MOHAMED KHIDER UNIVERSITY OF BISKRA.
FACULTY OF EXACT SCIENCES AND NATURAL AND LIFE SCIENCES
DEPARTMENT OF BIOLOGY

COURSE TITLE

Mathematics, statistics, computer science.

Intended for 1^{st} year undergraduate students
Field: Biology

Dr. AFROUN Faïrouz

Biskra university, 2025/2026

Contents

List of Figures

List of Tables

General introduction

Mathematics and statistics form the foundational pillars for analyzing and modeling phenomena across various scientific disciplines, particularly in the life sciences. They provide essential tools for understanding, describing, and predicting complex behaviors from experimental or observational data. In the study of functions and the analysis of biological data, two main areas are emphasized: **calculus** and **descriptive statistics**.

Calculus is a core component for studying functions and their variations. It enables the understanding of a function's behavior at a specific point or over an interval through concepts such as *continuity* and *differentiability*. Continuity ensures that the function does not exhibit breaks or discontinuities, which is crucial for modeling continuous biological phenomena, such as cell growth or population dynamics. Differentiation, on the other hand, measures the instantaneous rate of change of a function and plays a central role in analyzing rapid changes or optimizing biological processes.

The study of *integral calculus* complements this area by allowing the computation of the area under a curve or the accumulation of continuous quantities. This approach is particularly useful for determining global measures from local data, such as the total amount of a chemical substance produced in a biological reaction or the cumulative biomass of a population over time. The course also emphasizes practical methods for calculating integrals, providing students with tools to apply theoretical concepts to real-world problems.

In parallel, descriptive statistics is essential for the quantitative analysis of biological data. It allows for the *summarization, organization, and interpretation* of datasets, often large and complex, using measures such as mean, median, variance, and standard deviation. These metrics are fundamental for identifying trends, comparing groups, or assessing the variability of a biological phenomenon. An introduction to statistics in this context also prepares students for more advanced analyses, including statistical inference and probabilistic modeling.

Therefore, mastering **calculus tools** alongside the ability to **analyze and synthesize statistical data** equips students with a solid foundation to approach real scientific problems. This integrated approach is particularly suitable for studying biological systems, where the interaction between continuous variables and statistical measurements is common. Learning these fundamental concepts also lays the groundwork for advanced applications, ranging from mathematical modeling of biological processes to critical analysis of experimental data.

Chapter 1

Mathematical analysis

1.1 Functions of one variable, derivatives and integrals

In this chapter, we study two fundamental notions of calculus: **limits** and **continuity**. These concepts are essential for understanding how functions behave and change, which is useful in many areas of biology such as population dynamics, enzyme kinetics, and growth models.

What is a function?

A function is a relation that associates each element of a set called the starting set with an element of another set called the ending set.

$$f : D \to A$$
$$x \to f(x).$$

Définition 1.1 *Let f be a function on D_f.*

- *The function f is* ***odd*** *iff (if and only if) the following statements are correct.*
  1. *$\forall x \in D_f$ then $-x \in D_f$*
  2. *$\forall x \in D_f$ we have $f(-x) = -f(x)$.*
- *The function f is* ***even*** *iff the following statements are correct.*
  1. *$\forall x \in D_f$ then $-x \in D_f$*
  2. *$\forall x \in D_f$ we have $f(-x) = f(x)$.*
- *A* ***periodic*** *function is a function that repeats itself in regular intervals or periods. The function f is said to be periodic if $\forall\ x \in D_f\ \exists\ p \in \mathbb{R}^*$:*

$$f(x+p) = f(x)$$

.

Symbols	Explanation
$\forall$	for all
$\exists$	exists
$\in$	in
$\Rightarrow$	implies
$\Leftrightarrow$	equivalane
iff	if and only if

Limit of a Function

The **limit** of a function describes the value that the function approaches as the variable approaches a given number.

2.2 Formal Definition

Let $f(x)$ be defined around a. We say that $f(x)$ tends to a limit L as x tends to a, and we write:

$$\lim_{x \to a} f(x) = L$$

if for every $\varepsilon > 0$, there exists $\delta > 0$ such that:

$$|x - a| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

Left and Right Limits

- The **right limit** of f at a is $\lim\limits_{x \to a^+} f(x)$ (when $x > a$). - The **left limit** of f at a is $\lim\limits_{x \to a^-} f(x)$ (when $x < a$).

If both limits exist and are equal, then the limit at a exists.

2.4 Example

Let

$$f(x) = \begin{cases} x^2, & x < 2, \\ 3x - 2, & x \geq 2. \end{cases}$$

We have:

$$\lim_{x \to 2^-} f(x) = (2)^2 = 4, \quad \lim_{x \to 2^+} f(x) = 3(2) - 2 = 4$$

Therefore, $\lim_{x \to 2} f(x) = 4$.

Continuous Functions

Definition

A function f is said to be **continuous** at a point a if:

$$\lim_{x \to a} f(x) = f(a)$$

That is, the limit of $f(x)$ as x approaches a is equal to the actual value of f at that point.

Continuity on an Interval

A function f is continuous on an interval I if it is continuous at every point of I.

Examples

1. $f(x) = x^2$ is continuous everywhere on $\mathbb{R}$. 2. $f(x) = \frac{1}{x}$ is continuous on $\mathbb{R} \setminus \{0\}$.

Types of Discontinuity

- **Jump discontinuity:** when the left and right limits exist but are different.

$$f(x) = \begin{cases} 2x + 1, & x < 1 \\ x + 3, & x \geq 1 \end{cases}$$

  Then,

$$\lim_{x \to 1^-} f(x) = 3, \quad \lim_{x \to 1^+} f(x) = 4$$

  Since the two limits are not equal, f has a **jump discontinuity** at $x = 1$.

- **Infinite discontinuity:** when the function tends to infinity near a point. Example: $f(x) = \frac{1}{x}$ at $x = 0$.

- **Removable discontinuity:** when the limit exists but is not equal to $f(a)$. Example: $f(x) = \frac{x^2-1}{x-1}$ at $x = 1$.

The Intermediate Value Theorem (IVT)

Theorem Statement

If f is continuous on a closed interval $[a, b]$ and N is any number between $f(a)$ and $f(b)$, then there exists at least one $c \in [a, b]$ such that:

$$f(c) = N$$

Example

Let $f(x) = x^3 - x - 2$ on $[1, 2]$.

$$f(1) = -2, \quad f(2) = 4$$

Since 0 is between -2 and 4, by the IVT, there exists $c \in (1, 2)$ such that $f(c) = 0$. Numerically, $c \approx 1.52$.

Definition of Differentiability (One Variable)

Let f be a function defined in the neighborhood of x_0. We say that f is differentiable at a point x_0 if the limit.

$$f'(x) = \lim_{h \to 0} \frac{f(x) - f(x_0)}{x - x_0}. \tag{1.1}$$

If this limit exists and is finite, then the function f is differentiable at x_0.
we can also define the notion of differentiability of f at x_0 in the following way:

$$f'(x_0) = \lim_{h\to 0} \frac{f(x_0+h) - f(x_0)}{h}.$$

If this limit exists, we say that the function f is **differentiable at** x_0.
Notations:
We can use the notations $f'(x_0), Df(x_0), \frac{df}{dx}(x_0)$ to designate the derivative of f at x_0.

