People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

Echahid Hama Lakhdar El-Oued University

Faculty of Technology

ELECTRICAL ENGINEERING DEPARTMENT

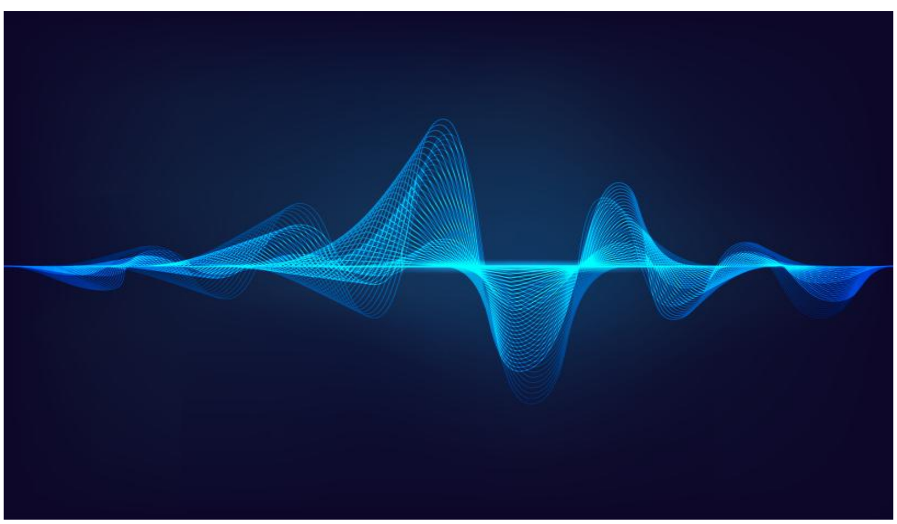
Field: Science and Technology

1st year Master, Specialization in Telecommunications systems

Course Handout for Teaching the Module:

RANDOM SIGNALS & STOCHASTIC PROCESSES

Dr. HADJADJI NARIMANE



2025 /2026

HADJADJI NARIMANE

Associate Professor (Class B)

in the Department of Electrical Engineering

Faculty of Technology

Echahid Hamma Lakhdar University,

El-Oued

E-mails: narimane.hadjadji@univ-eloued.dz,

COURSE INFORMATION

University: **Echahid Hamma Lakhdar EL-OUED**

Faculty: **of technology**

Department: **Electrical Engineering**

Target audience: **1st year Master, Specialty Telecommunications Systems**

Course title: **Random Signals and Stochastic Processes.**

Semester: **01**

Teaching unit: **Fundamental TUF**

Credit: **04**

Coefficient: **02**

Duration (Courses and TD): **15 weeks**

Teacher: **Dr. HADJADJI Narimane**

MODULE OBJECTIVES AND PRE-REQUISITES

Module Objectives

The competency targeted by this course, as a whole, is:

1. **Remember** the key concepts and properties related to random signals. The student is given multiple test questions and is asked to answer them. The goal is to recall his prior knowledge.
2. **Explain** the fundamental principles of random signals and stochastic processes.
3. **To Apply** advanced mathematical techniques for analyzing and processing random signal.
4. **Analysis** the statistical properties and performance.
5. **To design** and implement models for stochastic processes in telecommunication
6. **Evaluate** the effectiveness of methods based on practical scenarios.

Pre-requisites

To be able to successfully complete this course, it is recommended that learners have strong foundation in:

1. **Mathematics**: including basic calculus and linear Algebra such as, integration, vectors.
2. **Probability theory**: familiarity with probability distribution, expectations and random variables.
3. **Signal processing fundamentals**: as well as Fourier transform, filtering techniques.
4. **Programming skills**: proficiency in *MATLAB* and *Python* for simulation and data analysis (if required).

SUPPLEMENTARY RESOURCES

These condensed resources provide essential materials and online resources for mastering the prerequisite
topics needed for the module.

1. **Calculus and Linear Algebra**

- Textbook: "Introduction to Linear Algebra" by Gilbert Strang.
- Online Resource: Khan Academy - Calculus: https://www.khanacademy.org/math/calculus-1

2. **Probability and Statistics:**

- Textbook: "Introduction to Probability and Statistics" by William Mendenhall et al.
- Online Resource: Khan Academy - Probability and Statistics: https://www.khanacademy.org/math/statistics-probability

3. **Digital Signal Processing:**

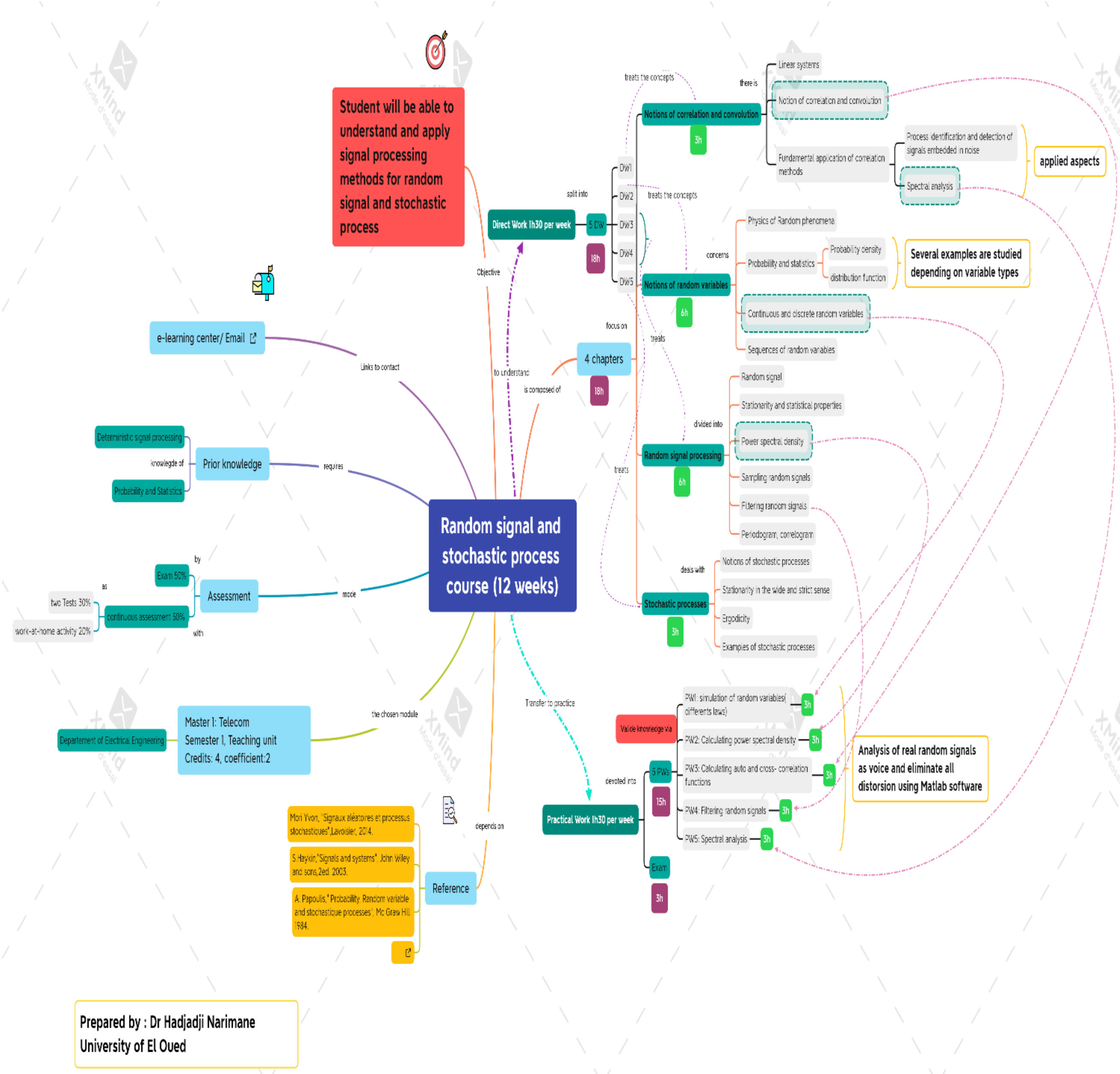
- Textbook: "Digital Signal Processing" by John G. Proakis and Dimitris G. Manolakis.
- Online Resource: DSP Related.com - Digital Signal Processing: https://www.dsprelated.com/tutorials.php , 11/10/2025.

4. **Signal Processing Software Tools:**

- MATLAB: https://www.mathworks.com/matlabcentral/fileexchange/27806-signal-processing. 11/10/2025.
- Python with NumPy and SciPy libraries: https://www.python.org/downloads/

5. [**cf.** Signaux_aléatoires_Mai_2015_résumé_Tous_Chapitres] : https://www.studypool.com/documents/17196643/signaux-al-atoires-mai-2015-r-sum-tous-chapitres

TECHNICAL CARD FOR THE MODULE



Conceptual Map of random signals and stochastic process module.

FOREWORD

Innovation in any field demands a constant pursuit of new knowledge and a deep understanding of the potential offered by emerging technologies. In the world of telecommunications, this is especially critical, given the sector's explosive and rapid development.

In signal processing, and particularly in random signal processing, digital techniques provide incredible possibilities for designing new systems. However, these techniques can be quite abstract, and applying them to real-world problems requires a solid theoretical foundation, which can be a significant hurdle.

This course module, ***Random signals and*** **stochastic process**, is designed for Master 1 students specializing in Telecommunications System at the University of El Oued. The primary goal is to provide a clear and concise overview of the key stochastic processing techniques, to compare their strengths and weaknesses, and to present the most useful results in a way that is immediately applicable.

The theoretical explanations have been kept to the essentials needed for a proper understanding and correct application of the concepts. For readers who wish to delve deeper, the bibliographic references provide a path to further study.

The handout is divided into **Four chapters**, organized and summarized as follows:

Chapter 1: Concepts of Correlation and Convolution

Chapter 2: Concepts of Random Variables

Chapter 3: Processing of Random Signals

Chapter 4: Stochastic Processes

TABLE OF CONTENTS

Chapter 1: Notions of correlation and convolution

Chapter 1: Notions of correlation and convolution

Course objective

At the end of this chapter, the student will be able to:

1. Understand the fundamentals of LTI systems and signal classification.

2. Master the mathematical operations of convolution and correlation.

3. Apply correlation methods for noise detection and system identification.

1.1 Recall Linear Systems

This section provides a fundamental review of Linear Time-Invariant (LTI) systems. A solid comprehension of these concepts is essential for understanding how random signals and stochastic processes are transformed, which is a core theme in telecommunications (e.g., signal filtering, channel modeling, and optimal reception).

The most important element of any communication system is to define the concepts of signal and system.

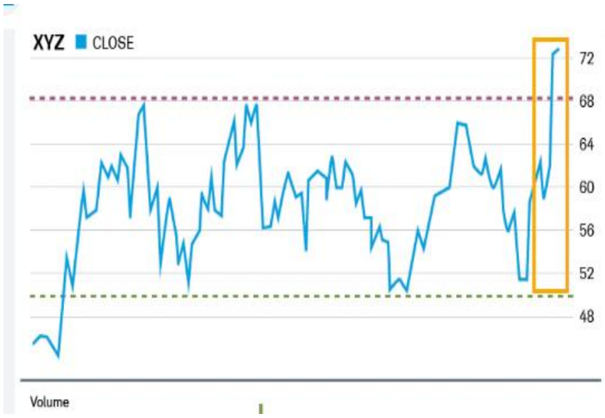
1.1.1 Definition of a Signal

A signal is a time-dependent function or quantity that conveys information or data. Signals are prevalent in various fields, including electronics, telecommunications, physics, engineering, biology, and more.

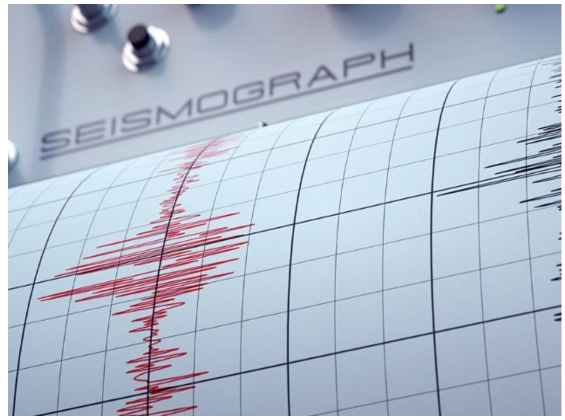
Examples:

1.1.2 Signal classification

- Analog or digital signals: Signals can be analog (continuous in amplitude) or digital (discrete in amplitude).
- Continuous or discrete signals: Signals can be continuous-time or discrete-time, depending on whether they are defined at every instant or only at



Finance: stock market price



Seismic vibration

Acoustic wave Sound

Figure 1.1: Different examples of random signals.

discrete time points.

- Random or deterministic signals: can be completely predictable or uncertainty.
- Periodic or non-periodic signals: repeats its pattern exactly after a specific time interval vs. no fixed period after which the signal repeats itself.
- Transient (finite energy) or permanent (finite power) signals: have a well-defined time duration (A single pulse in a radar signal or a camera flash) vs. periodic phenomena that persist indefinitely.

The purpose of the processing it undergoes is to extract information, modify the message it carries or adapt it to the means of transmission. The signal processing functions can be divided into two categories:

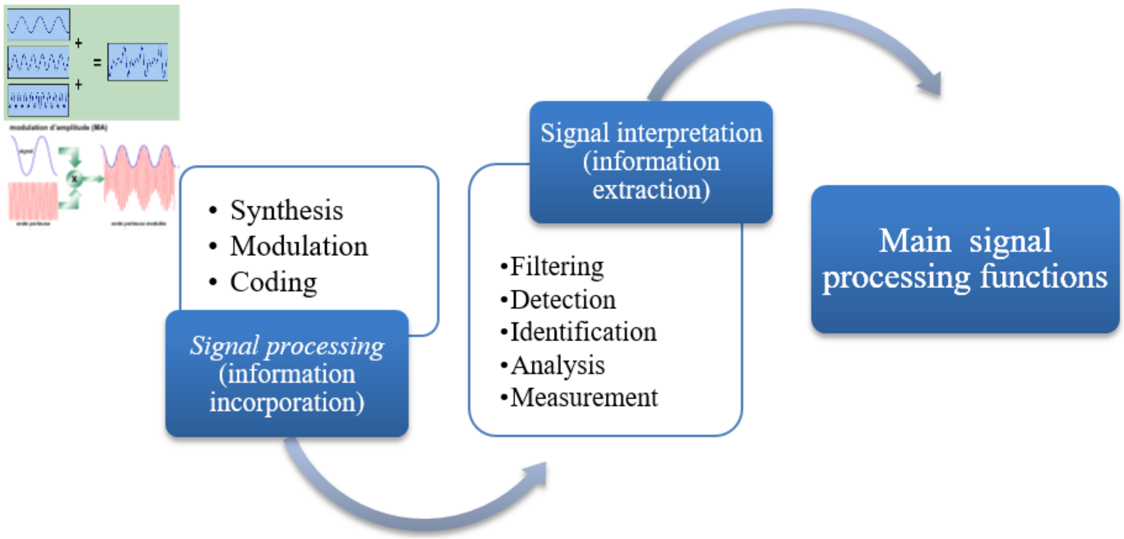


Figure 1.2: The signal processing functions.

Note: Every physical signal contains noise or a random component (external disturbance, noise, measurement error, etc.). **Noise** refers to any unwanted or random disturbance that interferes with the transmission, reception, or processing of signals or information. It is often characterized by its unpredictable and chaotic nature. Understanding and mitigating noise is crucial in many applications where signal quality and accuracy are essential.

Classic example of the telecom technician and the astronomer:

- For the telecom technician:
  - Waves from a satellite = signal.
  - signals from an astrophysical source = noise.
- For the astronomer:
  - Waves from a satellite = noise.
  - Signals from an astrophysical source = signal.

