurs on the 5th toss, so

$$X = 5.$$

Since the first success at trial k requires $(k-1)$ failures followed by 1 success,

$$p(X = k) = (1-p)^{k-1}p, \qquad k = 1, 2, 3, \ldots$$

Probability mass function of a geometric random variable

$$p_X(k) = (1-p)^{k-1}p, \qquad k = 1, 2, 3, \ldots$$

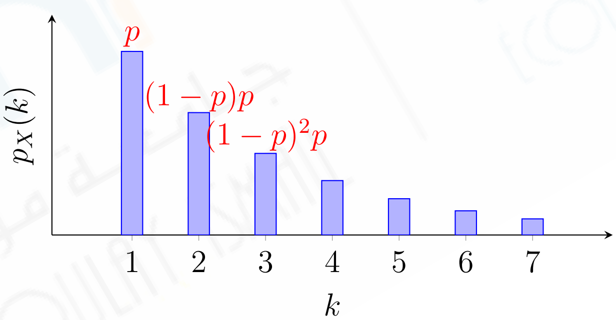


Figure 3.2 – Probability mass function of a geometric distribution.

Definition 1.8: Geometric Distribution

A random variable X follows a geometric distribution, written

$$X \hookrightarrow \mathcal{G}(p),$$

if independent Bernoulli(p) trials are performed and X counts the number of trials until the first success :

$$X(\Omega) = \mathbb{N}^\star$$

This is indeed a probability distribution, since :

- ✓ $p(X = k)$ is nonnegative,
- ✓ $\displaystyle\sum_{k\geqslant 1} p(X = k) = \sum_{k\geqslant 1} q^{k-1}p = \frac{p}{1-q} = 1$, where $q = 1 - p \in (0, 1)$.

Proposition 1.9: Expectation and Variance of the Geometric Distribution

$$\boxed{E(X) = \frac{1}{p}} \qquad \boxed{V(X) = \frac{q}{p^2} = \frac{1-p}{p^2}}$$

Indeed :

✔ $E(X) = \sum_{k\geqslant 1} kq^{k-1}p = p\sum_{k\geqslant 0} kq^{k-1} = \dfrac{p}{(1-q)^2} = \dfrac{1}{p}.$

✔ To compute $V(X)$, first compute $E(X(X-1))$:

$$\begin{aligned} E(X(X-1)) &= \sum_{k\geqslant 1} k(k-1)q^{k-1}p \\ &= 2\frac{pq}{(1-q)^3} \\ &= 2\frac{q}{p^2}. \end{aligned}$$

Moreover,

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ &= E(X(X-1)) + E(X) - (E(X))^2 \\ &= 2\frac{q}{p^2} + \frac{1}{p} - \frac{1}{p^2}, \end{aligned}$$

hence

$$V(X) = \frac{q}{p^2} = \frac{1-p}{p^2}.$$

Typical Example

Sampling with replacement from an urn containing two types of balls. We observe the index of the first appearance of a given type.

Example 3 : We sample with replacement from an urn containing 113 white balls and 7 black balls. The probability of drawing a black ball is

$$p = \frac{7}{120}.$$

Thus

$$E(X) = \frac{1}{p} = \boxed{\tfrac{120}{7}},$$

so on average one should expect between 17 and 18 draws to obtain a black ball for the first time.

2) The Poisson Distribution

Definition 1.10: Definition and Implementation

This distribution can model rare events. For example, it can model **the number of calls received by a switchboard**, **the number of passengers arriving at a ticket counter during a day**, etc. It is expressed using the exponential function and depends on a parameter $\lambda > 0$, which corresponds to the average number of occurrences of the observed phenomenon over the given time period. More formally :

Probability

A random variable X follows a Poisson distribution with parameter $\lambda > 0$, written $\mathcal{P}(\lambda)$, when $X(\Omega) = \mathbb{N}$ and for all $k \in \mathbb{N}$:

$$\boxed{P(X = k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}}$$

Here, k is the number of occurrences under consideration.

2- a) Mean and Variance

Proposition 1.11: Expectation and Variance of the Poisson Distribution

$$\boxed{E(X) = \lambda} \quad \boxed{V(X) = \lambda}$$

Proof:

✔

$$\begin{aligned} E(X) &= \sum_{k\geqslant 0} k\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} \\ &= \mathrm{e}^{-\lambda}\sum_{k\geqslant 1}\frac{\lambda^k}{(k-1)!} \\ &= \mathrm{e}^{-\lambda}\lambda\mathrm{e}^{\lambda} \\ &= \lambda. \end{aligned}$$

✔ To compute $V(X)$, first compute $E(X(X-1))$.

$$\begin{aligned} E(X(X-1)) &= \sum_{k\geqslant 0} k(k-1)P(X=k) \\ &= \sum_{k\geqslant 0} k(k-1)\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} \\ &= \mathrm{e}^{-\lambda}\lambda^2\mathrm{e}^{\lambda} \\ &= \lambda^2. \end{aligned}$$

Hence,

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ &= E(X(X-1)) + E(X) - (E(X))^2 \\ &= \lambda^2 + \lambda - \lambda^2, \end{aligned}$$

so $V(X) = \lambda$.

■

Example 4 :

At a hotel, an average of 1.25 people arrive every 10 minutes between 3 pm and 9 pm. Let X be the number of people arriving at this hotel in a 10-minute interval during that time window. Assume X follows a Poisson distribution.

1. Determine the probability that k people arrive in 10 minutes.
   Answer : $P(X = k) = \boxed{\frac{1.25^k e^{-1.25}}{k!}}$, since the average gives the parameter λ.
2. Determine the probability that 2 people arrive in 10 minutes.
   Answer : $P(X = 2) = \frac{1.25^2 e^{-1.25}}{2!} = 0.22\ldots$
3. Determine the probability that at most 4 people arrive in 10 minutes.
   Answer : $P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = \frac{e^{-1.25}}{0!} + \frac{1.25e^{-1.25}}{1!} + \frac{1.25^2 e^{-1.25}}{2!} + \frac{1.25^3 e^{-1.25}}{3!} + \frac{1.25^4 e^{-1.25}}{4!} = 0.99\ldots$
4. Determine the probability that at least 3 people arrive in 10 minutes.
   Answer : $1 - P(X = 0) - P(X = 1) - P(X = 2) = 1 - \frac{e^{-1.25}}{0!} - \frac{1.25e^{-1.25}}{1!} - \frac{1.25^2 e^{-1.25}}{2!} = 0.13\ldots$

Proposition 1.12: Sum of Two Poisson Distributions

If X_1 and X_2 are *independent* random variables with Poisson parameters λ_1 and λ_2, then $X_1 + X_2$ is Poisson with parameter $\lambda_1 + \lambda_2$.

2- b) Approximating the Binomial by the Poisson

When n is large and p is small, in such a way that the product $np = \lambda$ remains small relative to n, the binomial distribution $\mathcal{B}(n, p)$ can be approximated by the Poisson distribution $\mathcal{P}(\lambda)$. In practice, the approximation is valid if $n > 20$, $p \leqslant 0.1$, and $np \leqslant 5$.

$$\boxed{\text{Approximate } \mathcal{B}(n, p) \text{ by } \mathcal{P}(np) \text{ as soon as } n > 20,\ p \leqslant 0.1, \text{ and } np \leqslant 5.}$$

IMPORTANT RULE. When approximating one distribution by another, choose the parameter(s) of the approximating distribution so that its expectation (and variance, when enough parameters are available) matches the expectation (and variance) of the approximated distribution.