Example

Consider the function

$$f(x) = x^2$$

We compute the derivative at the point a:

$$f'(a) = \lim_{h\to 0} \frac{(a+h)^2 - a^2}{h}$$

$$= \lim_{h\to 0} \frac{a^2 + 2ah + h^2 - a^2}{h}$$

$$= \lim_{h\to 0} \frac{2ah + h^2}{h}$$

$$= \lim_{h\to 0} (2a + h) = 2a$$

Therefore,

$$f'(a) = 2a.$$

Definition (Left and right derivative): Let f be a function defined on an interval containing the point a. We say that f is right-differentiable at x_0 iff:

$$f'_+(a) = \lim_{h\to 0^+} \frac{f(a+h) - f(a)}{h},$$

exists in R. This limit is denoted by $f'_+(x_0)$ and is called the right derivative of f at x_0.

The **left derivative** of f at the point a is defined by

$$f'_-(a) = \lim_{h\to 0^-} \frac{f(a+h) - f(a)}{h},$$

if this limit exists. This limit is denoted by $f'_-(x_0)$ and is called the right derivative of f at x_0.

A function f is **differentiable at the point** a if

$$f'_-(a) = f'_+(a)$$

In this case, the common value is called the derivative of f at a and is denoted by $f'(a)$.

Example

Consider the function

$$f(x) = |x|$$

At $x = 0$ we compute the two derivatives.
Right derivative:

$$f'_+(0) = \lim_{h \to 0^+} \frac{|h| - 0}{h} = \lim_{h \to 0^+} \frac{h}{h} = 1$$

Left derivative:

$$f'_-(0) = \lim_{h \to 0^-} \frac{|h| - 0}{h} = \lim_{h \to 0^-} \frac{-h}{h} = -1$$

Since

$$f'_+(0) \neq f'_-(0)$$

the function $f(x) = |x|$ is **not differentiable at** $x = 0$.

Geometrical Interpretation

Let f be a function defined on an interval and let a be a point in its domain. If the function is differentiable at a, then the derivative $f'(a)$ represents the **slope of the tangent line** to the curve $y = f(x)$ at the point $(a, f(a))$.

The derivative is defined by

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

This limit corresponds to the slope of the **secant line** passing through the points $(a, f(a))$ and $(a+h, f(a+h))$. When h approaches 0, the secant line approaches the **tangent line** to the curve at the point $(a, f(a))$.

The equation of the tangent line to the curve $y = f(x)$ at the point a is

$$y = f(a) + f'(a)(x - a)$$

Example

Proposition. Let f be a function differentiable at a point x_0. Then f is continuous at x_0.

Proof

Since f is differentiable at x_0, the following limit exists:

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

This means that the quotient

The figure below shows the graph of a function $y = f(x)$:

The ratio $\dfrac{f(x_0+h)-f(x_0)}{h} = \tan(\theta)$ is the slope of the straight line joining point $A(x_0, f(x_0))$ to point $B(x_0+h, f(x_0+h))$ on the graph. When $h \to 0$, this line tends towards the tangent (AC) to the curve at a point $A(x_0, f(x_0))$. So we get:

$$f'(x_0) = \lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h} = \tan(\alpha) = \frac{CD}{AD}$$

is the slope of the tangent to the curve at point $A(x_0, f(x_0))$.

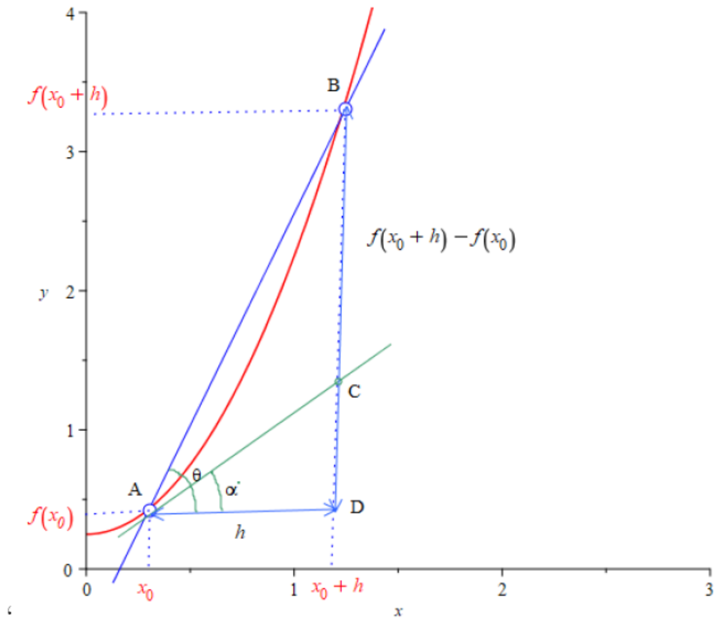

Figure 5.1: Geometrical Interpretation of Differentiability at a point x_0

$$\frac{f(x) - f(x_0)}{x - x_0}$$

has a finite limit as $x \to x_0$.
Now we write

$$f(x) - f(x_0) = (x - x_0)\frac{f(x) - f(x_0)}{x - x_0}.$$

Taking the limit as $x \to x_0$, we obtain

$$\lim_{x \to x_0} (f(x) - f(x_0)) = \left(\lim_{x \to x_0} (x - x_0)\right)\left(\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}\right).$$

Since

$$\lim_{x \to x_0} (x - x_0) = 0$$

and the second limit exists (because f is differentiable), we get

$$\lim_{x \to x_0} (f(x) - f(x_0)) = 0.$$

Therefore,

$$\lim_{x \to x_0} f(x) = f(x_0).$$

Hence, f is continuous at x_0.

□

1.1.0.0.1 Remark. The converse of this theorem is not true. A function can be continuous at a point x_0 without being differentiable at the same point.

For example, the function $x \mapsto |x|$ is continuous at $x_0 = 0$, but it is not differentiable at this point.

Mean Value Theorem

1) **Rolle's theorem**: Let f be a function such that

- f is continuous on the closed interval $[a, b]$,
- f is differentiable on the open interval (a, b).
- $f(a) = f(b)$

Then there exists a point $c \in (a, b)$ such that

$$f'(c) = 0.$$

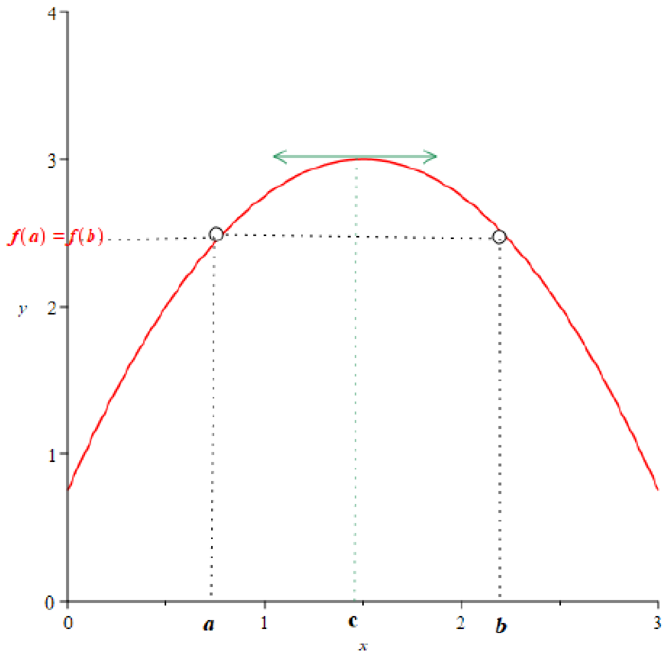

Figure 1.1: *Geometrical Interpretation of Rolle's theorem*

Geometrical Interpretation

2) **Mean Value Theorem**: Let f be a function such that

- f is continuous on the closed interval $[a, b]$,
- f is differentiable on the open interval (a, b).

Then there exists a point $c \in (a, b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Geometrical Interpretation

The theorem states that there exists at least one point c in the interval (a, b) where the tangent to the curve $y = f(x)$ is parallel to the secant line passing through the points $(a, f(a))$ and $(b, f(b))$.