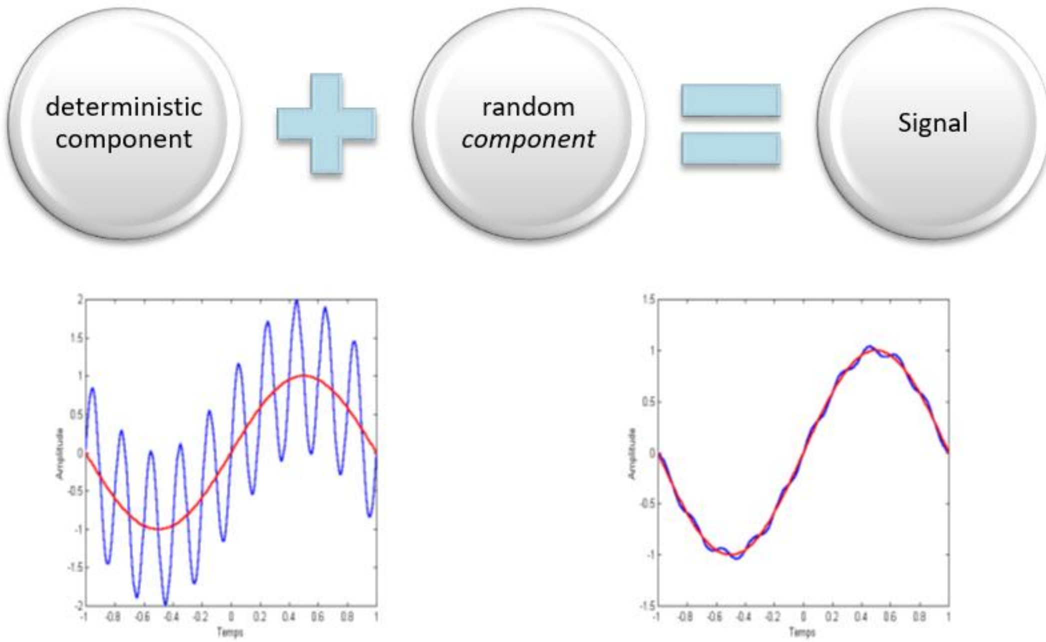


Figure 1.3: Noisy Signal example.

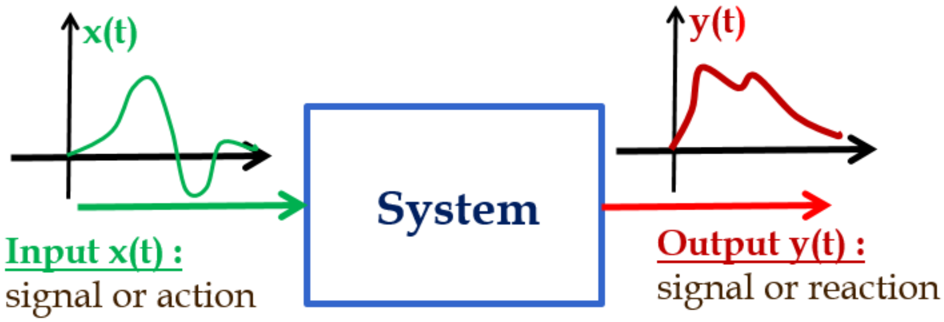
Those signals are often processed by diverse systems. So we can call a system a set of functional elements that interact with each other and establish a cause-and-effect relationship between its input signals and its output signals.

1.1.3 Definition of a System

In signal processing, a **system** is any physical device or algorithm that transforms an input signal x(t) (or x[n]) into an output signal y(t) (or y[n]) through a defined operation. We then say that y(t) is the effect of the system on x(t).

We represent this as: $\boldsymbol{y(t) = H\{x(t)\}}$

where H denotes the system operator.



$$y(t) = fct[x(t)]$$

Figure 1.4: System schema.

One can model (mathematically represent) a system with an equation linking the input x(t) to the output y(t). *fct* : designates a mathematical function.

1.1.4 Discrete and Continuous Linear Time-Invariant (LTI) System

A **Linear Time-Invariant (LTI)** system is a fundamental concept in signal processing and control theory, central to various engineering and scientific applications. It is defined by the combination of two essential properties: **Linearity** and **Time-Invariance**.

LTI systems can be broadly categorized into discrete-time and continuous-time systems, each with its own characteristics and mathematical representations. Here's a review of both types of LTI systems:

- Continuous-Time LTI Systems: are systems whose inputs and outputs are continuous-time signals, and are described using differential equations or transfer functions.
- Discrete-Time LTI Systems: are systems both the input and output signals are sequences defined at discrete time instances, and are described using the convolution sum or z-transform.

Practical Considerations: Continuous systems are typically associated with analog electronics, while discrete systems are central to digital computing and communication technologies.

1.1.5 Key Properties of Linear Time-Invariant (LTI) Systems

Linear systems are characterized by several key properties. These properties help define and understand how linear systems operate. Here are the main properties of linear systems:

a. **Linearity (Principle of Superposition)** : for a given system, $y_1(t)$ represents the output signal corresponding to the input signal $\alpha x_1(t)$, and $y_2(t)$ represents the output signal corresponding to the input signal $x_2(t)$, then for the input signal $x(t)= \alpha x_1(t) + \beta x_2(t)$, where α and β are constants, the system produces an output signal y(t) such that $y(t) = \alpha y_1(t) + \beta y_2(t)$.

$$\boldsymbol{H\{x_1(t) + x_2(t)\} = H\{x_1(t)\} + H\{x_2(t)\} = y_1(t) + y_2(t)}$$

This property of linearity allows us to break down complex systems into simpler parts and analyze them independently.

b. **Scaling (Homogeneity):** The response to a scaled input is the scaled version of the response to the original input. A system is homogeneous if it satisfies the property of scalar multiplication. In other words, if the input signal x(t) produces the output signal y(t), then $\lambda x(t)$ produces $\lambda y(t)$, where λ is a constant.

$$\boldsymbol{H\{a \cdot x(t)\} = a \cdot H\{x(t)\} = a \cdot y(t)}$$

Where a is a constant scalar. Combining these, for a linear system:

$$\boldsymbol{H\{a_1 x_1(t) + a_2 x_2(t)\} = a_1 x_1(t) + a_2 x_2 t)}$$

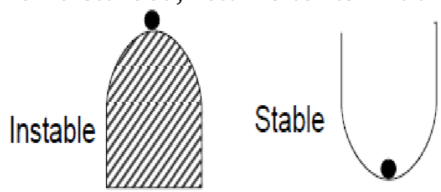
c. **Time-Invariance**: A system is time-invariant if a time shift or delay in the input signal results in an identical time shift in the output signal. The system's properties do not change over time.

Mathematically, if x(t) produces y(t), then $x(t - \tau)$ produces $y(t - \tau)$ for any time delay (τ). This property is crucial in understanding how systems respond to signals over time. This property expresses the fact that the system's characteristic does not depend on the origin of time, is still referred to as stationary.

If $\boldsymbol{y(t) = H\{x(t)\}}$, then for any time shift τ : $\boldsymbol{H\{x(t-\tau)\} = y(t-\tau)}$

d. **Zero Input, Zero Output (ZIZO) Response:** If the input signal is zero, i.e., x(t) = 0, the output response is also zero, i.e., y(t) = 0. This property allows us to explore and understand the inherent behavior of the system in the absence of external inputs.

e. **Stability**: A system is said to be stable if, in response to a bounded input, its output is bounded. A system which, when disturbed, returns to its initial state after the disturbance has disappeared.



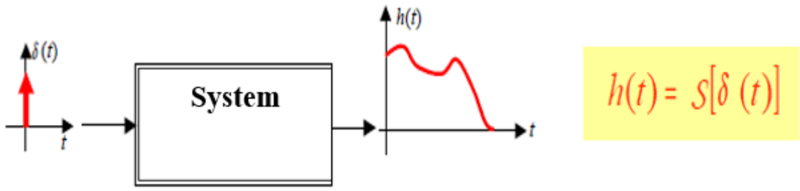
An **LTI system**, the bounded input and bounded output stability condition is equivalent to a requirement on the system's **impulse response**, *h(t)* or h[n].

- Continuous-Time LTI Systems: $\int_{-\infty}^{\infty} | \boldsymbol{h(t)} | \ \boldsymbol{dt} < \infty$
- Discrete-Time LTI Systems: $\sum_{\boldsymbol{n}=-\infty}^{\infty} | \boldsymbol{h[n]} | < \infty$

If an LTI system's impulse response decays quickly enough (e.g., is an exponentially decaying function), the system is stable. If the impulse response grows over time (e.g., an exponentially increasing function), the system is unstable.

1.1.6 The Impulse Response: A Fundamental Characterization

For an LTI system, the complete input-output relationship is uniquely defined by its response to a unit impulse.



- Continuous-Time: The impulse response, denoted h(t), is the output of the system when the input is a Dirac delta function δ(t).

$$\boldsymbol{h(t) = H\{\delta(t)\}}$$

- Discrete-Time: The impulse response h[n] is the output for a unit impulse δ[n].

$$\boldsymbol{h[n] = H\{\delta[n]\}}$$

Advantages of impulse response:

- Complete system characterization.
- Allows LTI system output to be calculated for other input signals.

1.1.7 Dynamic Filters and their Classification by Impulse Response

LTI systems are often called filters because they selectively attenuate or amplify different frequency components of the input signal. We classify them based on the duration and properties of their impulse response.

- **Finite Impulse Response (FIR) Filters** : The impulse response has a finite duration, i.e., it becomes exactly zero after a finite time. Typically implemented using a tapped-delay line (transversal filter). Always stable, and can be designed to have linear phase.
- **Infinite Impulse Response (IIR) Filters :** The impulse response has an infinite duration, theoretically never settling to zero exactly (though it may approach zero asymptotically). Example: Systems described by linear constant-coefficient differential/difference equations. The output depends on both past inputs and past outputs (recursive structure).

The function $H(\omega)$ is the **Transfer Function** (or **Frequency Response**) of the system, which is the Fourier Transform of the impulse response $h(t)$. It describes how the system scales and shifts the phase of different frequency components of the input signal.

Dynamic filters are typically classified based on their transfer function's magnitude response, $H(\omega)$:

- **Low-Pass Filter (LPF):** Passes low frequencies, attenuates high frequencies.
- **High-Pass Filter (HPF):** Passes high frequencies, attenuates low frequencies.
- **Band-Pass Filter (BPF):** Passes a specific band of frequencies, attenuates others.
- **Band-Stop Filter (BSF):** Attenuates a specific band of frequencies, passes others.

How to calculate the system output with any input signal from the **impulse response.!**

- notion of **Convolution**

1.2 Correlation and Convolution

Correlation and Convolution are basic operations that we will perform to extract information from signals. Moreover, because they are simple, they can be analyzed and understood very well, and they are also easy to implement and can be computed very efficiently.

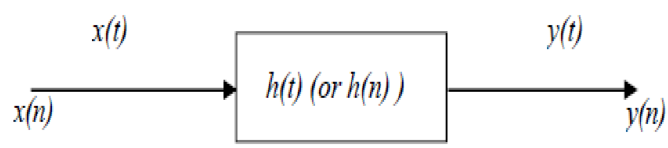
Our main goal is to understand exactly what correlation and convolution do, and why they are useful. We will also touch on some of their interesting theoretical properties. These operations have two key features: they are shift-invariant, and they are linear.

As seen before, Shift-invariant means that we perform the same operation at every point in the signal. Linear means that this operation is linear, that is, we replace every point with a linear combination of its neighbors.

1.2.1 Convolution

Convolution is a mathematical operation used to express the relation between input and output of an LTI system. Convolution is widely used in many fields like probability and statistics.

Convolution of the two signals produces a third signal. By using it, we can find the Zero state response. There are two types of convolutions:



- *Continuous Convolution*

It relates input, output and impulse response h(t) of an LTI system as:

$y(t) = x(t) * h(t)$, where y (t) = output of LTI, x (t) = input of LTI, h (t) = impulse response of LTI.

$$y(t) = x(t) * h(t) = \int_{-\infty}^{+\infty} x(\tau)h(t-\tau)d\tau \text{ or } \int_{-\infty}^{+\infty} x(t-\tau)h(\tau)d\tau$$

- *Discrete Convolution*

$$y(n) = x(n) * h(n) = \sum_{k=-\infty}^{+\infty} x(k)h(n-k) = \sum_{k=-\infty}^{+\infty} x(n-k)h(k)$$

1.2.1.1 Convolution Properties

If x(t), y(t), and z(t) are sequences, then the following useful properties of convolution is true:

- **Commutativity**

$$x(t) * y(t) = y(t) * x(t)$$

- **Distributivity**

$$x(t) * \big(y(t) + z(t)\big) = x(t) * y(t) + x(t) * z(t)$$

- **Associativity**

$$x(t) * y(t) * z(t) = \big(x(t) * y(t)\big) * z(t) = x(t) * \big(y(t) * z(t)\big)$$

- **Identity**

$$x(t) * \delta(t) = \delta(t) * x(t) = x(t)$$

- **Shifting (Delay)**

$$x(t) * \delta(t - t_0) = \delta(t - t_0) * x(t) = x(t - t_0)$$

$$x(t) * y(t - t_0) = z(t - t_0)$$

$$x(t - t_0) * y(t) = z(t - t_0)$$

$$x(t - t_0) * y(t - t_1) = z(t - t_0 - t_1)$$

- **Convolution of Unit Steps**

$$u(t) * u(t) = r(t)$$

$$u(t - T_1) * u(t - T_2) = R(t - T_1 - T_2)$$

$$u(n) * u(n) = [n + 1]u(n)$$

- **Scaling Property**

If $x(t) * y(t) = z(t)$ then $x(at) * y(at) = 1/|a|\, z(at)$

- **Differentiation of Output**

If $z(t) = x(t) * y(t)$ then $dz(t)dt = dx(t)dt * y(t)\ or\ dz(t)dt = x(t) * dy(t)dt$

Note:

- Convolution of two causal sequences is causal.
- Convolution of two anti causal sequences is anti causal.

- Convolution of two unequal length rectangles results a trapezium.
- Convolution of two equal length rectangles results a triangle.
- A function convoluted itself is equal to integration of that function.

Example: You know that $u(t) * u(t) = r(t)$

According to above note, $u(t) * u(t) = \int u(t)dt = \int 1dt = t = r(t)$

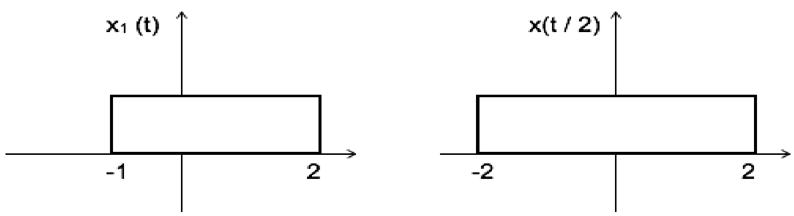
Here, you get the result just by integrating $u(t)$.

1.2.1.2 Limits of Convoluted Signal

If two signals are convoluted then the resulting convoluted signal has following range:

Sum of lower limits < t < sum of upper limits

Example: find the range of convolution of signals given below:



Here, we have two rectangles of unequal length to convolute, which results a trapezium.

The range of convoluted signal is: **Sum of lower limits < t < sum of upper limits**

$-1+(-2) < t < 2+2$ so $-3 < t < 4$, Hence the result is trapezium with period 7.

1.2.1.3 Area of Convoluted Signal

The area under convoluted signal is given by : $Ay = Ax \times Ah$

Where A_x is the area under input signal, A_h is the area under impulse response, and A_y is the area under output signal.

Proof: $y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau$ Take integration on both sides

$$\int y(t)dt = \int \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau = \int x(\tau)d\tau \int_{-\infty}^{\infty} h(t-\tau)dt$$

We know that area of any signal is the integration of that signal itself.

Example: what is area of the resultant convoluted signal given above?