3) The Hypergeometric Distribution

Typical Example

We draw simultaneously n balls from an urn containing $N_1 = pN$ "winning" balls and $N_2 = qN$ "losing" balls (with $q = 1 - p$, for a total of $pN + qN = N$ balls). We then count the number of winning balls drawn and denote by X the random variable giving this number.

Definition 1.13

X follows a hypergeometric distribution with parameters N, n, and p if we set $N_1 = pN$ (number of good elements) and $N_2 = (1-p)N$ (number of bad elements), and

$$P(X = k) = \frac{C_{N_1}^{k} C_{N_2}^{n-k}}{C_N^n}$$

with $X(\Omega) = [\![\max(0, n - N_2)\,;\min(N_1, n)]\!]$.
We write $X \hookrightarrow \mathcal{H}(n, p, N)$.

Proposition 1.14: Mean and Variance of the Hypergeometric Distribution

$$\boxed{E(X) = np} \qquad \boxed{V(X) = npq\frac{N-n}{N-1}}$$

Note that $V(X)$ tends to npq as $N \to \infty$.

Example 5 : A lake contains about one hundred fish, one quarter of which are pike. We catch 10 fish; the distribution of the number X of pike in the catch is $\mathcal{H}(10, 1/4, 100)$.
We find for the successive pairs
$(k, P(X = k))$: (0, 5%), (1, 18%), (2, 30%), (3, 26%), (4, 15%), (5, 5%), (6, 1%), (7, 0%), (8, 0%), (9, 0%), (10, 0%).
Hence the maximum likelihood is for 2 or 3 pike.
Moreover, the expected number of pike is $\boxed{\frac{10}{4} = 2.5}$.

4) The Negative Binomial Distribution

Definition 1.15

The negative binomial distribution describes the following situation : an experiment consists of a series of independent trials, yielding a "success" with probability p (constant during the whole experiment) and a "failure" with probability $q = 1 - p$. The experiment continues until a given number r of successes has been obtained. The random variable X represents **the number of failures before the r-th success**.

The probability mass function is :

$$P(X = k) = \binom{k+r-1}{k} p^r q^k = \binom{k+r-1}{r-1} p^r q^k, \qquad k = 0, 1, 2, \ldots$$

We write $X \hookrightarrow \mathcal{NB}(r, p)$.

Typical Example : Number of Failures Before the r-th Success

We toss a (biased) coin with probability p of landing Heads. Let X be the number of failures before the r-th Heads. Then X follows a negative binomial distribution with parameters r and p.

Property 1.16

If X follows the negative binomial distribution, then

$$E(X) = \frac{r(1-p)}{p}, \qquad V(X) = \frac{r(1-p)}{p^2}.$$

Remark 1.17

The negative binomial distribution can also be expressed using generalized binomial coefficients :

$$P(X = k) = \binom{-r}{k} p^r (-q)^k, \qquad k = 0, 1, 2, \ldots$$

where the generalized coefficient for a negative integer is defined by

$$\binom{-r}{k} := \frac{(-r)(-r-1)\cdots(-r-k+1)}{k!}.$$

This formulation explains the name *negative binomial distribution*. It is also known as the *Pascal distribution* and the *Pólya distribution*.

II - Continuous Probability Distributions

1) Uniform Distribution

Definition 2.1: Property and definition

Let $[a, b]$ be an interval with real numbers $a < b$. The *uniform distribution* on $[a, b]$ is the probability law whose density f is

$$f(x) = \begin{cases} 0, & \text{if } x < a, \\ \frac{1}{b-a}, & \text{if } a \leqslant x \leqslant b, \\ 0, & \text{if } x > b. \end{cases}$$

This continuous distribution is often denoted $\mathcal{U}(a, b)$.

Let us verify that f is indeed a probability density :

- ✔ f is piecewise continuous on $\mathbb{R}$.
- ✔ f is nonnegative on $\mathbb{R}$.
- ✔ The integral of f over $\mathbb{R}$ equals 1 :

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_a^b \frac{1}{b-a}\,dx = \frac{b-a}{b-a} = 1.$$

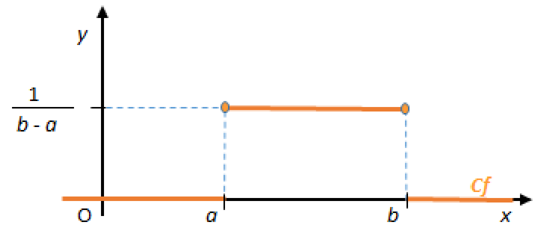


Figure 3.3 – Density of the uniform distribution $\mathcal{U}(a,b)$

Property 2.2

If X follows $\mathcal{U}(a,b)$, the cumulative distribution function of X is

$$F(x)=\begin{cases}0 & \text{if } x<a,\\ \dfrac{x-a}{b-a} & \text{if } a\leqslant x\leqslant b,\\ 1 & \text{if } x>b.\end{cases}$$

Proposition 2.3: Mean and Variance of the Uniform Distribution

Let X be a random variable with the uniform distribution on $[a,b]$. Then

$$\boxed{E(X)=\frac{a+b}{2}} \qquad \boxed{V(X)=\frac{(b-a)^2}{12}}.$$

Example 6 : Let $X \hookrightarrow \mathcal{U}[0,4]$. Compute $P(X=2)$, $P(2<X<3)$, and $E(X)$.
Since the law is continuous, the probability of any exact value is, by definition, zero.
Thus $P(X=2)=\boxed{0}$.
Also $P(2<X<3)=F(3)-F(2)=\frac{1}{4}(3-2)=\boxed{0.25}$.
And $E(X)=\dfrac{0+4}{2}=\boxed{2}$.

Example 7 : Ahmed and Fatima walk home from school. Their parents know they should arrive between 17 :00 and 18 :00. We model their arrival time by a random variable $X \hookrightarrow \mathcal{U}[17,18]$.

1. What is the probability they arrive between 17 :00 and 17 :15 ?

$$P\big(x\in[17,17.25]\big)=F(17.25)-F(17)=\frac{17.25-17}{18-17}=0.25=\boxed{25\%}.$$

This is natural : each quarter-hour has probability 25%, and over the full hour (17 :00–18 :00) the total probability is 100%.

2. When can their parents expect them to arrive ?

$$E(X)=\frac{17+18}{2}=17.5=\boxed{17{:}30}.$$

Example 8 : Markov inequality : uniform example

Let $X \hookrightarrow \mathcal{U}[-4,4]$. We want an upper bound for $p(X \geqslant 3)$ using Markov's inequality.
Since X takes negative values, we apply Markov's inequality to the non-negative random variable $|X|$. First compute

$$E(|X|) = \int_{-4}^{4} |x| \, \frac{1}{8} \, dx = \frac{2}{8} \int_{0}^{4} x \, dx = \frac{2}{8} \times \frac{4^2}{2} = 2.$$

For $a = 3$ we have

$$p(X \geqslant 3) \;\leqslant\; p(|X| \geqslant 3) \;\leqslant\; \frac{E(|X|)}{3} = \frac{2}{3}.$$

(The exact value is $p(X \geqslant 3) = \dfrac{4-3}{4-(-4)} = \dfrac{1}{8}$, so Markov's bound is valid but not tight.)

2) Exponential Distribution

Typical Example

These are cases where the random variable represents **the waiting time between two successive events** or the lifetime of a system not subject to aging (operation of electronic components, etc.), or the waiting time for an accidental event (earthquake, radioactive decay, ...).