Proof

Define the function

$$g(x) = f(x) - \left[\frac{f(b) - f(a)}{b - a}(x - a) + f(a)\right].$$

The function g is continuous on $[a, b]$ and differentiable on (a, b).
Moreover,

$$g(a) = f(a) - f(a) = 0$$

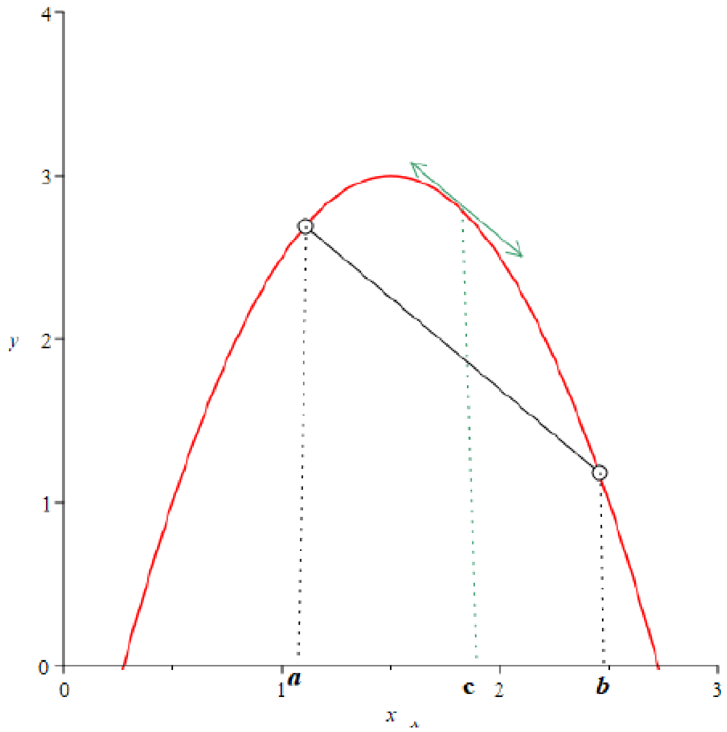

Figure 1.2: *Geometrical Interpretation of Mean Value Theorem.*

and

$$g(b) = f(b) - \left[\frac{f(b)-f(a)}{b-a}(b-a) + f(a)\right] = 0.$$

Thus

$$g(a) = g(b).$$

By Rolle's Theorem, there exists a point $c \in (a, b)$ such that

$$g'(c) = 0.$$

But

$$g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}.$$

Therefore,

$$f'(c) - \frac{f(b)-f(a)}{b-a} = 0$$

which gives

$$f'(c) = \frac{f(b)-f(a)}{b-a}.$$

□

Example

Consider the function

$$f(x) = x^2,$$

on the interval $[1, 3]$.

1)Verify the hypotheses of the Mean Value Theorem.
The function $f(x) = x^2$ is:

- continuous on $[1, 3]$,
- differentiable on $(1, 3)$.

Therefore, the Mean Value Theorem applies.
2) Compute the average rate of change:

$$\frac{f(3) - f(1)}{3 - 1} = \frac{9 - 1}{2} = 4$$

3) Compute the derivative:

$$f'(x) = 2x$$

4) Find c such that:

$$f'(c) = 4$$

$$2c = 4$$

$$c = 2$$

Conclusion
There exists a point $c = 2 \in (1, 3)$ such that

$$f'(2) = \frac{f(3) - f(1)}{3 - 1} = 4.$$

Thus, the Mean Value Theorem is verified.

Rules of Differentiation

Rule	**Formula**	**Examples**
Constant	$(c)' = 0$	$(3)' = 0$
Power	$(x^n)' = nx^{n-1}$	$(x^4)' = 4x^3$
Sum	$(f + g)' = f' + g'$	$(x^3 + x^2 + 3x)' = 3x^2 + 2x + 3$
Product	$(fg)' = f'g + fg'$	$(x^2 \cdot \sin x)' = 2x \cdot \sin x + x^2 \cdot \cos x$
Quotient	$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$	$\left(\frac{x^2}{\sin x}\right)' = \frac{2x \cdot \sin x - x^2 \cdot \cos x}{\sin^2 x}$
Chain Rule	$(f(g(x)))' = f'(g(x)) \cdot g'(x)$	$(\sin(x^2))' = \cos(x^2) \cdot 2x = 2x \cos(x^2)$

Common Derivatives:

Function	Derivative
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
e^x	e^x
$\ln x$	$\frac{1}{x}$

Hôpital's Rule

Theorem. Let f and g be functions differentiable on an open interval I containing a (except possibly at a itself), and suppose that

$$\lim_{x\to a} f(x) = 0 \quad \text{and} \quad \lim_{x\to a} g(x) = 0$$

or both limits are $\pm\infty$. If $g'(x) \neq 0$ on $I \setminus \{a\}$ and the limit

$$\lim_{x\to a} \frac{f'(x)}{g'(x)}$$

exists, then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

Example

Compute the following limit:

$$\lim_{x\to 0} \frac{\sin x}{x}.$$

1) Identify the indeterminate form:
As $x \to 0$, we have $\sin x \to 0$ and $x \to 0$, so the limit is of type $\frac{0}{0}$.

2) Apply Hôpital's Rule:

$$\lim_{x\to 0} \frac{\sin x}{x} = \lim_{x\to 0} \frac{(\sin x)'}{(x)'} = \lim_{x\to 0} \frac{\cos x}{1} = \cos 0 = 1.$$

Conclusion:

$$\lim_{x\to 0} \frac{\sin x}{x} = 1$$

L'Hôpital's Rule confirms the standard result.

Integration

Integration is one of the two main operations in calculus (the other is differentiation). While differentiation finds the *rate of change*, integration finds the *total quantity* or the *area under a curve.*

Integration is used in biology to calculate:

- Total bacterial growth over time,

- Total oxygen consumption,
- Accumulated concentrations of a substance.

If we know a function $f(x)$ that represents a rate (for example, growth rate), the **integral** of $f(x)$ between a and b gives the total change:

$$\text{Area under } y = f(x) \text{ between } x = a \text{ and } x = b.$$

The indefinite integral

In this section, we briefly present the definition of the indefinite integral and some of its properties.

1. The **indefinate integral** of $f(x)$ is a function $F(x)$ such that:
$$F'(x) = f(x),$$
and we write:
$$\int f(x)\,dx = \int F'(x)\,dx = F(x) + c,$$
where c is the constant of integration and $F(x) + c$ is called the antiderivative of $f(x)$.
2. for any real constant a we have
$$\int af(x)\,dx = a\int f(x)\,dx = aF(x) + c,$$
3. for any functions f and g, we have
$$\int (f(x) + g(x))\,dx = \int f(x)\,dx + \int g(x)\,dx$$
4. if we have
$$\int f(x)\,dx = F(x) + c,$$
then
$$\int f(ax+b)\,dx = \frac{1}{a}F(ax+b) + c, \text{ with } a \in \mathbb{R}^*.$$

The usual primitives

The following table groups together the usual antiderivatives that we need to know.

	function $f(x)$	anti derivative $F(x) + c$ (with c any real constant)
1.	x^n	$\frac{1}{n+1}x^{n+1} + c$ with $n \in \mathbb{R}$ and $n \neq -1$.
2.	$x^{-1} = \frac{1}{x}$	$\ln(\lvert x\rvert) + c$.
3.	e^x	$e^x + c$.
4.	a^x	$\frac{a^x}{\ln a} + c$, with $a \in \mathbb{R}_+$ and $a \neq 1$.
5.	$\sin(x)$	$-\cos(x) + c$.
6.	$\cos(x)$	$\sin(x) + c$.