Here area of $x_1(t)$ = length × width = 1 × 3 = 3, area of $x_2(t)$ = length × width = 1 × 4 = 4.

Area of convoluted signal = area of $x_1(t)$ × area of $x_2(t)$ = 3 × 4 = 12.

Duration of the convoluted signal = Period = 7.

1.2.1.4 Performing Convolution

Note: A useful fact to remember in performing the convolution of two finite-length sequences is that if x(n) is of length L_1, and h(n) is of length L_2, $x(n) * h(n) = y(n)$ will be of length of : $L = L_1 + L_2 - 1$.

L: number (no) of samples in output y(n).

L1: no of samples in output x(n).

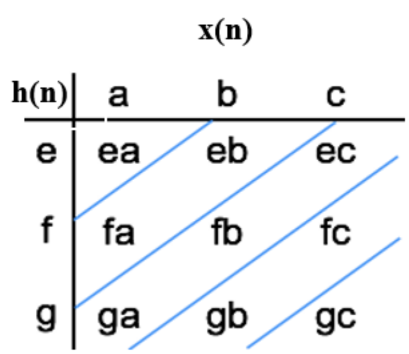
L2: no of samples in output h(n).

Let us see how to calculate ***discrete convolution***.

- **Convolution Table :**

Given x[n] = [a,b,c] and h[n] = [e,f,g], the convolution of the two signal scan be performed as the following procedure :

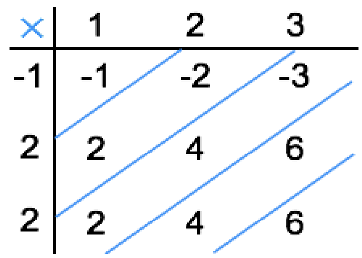
1. Write down the sequences x(n) and h(n) as shown in the table below.
2. Multiply each and every sample in h(n) with samples of x(n) and tabulate the values.
3. Group the elements in table by drawing diagonal lines as shown in table.
4. Starting from the left sum all the elements in each strip and write down in the same order.
5. Mark the symbol ↑ at time (n=0).



Convoluted output = [ea, eb+fa, ec+fb+ga, fc+gb, gc].

The resulting convoluted sequence is having [3+3-1] = 5 samples.

Example: Convolute two sequences x[n] = {1,2,3} & h[n] = {-1,2,2}.



Convoluted output y[n] = [-1, -2+2, -3+4+2, 6+4, 6] = [-1, 0, 3, 10, 6].

Here x[n] contains 3 samples and h[n] is also having 3 samples, so the resulting sequence having 3+3-1 = 5 samples.

Slide Rule Method

Another method for performing convolutions, which we call slide rule method, is practically convenient when both $x(n)$ and h(n) are finite in length. The steps involved in the slide rule method are as follows:

1- Placed the $x(n)$ signal.

2- Reverse the $h(n)$ signal.

3- Slide and shift the $h(-n)$ signal.

4- Multiply the $x(n)$ and the shifted $h(-n)$.

5- Sum all the values in same line.

………	x(-2)	x(-1)	x(0)	x(1)	x(2)	……
………	h(2)	h(1)	h(0)	h(-1)	h(-2)	……

Problem :

Use the slide rule to evaluate the convolution of x(n) and h(n) where x(n)=[1,1,1] and h(n)= [1,2,3].
↑

x(n)		1	1	1			
h(-n)	3	2	1				
3	2	1			y(-1)=1*1=1		
	3	2	1		y(0)=1*2+1*1=3		
		3	2	1	y(1)= 3*1+2*1+1*1=6		
			3	2	1	y(2)= 1*3+1*2=6	
				3	2	1	y(3)= 1*3=3

Y(t)=[1, 3, 6, 5, 3]
↑

1.2.1.5 Energy and power Signals

Any transmission of information is accompanied by a transfer of energy. Continuous or discrete signals or discrete signals are essentially characterized by the energy or power they carry. Many physical sensors measure energy or a quadratic quantity.

For example, optical sensors measure intensity, electricity meters measure electricity meters measure energy, etc.

The classification depends on whether the signal's total energy is finite or infinite.

- **Energy** is the total "effort" or "work" over all time.
- **Power** is the average "rate" of that effort.

a. Energy Signals

A signal is an **energy signal** if its total energy E is finite and non-zero ($0 < E < \infty$). Consequently, its average power *P*=0.

For a continuous-time signal $x(t)$:

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2 dt < \infty$$

$$\boldsymbol{P} = \lim_{T\to\infty} \frac{\mathbf{1}}{\mathbf{2T}} \int_{-\infty}^{+\infty} |x(t)|^2 dt = \lim_{T\to\infty} \frac{\mathbf{1}}{\boldsymbol{T}} \int_{-\frac{T}{2}}^{+\frac{T}{2}} |x(t)|^2 dt = 0$$

Let a discrete-time signal x(n), which $\sum_{-\infty}^{+\infty} |x(t)|^2$ exist and converge. Then the signal is said to have **finite energy** and the value of this sum is called the signal energy.

$$E_x = \sum_{-\infty}^{+\infty} |x(n)|^2 < \infty$$

$$\boldsymbol{P} = \lim_{N\to\infty} \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{N} + \mathbf{1}} \sum_{n=-N}^{+N} |x(n)|^2 = 0$$

Examples: Single pulse, exponentiel, often non-periodic signal.

b. Power Signals

A signal is a **power signal** if its average power *P* is finite and non-zero (0<*P*<∞). Consequently, its total energy *E*=∞.

For a continuous-time signal $x(t)$:

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2 dt < \infty$$

$$\boldsymbol{P} = \lim_{T\to\infty} \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{T}} \int_{-\infty}^{+\infty} |x(t)|^2 dt = \lim_{T\to\infty} \frac{\mathbf{1}}{\boldsymbol{T}} \int_{-\frac{T}{2}}^{+\frac{T}{2}} |x(t)|^2 dt = \infty$$

For a discrete-time signal $x(n)$:

$$E_x = \sum_{-\infty}^{+\infty} |x(n)|^2 < \infty$$

$$\boldsymbol{P} = \lim_{N\to\infty} \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{N} + \mathbf{1}} \sum_{n=-N}^{+N} |x(n)|^2 = \infty$$

Example*:* Continuous signal, constants (x(t)=a), A $\sin(2\pi f_0 t)$, periodic signals, unity step u(t), Dirac comb σ(t).

Note: There are signals that bare neither periodic nor of finite energy, for which the power cannot be defined such as the ramp x(n)=n. This signal type named **neither energy nor power.**

Finite energy signal ⇨ zero power.

Finite power signal ⇨ infinite energy.

1.2.2 Correlation

Correlation is a measure of similarity between two signals. If correlation is zero, then there is no similarity in between the two signals. If we have two signals $x_1(t)$ and $x_2(t)$, then the general formula of correlation is:

$$R(\tau) = \int_{-\infty}^{+\infty} x_1(t)x_2{}^*(t-\tau)dt = \int_{-\infty}^{+\infty} x_1(t-\tau)x_2{}^*(t)dt$$

There are two types of correlation:

- Auto correlation
- Cross correlation

1.2.2.1 Auto Correlation

a. Auto Correlation Function of Energy signals

Auto correlation function (ACF) gives a measure of match or similarity between a signal & its time delayed (shifted) version. It is represented with R(τ).

Consider a signals x(t) the auto correlation function of x(t) is given by:

$$R_{11}(\tau) = R(\tau) = \int_{-\infty}^{+\infty} x(t)x(t-\tau)dt = \int_{-\infty}^{+\infty} x(t)x(t+\tau)dt$$

Where τ = searching or scanning or delay parameter.

If the signal is complex then auto correlation function is given by

$$R_{11}(\tau) = R(\tau) = \int_{-\infty}^{+\infty} x(t)x^*(t-\tau)dt = \int_{-\infty}^{+\infty} x^*(t)x(t+\tau)dt$$

- **Auto-correlation Function properties of Energy Signal**

- Auto correlation exhibits conjugate symmetry i.e. $R\ (\tau)\ =\ R*(-\tau)$.
- Auto correlation function of energy signal at origin i.e. at τ=0 is equal to total energy of that signal, which is given as: $R(0) = E = \int_{-\infty}^{+\infty} |x(t)|^2 dt$
- Auto correlation function is maximum at τ=0 i.e. $|R\ (\tau)| \leq\ R\ (0)\ for\ all\ \tau.$
- Auto correlation function and energy spectral densities (ESD) are Fourier transform pairs. i.e.

$$F.T[R(\tau)] = \Psi(\omega)$$

$$\Psi(\omega) = \int_{-\infty}^{+\infty} R(\tau)e^{-j\omega\tau}d\tau$$

- $R(\tau) = x(\tau) * x(-\tau)$.

b. Auto Correlation Function of Power Signals

The auto correlation function of periodic power signal with period T is given by:

$$R(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-\infty}^{+\infty} x(t)x^*(t-\tau)dt$$

➢ **Auto-correlation Function properties of Power Signal**

- Auto correlation of power signal exhibits conjugate symmetry i.e. $R(\tau) = R^*(-\tau)$
- Auto correlation function at τ=0 (at origin) is equal to total power of that signal. i.e.

$$\boldsymbol{R(0) = P}$$

- Auto correlation function is maximum at τ= 0 and increase with decrease in τ and vice versa i.e.,

$$|R(\tau)| \leq R(0) \qquad \forall\tau$$

- Auto correlation function and power spectral densities (PSD) are Fourier transform pairs. i.e.,

$$F.T[R(\tau)] = S(\omega)$$

$$s(\omega) = \int_{-\infty}^{\infty} R(\tau)e^{-j\omega\tau}d\tau$$

- $R(\tau) = x(\tau) * x(-\tau)$

1.2.2.2 Cross Correlation Function

Cross correlation is the measure of similarity between one signal and the time delayed version of other signal. Consider two complex signals $x_1(t)$ and $x_2(t)$. The cross correlation of these two signals $R_{12}(\tau)$ is given by:

$$R_{12}(\tau) = \int_{-\infty}^{\infty} x_1(t){x_2}^*(t-\tau)dt = \int_{-\infty}^{\infty} x_1(t-\tau){x_2}^*(t)dt$$

$$R_{21}(\tau) = \int_{-\infty}^{\infty} {x_1}^*(t)x_2(t-\tau)dt = \int_{-\infty}^{\infty} {x_1}^*(t-\tau)x_2(t)dt$$

If we have two real signals $x_1(t)$ and $x_2(t)$, then

$$R_{12}(\tau) = \int_{-\infty}^{\infty} x_1(t)x_2(t-\tau)dt = \int_{-\infty}^{\infty} x_1(t-\tau)x_2(t)dt = R_{21}(\tau)$$

➢ **Cross Correlation Function properties of Energy and Power Signals**

- Auto correlation exhibits conjugate symmetry i.e. $R_{12}(\tau) = {R_{21}}^*(-\tau)$
- Cross correlation is not commutative like convolution i.e.

$$R_{12}(\tau) = R_{21}(-\tau)$$

- If $R_{12}(0) = 0$, then the two signals are orthogonal to each other.

$$R_{12}(0) = \int_{-\infty}^{\infty} x_1(t)x_2(t)dt$$

- Cross correlation function corresponds to the multiplication of spectrums of one signal to the complex conjugate of spectrum of another signal. i.e.

$$R_{12}(\tau) \xrightarrow{FT} X_1(\omega){X_2}^*(\omega)$$
$$R_{21}(\tau) \xrightarrow{FT} {X_1}^*(\omega)X_2(\omega)$$

This also called as ***correlation theorem***.

1.3 Application of the Correlation Concept to Physical Quantities

The mathematical definitions of autocorrelation and cross-correlation, while theoretically sound, find their true power and utility when applied to physical signals and systems. This chapter bridges the gap between abstract theory and practical application, demonstrating how correlation concepts become indispensable tools for measurement, estimation, and system analysis in telecommunications and signal processing. We will explore how these statistical measures reveal hidden information about signals, enable precise measurements, and facilitate the understanding of complex systems.

Field	Application	Function Used	Physical Quantity Analyzed
Signal Processing & Communica-tions	**Signal Detection/ Extraction**	Cross-Correlation	Extracting a known signal (like a radar pulse or a GPS code) from severe noise.
	Time-Delay Estimation	Cross-Correlation	Locating the source of an acoustic signal using two microphones, or measuring distances in sonar/radar.
	Synchronization	Cross-Correlation	Establishing symbol timing in digital communication systems.

Physics & Spectroscopy	**Fluorescence Correlation Spectroscopy (FCS)**	Autocorrelation	Measuring the concentration, diffusion coefficient, and molecular dynamics of fluorescent particles in a solution.
	Fluid Dynamics/Turbulence	Autocorrelation	Characterizing the coherence length and time scales of velocity fluctuations in turbulent flow.
Geophysics & Seismology	**Seismic Noise Correlation**	Cross-Correlation	Using background seismic noise recordings at two different stations to extract the impulse response of the Earth between them, aiding subsurface imaging.
Biomedical Engineering	**Noise Filtering/ Analysis**	Autocorrelation	Detecting and quantifying the presence of periodic noise (like 60-Hz line noise) in physiological signals (e.g., EMG, EEG).
	Movement Analysis	Cross-Correlation	Quantifying the delay in muscle activation (co-activation timing) between different muscles during movement or gait.
Finance/Econometrics	**Time Series Analysis**	Autocorrelation/ Cross-Correlation	Analyzing the dependence of a stock price on its past values, or the lead/lag relationship between two different markets.

1.4 Fundamental Applications of Correlation Methods

1.4.1 Identification of processes and detection of signals buried in noise

1.4.1.1. Radar and Sonar: Detecting Faint Echoes

Concept: In radar (Radio Detection and Ranging) and sonar (Sound Navigation and Ranging), a known signal (a "pulse") is transmitted. This pulse travels, reflects off a target, and an attenuated, delayed version (the "echo") returns to the receiver. The challenge is that the echo can be extremely weak, often buried deep within background noise. Cross-correlation is used to "match" the received signal with a template of the transmitted pulse, enhancing the signal and pinpointing the echo's arrival time.

Method:

1. A known radar pulse s(t) is transmitted.
2. The receiver picks up a signal x(t), which is primarily noise n(t) plus a very weak, delayed, and possibly attenuated version of the transmitted pulse, $A\,s(t-T_d)$, where A is the attenuation and T_d is the round-trip delay.
3. The received signal x(t) is cross-correlated with a stored copy (template) of the original transmitted pulse s(t).
4. Since the noise is uncorrelated with s(t), the cross-correlation of n(t) and s(t) will average to near zero.
5. However, the cross-correlation of $A\,s(t-T_d)$, and s(t) will produce a strong peak at a time shift equal to T_d. This peak indicates the presence of the echo and precisely measures the time delay.

Figure Explanation:

- **Transmitted Pulse s(t):** A clean, well-defined pulse (e.g., a chirp or a coded pulse). This is what the radar sends out.
- **Received Signal x(t) = Echo+ Noise :** This signal is dominated by random noise. A very faint, delayed version of the transmitted pulse (the "echo") is barely visible, if at all, to the naked eye within the noisy background.
- **Cross-Correlation x(t)*s(t):** This panel conceptually shows the process of "sliding" the template s(t) across the received signal x(t), multiplying point-by-point, and integrating.

- **Correlation Output:** The result of the cross-correlation. A very clear and strong peak emerges at the specific time delay T_d, indicating the precise arrival time of the echo. This peak's location is used to calculate the target's distance.