Definition 2.4: Definition and Property

Let $\lambda > 0$. The exponential distribution with parameter λ is the probability law with density

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0,\infty)}(x) = \begin{cases} 0 & \text{if } x < 0, \\ \lambda e^{-\lambda x} & \text{if } x \geqslant 0. \end{cases}$$

We write $\mathcal{E}(\lambda)$.

Property 2.5

If $T \hookrightarrow \mathcal{E}(\lambda)$, then the cumulative distribution function of T is

$$F(t) = \begin{cases} 0 & \text{if } t < 0, \\ 1 - e^{-\lambda t} & \text{if } t \geqslant 0. \end{cases}$$

Proposition 2.6: Expectation and Variance of the Exponential Distribution

If $T \hookrightarrow \mathcal{E}(\lambda)$, then

$$\boxed{E(T) = \frac{1}{\lambda}} \quad \boxed{V(T) = \frac{1}{\lambda^2}}.$$

Example 9 : Suppose the waiting time in minutes at a customer service desk is modeled by $T \hookrightarrow \mathcal{E}(0.2)$.

1) What is the probability of waiting at most 5 minutes ?

$$P(0 \leqslant T \leqslant 5) = F(5) - F(0) = 1 - \mathrm{e}^{-0.2\times 5} = \boxed{1 - \tfrac{1}{e}} \approx 0.632.$$

2) What is the probability of waiting more than 10 minutes ?
$P(T \geqslant 10) = 1 - P(0 \leqslant T < 10) = \mathrm{e}^{-0.2\times 10} = \boxed{\mathrm{e}^{-2}} \approx 0.135.$

3) How long should a customer expect to wait, on average ?

$$E(T) = \frac{1}{0.2} = \boxed{5 \text{ min}}.$$

Example 10 : [Markov inequality : exponential example]

Let X be an exponential random variable with parameter $\lambda = 1$, i.e.

$$f_X(x) = e^{-x}\,\mathbf{1}_{[0,\infty)}(x).$$

Then $X \geqslant 0$ and

$$E(X) = \int_0^\infty x e^{-x}\,dx = 1.$$

For any $a > 0$, Markov's inequality yields

$$P(X \geqslant a) \leqslant \frac{E(X)}{a} = \frac{1}{a}.$$

For comparison, the exact tail probability is

$$P(X \geqslant a) = \int_a^\infty e^{-x}\,dx = e^{-a}.$$

Therefore, for all $a > 0$,

$$e^{-a} \leqslant \frac{1}{a}.$$

The bound given by Markov's inequality can be improved by using **Chebyshev's inequality**. Recall that if a random variable Y has mean μ and variance σ^2, then for any $c > 0$,

$$p\big(|Y - \mu| \geqslant c\big) \leqslant \frac{\sigma^2}{c^2}.$$

Example 11 : Chebyshev inequality : exponential example

Let X be as above, $X \hookrightarrow \mathcal{E}(1)$. We know that

$$E(X) = 1, \qquad V(X) = 1.$$

For $a > 1$, write

$$P(X \geqslant a) = P(X - 1 \geqslant a - 1) \ \leqslant\ P\big(|X - 1| \geqslant a - 1\big),$$

since $\{X - 1 \geqslant a - 1\} \subset \{|X - 1| \geqslant a - 1\}$.

Now apply Chebyshev's inequality to the centred variable $X - 1$:

$$P\big(|X - 1| \geqslant a - 1\big) \ \leqslant\ \frac{V(X)}{(a-1)^2} = \frac{1}{(a-1)^2}.$$

Hence, for all $a > 1$,

$$P(X \geqslant a) \ \leqslant\ \frac{1}{(a-1)^2}.$$

That is

$$e^{-a} \ \leqslant\ \frac{1}{(a-1)^2}.$$

which is a much tighter bound than the Markov estimate $P(X \geqslant a) \leqslant 1/a$.

III - Moment Generating Functions

In this section, we introduce *moment generating functions*, which provide a more convenient method for deriving several fundamental results, such as the fact that the sum of independent binomial random variables is again binomial, and that the sum of independent Poisson random variables is Poisson.
Beyond simplifying the study of sums of independent random variables, moment generating functions play a central role in probability theory. In particular, they will be used to prove the **Central Limit Theorem**, one of the most important results in statistics, and to derive important probabilistic inequalities.
Although MGFs are primarily theoretical tools, they are essential for understanding the structure and behavior of probability distributions.

1) Moments

Definition 3.1: Moments

Let X be a random variable and let $c \in \mathbb{R}$ be a real number. For any integer $k \geqslant 1$:

☞ The *kth moment* of X is defined by

$$E\big(X^k\big)\,.$$

☞ The *kth moment of X about c* is defined by

$$E\big((X-c)^k\big)\,.$$

Remark 3.2

The first four moments of a random variable are commonly used in probability and statistics, although we mainly focus on the first two.

✔ The *first moment* of X is the *mean* of the distribution, defined by

$$\mu = E(X).$$

It describes the center or average value of the distribution.

✔ The *second moment of X about μ* is the *variance* of the distribution, defined by

$$\sigma^2 = V(X) = E\big((X-\mu)^2\big)\,.$$

It measures the spread of the distribution, that is, how much the values of X vary around the mean.

2) Moment Generating Functions (MGFs)

Definition 3.3: Moment Generating Function (MGF)

Let X be a random variable. The *moment generating function* (MGF) of X is the function defined, for a real parameter t, by

$$M_X(t) = E\left(e^{tX}\right),$$

whenever this expectation exists.
If X is a discrete random variable, then,

$$M_X(t) = \sum_{x \in X(\Omega)} e^{tx}\, p_X(x).$$

☞ If X is a continuous random variable, then,

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx}\, f_X(x)\, dx.$$

We say that the MGF of X *exists* if there exists $\varepsilon > 0$ such that $M_X(t)$ is finite for all

$$t \in (-\varepsilon, \varepsilon),$$

since the above sum or integral may diverge.

Remark 3.4

The restriction $t \in (-\varepsilon, \varepsilon)$ is necessary because the exponential function e^{tX} may grow too fast, causing the defining sum or integral of the MGF to diverge for large values of $|t|$.
Requiring finiteness on an open interval containing 0 guarantees that the MGF is well defined and that **all derivatives at $t = 0$ exist**, which is crucial because (see Property 3.7)

$$M_X^{(k)}(0) = E(X^k),$$

so MGFs truly *generate moments*.

Example 12 :

Find the moment generating function (MGF) of the following random variables :

a) X is a discrete random variable with probability mass function

$$p_X(k) = \begin{cases} \frac{1}{3}, & k = 1, \\ \frac{2}{3}, & k = 2. \end{cases}$$

b) Y is a continuous random variable uniformly distributed on $(0, 1)$, i.e.

$$Y \hookrightarrow \text{Unif}(0, 1).$$

Solution :

a) Since X is discrete, we use :

$$M_X(t) = E\left(e^{tX}\right) = \sum_k e^{tk} p_X(k).$$

Thus,

$$M_X(t) = \frac{1}{3}e^t + \frac{2}{3}e^{2t}, \quad t \in \mathbb{R}.$$

b) Since Y is continuous with density

$$f_Y(y) = \begin{cases} 1, & 0 \leqslant y \leqslant 1, \\ 0, & \text{otherwise}, \end{cases}$$

its MGF is given by

$$M_Y(t) = E\big(e^{tY}\big) = \int_0^1 e^{ty}\, dy.$$

A direct computation gives

$$M_Y(t) = \begin{cases} \dfrac{e^t - 1}{t}, & t \neq 0, \\ 1, & t = 0. \end{cases}$$

3) Properties and uniqueness of moment generating functions

There are several useful properties of moment generating functions (MGFs).
Let X and Y be independent random variables, and let $a, b \in \mathbb{R}$ be real numbers.
Recall that the moment generating function of X is

$$M_X(t) = E\big(e^{tX}\big),$$

whenever it exists.