7.	$\frac{1}{\sin^2(x)}$	$-ctan(x)+c.$
8.	$\frac{1}{\cos^2(x)}$	$\tan(x)+c.$
9.	$\frac{1}{\sqrt{1-x^2}}$	$\arcsin(x)+c=-\arccos(x)+c.$
10.	$\frac{1}{1+x^2}$	$\arctan(x)+c.$
11.	$\frac{1}{\sqrt{a^2-x^2}}$	$\arcsin\left(\frac{x}{a}\right)+c.$
12.	$\frac{1}{a^2+x^2}$	$\frac{1}{a}\arctan(\frac{x}{a})+c,$ with $a\in\mathbb{R}^*$
13.	$\frac{1}{a^2-x^2}$	$\frac{1}{2a}\ln\left(\left\lvert\frac{a-x}{a+x}\right\rvert\right)+c.$ with $a\in\mathbb{R}^*$
14.	$\frac{1}{\sqrt{x^2\pm a^2}}$	$\ln\left(\left\lvert x+\sqrt{a^2\pm x^2}\right\rvert\right)+c,$ with $a\in\mathbb{R}^*$.
15.	$\ln(x)$	$x\ln(x)-x+c.$
16.	$\log_a(x)$	$x\log_a(x)-x\log_a(e)+c,$ with $a>0.$
17.	$\tan(x)$	$-\ln(\lvert\cos(x)\rvert)+c.$
18	$\sinh(x)$	$\cosh(x)+c$
19	$\cosh(x)$	$\sinh(x)+c.$
20	$\tanh(x)$	$\ln(\cosh(x))+c.$
21.	$\frac{u'(x)}{u(x)}$	$\ln(\lvert u(x)\rvert)+c$
22.	$u'(x)e^{u(x)}$	$e^{u(x)}+c$
23.	$\frac{u'(x)}{\sqrt{u(x)}}$	$2\sqrt{u(x)}+c.$
24.	$u'(x)(u(x))^n$	$\frac{1}{n+1}u(x)^{n+1}+c$ with $n\in\mathbb{R}$ and $n\neq-1.$
25.	$u'(x)\sin(u(x))$	$-\cos(u(x))+c.$
26.	$u'(x)\cos(u(x))$	$\sin(u(x))+c.$

Table 1.1: *List of common primitives*

Integration methods

Integration by substitution

The substitution method consists of choosing a new variable which allows the initial function, whose form is complex, to be transformed into a new form close to those listed in the Table 1.1.

Let $\int g(x)\,dx$ be an integral, of which g is not listed in Table 1.1.

Suppose that, we propose to calculate this integral using the substitution method. Then the first and most vital step is to be able to rewrite our integral in the following form:

$$\int g(x)\,dx = a\int f(h(x))h'(x)\,dx,\ \text{(with a is a real constant;} \tag{1.2}$$

and that the function f appears in Table 1.1.

In this case, we set $u=h(x)$, which gives $du=h'(x)dx$. Hence,

$$\int g(x)\,dx = a\int f(u)\,du = aF(u)+c. \tag{1.3}$$

At this stage, all that remains is to replace the variable u with $h(x)$ in the antiderivative of f. This allows us to obtain the expression of the integral we are looking for as a function of the initial variable, namely the variable x.

$$\int g(x)\,dx = aF(h(x))+c. \tag{1.4}$$

Integration by parts

The integration by parts, called also partial integration, is a process that finds the integral of a product of functions in terms of the integral of the product of their derivative and antiderivative. Mathematically can be expressed as follows:

$$\int u'(x)v(x)\,dx = u(x)v(x) - \int u(x)v'(x)\,dx.$$

It should be noted that this last equation is easy to prove. Indeed, it follows directly from combining the definition of the derivative of the product of two functions with the definition of an integral.

$$\begin{aligned}(u(x)v(x))' = u'(x)v(x) + u(x)v'(x) \quad &\Longrightarrow \int (u(x)v(x))'\,dx = \int u'(x)v(x) + u(x)v'(x)\ dx\\ &\Longrightarrow u(x)g(x) = \int u'(x)v(x)\ dx + \int u(x)v'(x)\ dx\\ &\Longrightarrow \int u'(x)v(x)\,dx = u(x)v(x) - \int u(x)v'(x)\,dx.\end{aligned}$$

Integration of Rational Functions

Note that a rational function is a function that is the ratio of two polynomials. Any function of one variable, x, is called a rational function if it can be represented by $f(x) = p_n(x)/q_m(x)$, where $p_n(x)$ and $q_m(x)$ are polynomials of order $n \in \mathbb{N}$ and $m \in \mathbb{N}^*$ respectively, and that $q_m(x) \neq 0$.

In practice, many integrals involve rational expressions, but only a few can be solved directly. Indeed, calculating the integral of a rational function generally requires first decomposing the integrand into partial fractions.

Note that the partial fractions decomposition is a process that consists of taking a rational expression and decomposing it into simpler rational expressions that can be added together to obtain the original rational expression.

Let us analyze some specific cases of the rational function f. **Example 01**: To evaluate the following integral:

$$\int \frac{1}{(x^2+1)(x-1)}\,dx$$

we need to find constants A, B and C such that:

$$\frac{1}{(x^2+1)(x-1)} = \frac{A}{x-1} + \frac{Bx+C}{x^2+1}.$$

By multiplying both sides by$(x^2+1)(x-1)$, we obtain :

$$1 = A(x^2+1) + (Bx+C)(x-1).$$

Let's develop:

$$1 = A(x^2+1) + Bx(x-1) + C(x-1) = Ax^2 + A + Bx^2 - Bx + Cx - C.$$

Let's group similar terms together:

$$1 = (A+B)x^2 + (-B+C)x + (A-C).$$

By identifying the coefficients :

$$\begin{cases} A + B = 0, \\ -B + C = 0, \\ A - C = 1. \end{cases}$$

We solve the system :

$$C = B, \quad A = -B, \quad \Rightarrow A - C = -B - B = -2B = 1 \Rightarrow B = -\frac{1}{2}.$$

So :

$$A = \frac{1}{2}, \quad B = -\frac{1}{2}, \quad C = -\frac{1}{2}.$$

The decomposition into partial fractions is therefore:

$$\frac{1}{(x^2+1)(x-1)} = \frac{1/2}{x-1} - \frac{1}{2} \cdot \frac{x+1}{x^2+1}.$$

So:

$$\int \frac{1}{(x^2+1)(x-1)}\, dx = \frac{1}{2} \int \frac{dx}{x-1} - \frac{1}{2} \int \frac{x+1}{x^2+1}\, dx.$$

We separate the second integral:

$$\int \frac{x+1}{x^2+1}\, dx = \int \frac{x}{x^2+1}\, dx + \int \frac{1}{x^2+1}\, dx = \frac{1}{2} \ln(x^2+1) + \tan^{-1}(x).$$

Substitution in the integral :

$$\int \frac{1}{(x^2+1)(x-1)}\, dx = \frac{1}{2} \ln|x-1| - \frac{1}{2} \left(\frac{1}{2} \ln(x^2+1) + \arctan(x) \right) + C.$$

Let's simplify :

$$\boxed{\int \frac{1}{(x^2+1)(x-1)}\, dx = \frac{1}{2} \ln|x-1| - \frac{1}{4} \ln(x^2+1) - \frac{1}{2} \arctan(x) + C.}$$

Example 02: To evaluate the following integral:

$$\int \frac{x^3 + 2x^2 + 3}{x^2 + 1}\,dx$$

Perform long division:

$$\frac{x^3 + 2x^2 + 3}{x^2 + 1} = x + 2 + \frac{-x + 1}{x^2 + 1}.$$

Then integrate term by term:

$$\int \left(x + 2 + \frac{-x + 1}{x^2 + 1} \right) dx = \frac{x^2}{2} + 2x + \int \frac{-x}{x^2 + 1}\,dx + \int \frac{1}{x^2 + 1}\,dx.$$

$$\int \frac{-x}{x^2 + 1}\,dx = -\frac{1}{2}\ln(x^2 + 1), \qquad \int \frac{1}{x^2 + 1}\,dx = \arctan(x).$$

$$\boxed{\int \frac{x^3 + 2x^2 + 3}{x^2 + 1}\,dx = \frac{x^2}{2} + 2x - \frac{1}{2}\ln(x^2 + 1) + \arctan(x) + C.}$$

1.2 Approximation Methods

In biology, we often encounter functions whose exact values or integrals are difficult or impossible to compute analytically. **Approximation methods** allow us to replace exact values with approximate ones while keeping the error under control.