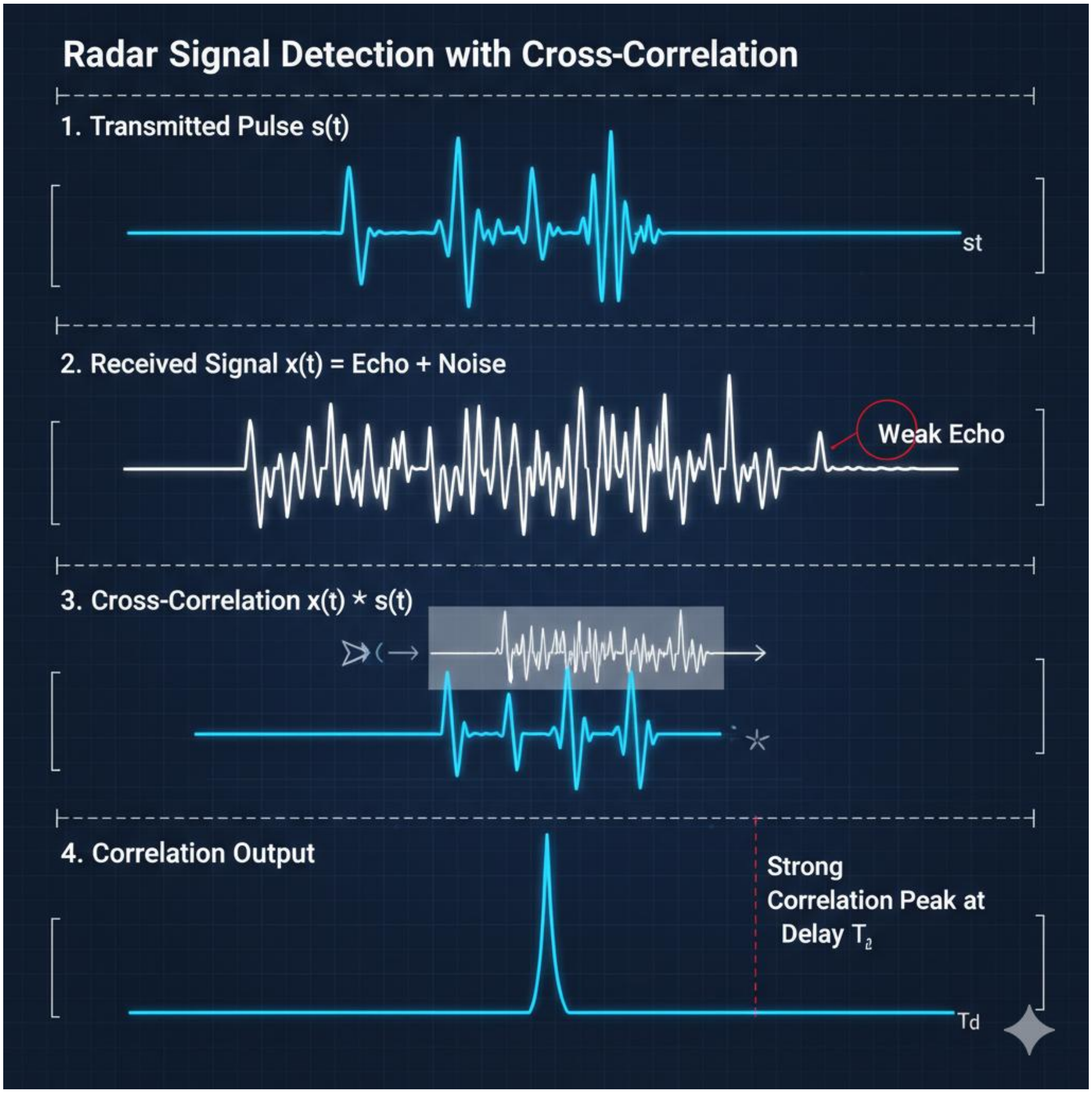


Figure 1.5: Cross-correlation for Radar signal detection.

1.4.2 Spectral analysis (by filtering, Fourier transform, correlation, spectral densities)

Concept: Spread Spectrum techniques deliberately "spread" a narrowband signal over a much wider bandwidth using a pseudo-random (PN) code. This makes the signal robust to interference and jamming, and also enables Code Division Multiple Access (CDMA), where multiple users can share the same frequency band simultaneously, each using a unique PN code. Correlation is the key to recovering the original data and distinguishing users.

Method (for Direct Sequence Spread Spectrum - DSSS):

1. **Spreading:** A narrow band data signal is multiplied by a high-rate PN code, "spreading" its energy across a wide bandwidth.
2. **Transmission:** The spread signal is transmitted along with other users' spread signals (in CDMA) and interference.
3. **De-spreading (Correlation):** At the receiver, the incoming wideband signal is multiplied by an identical, synchronized copy of the PN code used by the desired user.

Result:

- The desired user's signal is "de-spread" back into its original narrowband form, with a significant gain.
- Signals from other users (using different PN codes) and narrowband interference are **not correlated** with the desired user's PN code. Their energy remains spread, effectively being pushed into the background noise relative to the de-spread desired signal.

Figure Explanation:

1. **Spreading: Data x PN Code (Wideband Signal):**
    - **Left (Green):** A low-rate digital data bit (e.g., a "1").
    - **Middle (Blue):** A high-rate pseudo-random (PN) code (often called a "chipping sequence"). This code determines the spreading.
    - **Right (Blue, Spread Spectrum):** The result of multiplying the data bit by the PN code. The original data's energy is now spread across a much wider bandwidth, appearing as a complex, "noise-like" signal.
2. **Transmission/Reception: Mixed Signals + Interference:**

- **Left (Green):** The desired user's spread signal (User A).
- **Middle (Orange/Red):** Other users' spread signals (if CDMA) and/or strong narrowband interference (e.g., a strong tone) that might attempt to jam the communication. All these signals are mixed together, making it impossible to distinguish the desired signal without the correct PN code.
- **Right (Red):** A strong narrowband interferer.

3. **De-spreading: Received Signal x *Desired* PN Code (Correlation):**
    - **Left (Blue):** The received mixed signal.
    - **Middle (Blue/Orange/Red, "Correlation"):** The received signal is multiplied by a synchronized copy of the *desired user's* PN code (User A's PN code).
    - **Right (Green):** The result of this multiplication. The desired user's signal is "de-spread" back to its original narrowband form, appearing as a clean signal (a spike representing the data bit). The other users' signals and interference remain spread out, effectively becoming background noise to the desired signal.
4. **Filtering: Recovered Narrowband Data:**
    - A simple narrowband filter (e.g., a low-pass filter) is applied to the de-spread signal.
    - This filter easily passes the now-narrowband desired signal while filtering out the remaining spread energy from other users and interference.
    - The original data bit is recovered cleanly, with a high SNR.

These figures illustrate how correlation acts as a powerful "matched filter" to specifically detect and extract desired signals from a complex and noisy environment.

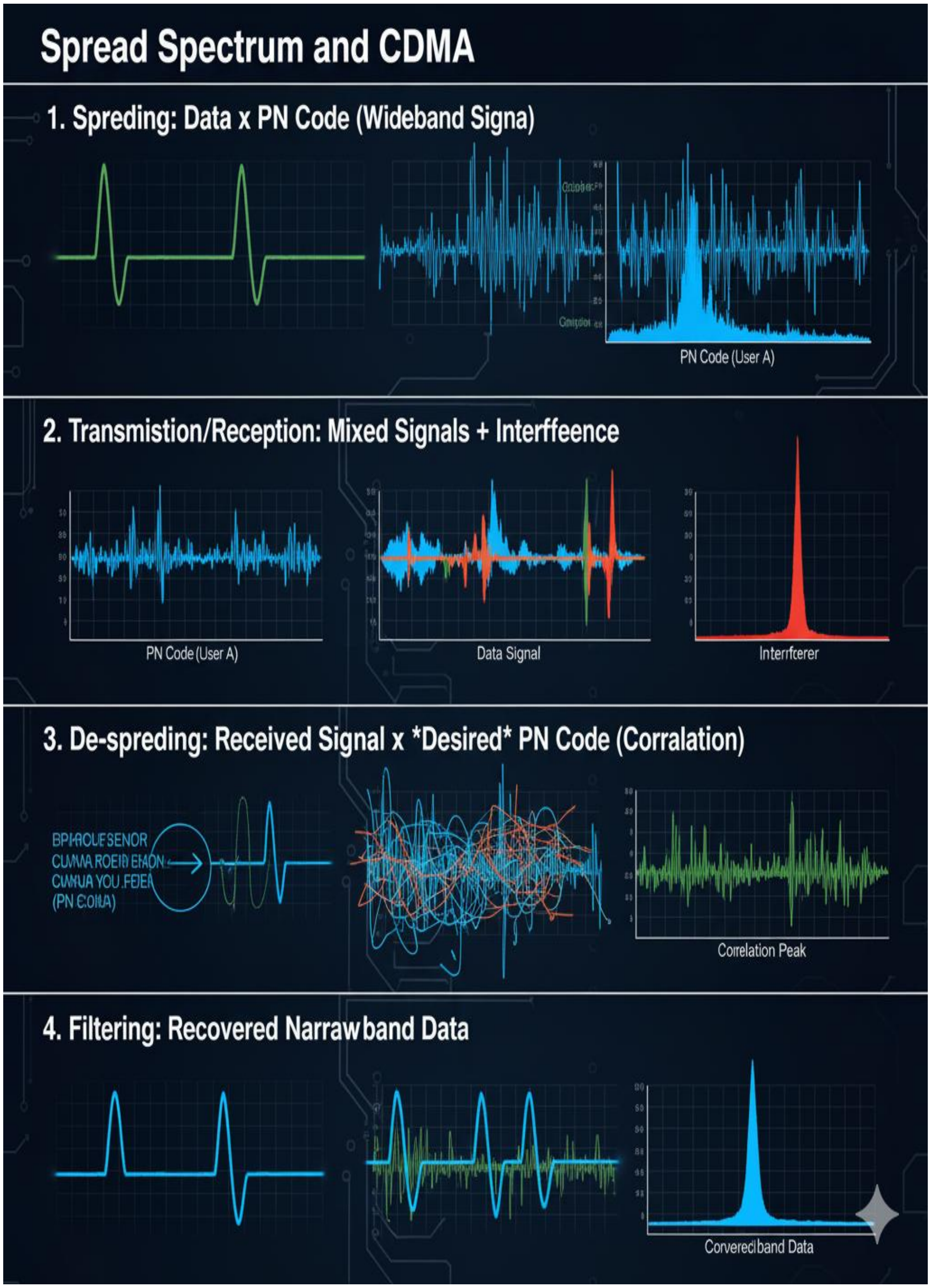


Figure 1.6: Spectrum signal.

EXERCISES

Exercise 1.1: Basic signals

Show graphically the functions: Unit ramp *r(t)*, Dirac impulse *σ(t)*, Unit step *u(t)*, causal sine *s(t)*.

Exercise 1.2: linear and non-linear systems

Check whether the following systems are linear or not.

- $y(t) = x(t^2)$, $y(t) = 2x(t)^2$, $y(t) = e^{x(t)}$, $y(t) = \int_{-\infty}^{t} x(\tau)d\tau$,
- $2\frac{dy(t)}{dt} + 5y(t) = x(t)^2$, $\frac{dy(t)}{dt} + y(t) = x(t)\frac{dx(t)}{dt}$, $\frac{d^2y(t)}{d^2t} + 2ty(t) = t^2x(t)$

Exercise 1.3: periodic and no periodic signal

Find whether the following signals are periodic or not, if periodic find the fundamental period.

- $x(t) = sin^2(400\pi t)$, $x(t) = \cos(2t) + \sin(3t)$

Exercise 1.4: Time invariant or variant.

Check whether the following systems are time invariant or variant.

- $y(t) = tx(t)$, $y(t) = t^2x(t)$,$y(t) = x(t)\sin(10\pi t)$, $y(t) = x(t)\cos(200\pi t)$, $\text{y(t)} = x(t)^2$, $y(t) = x(-2t)$, $y(t) = e^{2x(t)}$, $y(t) = \sin(x(t))$,
- $y(t) = \int_{-\infty}^{t} x(\tau)d\tau$

Exercise 1.5: Energy or power signals

Determine whether the following signals are energy or power.

- $x(t) = sin^2(w_0t)$, $x(t) = rect(\frac{t}{\tau})$, $x(t) = \sin(w_0t)rect(\frac{t}{\tau})$, $x(t) = A$
- $x(t) = u(t)$, $x(t) = tu(t)$, $x(t) = e^{j(3t+\frac{\pi}{2})}$

Exercise 1.6: convolution

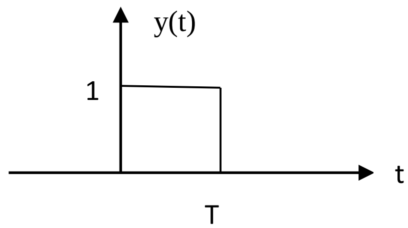
Calculate the convolution product of the functions:

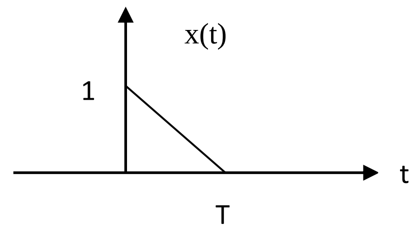
$$y_1(t) = rect(t) * rect(t)$$

$$y_2(t) = e^{-t}u(t) * e^{-2t}u(t)$$

Exercise 1.7:

1. Give the analytical expressions for convolution ($z(t) = x(t) * y(t)$) in the different regions of definition.





2. Represent the signal $z(t)$.

Exercise 1.8:

Find z(t) the convolution product of the signals using Slide and Shift method:

$x(t) = t^2 + 2t + 1, y(t) = t^2 + 3t + 4$

Exercise 1.9:

Find the convolution of the sequence $x(n) = [4,2,1,3], h(n) = [1,\underset{\uparrow}{2},2,1]$

Find the convolution of the sequence $x(n) = \begin{cases} 1 & 1 \leq n \leq 4 \\ 0 & otherwise \end{cases}$

With $y(n) = \begin{cases} n & -2 \leq n \leq 2 \\ 0 & otherwise \end{cases}$

Exercise 1.10:

Calculate autocorrelation functions of the signals:

$$x(t) = rect(t)$$

$$x(t) = \mu(t)$$

Calculate intercorrelation functions of the signals:

1. $x_1(t) = A_1 e^{2j\pi f_1 t}$ and $x_2(t) = A_2 e^{2j\pi f_2 t}$

Discuter by relative values of f_1 and f_2.

2. $y_1(t) = e^{-at}$ and $y_2(t) = A_2 e^{-2at}$

Where a is a positive real

Exercise 1.11:

Use the tabular method to evaluate the cross-correlation between x(n) and h(n) where

$x(n) = [1,1,1], h(n) = [1,\underset{\uparrow}{2},3]$

1.5 Conclusion

In summary, this chapter has established that correlation is far more than a theoretical statistic; it is a fundamental operational tool. The methods of cross-correlation and autocorrelation provide the definitive means to solve two of the most critical problems in signal processing: identifying the unknown and detecting the imperceptible.

By using white noise and cross-correlation, we can characterize any LTI system, extracting its impulse response from a noisy output. Furthermore, through the principle of the matched filter (a direct application of correlation) we can optimally discern known signals deeply buried in noise. Also, it is maximizing the signal-to-noise ratio and forming the bedrock of reliable digital communication, radar, and sonar systems.

Ultimately, these applications demonstrate that in a stochastic world, correlation provides the essential framework for transforming uncertainty into measurable, actionable information.

Chapter 2: Notions of random variables

Chapter 2: Notions of random variables

Course objective

Upon completion of this chapter, the student will be able to:

1. **Define** probability and explain its role as a formalism for modeling random experiments and quantifying uncertainty.
2. **Identify** and describe the concept of a random variable as a numerical representation of outcomes from random phenomena.
3. **Classify** random variables into their primary types: discrete and continuous.
4. **Understand** and utilize the core functions describing random variables: Probability Distribution Functions (PDFs) and Cumulative Distribution Functions (CDFs).
5. **Calculate** and interpret key statistical measures of random variables, including expected value and variance.
6. **Recognize** and apply common probability distributions (e.g., Binomial, Poisson, Normal) to relevant practical scenarios.
7. **Relate** the fundamental concepts of probability and random variables to their crucial applications in fields such as statistics, finance, engineering, and physics.