Property 3.5: MGF of a linear transformation

Let X be a random variable and let $a, b \in \mathbb{R}$. If the MGF of X exists, then the MGF of $aX + b$ is given by

$$M_{aX+b}(t) = e^{tb}\, M_X(at).$$

Proof:
We compute directly :

$$M_{aX+b}(t) = E\big(e^{t(aX+b)}\big) = E\big(e^{tb}e^{taX}\big) = e^{tb}E\big(e^{taX}\big) = e^{tb}M_X(at).$$

■

Property 3.6: MGF of a sum of independent random variables

Let X and Y be independent random variables whose MGFs exist. Then the MGF of $X + Y$ is

$$M_{X+Y}(t) = M_X(t)\, M_Y(t).$$

Proof:
Using independence, we obtain

$$M_{X+Y}(t) = E\big(e^{t(X+Y)}\big) = E\big(e^{tX}e^{tY}\big) = E\big(e^{tX}\big)\, E\big(e^{tY}\big) = M_X(t)\, M_Y(t).$$

■

Property 3.7: Generating moments using MGFs

Assume that the MGF $M_X(t)$ exists in a neighborhood of $t=0$. Then, for any integer $k \geqslant 1$,

$$M_X^{(k)}(0) = E\left(X^k\right).$$

Proof:
We illustrate the idea for the first two moments. Assume that X is a discrete random variable with support $X(\Omega)$. By definition of the MGF,

$$M_X(t) = E\left(e^{tX}\right) = \sum_{x \in X(\Omega)} e^{tx} p_X(x).$$

Taking the derivative with respect to t, we obtain

$$M_X'(t) = \frac{d}{dt} E\left(e^{tX}\right) = \frac{d}{dt} \sum_{x \in X(\Omega)} e^{tx} p_X(x) = \sum_{x \in X(\Omega)} \frac{d}{dt}\left(e^{tx} p_X(x)\right).$$

Since x is constant with respect to t, we have

$$\frac{d}{dt} e^{tx} = x e^{tx},$$

and therefore

$$M_X'(t) = \sum_{x \in X(\Omega)} x e^{tx} p_X(x).$$

Evaluating at $t=0$, we obtain

$$M_X'(0) = \sum_{x \in X(\Omega)} x e^{0} p_X(x) = \sum_{x \in X(\Omega)} x p_X(x) = E(X).$$

Now, consider the second derivative :

$$M_X''(t) = \frac{d}{dt} M_X'(t) = \frac{d}{dt} \sum_{x \in X(\Omega)} x e^{tx} p_X(x) = \sum_{x \in X(\Omega)} \frac{d}{dt}\left(x e^{tx} p_X(x)\right).$$

Again, since x is constant with respect to t, we get

$$M_X''(t) = \sum_{x \in X(\Omega)} x^2 e^{tx} p_X(x).$$

Evaluating at $t=0$ yields $M_X''(0) = \sum_{x \in X(\Omega)} x^2 e^0 p_X(x) = \sum_{x \in X(\Omega)} x^2 p_X(x) = E(X^2)$.
This reveals a clear pattern : taking the nth derivative of $M_X(t)$ and evaluating at $t=0$ produces the nth moment $E(X^n)$. ■

Remark 3.8

If two random variables have the same MGF on an open interval containing 0, then they have the same distribution. This property is called the *uniqueness of the MGF* and makes MGFs a powerful tool for identifying distributions.

Theorem 3.9: Properties and Uniqueness of Moment Generating Functions

Let $f : \mathbb{R} \to \mathbb{R}$ be a function and denote by $f^{(n)}(x)$ its nth derivative. Let X and Y be independent random variables, and let $a, b \in \mathbb{R}$. Assume that the corresponding MGFs exist in a neighborhood of 0.

1. **Moments from the MGF.** For every integer $n \geqslant 1$,

$$M_X^{(n)}(0) = E(X^n).$$

In particular,

$$M_X'(0) = E(X), \qquad M_X''(0) = E(X^2).$$

This explains why M_X is called a *moment generating function.*

2. **Linear transformation.** The MGF of $aX + b$ is given by

$$M_{aX+b}(t) = e^{tb} M_X(at).$$

3. **Sum of independent variables.** If X and Y are independent, then

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

4. **Uniqueness.** The following statements are equivalent :
   (a) X and Y have the same distribution ;
   (b) $f_X(x) = f_Y(x)$ for all $x \in \mathbb{R}$;
   (c) $F_X(x) = F_Y(x)$ for all $x \in \mathbb{R}$;
   (d) There exists $\varepsilon > 0$ such that

$$M_X(t) = M_Y(t) \quad \text{for all } t \in (-\varepsilon, \varepsilon).$$

Consequently, the MGF uniquely **characterizes** the distribution of a random variable, just like the PDF.

Proof of uniqueness in a special case:
Assume that X and Y are discrete random variables with finite support $\Omega = \{0, 1, \ldots, m\}$ and that

$$M_X(t) = M_Y(t) \quad \text{for all } t \in \mathbb{R}.$$

By definition of the MGF,

$$\sum_{k=0}^{m} e^{tk} p_X(k) = \sum_{k=0}^{m} e^{tk} p_Y(k).$$

Subtracting both sides gives

$$\sum_{k=0}^{m} e^{tk} \big(p_X(k) - p_Y(k)\big) = 0.$$

Let $a_k = p_X(k) - p_Y(k)$ and write $e^{tk} = (e^t)^k$. Then

$$\sum_{k=0}^{m} a_k (e^t)^k = 0 \quad \text{for all } t.$$

This is a polynomial in e^t that vanishes identically, hence all coefficients a_k must be zero. Therefore $p_X(k) = p_Y(k)$ for all k, and X and Y have the same distribution. ■

Example 13 :

Suppose that $X \hookrightarrow \mathcal{P}(\lambda)$, meaning that X has support

$$X(\Omega) = \{0, 1, 2, \dots\}$$

and probability mass function

$$p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in X(\Omega).$$

Compute the moment generating function $M_X(t)$.

Solution : First, recall the Taylor series expansion of the exponential function :

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Using the definition of the MGF, we obtain

$$M_X(t) = E\big(e^{tX}\big) = \sum_{k=0}^{\infty} e^{tk} p_X(k).$$

Substituting the expression of the PMF gives

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty} \frac{(e^t)^k \lambda^k}{k!}.$$

This can be rewritten as

$$M_X(t) = e^{-\lambda}\sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!}.$$

We conclude that

$$M_X(t) = e^{-\lambda} e^{\lambda e^t} = \exp\big(\lambda(e^t - 1)\big).$$

Example 14 :

If $Y \hookrightarrow \mathcal{P}(\gamma)$ and $Z \hookrightarrow \mathcal{P}(\mu)$ are independent random variables, show that

$$Y + Z \hookrightarrow \mathcal{P}(\gamma + \mu)$$

using the uniqueness property of moment generating functions.