Examples:

- Estimating the growth of a population over a given time period.
- Calculating the total quantity of a substance secreted during a biological process.
- Approximating function values that cannot be expressed in a simple form.

Function Approximation Using Taylor Expansion

Principle

If a function f is differentiable several times near a point a, it can be approximated by a polynomial, called the **Taylor polynomial**:

$$f(x) \approx f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots$$

Example

For $f(x) = e^x$ around $a = 0$:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

Thus, if x is small:

$$e^x \approx 1 + x + \frac{x^2}{2}$$

Numerical Approximation of a Derivative

Finite Difference Formula

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$

where h is a small number.

Example

If $f(x) = \ln(x)$, for $x = 2$ and $h = 0.001$:

$$f'(2) \approx \frac{\ln(2.001) - \ln(2)}{0.001} \approx 0.5$$

The exact value is $f'(x) = \frac{1}{x} = 0.5$.

Numerical Integration Methods

Rectangle (Riemann Sum) Method

Divide the interval $[a, b]$ into n subintervals of width $h = \frac{b-a}{n}$ and approximate the area under the curve:

$$\int_a^b f(x)\,dx \approx h \sum_{i=0}^{n-1} f(x_i)$$

Trapezoidal Rule

$$\int_a^b f(x)\,dx \approx \frac{h}{2} \left[f(a) + 2 \sum_{i=1}^{n-1} f(x_i) + f(b) \right]$$

Simpson's Rule

$$\int_a^b f(x)\,dx \approx \frac{h}{3} \left[f(a) + 4 \sum_{i \text{ odd}} f(x_i) + 2 \sum_{i \text{ even}} f(x_i) + f(b) \right]$$

Biological Application

The growth rate of a bacterial population is given by $v(t) = 3t^2 + 2t$ (mg/h).

We want to estimate the total biomass produced between $t = 0$ and $t = 2$ hours.

Using the trapezoidal rule with $n = 2$:

$$h = 1, \quad t_0 = 0,\ t_1 = 1,\ t_2 = 2$$

$$v(0) = 0,\ v(1) = 5,\ v(2) = 16$$

$$\int_0^2 v(t)\,dt \approx \frac{1}{2}[0 + 2(5) + 16] = 13$$

Result: total biomass ≈ 13 mg.

1.3 Series, Positive-Term Series, and Riemann Series

Introduction to Numerical Series

A **numerical series** is the sum of the terms of a sequence (u_n):

$$\sum_{n=0}^{\infty} u_n = u_0 + u_1 + u_2 + \dots$$

Définition 1.2 *If the sequence of partial sums*

$$S_n = \sum_{k=0}^{n} u_k$$

*has a finite limit as $n \to \infty$, then the series **converges**:*

$$\sum_{n=0}^{\infty} u_n = \lim_{n\to\infty} S_n$$

*Otherwise, it **diverges**.*

Examples

- Geometric series: $\sum_{n=0}^{\infty} q^n$, $|q| < 1$: converges

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}$$

- Harmonic series: $\sum_{n=1}^{\infty} \frac{1}{n}$:diverges.

Positive-Term Series

A series $\sum u_n$ has **positive terms** if $u_n \geq 0$ for all n.

Convergence Criterion

For positive-term series:

$$\sum u_n \text{ converges} \iff \text{the partial sums } (S_n) \text{ are bounded.}$$

Comparison Method

- If $0 \leq u_n \leq v_n$ for all n and $\sum v_n$ converges, then $\sum u_n$ converges.
- If $\sum u_n$ diverges and $u_n \geq v_n \geq 0$, then $\sum v_n$ diverges.

Riemann Series

A **Riemann series** has the form:

$$\sum_{n=1}^{\infty} \frac{1}{n^\alpha}, \quad \alpha \in \mathbb{R}.$$

Convergence

- $\alpha > 1$: series converges.
- $\alpha \leq 1$: series diverges.

Examples

- $\sum_{n=1}^{\infty} \frac{1}{n^2}$: converges.
- $\sum_{n=1}^{\infty} \frac{1}{n}$: diverges.

1.4 Functions of Several Variables, Partial Derivatives, and Differentials

Functions of Several Variables

Definition

A **function of several real variables** assigns a unique value to each point in a domain of $\mathbb{R}^n$:

$$f : \mathbb{R}^2 \supset D_f \to \mathbb{R}, \quad (x, y) \mapsto f(x, y)$$

Domain and Image

- **Domain** D_f: all points (x, y) where $f(x, y)$ is defined.
- **Image**: all values $f(x, y)$ takes.

3. Examples

- $f(x, y) = x^2 + y^2$, domain: $\mathbb{R}^2$
- $f(x, y) = \sqrt{x - y}$, domain: $x \geq y$

Graphical Representation

- Graph is a surface in $\mathbb{R}^3$.
- Contour lines (level curves) visualize surfaces in biology (e.g., concentration levels, population density).

Partial Derivatives

Definition

Partial derivative with respect to x at (x_0, y_0):

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

Partial derivative with respect to y:

$$\frac{\partial f}{\partial y}(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}$$

Notation

$$f_x = \frac{\partial f}{\partial x}, \quad f_y = \frac{\partial f}{\partial y}, \quad f_{xx}, f_{yy}, f_{xy}, f_{yx} \text{ for higher-order derivatives.}$$

Examples

- $f(x, y) = x^2y + 3y^2 \Rightarrow f_x = 2xy,\ f_y = x^2 + 6y$
- $f(x, y) = e^{xy} \Rightarrow f_x = ye^{xy},\ f_y = xe^{xy}$

Differentials

1. Differential of a Function of Two Variables

If $f(x, y)$ is differentiable, the differential df gives an approximate change:

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy$$

Interpretation

- df gives a linear approximation of change in f near (x_0, y_0).
- In biology, it estimates small changes in population, concentration, or growth rate due to environmental changes.

Example

For $f(x, y) = x^2y$:

$$df = 2xy\,dx + x^2\,dy$$

If $x = 1$, $y = 2$, $dx = 0.1$, $dy = 0.05$:

$$df = 2(1)(2)(0.1) + (1)^2(0.05) = 0.4 + 0.05 = 0.45$$

1.5 Double and Triple Integrals

Double and triple integrals extend the concept of integration to functions of two or three variables, allowing the computation of:

- Areas under surfaces (double integrals)
- Volumes of 3D regions (triple integrals)

Biological applications:

- Estimating total biomass in a spatial region
- Calculating concentration of a substance in a volume
- Modeling spatial distribution of populations or molecules

Double Integrals

Definition

For a function $f(x, y)$ defined on a region $D \subset \mathbb{R}^2$, the double integral is:

$$\iint_D f(x, y)\, dA$$

where dA represents an infinitesimal area element.