2.1 Introduction

In this chapter, the formalism of probability is introduce, where we explore the concept of **random variables**. These are numerical representations of outcomes from random experiments, allowing us to quantify uncertainty and make informed predictions.

Why Study Random Variables?

Understanding random variables is crucial for a variety of fields, including:

- **Statistics:** Analyzing data and drawing conclusions.
- **Finance:** Modeling stock prices and risk.
- **Insurance:** Assessing risk and setting premiums.
- **Engineering:** Designing reliable systems.
- **Physics:** Describing quantum mechanics.

So, let's dive into the world of random variables and unlock the secrets of probability!

What is probability?

Probability can be used to develop mathematical models of random and stochastic phenomena.

2.2 Physical Notion of Random Phenomena

This field uses the mathematical tools of probability and statistics to model and understand unpredictable physical systems, such as Brownian motion, thermal noise in circuits, or quantum measurements.

2.2.1 Deterministic vs. Random

In classical mechanics, if you know the initial conditions and forces, you can predict the future state exactly (deterministic). In random phenomena (e.g., radioactive decay, thermodynamics), this is impossible. We can only assign probabilities to different outcomes.

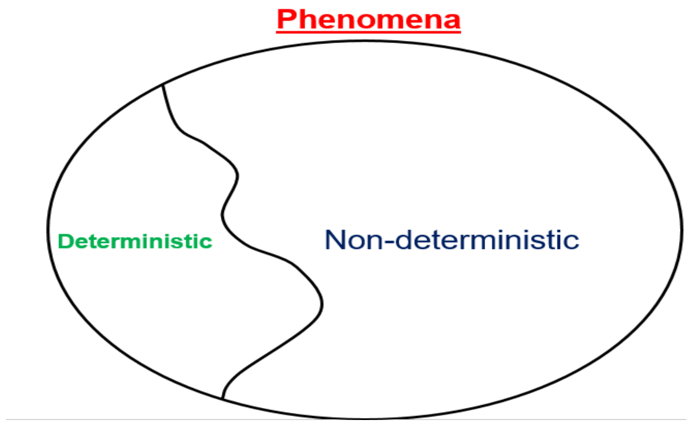


Figure 2.1: Phenomena types.

2.2.1.1.Deterministic phenomena

There is a mathematical model that "perfectly" predicts the evolution and future of a phenomenon. Numerous examples, particularly of deterministic systems or models, exist in physics and chemistry. A deterministic model will therefore always produce the same output from a given starting condition or initial state.

For example: physical systems, such as linear filters, described by differential equations represent deterministic systems, even if the state of the system at a given point in time may be difficult to describe explicitly.

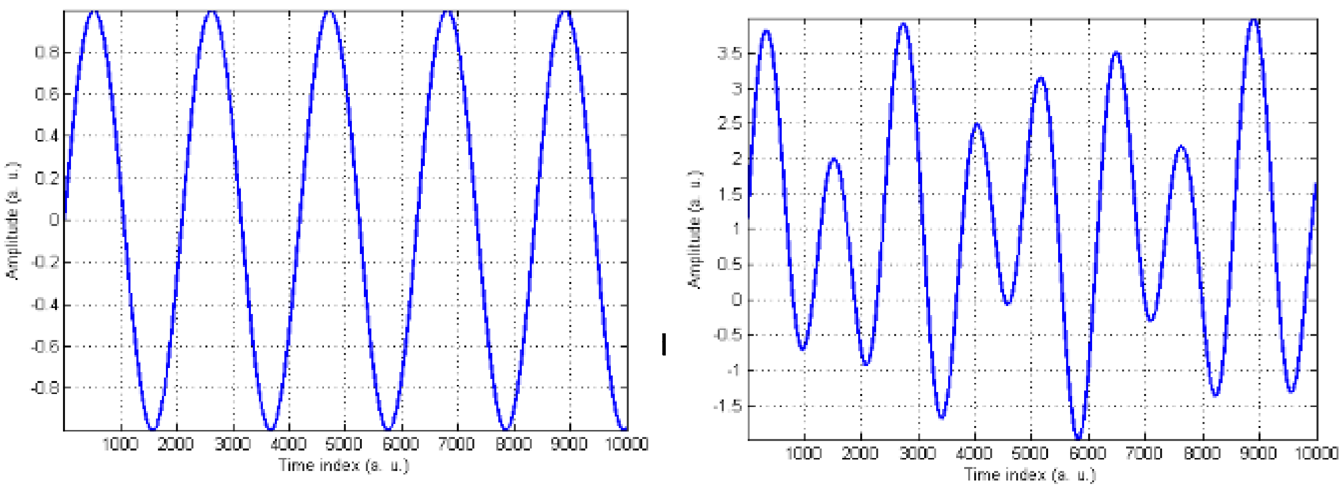


Figure 2.2: Different signals.

2.2.1.2.Non-deterministic phenomena

There is no mathematical model that can "perfectly" predict the evolution and future of a phenomenon. This applies to almost all natural phenomena (sound, image, temperature, etc.).

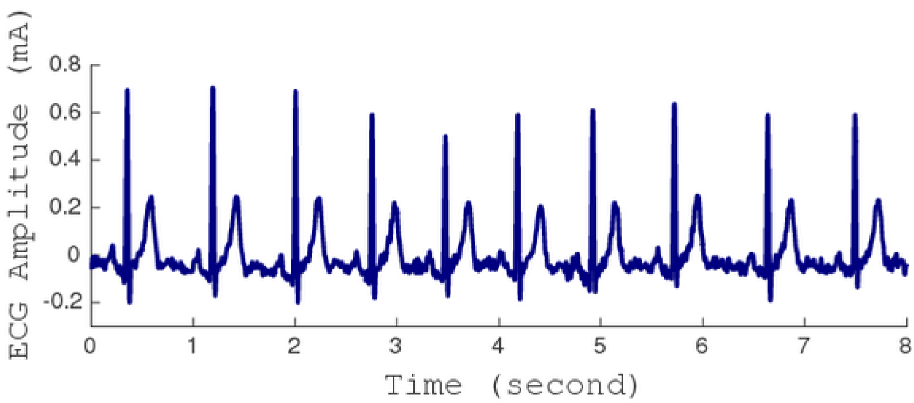


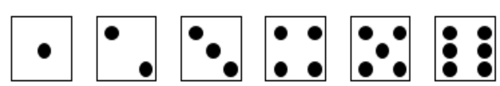
Figure 2.3: Example d'un signal ECG.

Non-deterministic phenomena can be consisted into two groups. The main one is the **Random phenomena.** It is unable to predict results, but in the long term, results show statistical regularity.

Example: **<u>Toss a coin</u>**

Impossible to predict at each toss whether it will be heads or tails.

In the long term, we can predict that 50% of "heads" will come up and 50% of "tails" will come up.



<u>A roll of the die</u>

Impossible to predict on each roll whether the die face will be "1", "2", "3", "4", "5" or "6".

In the long term, we can predict that 16.66% of each of the six-die faces s "faces" will occur.

2.2.2 Experiments, samples space, and events

2.2.2.1. Definition of random Experiment (E)

It is a procedure we perform that produces some result, e.g., the experiment E5 consist of tossing a coin five times.

2.2.2.2. Definition of an outcome (ξ)

It is a possible result of an experiment, e.g., the outcomes **ξ1** of the experiment E5 is ***HHTHT*** sequences tosses of Heads (H) or Tails (T).

2.2.2.3. Definition of an event (C)

It is a certain set of outcomes of an experiment, e.g., the event **C1** associated with the experiment E5 might be **C1**= {all outcomes consisting of an even number of heads}.

2.2.2.4. Definition of Sample Space (s)

It is the collection or set of all possible outcomes of random experiment E. Elements of s called sample points (x values).

Example: **E**: Tossing a coin twice

Sample space: {HH, HT, TH, TT}.

$\{x1, x2, x3, x4\}.$

Probability formula: By convention, probabilities are real numbers between 0 and 1. A probability of 0 refers to something that never occurs; a probability of 1 refers to something that always occurs. Probabilities between 0 and 1 refer to things that sometimes occur.

Probability= $\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$.

2.2.3 Continuous and Discrete Random Variables

2.2.3.1. Definition of Random Variable

A random variable ***(RV)*** is a numerical representation of the outcome of a random experiment. In addition, it is real values function, which assign a real number to each sample point in the sample space.

Example:

- **E:** Flipping three coins at same time.
- **Outcomes:** Heads (H) or Tails (T).

- **S= 2^3={HHH,HHT,HTH,THH,TTT,TTH,THT,HTT}**
- **Random Variable:** let's suppose the number of tails is the random variable X:
  - T= 1 (for Head)
  - H = 0 (for tails)

  X= { 0, 1, 1, 1, 3, 2, 2, 2}.

2.2.3.2. Types of Random Variables

It can be classified by:
- Discrete Random Variable (DRV)
- Continuous Random Variable (CRV)

a. **Discrete Random Variable (DRV)**

These are the random variables, which can take only a finite number of values in a finite observation interval. Therefore, we can say that DRV has distinct value.

Example, Number of heads obtained after two coin are tossed, the number of cars passing a certain point in an hour.

b. **Continuous Random Variable (CRV)**

A random variable that takes on an infinite number is known as continuous random variable. Ex, a RV that measures the time taken in completing a job is CRV.

2.3 Review of Probability and Statistics (probability density, cumulative distribution function, etc.)

2.3.1. Cumulative Distribution Functions (CDF)

2.3.1.1. Definition

The cumulative distribution function (CDF) of random variable 'X' defined as the probability that the random variable X takes a value ''less than or equal to x'' {X ≤x} for all values of $-\infty < x < \infty$.

- Mathematically CDF : $F_X(x) = P(X \le x)$.

Other names of CDF:

- Probability distribution function of the random variable.
- Cumulative probability distribution function.
- Distribution function of the random variable.

2.3.1.2. Properties of CDF:

1. The distribution function is always between 0 and 1; $0 \leq F_X(x) \leq 1$
2. $F_X(-\infty) = 0 \; and \; F_X(+\infty) = 1$
3. $F_X(x_1) \leq F_X(x_2)$ if $(x_1) \leq (x_2)$, so according to this property, $F_X(x)$ i.e. the distribution function is **monotone nondecreasing** function of *x*.

2.3.1.3. CDF for discrete random variable

If 'X' is a discrete random variable, then it takes values at discrete points.

- CDF can be defined for this case as: $F_X(x) = P(X \leq x)$

  Suppose $X = \{x_1, x_2, x_3, x_4, x_5, \dots \dots \dots \dots . . x_1\}$

Therefore, the CDF for a discrete variable at the complete range of x can be defined as:

$$F_X(x) = \begin{cases} 0 & for \quad -\infty < x < x_1 \\ \sum_{j=1}^{n} P(X = x_i) & for \quad x_1 < x < x_n \\ 1 & for \quad x_n < x < +\infty \end{cases}$$

CDF of a discrete variable at any event is equal to the summation of the probabilities of random variable up to that certain event.

As x varies from $-\infty$ to $+\infty$, the graph of CDF i.e. $F_X(x)$ resembles a staircase with upward steps having height $P(X = x_i)$ at each $x = x_j$. Figure 2.4 shows the general form of the CDF.

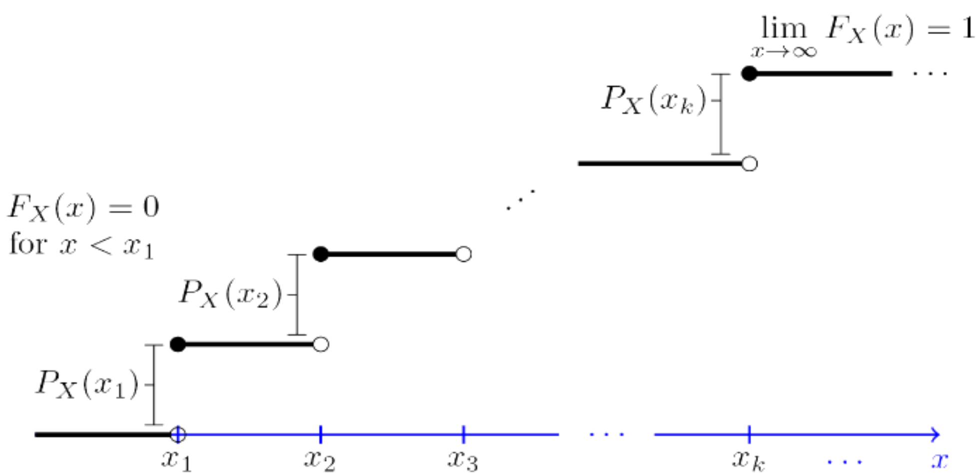


Figure 2.4: CDF of a discrete random variable.

However, note one thing that the graph of $F_X(x)$ remains constant btw the two steps or events.

Example: Tossing a coin twice, find the CDF?

S= {HH, HT, TH, TT}, X= {0, 1, 2, 3}, P(x) = $\left\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right\}$

First, note that if *x<0*, then: $F_X(x=0) = F_X(0) = P(X \le 0) = P(x=0) + P(x<0)$

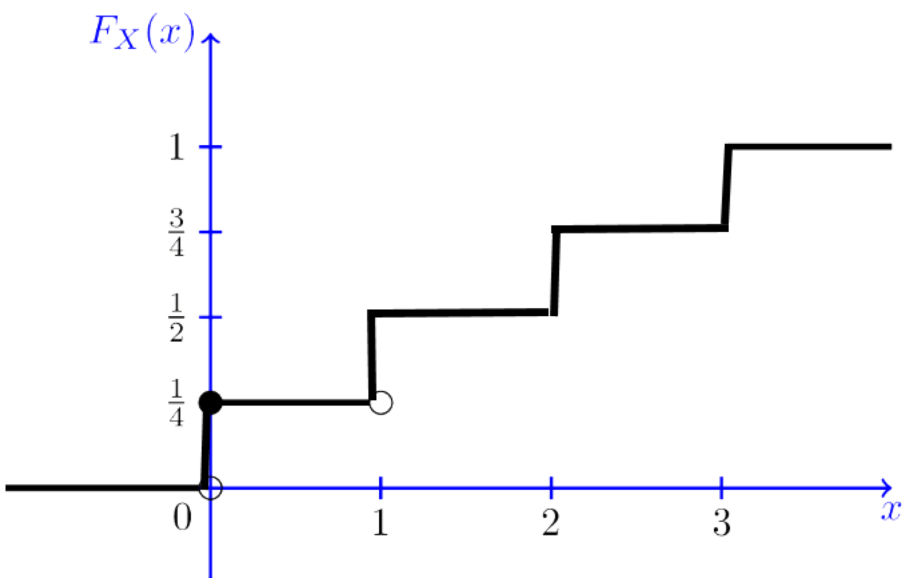
$$F_X(0) = P(x=0) = \frac{1}{4}$$

$$F_X(x=1) = F_X(1) = P(X \le 1) = P(x=1) + P(x=0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$F_X(x=2) = F_X(2) = P(X \le 2) = P(x=2) + P(x=1) + P(x=0) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

$$F_X(x=3) = F_X(3) = P(X \le 3) = P(x=3) + P(x=2) + P(x=1) + P(x=0)$$

$$= 4 \times \frac{1}{4} = 1$$



To summarize, we have:

$$F_X(x) = \begin{cases} 0 & for\ x < 0 \\ \frac{1}{4} & for\ 0 \le x < 1 \\ \frac{3}{4} & for\ 1 \le x < 2 \\ 1 & for\ x \ge 2 \end{cases}$$

$$F_X(x) = \frac{1}{4}\mu(x-0) + \frac{1}{4}\mu(x-1) + \frac{1}{4}\mu(x-2) + \frac{1}{4}\mu(x-3) =$$

$$= P(x=0)\mu(x-0) + P(x=1)\mu(x-1) + P(x=2)\mu(x-2) + P(x=3)\mu(x-3)$$

$$F_X(x) = \sum_{i=1}^{4} P(x_i)\mu(x-x_i)$$

For discrete Random variable:

$$F_X(x) = \sum_{i=1}^{N} P(x_i)\mu(x-x_i)$$

2.3.2. Joint Cumulative Distribution Functions (CDF)

Here we will discuss the CDF for two random variables X and Y.