Solution : First, note that a $\mathcal{P}(\gamma + \mu)$ random variable has moment generating function

$$M(t) = \exp\big((\gamma + \mu)(e^t - 1)\big),$$

obtained by substituting $\gamma + \mu$ for the parameter.

Since Y and Z are independent, by the MGF product property we have

$$M_{Y+Z}(t) = M_Y(t) M_Z(t).$$

Using the MGF of a Poisson random variable, we obtain

$$M_{Y+Z}(t) = \exp\big(\gamma(e^t - 1)\big)\exp\big(\mu(e^t - 1)\big) = \exp\big((\gamma + \mu)(e^t - 1)\big).$$

The MGF of $Y + Z$ is therefore the same as that of a $\mathcal{P}(\gamma + \mu)$ random variable.

By the uniqueness of moment generating functions, it follows that

$$Y + Z \hookrightarrow \mathcal{P}(\gamma + \mu).$$

Definition 3.10: Convergence in probability

A sequence of random variables $(Y_n)_{n\geqslant 1}$ is said to *converge in probability* to a random variable Y if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\big(|Y_n - Y| \geqslant \varepsilon\big) = 0.$$

In this case, we write

$$Y_n \xrightarrow{P} Y \quad \text{as } n \to \infty.$$

Example 15 :

Let $(X_n)_{n\geqslant 1}$ be a sequence of random variables such that

$$X_n \hookrightarrow \mathcal{E}(n),$$

that is, X_n follows an exponential distribution with parameter n. Prove that

$$X_n \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

Fix $\varepsilon > 0$. We have

$$P\big(|X_n - 0| \geqslant \varepsilon\big) = P(X_n \geqslant \varepsilon).$$

Using the tail probability of the exponential distribution,

$$P(X_n \geqslant x) = e^{-nx}, \quad x \geqslant 0,$$

we obtain

$$P(X_n \geqslant \varepsilon) = e^{-n\varepsilon} \xrightarrow[n\to\infty]{} 0.$$

Hence,

$$X_n \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

Example 16 :

A classical example is defined as follows. For each $n \geqslant 1$, the random variable Y_n is defined by

$$Y_n(\Omega) = \{0, n^2\}.$$

Its probability distribution is given by

$$p_{Y_n}(y) = \begin{cases} 1 - \dfrac{1}{n}, & \text{if } y = 0, \\ \dfrac{1}{n}, & \text{if } y = n^2. \end{cases}$$

Fix $\varepsilon > 0$. For n large enough, we have $n^2 \geqslant \varepsilon$, and therefore

$$P\big(|Y_n - 0| \geqslant \varepsilon\big) = P(Y_n = n^2) = \frac{1}{n} \xrightarrow[n\to\infty]{} 0.$$

This verifies that

$$Y_n \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

However,

$$E(Y_n) = n \xrightarrow[n\to\infty]{} +\infty.$$

This shows that the sequence $(E(Y_n))$ does not converge, **but** the sequence (Y_n) converges in probability to 0.

Hence, **convergence in probability does not imply convergence of expectations**.

Figure 3.4 – Illustration of convergence in probability without convergence of expectations

IV - Normal Distribution (Laplace–Gauss Distribution)

From a historical perspective, the nature and exceptional importance of this law were perceived in 1773 by Abraham de Moivre when he considered the limiting form of the binomial distribution.
In 1772, Pierre-Simon Laplace studied it in his theory of errors. But it was only in 1809 for Carl Friedrich Gauss and in 1812 for Laplace that it took its definitive form. Hence it is sometimes called the Laplace law, sometimes the Gauss law, and sometimes the Laplace–Gauss law. It is also known as the "normal law," which does not imply that other laws are "abnormal."
It is fundamentally important because many statistical methods are based on it. This is due to the fact that it acts as a limit distribution under very general conditions. In particular, **the normal distribution models noise that results from the accumulation of many small, independent perturbations.**

Concrete situation

We often encounter complex phenomena that result from many causes with small, more-or-less independent effects. A typical example is the error made when measuring a physical quantity. This error results from a large number of factors, such as uncontrollable variations in temperature or pressure, atmospheric turbulence, vibrations of the measuring device, etc. Each factor has a small effect, yet the resulting error may not be negligible. Two measurements carried out under conditions the experimenter considers identical can therefore yield different results.
This occurs, for example, in :

- Meteorology, for the distribution of random phenomena such as temperature and pressure.
- Biology, for the distribution of biometric characteristics such as height or weight in a homogeneous population.
- Engineering, for the distribution of the dimensions of machined parts.
- Economics, for accidental fluctuations of an economic quantity (production, sales, ...).

Definition 4.1

- A continuous random variable follows the *standard normal distribution* if its probability density function is given by
$$f(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$
- A continuous random variable follows a *normal distribution* if its probability density function has the form
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}, \quad x \in \mathbb{R},$$
where $m \in \mathbb{R}$ and $\sigma > 0$. We write $\mathcal{N}(m, \sigma^2)$.

Remark 4.2

1. One can show that f is indeed a probability density function. Indeed, by the **Gaussian integral**,

$$\int_{\mathbb{R}} \mathbf{e}^{-x^2/2}\, dx = \sqrt{2\pi},$$

and therefore

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}\, dx = 1,$$

as it should be. This proves that the standard normal distribution defines a valid probability density function.

2. Since a probability density function is always nonnegative, the parameter σ must be strictly positive.

Property 4.3: Expectation and variance of the normal distribution

$$\boxed{E(X) = m} \quad \boxed{V(X) = \sigma^2,} \quad \boxed{\sigma(X) = \sigma.}$$

Proof:
One can compute these directly from the Gaussian integral. ∎

0- a) Shape of the Normal Distribution

The probability density of the normal distribution has the shape of a "**bell curve**." In fact, it is not a single curve but rather a family of curves depending on m and σ.

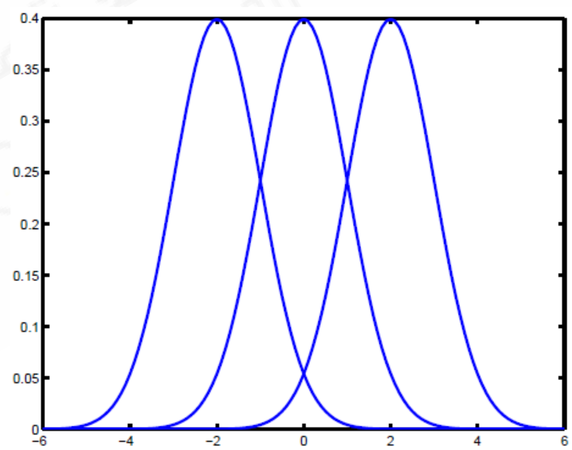


Figure 3.5 – Same standard deviation ; different means −2, 0, 2

We can make a few remarks about these curves.

a) The distribution is symmetric about the line $x = m$. Hence the area under the curve on each side of this line equals 0.5.

b) The distribution spreads out more as σ increases.