Rectangular Region

If $D = [a, b] \times [c, d]$:

$$\iint_D f(x, y)\, dA = \int_a^b \int_c^d f(x, y)\, dy\, dx$$

Example

$f(x, y) = x + y$ over $[0, 1] \times [0, 2]$:

$$\iint_D (x + y)\, dA = \int_0^1 \int_0^2 (x + y)\, dy\, dx = \int_0^1 (2x + 2)dx = [x^2 + 2x]_0^1 = 3$$

Triple Integrals

Definition

For $f(x, y, z)$ in a 3D region $V \subset \mathbb{R}^3$:

$$\iiint_V f(x, y, z)\, dV$$

where dV is an infinitesimal volume element.

Rectangular Box

If $V = [a, b] \times [c, d] \times [e, f]$:

$$\iiint_V f(x, y, z)\, dV = \int_a^b \int_c^d \int_e^f f(x, y, z)\, dz\, dy\, dx$$

Example

$f(x, y, z) = 1$ over $V = [0, 1] \times [0, 1] \times [0, 2]$:

$$\iiint_V 1\, dV = \int_0^1 \int_0^1 \int_0^2 1\, dz\, dy\, dx = 2$$

Biological Application

- Total mass or concentration in a 3D tissue volume - Total number of molecules in a cellular volume

Iterated Integrals and Fubini's Theorem

Double and triple integrals can often be computed as nested single integrals:

$$\iint_D f(x, y)\, dA = \int_a^b \left(\int_{g_1(x)}^{g_2(x)} f(x, y)\, dy \right) dx$$

$$\iiint_V f(x, y, z)\, dV = \int_a^b \int_{g_1(x)}^{g_2(x)} \int_{h_1(x,y)}^{h_2(x,y)} f(x, y, z)\, dz\, dy\, dx$$

1.6 Surface and Volume Calculation

Surface Area Calculation

Surface of a Parametric or Graph Surface

Let a surface $z = f(x, y)$ over a region $D \subset \mathbb{R}^2$. The surface area S is:

$$S = \iint_D \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}\, dA$$

Example

For $f(x, y) = x^2 + y^2$ over $D = [0, 1] \times [0, 1]$:

$$f_x = 2x, \quad f_y = 2y$$

$$S = \iint_D \sqrt{1 + (2x)^2 + (2y)^2}\, dx\, dy$$

Biological Applications

- Estimation of cell membrane surface
- Surface area of tissues or organs

Volume Calculation

Volume Under a Surface

For $z = f(x, y)$ over D:

$$V = \iint_D f(x, y)\, dA$$

Example

Volume under $z = x + y$ over $[0, 1] \times [0, 2]$:

$$V = \int_0^1 \int_0^2 (x + y)\, dy\, dx = 3$$

Volume of a 3D Region

For a 3D region $V \subset \mathbb{R}^3$:

$$\text{Volume} = \iiint_V 1\, dV$$

Example: Box $[0, 1] \times [0, 1] \times [0, 2] \Rightarrow V = 2$

Biological Applications

- Total mass or concentration in tissue volume
- Total number of molecules in a cellular volume

Cylindrical and Spherical Coordinates

Cylindrical Coordinates

$$x = r\cos\theta,\ y = r\sin\theta,\ z = z, \quad dV = r\, dr\, d\theta\, dz$$

Spherical Coordinates

$$x = \rho\sin\phi\cos\theta,\ y = \rho\sin\phi\sin\theta,\ z = \rho\cos\phi, \quad dV = \rho^2\sin\phi\, d\rho\, d\phi\, d\theta$$

Calculation	Formula	Integral Type	Biological Use
Surface area	$S = \iint_D \sqrt{1 + f_x^2 + f_y^2}\, dA$	Double integral	Cell membrane, tissue surface
Volume under surface	$V = \iint_D f(x, y)\, dA$	Double integral	Nutrient volume, fluid content
Volume in 3D	$V = \iiint_V 1\, dV$	Triple integral	Tissue, cellular volume
Cylindrical	$dV = r\, dr\, d\theta\, dz$	Triple integral	Cylindrical organs
Spherical	$dV = \rho^2\sin\phi\, d\rho\, d\phi\, d\theta$	Triple integral	Spherical cells or organs

Chapter 2

Probabilities

This chapter will present the fundamental concepts of probability theory (random variables, probability density functions, cumulative distribution functions, expected value, etc.). It concludes with the presentation of some commonly used probability distributions.

2.1 Random Variable and Bernoulli Variable

Probability theory deals with random experiments and random phenomena, that is, experiments or natural phenomena which, under determined and stable conditions, do not always lead to the same outcome. However, a certain statistical regularity can be observed. The study of this regularity is the subject of a mathematical theory. **Examples:**

- Tossing a coin ⇒ outcomes: *Heads* or *Tails*.
- Testing for the presence of a gene ⇒ *Present* or *Absent*.
- Measuring the weight of bacteria ⇒ a numeric value that varies.

Random Variable

A random variable ($r.v.$) is a real number associated with the outcome of an experiment, and thus a random number. If the experiment is repeated, this number generally changes.

Exemple 1 *The height of an individual randomly selected from a population, or the number of "heads" in a series of 10 coin tosses.*

Types of Random Variables

Discrete random variable : Takes a finite or countable number of values. **Example**: Number of living cells in a sample.
Continuous random variable: Takes any real value in an interval. **Example**: Weight of an organ, glucose concentration.

Discrete random variable

In general, if a discrete random variable X can take the values $x_1,\ x_2, \ldots, x_n$ with respective probabilities $p_1,\ p_2, \ldots, p_n$, we say that X has the *probability distribution* given by the set of pairs:

$$(x_1, p_1),\ (x_2, p_2), \ldots, (x_n, p_n)$$

Note that a probability distribution satisfies the following properties:

- $p_i \geq 0$ for all i;
- $\sum_i p_i = 1$.

A discrete probability distribution can be graphically represented using a bar chart or column diagram (similar to a frequency distribution in descriptive statistics).

Continuous random variable

To compute probabilities associated with a continuous random variable, we use its *probability density function*, which allows the calculation of the probability that X lies in an interval $[a, b]$. More precisely, the probability density (or density) of X is a function f_X such that:

1. $f_X(x) \geq 0$ for all x,
2. $\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$,
3. $P(a \leq X \leq b) = \int_a^b f_X(x)\, dx$.

When there is no ambiguity, we use the symbol f instead of f_X.

The Bernoulli Variable

A **Bernoulli variable** is a random variable that takes only two possible values:

$$X = \begin{cases} 1 & \text{if success occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

We say that X follows a Bernoulli distribution with parameter p:

$$X \sim \mathcal{B}(p)$$

where $p = P(X = 1)$ and $1 - p = P(X = 0)$.

Biological Example

Testing for a specific gene:

$$X = \begin{cases} 1 & \text{if the gene is present,} \\ 0 & \text{if the gene is absent.} \end{cases}$$

If $p = 0.3$, then:

$$P(X = 1) = 0.3, \quad P(X = 0) = 0.7$$

2.2 Statistical laws and biostatistical applications

A **statistical law** (or probability distribution) describes how the values of a random variable are distributed. It indicates which values are likely to occur and their probabilities.

Two main types:

- **Discrete laws:** variables taking countable values.
- **Continuous laws:** variables taking real values in an interval.

2.2.1 Discrete Probability Laws

A discrete random variable X takes a finite or countable number of values.

$$\sum_i P(X = x_i) = 1$$

Main Discrete Distributions

Distribution	Description	Example
Bernoulli	Two outcomes: success/failure.	Gene presence.
Binomial	Number of successes in n trials.	Infected animals in a group.
Poisson	Number of rare events in a time/space unit.	Mutations per cell.