2.3.2.1. Definition

Joint CDF (or combined CDF) $F_{XY}(x,y)$ of two random variables X and Y is defined as the probability that the random variable X is ≤ a specified value x and that the random variable Y is ≤ the specified value y, then:

$$F_{XY}(x,y) = P(X \le x, Y \le y)$$

2.3.2.2. Properties of joint CDF

1. Joint CDF is nondecreasing function of both *x* and *y*.
2. The combined CDF is non-negative function $F_{XY}(x,y) \ge 0$.
3. Joint CDF is always continuous everywhere.

2.3.3. Probability Density Function (PDF)

2.3.3.1. Definition

The derivative of cumulative distribution function with respect (w.r.t) some dummy variable is called as probability density function (PDF). Mathematically we can define is as:

PDF: $f_X(x) = \frac{d}{dx}F_X(x)$.

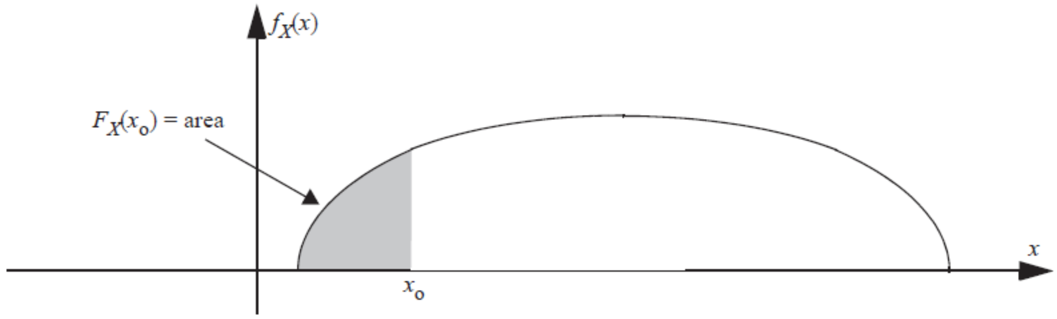


Figure 2.5: Relationship between the PDF and CDF of a random variable.

Now let's discuss some important properties.

2.3.3.2. PDF properties

1. PDF is always nonzero for all values of *x:* $f_X(x) \geq 0$.

Proof: CDF increases monotonically, derivative of CDF is positive always.

2. The area under the PDF curve is always equal to unity: $\int_{-\infty}^{+\infty} f_X(x)dx = 1$.
3. It is possible to get CDF by integrating PDF: $F_X(x) = \int_{-\infty}^{+\infty} f_X(x)\, dx$.
4. Probability of the event $\{x_1 \leq x \leq x_2\}$ is given by the area under the PDF curve in $x_1 \leq x \leq x_2$ range. $P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f_X(x)\, dx$.

2.3.4. Joint Probability Density Function (PDF)

2.3.4.1. Definition

Joint probability density function is simply the PDF of two or more random variables. The joint PDF of any two random variables X and Y can be defined as the partial derivative of the joint CDF, with respect (w.r.t) dummy variables *x* and *y*.

$$\textbf{Joint-PDF: } f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y).$$

2.3.4.2. Properties of joint PDF

1. PDF is non-negative *x;* $f_{XY}(x,y) \geq 0$.
2. Joint PDF is continuous everywhere as joint CDF.
3. The total volume under the surface of joint PDF is equal to unity

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x,y)dxdy = 1.$$

4. For two statistically independent random variables X and Y:

$$f_{XY}(x,y) = f_X(x) f_Y(y)$$

2.3.5. Relation between probability and joint PDF

$$P(x_1 \leq x \leq x_1) = \int_{x_1}^{x_2} f_X(x) dx$$

On extending this relation to the two random variables X and Y:

$$P(x_1 \leq x \leq x_1, y_1 \leq y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{XY}(x, y) dx\, dy$$

For independents random variables X and Y, the joint PDF is given as the product of two separate PDF.

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

$$P(x_1 \leq x \leq x_1, y_1 \leq y \leq y_2) = \int_{x_1}^{x_2} f_X(x) dx + \int_{y_1}^{y_2} f_Y(y) dy$$

It is possible to get CDF by integrating PDF: $F_X(x) = \int_{-\infty}^{+\infty} f_X(x)\, dx$.

Probability of the event $\{x_1 \leq x \leq x_2\}$ is given by the area under the PDF curve in $x_1 \leq x \leq x_2$ range. $P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f_X(x)\, dx$.

Example: the joint PDF of two random variables X and Y is given as:

$$f_{XY}(x, y) = \begin{cases} c(2x + y) & 0 \leq x \leq 2\ and\ 0 \leq y \leq 3 \\ 0 & elsewhere \end{cases}$$

Then find the value of constant ‘c’.

<u>*Solution:*</u>

For $f_{XY}(x, y)$ to be a probability density function:

$$\int_{y=0}^{3} \int_{x=0}^{2} f_{XY}(x, y)\, dx\, dy = 1$$

Substitute $f_{XY}(x, y) = c(2x + y)$ in the region:

$$\int_{y=0}^{3} \int_{x=0}^{2} c(2x + y)\, dx\, dy = 1$$

$$\int_{x=0}^{2} (2x + y)\, dx = \left[x^2 + yx\right]_{x=0}^{2}$$

At $x = 2$: $4 + 2y$
At $x = 0$: 0

So:

$$\int_{x=0}^{2} (2x + y)\, dx = 4 + 2y$$

$$\int_{y=0}^{3} (4 + 2y)\, dy = \left[4y + y^2\right]_{y=0}^{3}$$

At $y = 3$: $12 + 9 = 21$
At $y = 0$: 0

So:

$$\int_{y=0}^{3} (4 + 2y)\, dy = 21$$

$$c \times 21 = 1$$

$$c = \frac{1}{21}$$

2.3.6. Conditional probability density function (conditional PDF)

2.3.6.1. Definition

A probability density function is known as conditional PDF, when one random variable out of two random variables, has two random variables, has a fixed value.

Here suppose we have two random variables X and Y, and X has a value =*x*. In this case, the conditional PDF of Y when X=*x* is:

$$f_{X/Y=y}(y/x) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

Here $f_X(x)$ is the marginal density of X.

In the same way, conditional PDF of X when random variable Y takes a fixed value of y is given as:

$$f_{Y/X=x}(x/y) = \frac{f_{XY}(x,y)}{f_X(x)}$$

Here $f_Y(y)$ is the marginal density of Y.

For discrete random variable, it is call Conditional Probability Mass Function (PMF).

The conditional PMF of X given Y=*y* is:

$$P_{X/Y}(x/y) = \mathrm{P}(X = x/Y = y) = \frac{P_{XY}(x,y)}{P_Y(y)}$$

Similarly, the conditional PMF of Y given X=*x* is:

$$P_{Y/X}(y/x) = \frac{P_{XY}(x,y)}{P_X(x)}, \text{ provided } P_X(x) > 0$$

2.3.7. Properties of conditional PDF

1. Conditional PDF is non-negative $f_Y(y/x) \geq 0$ and $\boldsymbol{f_X(x/y)} \geq 0$.
2. $\int_{-\infty}^{+\infty} f_X(\boldsymbol{x/y})dx = 1$ and $\int_{-\infty}^{+\infty} f_Y(y/x)dy = 1$.
3. If the random variables are independents: $f_Y(y/x) = f_Y(y)$ and $f_X(x/y) = f_X(x)$.

2.3.8. Independence of random variable

Independence is a fundamental concept in probability theory that describes situations where knowledge of one random variable provides no information about another.

Two random variables X and Y are **independent** if:

$$P(X \in \mathrm{A}, Y \in \mathrm{B}\ ) = P(X \in \mathrm{A}).\mathrm{P}(Y \in \mathrm{B})$$

For Continuous Random Variables:

$$f_{XY}(\mathrm{x,y}) = f_X(x)f_Y(y)\ for\ all\ x,y$$

where f_{XY} is the joint PDF, and f_X, f_Y are marginal PDFs.

For Discrete Random Variables:

$$P_{XY}(\mathrm{x,y}) = P_X(x)P_Y(y)\ for\ all\ x,y$$

Where P_{XY} is the joint PMF, and P_X, P_Y are marginal PMFs

2.3.9. Real distribution and density function

The following are the most generally used distribution and density function.

2.3.9.1. Common Continuous Probability Distribution

Continuous random variables are characterized by probability density functions (PDFs). Here are the main probability models:

a. Uniform Distribution

The uniform probability density function is constant over an interval [a, b). The PDF and its corresponding CDF are:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \le x < b, \\ 0 & \text{elsewhere.} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & a \le x < b, \\ 1 & x \ge b. \end{cases}$$

Mean: $\frac{a+b}{2}$

Variance: $\frac{(b-a)^2}{12}$

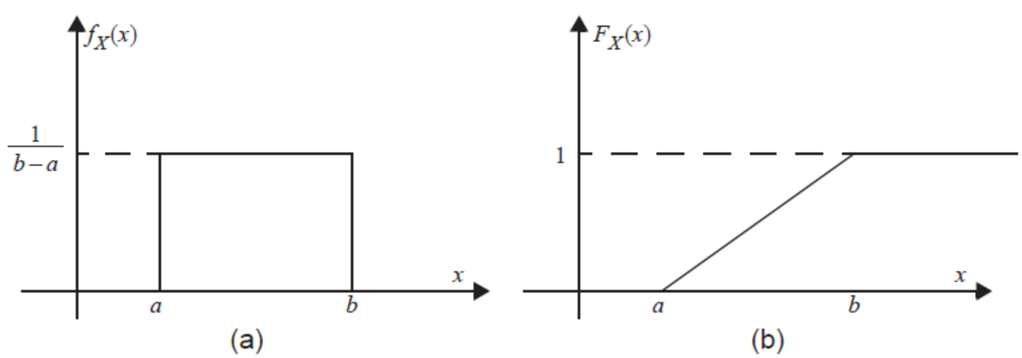


Figure 2.6: Probability density function (a) and cumulative distribution function (b) of a Uniform random variable.

b. Normal (Gaussian) Distribution

A Gaussian random variable X is one whose probability density function can be written in the general form:

$$\textbf{PDF: } f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The PDF of the Gaussian random variable has two parameters, m (μ) and σ, which have the interpretation of the mean and standard deviation respectively.1 The parameter σ^2 is referred to as the variance.

The plot of the Gaussian density function is bell-shaped curve and symmetrical about its mean value. The total area under the density function is one. i.e.

$$\frac{1}{\sqrt{2\pi\sigma_x^2}}\int_{-\infty}^{+\infty}\frac{e^{-(x-\mu)^2}}{2\sigma_x^2}dx = 1$$

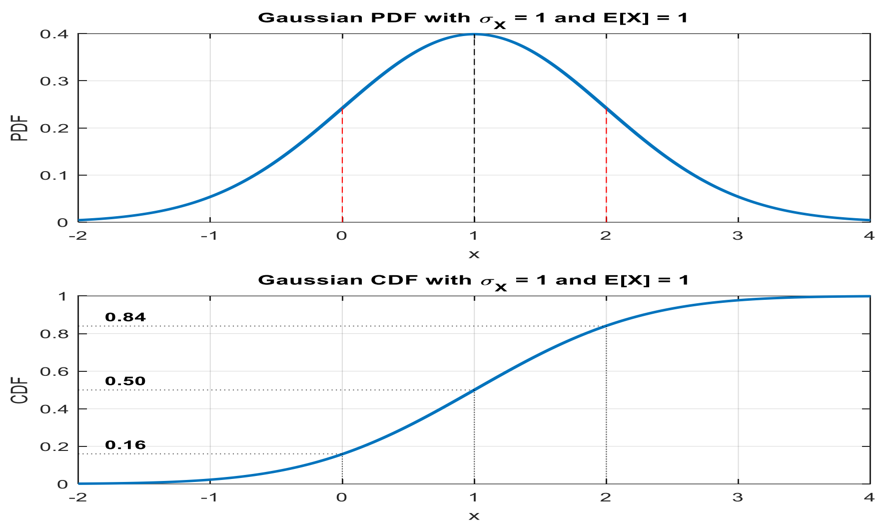


Figure 2.7: PDF (a) and CDF (b) of a Gaussian random variable.

Applications : Measurement errors, natural phenomena, Central Limit Theorem.

c. Exponential Distribution

The exponential random variable has a probability density function and cumulative distribution function given (for any b > 0) by :

$$f_X(x) = \frac{1}{b}e^{-\left(\frac{x}{b}\right)}u(x)$$

$$F_X(x) = [1 - e^{-\left(\frac{x}{b}\right)}]u(x)$$

Mean: b, variance: b^2

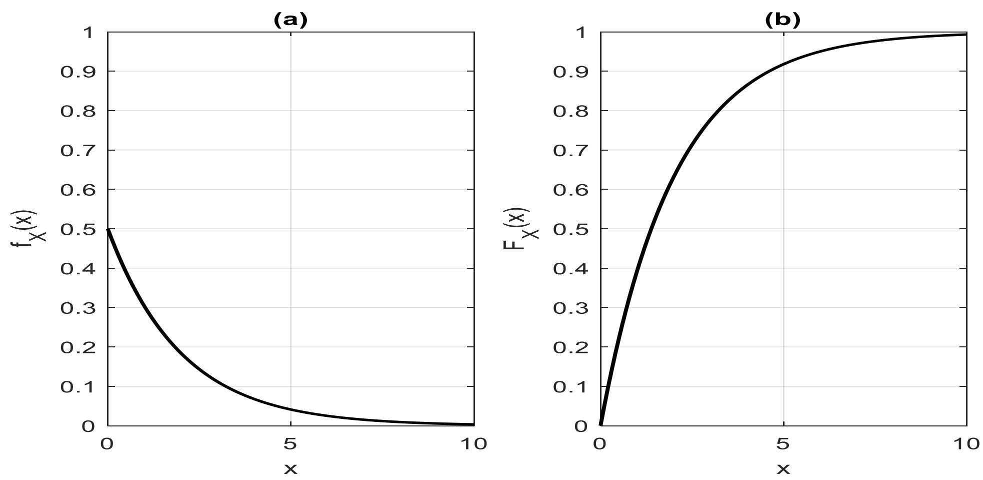


Figure 2.8: Probability density function (a) and cumulative distribution function (b) of an exponential random variable, b = 2.

2.4. Moments and Conditional Statistics

Numerical characteristics, also known as descriptive statistics, provide insights into the central tendency, dispersion, and shape of a random variable's distribution. They help to understand and quantify various aspects of the data.

2.4.1. Mean value

It is also called **expected** or **average value**. It is calculated by summing all values and dividing by the total number of observations.

$$x_m = E[x] = \int_{-\infty}^{+\infty} x f_X(x)dx$$

Where **E** is expectation of random variable X.