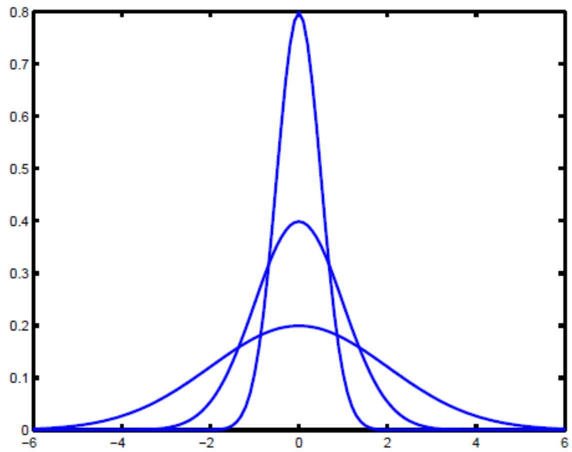


Figure 3.6 – Same mean ; different standard deviations 0.5, 1, 2

Lemma 4.4: MGF of a normal random variable

If $X \hookrightarrow \mathcal{N}(m, \sigma^2)$, then its moment generating function exists for all $t \in \mathbb{R}$ and is given by

$$M_X(t) = E\left(e^{tX}\right) = \exp\left(mt + \frac{\sigma^2 t^2}{2}\right).$$

Proof:
Let $X \hookrightarrow \mathcal{N}(m, \sigma^2)$. Its density is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \qquad x \in \mathbb{R}.$$

For any $t \in \mathbb{R}$, by definition,

$$M_X(t) = E\left(e^{tX}\right) = \int_{-\infty}^{+\infty} e^{tx} f_X(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(tx - \frac{(x-m)^2}{2\sigma^2}\right) dx.$$

We now complete the square in the exponent. Write

$$tx - \frac{(x-m)^2}{2\sigma^2} = -\frac{1}{2\sigma^2}\Big[(x-m)^2 - 2\sigma^2 t\, x\Big].$$

Since

$$(x-m)^2 - 2\sigma^2 t\, x = x^2 - 2(m+\sigma^2 t)x + m^2 = \big(x - (m+\sigma^2 t)\big)^2 - (m+\sigma^2 t)^2 + m^2,$$

we have

$$tx - \frac{(x-m)^2}{2\sigma^2} = -\frac{\big(x-(m+\sigma^2 t)\big)^2}{2\sigma^2} + \frac{(m+\sigma^2 t)^2 - m^2}{2\sigma^2}.$$

Hence,

$$M_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{(m+\sigma^2 t)^2 - m^2}{2\sigma^2}\right) \int_{-\infty}^{+\infty} \exp\left(-\frac{\big(x-(m+\sigma^2 t)\big)^2}{2\sigma^2}\right) dx.$$

Make the change of variable $u = x - (m+\sigma^2 t)$. Then $du = dx$ and

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{\big(x-(m+\sigma^2 t)\big)^2}{2\sigma^2}\right) dx = \int_{-\infty}^{+\infty} \exp\left(-\frac{u^2}{2\sigma^2}\right) du = \sigma\sqrt{2\pi},$$

by the Gaussian integral. Thus,

$$M_X(t) = \exp\left(\frac{(m+\sigma^2 t)^2 - m^2}{2\sigma^2}\right).$$

Finally,

$$\frac{(m+\sigma^2 t)^2 - m^2}{2\sigma^2} = \frac{m^2 + 2m\sigma^2 t + \sigma^4 t^2 - m^2}{2\sigma^2} = mt + \frac{\sigma^2 t^2}{2}.$$

Therefore,

$$M_X(t) = \exp\left(mt + \frac{\sigma^2 t^2}{2}\right), \qquad \forall t \in \mathbb{R}.$$

In particular, the integral is finite for every $t \in \mathbb{R}$, so the MGF exists for all real t. ■

Proposition 4.5: Stability properties of the normal distribution

1 **Sum of two normal variables.** Let X_1 and X_2 be *independent* random variables. If

$$V \hookrightarrow \mathcal{N}\left(\mu, \sigma^2\right) \quad \text{and} \quad W \hookrightarrow \mathcal{N}\left(\nu, \gamma^2\right),$$

then

$$V + W \hookrightarrow \mathcal{N}\left(\mu + \nu, \sigma^2 + \gamma^2\right).$$

2 **Affine transformation.** Let $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$ and let $a, b \in \mathbb{R}$. Then

$$aX + b \hookrightarrow \mathcal{N}\left(a\mu + b, a^2\sigma^2\right).$$

Proof:

Thanks to Lemma 4.4, if $X \hookrightarrow \mathcal{N}(m, \sigma^2)$, then its moment generating function exists for all $t \in \mathbb{R}$ and is given by

$$M_X(t) = E\left(e^{tX}\right) = \exp\left(mt + \frac{\sigma^2 t^2}{2}\right).$$

Case 1 : Sum of two independent normal random variables.

Let $V \hookrightarrow \mathcal{N}(\mu, \sigma^2)$ and $W \hookrightarrow \mathcal{N}(\nu, \gamma^2)$ be independent random variables, and set $S = V + W$. Since V and W are independent, we have

$$M_S(t) = M_V(t) M_W(t).$$

Using the MGF of a normal distribution, we obtain

$$M_S(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \exp\left(\nu t + \frac{\gamma^2 t^2}{2}\right) = \exp\left((\mu+\nu)t + \frac{(\sigma^2+\gamma^2)t^2}{2}\right).$$

This is the MGF of a normal distribution with mean $\mu+\nu$ and variance $\sigma^2+\gamma^2$. By the uniqueness of moment generating functions, we conclude that

$$V + W \hookrightarrow \mathcal{N}(\mu+\nu, \sigma^2+\gamma^2).$$

Case 2 : Linear transformation of a normal random variable.

Let $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$ and let $a, b \in \mathbb{R}$. Consider the random variable $Y = aX + b$. Using the MGF transformation property, we have

$$M_Y(t) = M_{aX+b}(t) = e^{bt} M_X(at).$$

Substituting the MGF of X, we obtain

$$M_Y(t) = e^{bt} \exp\left(\mu(at) + \frac{\sigma^2 (at)^2}{2}\right) = \exp\left((a\mu+b)t + \frac{a^2\sigma^2 t^2}{2}\right).$$

This is the MGF of a normal distribution with mean $a\mu + b$ and variance $a^2\sigma^2$. By the uniqueness of moment generating functions, we conclude that

$$aX + b \hookrightarrow \mathcal{N}(a\mu + b, a^2\sigma^2).$$

■

1) Standard Normal Distribution

To any random variable X one can associate a standardized variable $\dfrac{X-m}{\sigma}$ with mean 0 and variance 1.

It is fairly easy to show that if we carry out this transformation on a normally distributed variable, the standardized variable is still normal but now with parameters 0 and 1.

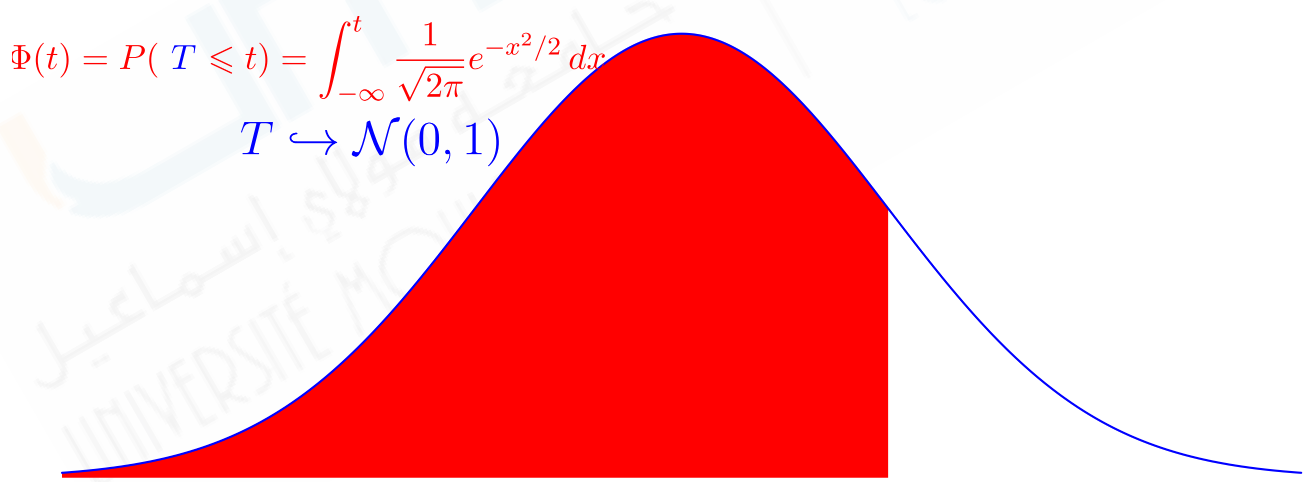
The standardized law is called the **standard normal distribution**, denoted $\mathcal{N}(0,1)$.