Examples

- Bernoulli: $P(X = 1) = p$, $P(X = 0) = 1 - p$.
- Binomial: $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.
- Poisson: $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$.

2.2.2 Continuous Probability Laws

A continuous random variable X is defined by its density function $f(x)$ such that:

$$P(a \leq X \leq b) = \int_a^b f(x)dx, \quad \int_{-\infty}^{+\infty} f(x)dx = 1$$

Main Continuous Distributions

In the remainder of this section, we will briefly define the most frequently used univariate distribution models as approximate descriptions of real-world distributions in statistics. These models depend on parameters that must be estimated from the data; therefore, they are called parametric models.

The Gaussian (Normal) Distribution

A random variable X is said to have (or to follow) a standard normal distribution, also called the standard Gaussian distribution, denoted by $X \rightsquigarrow N(0,1)$, if its probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{2.1}$$

The graph of f is a "bell-shaped" curve (see Figure 2.1).

If X follows a $N(0,1)$ distribution, then $E(X) = 0$ and $Var(X) = 1$.

The cumulative distribution function of X is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt. \tag{2.2}$$

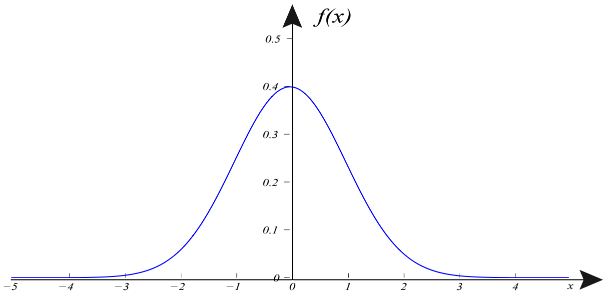

Figure 2.1: *Standard Gaussian distribution*

To determine the values of $F(x)$, numerical tables (see the Gaussian distribution table) or numerical integration methods are commonly used.

If $X \rightsquigarrow N(0,1)$, then the random variable $Y = \sigma X + \mu$ follows a Gaussian distribution with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$, and its density is given by

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)}. \tag{2.3}$$

Applying the previous transformation in reverse allows one to transform a random variable Y with distribution $N(\mu, \sigma^2)$ into the standardized variable

$$X = \frac{Y - \mu}{\sigma}, \tag{2.4}$$

which follows a $N(0,1)$ distribution. This transformation allows the computation of probabilities related to Y using the cumulative distribution function and tables of the $N(0,1)$ distribution.

One of the main results concerning the normal distribution is the Central Limit Theorem, summarized as follows:

Théorème 2.1 *(Central Limit Theorem, CLT)*
Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with an unknown distribution F_X, such that $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Then, as $n \to \infty$,

$$\frac{\left(\sum_{i=1}^{n} X_i/n\right) - \mu}{\sigma/\sqrt{n}} \rightsquigarrow N(0,1). \tag{2.5}$$

This theorem can be interpreted as follows: the standardized arithmetic mean is approximately Gaussian $N(0,1)$, regardless of the distribution F_X, provided that n is sufficiently large. The distribution of the arithmetic mean is approximately normal with mean μ and variance σ^2/n, and these parameters can be estimated. Unfortunately, there is generally no simple rule to determine the minimum value of n required for the approximation to be accurate, as it depends on the shape of F_X. However, in practice, one often assumes that $n \geq 30$ is sufficient.

The χ^2 (Chi-square) Distribution

Let $X_1, \ldots, X_n$ be n independent and identically distributed (*i.i.d.*) random variables following a standard normal distribution. The random variable

$$Z = X_1^2 + X_2^2 + \cdots + X_n^2$$

is said to follow a *chi-square distribution* with n *degrees of freedom*, denoted by χ_n^2. The probability density function of this distribution is

$$f(z) = \frac{z^{(n/2-1)}}{2^{n/2}\Gamma(n/2)} e^{-z/2}, \qquad z \geq 0, \tag{2.6}$$

where $\Gamma(\cdot)$ denotes the Gamma function, defined by

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx, \qquad p > 0. \tag{2.7}$$

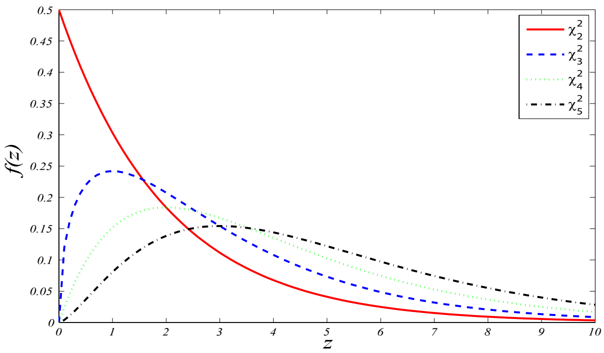

Figure 2.2: *Chi-square distribution for different degrees of freedom*

The cumulative distribution function is usually computed using computer software or chi-square distribution tables (see Tables). The mean and variance of the chi-square distribution are $E(Z) = n$ and $\sigma^2(Z) = 2n$, respectively.

Remarque 2.1 *Let Z_1 and Z_2 be two chi-square random variables with n and m degrees of freedom, respectively. Then the random variable $Z = Z_1 + Z_2$ also follows a chi-square distribution with $n + m$ degrees of freedom, i.e., $Z \rightsquigarrow \chi_{n+m}^2$.*

Student's t Distribution

Suppose that $X_0, X_1, \ldots, X_n$ are $n + 1$ independent and identically distributed random variables following a standard normal distribution. The random variable

$$T = \frac{X_0}{\sqrt{\frac{1}{n}(X_1^2 + \cdots + X_n^2)}} = \frac{X_0}{\sqrt{Z/n}}, \quad (Z \rightsquigarrow \chi_n^2) \tag{2.8}$$

is said to follow a t distribution (or Student's t distribution) with n degrees of freedom, denoted by t_n. Its density function is

$$f(t) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{n\pi}} (1 + t^2/n)^{-(n+1)/2}. \tag{2.9}$$

The cumulative distribution function is usually computed using computer software or t-distribution tables (see Tables). The mean and variance of the t distribution are $E(T) = 0$ and $\sigma^2(T) = n/(n-2)$ for $n > 2$.

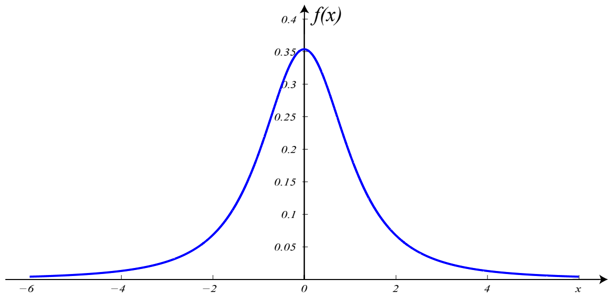

Figure 2.3: *Student's t distribution*

The Fisher (Fisher–Snedecor) F Distribution

Let $X_1, \ldots, X_{n+m}$ be $n+m$ independent random variables following a standard normal distribution. The random variable

$$Y = \frac{\frac{1}{n}(X_1^2 + \cdots + X_n^2)}{\frac{1}{m}(X_{n+1}^2 + \cdots + X_{n+m}^2)} = \frac{Z_1/n}{Z_2/m}, \quad (Z_1 \rightsquigarrow \chi_n^2,\ Z_2 \rightsquigarrow \chi_m^2) \tag{2.10}$$

follows an F distribution with n degrees of freedom in the numerator and m degrees of freedom in the denominator, denoted by $F_{(n,m)}$.