Example: Consider the random variable X with *PMF* as tabulated below:

Value of the random variable x	0	1	2	3
$p_X(x)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$

$$\therefore\ \mu_X = \sum_{i=1}^{N} x_i p_X(x_i)$$

$$= 0\times\frac{1}{8} + 1\times\frac{1}{8} + 2\times\frac{1}{4} + 3\times\frac{1}{2}$$

$$= \frac{17}{8}$$

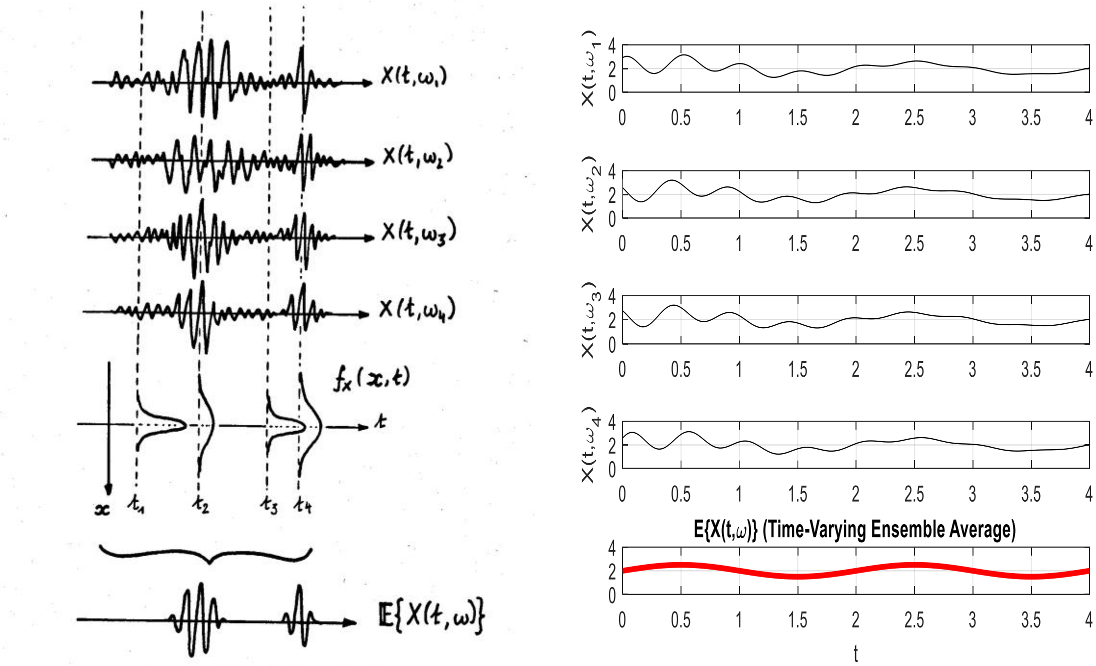


Figure 2.9: Mean value.

2.4.2. Second moment

It is called mean squared value of random variable X.

$$E[x^2] = \int_{-\infty}^{+\infty} x^2 f_X(x)dx.$$

- **Larger second moment:** Indicates that the values of the random variable are more spread out from the mean.
- **Smaller second moment:** Indicates that the values are more concentrated around the mean.

2.4.3. Variance (var)

It is the second central moment. Variance measures the average magnitude of the random variable fluctuation from its expectation. A smaller variance implies that the random values are more clustered about the mean, Similarly, a bigger variance means that the random values are more scattered.

$$\sigma^2 = E[x^2] - {x_m}^2.$$

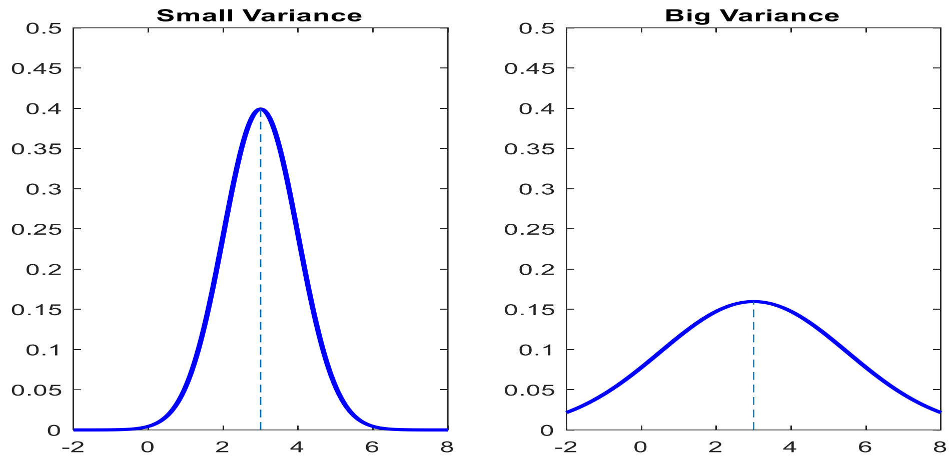


Figure 2.10: Variation in Gaussian PDF.

Example: Find the variance of the random variable discussed in previous Example.

$$\sigma_X^2 = E(X-\mu_X)^2$$
$$= (0-\frac{17}{8})^2 \times \frac{1}{8} + (1-\frac{17}{8})^2 \times \frac{1}{8} + (2-\frac{17}{8})^2 \times \frac{1}{4} + (3-\frac{17}{8})^2 \times \frac{1}{2}$$
$$= \frac{117}{128}$$

2.4.4. Standard deviation (std)

It provides a measure of dispersion in the same units as the data. It is the PDF width, $\sigma = \sqrt{\sigma^2}$.

Example: find the mean value, variance and std of random variable X, if it is uniformly distributed between the range $0 \le x \le 2$.

Solution:

Mean x_m:

$$\mu = \frac{a+b}{2} = \frac{0+2}{2} = \frac{2}{2} = 1$$

Variance (σ^2):

$$\sigma^2 = \frac{(b-a)^2}{12} = \frac{(2-0)^2}{12} = \frac{4}{12} = \frac{1}{3}$$

Standard Deviation (σ):

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{3}} = \frac{1}{\sqrt{3}}$$

➢ **Properties of Expectation**

- **Sum of Random Variables** : $E[X + Y] = E[X] + E[Y]$
- **Scalar Multiplication** : $E[aX] = aE[X]$
- **General Linear Combination** : $E[aX + bY] = aE[X] + bE[Y]$
- $E[c] = c$, c is a constant.
- $E[XY] = E[X]E[Y]$, if X and Y are **independent** random variable.
- $Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$

2.5. Sequences of Random Variables - Functions of Random Variables – Covariance

The Connection between these concepts:

1. **Sequences of RVs** → Studying **multiple** random variables at once (infinite or finite), possibly dependent. Important for laws of large numbers, central limit theorem, stochastic processes.
2. **Functions of RVs** → Tools to transform/combine RVs (e.g., sums, products, $g(X_1,X_2)$), needed for deriving distributions and moments.
3. **Covariance** → A measure of *linear* relationship between two RVs; builds on joint behavior, linking sequences and functions (e.g., covariance of X_i and X_j in a sequence).

2.5.1. Sequences of Random Variables

Definition : Let S be a sample space. A sequence of random variables Xn is a mapping that assigns a real-valued function to each index n:

$$X_n: S \to R \; for \; n = 1,2,3, \dots \dots \dots ....$$

This means that for every outcome $s \in S$, the sequence $X_1(s)$, $X_2(s)$, $X_3(s)$, ……is a sequence of real numbers.

Main Characteristics :

- Joint Distribution : For any finite n, the variables X_1, X_2,……., X_n have a joint cumulative distribution function (CDF).
- Independence : We often study sequences where the variables are independent, meaning the outcome of X_i does not affect X_j.

$$P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) = \prod_{i=1}^{n} P(X_i \leq x_i).$$

- Identical Distribution : If every X_n in the sequence follows the exact same probability distribution, they are called identically distributed.

$$F_{X_1}(x) = F_{X_2}(x) = \cdots \ldots \ldots \ldots . = F_{X_n}(x) \; for \; all \; x$$

Sequences are the foundation for asymptotic theory. In engineering and data science, we rarely have "infinite" data, but we use sequences to understand what happens to our estimates (like a sample mean) as the number of observations n grows very large. This leads directly into the study of **convergence**.

2.5.2. Types of Convergence

There are four primary ways to define how a sequence of random variables approaches a limit X.

2.5.2.1. Convergence in Distribution (Weakest)

Also called **convergence in law**, this is the most common form of convergence used in the **Central Limit Theorem**. It doesn't require the variables to be close to each other, only that their *probability distributions* become identical in the limit.

Definition: $X_n \overset{d}{\to} X$ if the cumulative distribution functions (CDFs) satisfy:

$$\lim_{n \to \infty} F_n(x) = F(x)$$

For every point x where F(x) is continuous.

2.5.2.2. Convergence in Probability

This is a stronger form of convergence. It means the probability that Xn is "far" from X goes to zero. This is the foundation for the **Weak Law of Large Numbers (WLLN)**.

Definition: $X_n \xrightarrow{P} X$ ε > 0:

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

2.5.2.3. Almost Sure Convergence (Strongest)

Also known as **convergence with probability 1**. This is the probabilistic equivalent of pointwise convergence in calculus. It implies that for almost every outcome in the sample space, the sequence of numbers produced will converge. This underpins the **Strong Law of Large Numbers (SLLN)**.

- **Definition:** $X_n \xrightarrow{\text{a.s.}} X$ if:

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

2.5.2.4. Convergence in Mean (: L^p Convergence)

This focuses on the "average" distance between the variables. The most common case is **mean-square convergence** (p=2).

- **Definition:** $X_n \xrightarrow{L^p} X$ if:

$$\lim_{n\to\infty} E[|X_n - X|^p] = 0$$

The relationship between these types is directional. If a sequence converges in a "stronger" sense, it automatically converges in the "weaker" senses below it.

Implication Path	Description
Almost Sure →Probability	If it converges almost surely, it converges in probability.
Mean Square →Probability	If the average squared distance goes to zero, it converges in probability.
Probability →Distribution	If it converges in probability, it converges in distribution.

2.5.3. Covariance

2.5.3.1. Definition

Covariance measures how two random variables change together:

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

2.5.3.2. How to Interpret the Value

The sign (positive or negative) of the covariance is what's most immediately useful:

- **Positive Covariance:** Indicates a **direct relationship**. When one variable is above its mean, the other tends to be above its mean as well.

  Example: Hours Studied (X) and Exam Score (Y). Generally, more hours studied correlates with a higher score.

- **Negative Covariance:** Indicates an **inverse relationship**. When one variable is above its mean, the other tends to be below its mean.

  Example: Time Spent on Video Games (X) and Exam Score (Y). Generally, more time gaming might correlate with a lower score.

- **Covariance near Zero:** Suggests **no linear relationship**. The variables don't show a consistent pattern of moving together.

2.5.3.3. Properties

1. $Cov(X, Y) = Var(X)$
2. $Cov(X, Y) == Cov(Y, X)$ symmetric
3. $Cov(aX + b, cY + d) = ac. Cov(X, Y)$
4. $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$

Variance of Sums:

$$Var(X + Y) = Var(X + Y) + Var(X) + Var(Y) + 2\,Cov(X, Y)$$

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$$

2.5.4. Correlation Coefficient (Pearson)

Consider the two random variables X and Y, the second order joint moment ρ_{XY} is called the Correlation of X and Y.

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_x \sigma_y}, -1 \leq \rho \leq 1$$

- If two random variables X and Y are statistically independent then X and Y are said to be uncorrelated. That is $E[XY] = E[X]\,E[Y]$.

- If the Random variables X and Y are orthogonal then their correlation is zero. i.e. $\rho_{XY} = 0$.

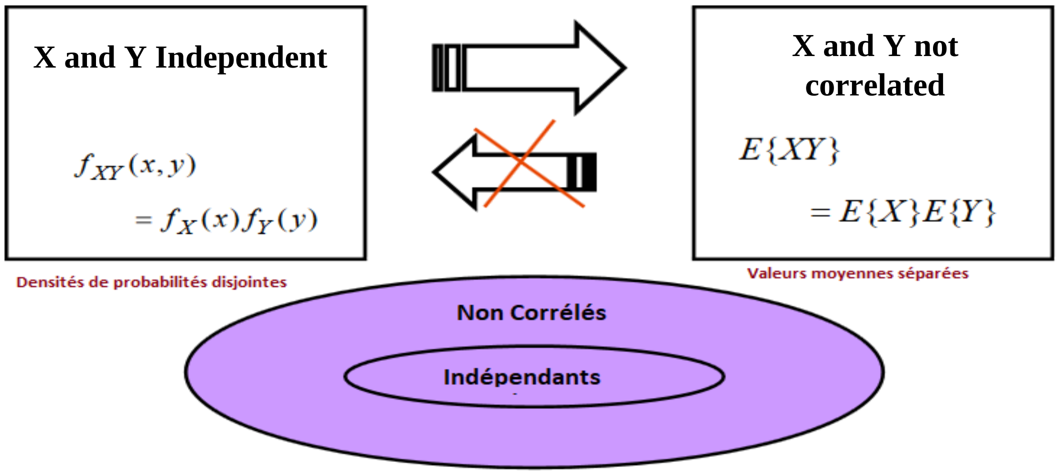


Figure 2.11: Correlation schema.

2.5.4.1. How to Interpret the Value

The correlation coefficient is always between **-1 and +1**. This makes its interpretation universal.

- $r = +1$: **Perfect positive linear relationship.** All data points lie exactly on an upward-sloping straight line.
- $r > 0$: **Positive correlation.** As one variable increases, the other tends to increase. The closer to +1, the stronger the relationship.
- $r = 0$: **No linear relationship.** There is no straight-line trend between the variables.
- $r < 0$: **Negative correlation.** As one variable increases, the other tends to decrease. The closer to -1, the stronger the (inverse) relationship.
- $r = -1$: **Perfect negative linear relationship.** All data points lie exactly on a downward-sloping straight line.

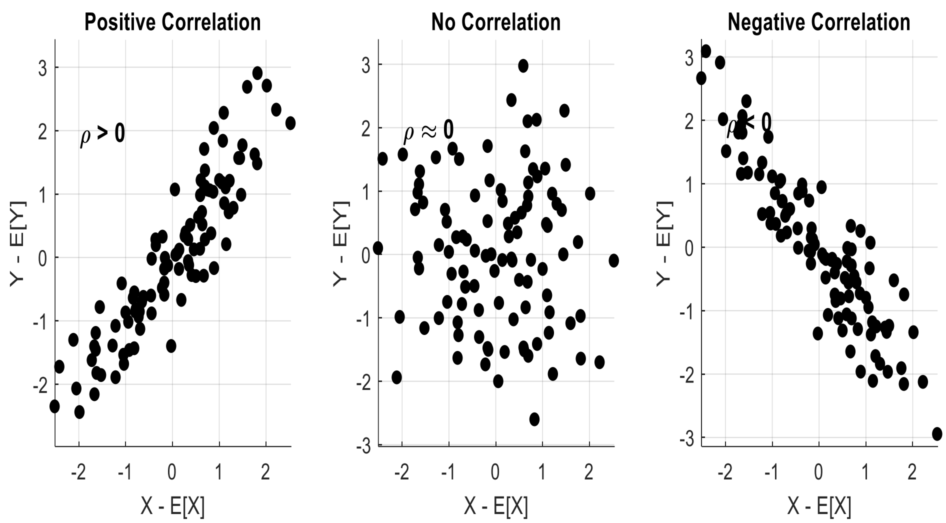


Figure 2.12: Correlation Types.

EXERCISES

Exercice 2.1 fair coin is tossed four times, and the sequence of heads and tails is observed.

a) List each of the 16 sequences in the sample space S.
b) Let the events A, B, C, and D by :
A= {at least 3 heads}, B= {at least 3 heads}, C= {heads on the third toss}, and D= {1 head and 3 tails}.

In the probability set function assigns $\frac{1}{16}$ to each outcome in the sample space, find:

P(A), P(B), P(C), P(D), P(A∩B), P(A∩C), P(A∪C), P(B∩D).

Exercise 2.2

Part 1/ a continuous random variable X has probability density function:

a) $f(x) = 6(\sqrt{x} - x) \qquad 0 < x < 1$
b) Find distribution function of X.