Thus, if $X \hookrightarrow \mathcal{N}(m, \sigma^2)$, set $T = \dfrac{X-m}{\sigma}$ and then $T \hookrightarrow \mathcal{N}(0,1)$.

We can summarize the correspondence as follows :

$X \hookrightarrow \mathcal{N}(m,\sigma^2)$ $E(X) = m$ $V(X) = \sigma^2$	$T = \dfrac{X-m}{\sigma}$	$T \hookrightarrow \mathcal{N}(0,1)$ $E(T) = 0$ $V(T) = 1$

The distribution $\mathcal{N}(0,1)$ is tabulated via the cumulative distribution function for positive arguments. It gives the values of $\Phi(t) = P(T \leqslant t)$ for $t > 0$. This number represents the area under the curve of the density above the interval $]-\infty, t]$.



1- a) Using the Standard Normal Table

The first column of Table 3.1 gives the units and tenths of the values of T, while the hundredths of T are read along the top row. The entry at the intersection of the appropriate row and column gives the desired area.

- Find $P(T \leqslant 0.5)$:
  at the intersection of row "0.5" and column "0.00", we read 0.6915.
- Find $P(-0.5 \leqslant T \leqslant 0)$.
  Using symmetry of the curve about the y–axis, we get $P(-0.5 \leqslant T \leqslant 0) = P(0 \leqslant T \leqslant 0.5) = \Phi(0.5) - \Phi(0) = 0.6915 - 0.5 = 0.1915$.
  And what about $P(-0.5 < T < 0)$?
- Find $P(-2.24 \leqslant T \leqslant 1.12)$.
  The required area is the following sum :

$$\begin{aligned} P(-2.24 \leqslant T \leqslant 1.12) &= P(-2.24 \leqslant T \leqslant 0) + P(0 < T \leqslant 1.12) \\ &= 0.4875 + 0.3686 = 0.8561. \end{aligned}$$

  Or equivalently :

$$\begin{aligned} P(-2.24 \leqslant T \leqslant 1.12) &= \Phi(1.12) - \Phi(-2.24) \\ &= \Phi(1.12) - 1 + \Phi(2.24) \\ &= 0.8686 + 0.9875 - 1 = 0.8561. \end{aligned}$$

  where we used $\boxed{\Phi(-t) = 1 - \Phi(t)}$.
- Find $P(1 \leqslant T \leqslant 2)$.
  The required area is the difference :

$$P(1 \leqslant T \leqslant 2) = \Phi(2) - \Phi(1) = 0.9772 - 0.8413 = 0.1359.$$

- Find t such that $P(0 \leqslant T \leqslant t) = 0.4750$.
  This is the inverse problem of the previous examples. Locate the given area in Table 3.1 and read off the corresponding value of T. We find $t = 1.96$. If the area is not listed exactly, use linear interpolation between adjacent entries or take the nearest value.

Table 3.1 – Table of the Standard Normal Distribution

t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

From the table, for $X \hookrightarrow \mathcal{N}(0,1)$, we read $\Phi(0.56) \overset{\text{def}}{=} P(X \leqslant 0.56) = 0.7123$.

V - Convolution

Convolution is a mathematical operation that allows us to determine the distribution of the sum of two independent random variables. For example, suppose that the amount of gold a company can mine in country A is X tons per year, and the amount of gold it can mine in country B is Y tons per year, independently.
Each quantity is modeled by its own probability distribution.
What is the distribution of the total amount of gold mined,

$$Z = X + Y\,?$$

Moreover, if the company's profit is given by a function of the total production, for instance

$$g(Z) = \sqrt{X+Y},$$

then convolution allows us to determine the probability density function of the profit as well.

This topic is best understood through examples.

Example 17 :

Let X and Y be independent random variables uniformly distributed on $\{1, 2, 3, 4\}$, corresponding to independent rolls of a fair four-sided die (see Example 7 in Chapter 2).
Determine the probability mass function of

$$Z = X + Y.$$

Solution :
We first determine the range of $Z = X + Y$. Since both X and Y take values in

$$X(\Omega) = Y(\Omega) = \{1, 2, 3, 4\},$$

it follows that

$$Z(\Omega) = \{2, 3, 4, 5, 6, 7, 8\}.$$

Are the probabilities of Z uniform on this set ?
In other words, are we equally likely to obtain $Z = 2$ and $Z = 5$?
The answer is no, because there is only one way to obtain $Z = 2$ (namely $(X, Y) = (1, 1)$), whereas there are several ways to obtain $Z = 5$.
For instance, to compute $P(Z = 3)$, we sum over all possible values of X :

$$P(Z = 3) = P(X = 1, Y = 2) + P(X = 2, Y = 1).$$

Using independence, this becomes

$$P(Z = 3) = P(X = 1)P(Y = 2) + P(X = 2)P(Y = 1)$$

Since X and Y are uniformly distributed on $\{1, 2, 3, 4\}$, we have

$$P(Z = 3) = \frac{1}{4} \times \frac{1}{4} + \frac{1}{4} \times \frac{1}{4}.$$

More generally, for any $z \in Z(\Omega)$, the probability mass function of Z is given by

$$p_Z(z) = P(Z = z) = \sum_{x \in X(\Omega)} P(X = x, Y = z - x) = \sum_{x \in X(\Omega)} P(X = x)P(Y = z - x),$$

that is,

$$\boxed{p_Z(z) = \sum_{x \in X(\Omega)} p_X(x)\, p_Y(z - x).}$$

The intuition is simple : to obtain the value $Z = z$, we sum over all possibilities $X = x$ such that $Y = z - x$. It is possible that $p_Y(z - x) = 0$ for some values of x, as seen above.
This formula is extremely general and applies to the sum of any two independent discrete random variables.
In the continuous case, the idea is exactly the same. If X and Y are independent continuous random variables with densities f_X and f_Y, and if $Z = X + Y$, then the density of Z is obtained by replacing probabilities with densities :

$$\boxed{f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)\, f_Y(z - x)\, dx.}$$

This integral formula is called the *convolution* of f_X and f_Y.

Theorem 5.1: Convolution

Let X and Y be independent random variables, and define

$$Z = X + Y.$$

1 **Discrete case.** If X and Y are discrete random variables with probability mass functions p_X and p_Y, then for any z,

$$p_Z(z) = \sum_{x \in X(\Omega)} p_X(x)\, p_Y(z - x).$$

2 **Continuous case.** If X and Y are continuous random variables with densities f_X and f_Y, then for any z,

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)\, f_Y(z - x)\, dx.$$

The roles of X and Y may be interchanged. Note the strong similarity between the discrete and continuous formulas.