The probability density function of this distribution is

$$f(y) = \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} n^{n/2} m^{m/2} y^{(n/2)-1} (m+ny)^{-(n+m)/2}, \qquad y \geq 0. \tag{2.11}$$

Calculations involving the F distribution are generally performed using computer software or F distribution tables (see Tables). The mean of the F distribution is $E(Y) = \frac{m}{m-2}$ for $m > 2$, and its variance is

$$\sigma^2(Y) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}, \quad \text{for } m > 4.$$

Applications in Biostatistics

Domain	**Use of Probability Laws**
Genetics	Presence or absence of a gene (Bernoulli).
Epidemiology	Number of infected individuals (Poisson).
Physiology	Biological measurements (Normal).
Cell biology	Waiting times (Exponential).

2.3 Statistical Parameters and Properties

In biostatistics, it is important to describe a set of data (for example, the weight of individuals or the number of cells in a sample). To summarize and interpret these data, we use **statistical parameters** that describe:

- The **central tendency** (where the data are located),
- The **dispersion** (how spread out they are),
- The **shape** (how the distribution looks).

Parameters of Position

These parameters describe the central value around which data are distributed.

Mean (Arithmetic Mean)

The **mean** is the sum of all observed values divided by the number of observations:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

It represents the center of gravity of the data.

Example: Heights of 5 plants (cm): 12, 15, 18, 20, 25

$$\bar{x} = \frac{12 + 15 + 18 + 20 + 25}{5} = 18$$

Expectation (mean) of a random variable

Often, it is sufficient to characterize a distribution using a few numbers rather than the complete distribution. The most commonly used measures are the mean (or expectation), variance, standard deviation, and fractiles.

Let X be a discrete random variable with distribution (x_i, p_i), $i = 1, 2, \ldots, n$. The mathematical expectation (or expectation) of X is defined as:

$$\mu(X) = x_1 p_1 + x_2 p_2 + \cdots = \sum_i x_i p_i.$$

The notation $E(X)$ is also used instead of $\mu(X)$.

If X is a continuous random variable with density f, then the expectation of X is:

$$\mu(X) = \int_{-\infty}^{+\infty} x f(x)\, dx.$$

Proposition 2.1 *1. A very important property is that expectation is a linear operator. Let X and Y be two random variables and a and b be constants. Then:*

$$E(aX + bY) = aE(X) + bE(Y).$$

2. Another useful property concerns the expectation of a transformation of a random variable. Let g be a real-valued function and $Y = g(X)$. Then:

$$E(Y) = E(g(X)) = \sum_i g(x_i) P(X = x_i), \quad \textit{(discrete case)}, \tag{2.12}$$

$$E(Y) = E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x)\, dx, \quad \textit{(continuous case)}. \tag{2.13}$$

Median

The **median** is the middle value that divides the data set into two equal halves. It is not influenced by extreme values.

Example: Data: 10, 12, 15, 20, 60 ? Median = 15.

Mode

The **mode** is the most frequent value in the data set.

Example: Data: 2, 4, 4, 5, 7 ? Mode = 4.

Parameters of Dispersion

These parameters measure how much the data vary around the mean.

Range

$$\text{Range} = x_{\max} - x_{\min}$$

It is a simple but rough indicator of variability.

Variance

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

It measures how far each value is from the mean.

Standard Deviation

$$s = \sqrt{s^2}$$

It is expressed in the same units as the data and gives a direct idea of variability.

Example: If the standard deviation of plant heights is $s = 3$ cm, most plants are within $\bar{x} \pm 3$ cm of the mean.

Variance and standard deviation of a random variable

The expectation provides information about the average value of a random variable but does not account for variability. To measure variability, variance or standard deviation is used.

Let X be a random variable. The (population) variance of X is defined as:

$$\sigma^2(X) = E\big([X - E(X)]^2\big).$$

The (population) standard deviation of X is defined by:

$$\sigma(X) = \sqrt{\sigma^2(X)}.$$

Coefficient of Variation

$$CV = \frac{s}{\bar{x}} \times 100$$

It expresses variability relative to the mean (useful for comparing different variables).

Example: Mean = 20, $s = 4$? $CV = 20\%$.

Biological Application: The coefficient of variation is often used to evaluate variability in biological samples such as weight, enzyme activity, or measurement precision.

2.4 Cumulative distribution function and density function

In probability and biostatistics, random variables are used to represent uncertain biological quantities (such as cell size, weight, or enzyme concentration). To describe the behavior of these variables, two key functions are used:

- The **Cumulative Distribution Function (CDF)**, noted $F(x)$
- The **Probability Density Function (PDF)**, noted $f(x)$

These functions allow us to calculate probabilities and understand how data are distributed.

2.4.1 Cumulative Distribution Function

The **Cumulative Distribution Function** of a random variable X is defined by:

$$F(x) = P(X \leq x)$$

It gives the probability that X takes a value less than or equal to x.

Properties of the CDF

- $0 \leq F(x) \leq 1$
- $F(x)$ is always an **increasing function**
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to +\infty} F(x) = 1$

Example for a Discrete Variable

Let X be the number of living bacteria observed in a microscopic field, with:

$$P(X = 0) = 0.2, \quad P(X = 1) = 0.5, \quad P(X = 2) = 0.3$$

Then:

$$F(0) = 0.2, \quad F(1) = 0.7, \quad F(2) = 1$$

Interpretation: The probability of observing at most one bacterium is $F(1) = 0.7$.

2.4.2 Probability Density Function

Definition (for continuous variables)

The **Probability Density Function** (PDF) of a continuous random variable X is a non-negative function $f(x)$ such that:

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

Relation between PDF and CDF:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

and conversely:

$$f(x) = \frac{dF(x)}{dx}$$

Properties of the PDF

- $f(x) \geq 0$
- The total probability is 1:

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$

- The PDF gives a **density**, not a direct probability:

$$P(X = x) = 0 \text{ (for continuous variables)}$$

Example: Normal (Gaussian) Distribution

The **normal distribution** is one of the most common in biology (e.g., body weight, enzyme concentration).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- μ: mean (average)
- σ: standard deviation (spread)

Biological Applications

- The CDF allows us to calculate the probability that a biological measurement is below a threshold (e.g., glucose level under 1.2 g/L).
- The PDF allows us to model how a quantitative biological characteristic is distributed (e.g., weights of organisms, enzyme activity levels).
- Both are essential tools in biostatistics for modeling and interpreting biological data.

Bibliography

[1] James Stewart, *Calculus: Early Transcendentals*, 8th Edition, Cengage Learning, 2015.

[2] Sheldon M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, 5th Edition, Academic Press, 2014.

[3] Ron Larson, Bruce Edwards, *Calculus*, 10th Edition, Cengage Learning, 2017.

[4] John A. Rice, *Mathematical Statistics and Data Analysis*, 3rd Edition, Cengage Learning, 2007.

[5] Dennis Wackerly, William Mendenhall, Richard L. Scheaffer, *Mathematical Statistics with Applications*, 7th Edition, Cengage Learning, 2014.

[6] Howard Anton, Irl Bivens, Stephen Davis, *Calculus*, 10th Edition, Wiley, 2012.

[7] William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd Edition, Wiley, 1968.

[8] George Casella, Roger L. Berger, *Statistical Inference*, 2nd Edition, Duxbury, 2002.

[9] G. H. Hardy, *A Course of Pure Mathematics*, 10th Edition, Cambridge University Press, 2008.

[10] Jay L. Devore, *Probability and Statistics for Engineering and the Sciences*, 8th Edition, Cengage Learning, 2012.