Démonstration.

1 **Discrete case.** For any z,

$$p_Z(z) = P(Z = z) = \sum_{x \in X(\Omega)} P(X = x, Z = z)$$

by the law of total probability.
Since $Z = X + Y$, the event $\{X = x, Z = z\}$ is equivalent to $\{X = x, Y = z - x\}$, hence

$$p_Z(z) = \sum_{x \in X(\Omega)} P(X = x, Y = z - x).$$

Using independence,

$$p_Z(z) = \sum_{x \in X(\Omega)} P(X = x)P(Y = z - x) = \sum_{x \in X(\Omega)} p_X(x)\, p_Y(z - x).$$

2 **Continuous case.** Assume X and Y are independent continuous random variables with densities f_X and f_Y, and define $Z = X + Y$.

Fix $x \in \mathbb{R}$. Conditionally on $\{X = x\}$ we have

$$Z \mid (X = x) = x + Y.$$

A translation by x shifts the density, hence

$$f_{Z|X}(z \mid x) = \underbrace{f_{x+Y}(z)}_{\text{since } Y \text{ are independent}} = f_Y(z-x).$$

Using the product rule for densities, the joint pdf of (X, Z) is given by

$$f_{X,Z}(x,z) = \underbrace{f_X(x)}_{\text{marginal pdf of } X} \underbrace{f_{Z|X}(z \mid x)}_{\text{conditional pdf of } Z \text{ given } X},$$

and therefore

$$f_{X,Z}(x,z) = f_X(x)\, f_Y(z-x).$$

Finally, the marginal density of Z is obtained by integrating out x :

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{X,Z}(x,z)\, dx = \int_{-\infty}^{+\infty} f_X(x)\, f_Y(z-x)\, dx.$$

This proves the convolution formula on the continuous case.

$\square$

Example 18 : Sum of Two Poisson Random Variables (see Proposition 1.12)

Let $X \hookrightarrow \mathcal{P}(\lambda_1)$ and $Y \hookrightarrow \mathcal{P}(\lambda_2)$ be independent. Show that $Z = X + Y \hookrightarrow \mathcal{P}(\lambda_1 + \lambda_2)$.

Solution : For $n \in \mathbb{N}$, the convolution formula gives

$$p_Z(n) = \sum_{k=0}^{n} p_X(k) p_Y(n-k).$$

Substituting the Poisson PMFs,

$$p_Z(n) = \sum_{k=0}^{n} \frac{e^{-\lambda_1}\lambda_1^k}{k!} \frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}.$$

After algebraic simplification,

$$p_Z(n) = e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!},$$

which is the PMF of a $\mathcal{P}(\lambda_1 + \lambda_2)$ distribution.

Example 19 : Sum of Two Uniform$(0,1)$ Variables

Let X and Y be independent Unif$(0,1)$ random variables. Find the density of $Z = X + Y$.

Solution :

To be done, Incha'Allah, during the session of 7 January 2026.

Example 20 : Sum of Exponential and Uniform Variables (to be explained, Incha'Allah, during the session of 7 January 2026.)

Let $X \hookrightarrow \text{Exp}(\lambda)$ and $Y \hookrightarrow \text{Unif}(0,1)$ be independent.

Find the density of $Z = X + Y$.

Solution :

To be done, Incha'Allah, during the session of 7 January 2026.

VI - Weak Law of Large Numbers (WLLN)

In probability and statistics, we often observe random phenomena repeatedly and compute their empirical average. A fundamental question naturally arises :

Does the average of many independent observations stabilize around the true mean of the underlying distribution ?

For instance, when analyzing experimental data, simulating random processes, or estimating unknown parameters, we expect that averaging many observations reduces randomness and reveals the true expected value.

The **Weak Law of Large Numbers** provides a precise mathematical answer to this question. It states that the sample mean of independent and identically distributed random variables converges *in probability* to their common mean.

Definition 6.1: Sample mean

et $X_1, X_2, \ldots, X_n$ be random variables. The *sample mean* is defined by

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

We now state the theorem formally.

Theorem 6.2: Weak Law of Large Numbers

et $(X_n)_{n\geqslant 1}$ be a sequence of independent and identically distributed (iid) random variables with finite mean

$$E(X_i) = \mu \quad \text{and finite variance} \quad V(X_i) = \sigma^2 < \infty.$$

Let

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

denote the sample mean.
Then

$$\overline{X}_n \xrightarrow{P} \mu \quad \text{as } n \to \infty.$$

Proof (using Chebyshev's inequality)

Since the random variables X_i are iid, we have

$$E(\overline{X}_n) = \mu, \qquad V(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{\sigma^2}{n}.$$

Let $\varepsilon > 0$. By Chebyshev's inequality (see Proposition 5.1 in Chapter 2),

$$P\big(|\overline{X}_n - \mu| \geqslant \varepsilon\big) \leqslant \frac{V(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Letting $n \to \infty$, we obtain

$$\lim_{n\to\infty} P\big(|\overline{X}_n - \mu| \geqslant \varepsilon\big) = 0.$$

Therefore,

$$\overline{X}_n \xrightarrow{P} \mu.$$

□

Example 21 : Coin toss and the Weak Law of Large Numbers

Consider a fair coin toss with sample space

$$\Omega_0 = \{H, T\}, \qquad P_0(\{H\}) = P_0(\{T\}) = \frac{1}{2}.$$

Repeat the experiment infinitely many times. The resulting probability space is

$$\Omega = \Omega_0 \times \Omega_0 \times \cdots, \qquad P = \text{product measure}.$$

For each toss $k \geqslant 1$, define

$$X_k : \Omega \to \mathbb{R}, \qquad X_k(\omega) = \begin{cases} 1, & \text{if the } k\text{-th toss is } H, \\ 0, & \text{if the } k\text{-th toss is } T. \end{cases}$$

Thus, X_k records whether a head occurs at the k-th toss.

For the first n tosses, define

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k,$$

which represents the relative frequency of heads among the first n tosses.

The random variables X_k are independent and identically distributed, with

$$E(X_k) = \frac{1}{2}, \qquad V(X_k) = \frac{1}{4}.$$

By the Weak Law of Large Numbers,

$$\overline{X}_n \xrightarrow{P} \frac{1}{2} \quad \text{as } n \to \infty.$$

Interpretation. Although each coin toss is random, the proportion of heads stabilizes around $1/2$ when the number of tosses becomes large.

What to keep in mind

Consider repeatedly observing independent realizations $X_1, X_2, \ldots, X_n$ of the same random phenomenon. Each observation is random and may vary significantly. However, when we compute the sample mean

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

the fluctuations tend to cancel out.

As n increases :

- the distribution of $\overline{X}_n$ becomes increasingly concentrated around the true mean μ ;
- large deviations of $\overline{X}_n$ from μ become less and less likely, since

$$\frac{V(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0 \quad \text{as } n \to \infty;$$

- the Weak Law of Large Numbers formalizes this intuition by stating that

$$P\big(|\overline{X}_n - \mu| \geqslant \varepsilon\big) \longrightarrow 0 \quad \text{for every } \varepsilon > 0.$$

In other words, although individual observations remain random, their average becomes increasingly stable.

The lecture session of **7 January**, Incha'Allah, will have the following objectives.

- Proof of the **convolution theorem** ;
- Answering questions on the examples related to the convolution theorem and the WLLN ;
- Introduction to the **Central Limit Theorem (CLT)**.