Part 2/ the table below describes the probability distribution for a discrete random variable X.

x	4	5	6	8	9
$P(X = x)$	0.2	**a**	**b**	0.5	0.15

Determine the value of **a** and **b** if $P(x \leq 5) = 0.3$

Exercise 2.3

Consider the function:

$$f(x) = \begin{cases} k - \frac{x}{4} & ,1 \leq x \leq 3 \\ 0 & otherwise \end{cases}$$

Which used as probability density function for a continuous random variable.

a) Find the value of K.
b) Find the value of if $P(x \leq 2.5)$

Exercice 2.4

c) Find the value of constant K, such that the PDF given by: $f_X(x) = \begin{cases} \frac{1}{K} & a \leq x \leq b \\ 0 & otherwise \end{cases}$
d) Find the CDF of the random variable X if K satisfies the condition for $f_X(x)$ to be PDF.

***Exercice* 2.5** 3 digital message is transmitted over a noisy channel having probability of error $P(E) = \frac{2}{5}$

- ✓ Show the mapping on real axis and find CDF.

***Exercice* 2.6** of Random Variables is given as:

$$f_X(x) = e^{-x} \ \ for \ \ x \geq 0$$

a) Find mean (m_x)
b) Find variance (v)
c) Standard deviation (std)

***Exercice* 2.7 :** probability density function of random variables X and Y is given by:

$$f_{XY}(x, y) = \begin{cases} Kxy & 0 < y < x < 1 \\ 0 & otherwise \end{cases}$$

a) Find K.
b) Find $E[X/ (Y = y)]$

***Exercice* 2.8**

$$f_{XY}(x,y) = \begin{cases} \frac{1}{18} & 0 < x < 3\,, x^2 < y < 9 \\ 0 & otherwise \end{cases}$$

a) $E[X]$
b) $E[Y]$
c) $E[X+Y]$
d) $E[XY]$

***Exercice* 2.9**

$$f_{XY}(x,y) = \begin{cases} 6y & 0 \le y \le x \le 1 \\ 0 & otherwise \end{cases}$$

Are X and Y independent?

2.4 Conclusion

This chapter on random variables has provided the essential mathematical background for quantifying uncertainty. We began by defining random variables as numerical representations of random experiments, classifying them into discrete and continuous types, each with their respective probability and cumulative distribution functions (PDFs and CDFs). The core concepts of expected value and variance were introduced as the primary tools for understanding a variable's central tendency and variability, forming the basis of moment analysis.

Furthermore, we explored how covariance and correlation measure the relationship between two variables, extending our analysis to multivariate contexts. Finally, by examining sequences and functions of random variables, we laid the basis for understanding more complex phenomena and convergence principles like the Law of Large Numbers.

Chapter 3: Processing random signals

Chapter 3: Processing random signals

Course objective

At the end of this chapter, the student will be able to:

1. **Identify** the fundamental role of **random signals** in modern communication systems, specifically concerning the transfer of information.
2. **Explain** why useful message signals must be considered **random**, recognizing that the receiver lacks *a priori* knowledge of the specific waveform being transmitted.
3. **Recognize** the source and nature of **noise** in a communication channel, understanding that it is composed of random electrical signals.
4. **State** the necessity for having **good mathematical and statistical descriptions** of random signals for the effective design and analysis of communication systems.

3.1 Introduction

In telecommunications, all real-world signals are considered **random signals** or **stochastic processes** (which is the mathematical term for a random signal).

- **Signals of Interest:** Even intended signals, like speech, video, or data packets, have an unpredictable, or random, component when viewed over time. For example, you can't predict the next data bit in a stream.
- **Noise:** The most critical random signal is noise (e.g., thermal noise, atmospheric noise, interference). Noise corrupts all transmitted and received signals.
- **The Challenge:** Unlike deterministic signals (which can be described by a fixed mathematical function, like $x(t) = A\,cos(\omega t)$, a random signal is a collection of possible functions, each with a certain probability. We cannot analyze a single random signal realization; we must analyze its **statistical properties**.

Processing a random signal typically involves passing it through a system (a filter, a detector, an estimator) to separate the intended signal from the noise or to extract meaningful information. This requires understanding how the system changes the signal's statistics.

3.2 Random signals (statistical and temporal representations)

3.2.1 Defining the Random Signal

A **Random Signal** (or **Stochastic Process**), $X(t)$ or $X(\omega)$, is an indexed collection of random variables, where the index t represents time.

- **Ensemble:** The process is a collection of infinitely many possible time functions, called realizations or sample functions, $x_i(t)$. Each realization is the result of a single, non-repeatable experiment (like recording noise in a circuit).
- **Time Point:** At any fixed time t_1, the signal value $X(t_1)$ is a **Random Variable** (RV).
- **Significance:** Because we cannot predict the exact value $x_i(t)$ at any time, we must describe the signal using its average properties (statistics), which leads to the two primary forms of representation.

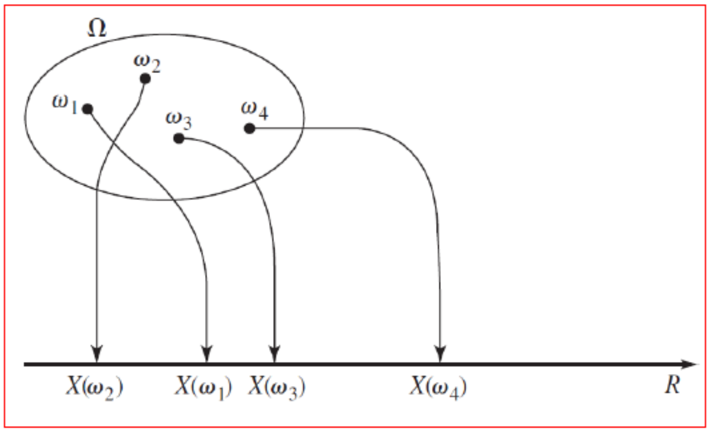


Figure 3.1: Real mapping of random variable.

3.2.2 Statistical (Ensemble) Representation

The statistical representation describes the distribution of the signal's amplitude values across the entire ensemble of realizations at one or more fixed time instants. This requires the use of Probability Density Functions (PDFs) and Statistical Moments (as mention the chapter 2).

3.2.3 Temporal (Time-Domain) Representation

The temporal representation describes the structure of the signal's variation over time, particularly the Autocorrelation Function.

3.2.3.1. Autocorrelation Function (ACF)

The Autocorrelation Function (ACF) is the primary temporal measure for system analysis. The autocorrelation function of the random process X(t),denoted as $R_{XX}(t_1, t_2)$, is defined as the statistical expectation of the product of two samples by $R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)]$. It is usually denoted by $R_X(t_1, t_2)$ for short.

$$R_{XX}(t_1,\ t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2) dx_1 dx_2$$

- If $R_{XX}(t_1,\ t_2)$ is large, the samples at t_1 and t_2 are strongly **correlated** (i.e., the value at t_2 is highly predictable from the value at t_1).

- If $R_{XX}(t_1, t_2)$is close to zero, the samples are nearly **uncorrelated**.

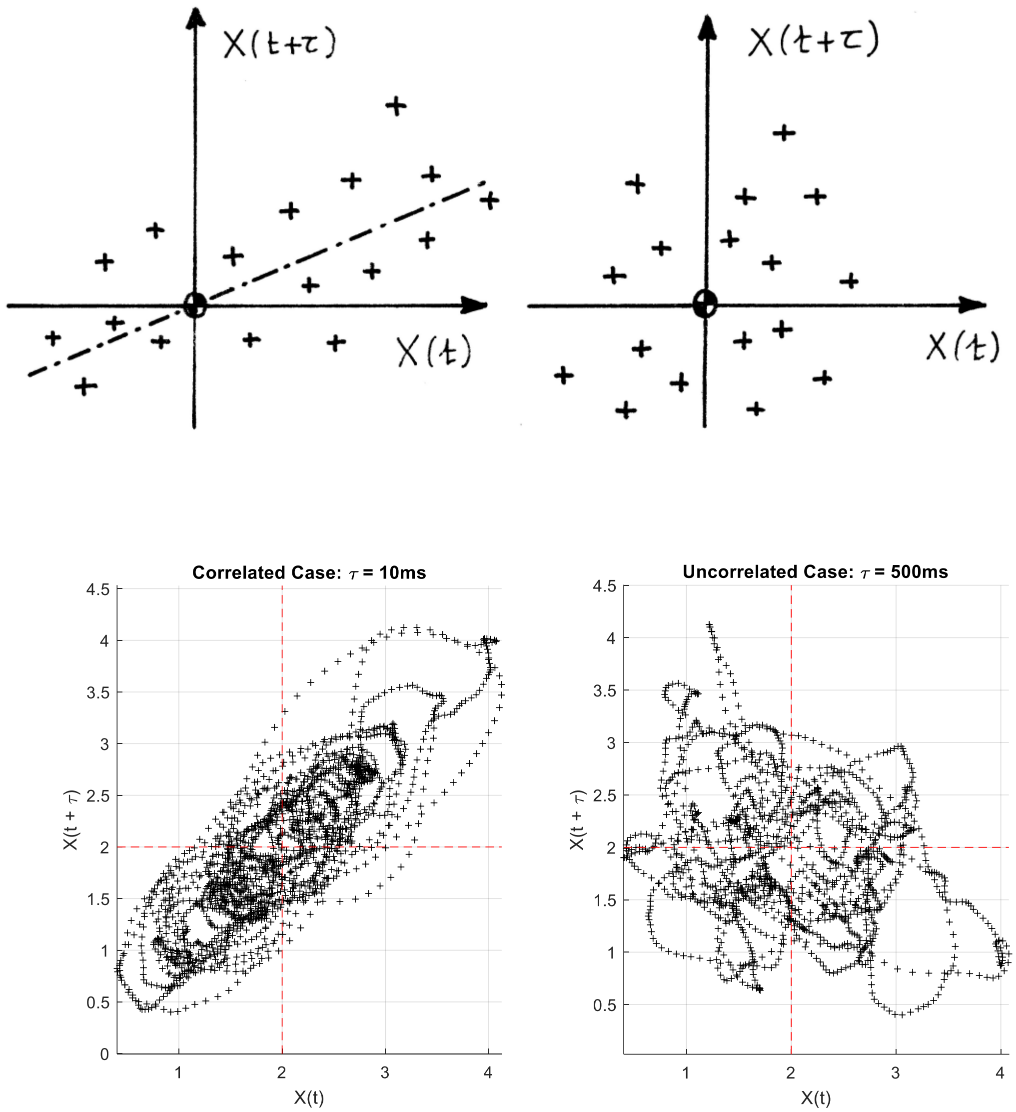


Figure 3.2: Autocorrelation Function.

This general definition is valid for both **stationary** and **non-stationary random processes**.

Example: For $X(t) = A\,cos(2\pi f_0 t + \Theta)$

$$
\begin{aligned}
R_X(t_1, t_2) &= E[A\cos(2\pi f_0 t_1 + \Theta) A\cos(2\pi f_0 t_2 + \Theta)] \\
&= A^2 E\left[\frac{1}{2}\cos 2\pi f_0(t_1 - t_2) + \frac{1}{2}\cos(2\pi f_0(t_1 + t_2) + 2\Theta)\right] \\
&= \frac{A^2}{2}\cos 2\pi f_0(t_1 - t_2)
\end{aligned}
$$

$$
E[\cos(2\pi f_0(t_1 + t_2) + 2\Theta)] = \int_0^{2\pi} \cos[2\pi f_0(t_1 + t_2) + 2\theta]\frac{1}{2\pi}\, d\theta = 0
$$

3.2.3.2. The Special Case: Wide-Sense Stationary (WSS)

However, our main interest is focused on stationary processes, for which a further simplification of $R_{XX}(t_1,\, t_2)$ is possible. We can recall that for a Wide-Sense Stationary (WSS), all ensemble averages are independent of the time origin. Consequently, for a wide-sense stationary process, shifting the representation from two time variables $(t_1,\, t_2)$ to a single time T:

$$R_X(t_1,\, t_2) = R_x(t_1 + T,\, t_2 + T)$$

$$R_X(t_1,\, t_2) = E(\mathrm{X}(t_1 + \mathrm{T})\mathrm{X}(\, t_2 + T))$$

Since this expression is independent of the choice of the time origin, we can for example define $T = -t_1$, which leads us to write:

$$R_X(t_1,\, t_2) = R_x(t_1 + T,\, t_2 + T)$$

$$R_X(t_1,\, t_2) = E(\mathrm{X}(0)\mathrm{X}(t_1 - \, t_2))$$

We see that this expression depends only on the time difference $t_2 - t_1$. By setting this time difference equal to $\tau = t_2 - t_1$ and by removing the zero from the argument of R_X, we can rewrite the previous equation as:

$$R_X(\tau) = E(\mathrm{X}(t_1)\mathrm{X}(\, t_1 + \tau))$$

It is also possible to define a time autocorrelation function for a particular sample function as:

$$\Re_x(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t+\tau)\, dt = \langle x(t)x(t+\tau) \rangle$$

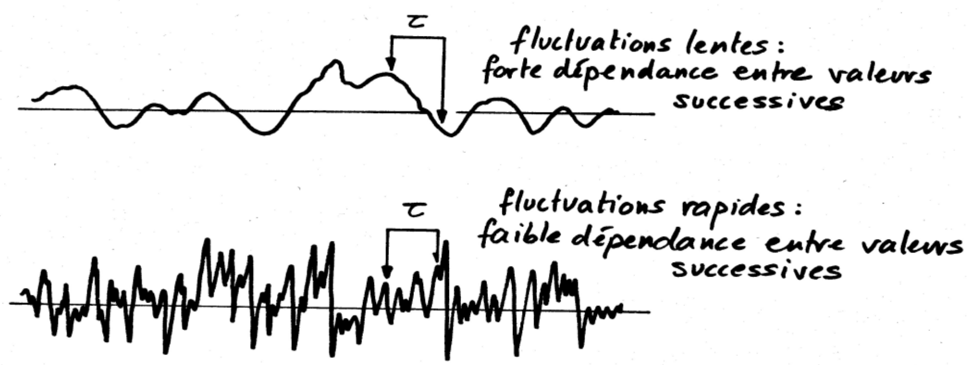


Figure 3.3: Wide-Sense Stationary (WSS) processes.

1.2.3.1. Autocovariance Function (ACVF)

The ACVF, $C_X(t_1, t_2)$, measures the correlation of the **AC components** of the signal:

$$C_X(t_1, t_2) = E[(X(t_1) - \mu_x(t_1))(X(t_2) - \mu_x(t_2))]$$

The relation to ACF is :

$$R_X(t_1, t_2) = C_X(t_1, t_2) + \mu_x(t_1)\mu_x(t_2)$$

For a **WSS** process with mean μ_x:

$$C_X(\tau) = R_X(\tau) - \mu_X{}^2$$

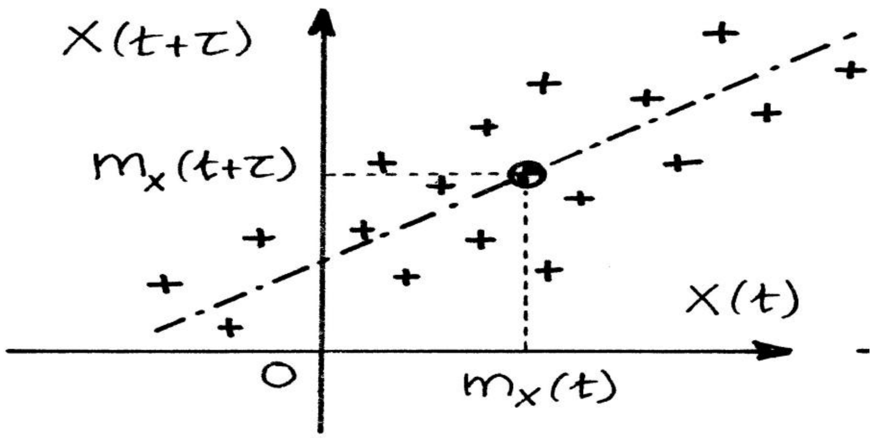


Figure 3.4: Autocovariance Function (ACVF).

3.3 Stationarity and statistical properties (mean, variance, standard deviation, etc.)

Stationarity means that the statistical properties of a random signal do not change over time. Since a random signal X(t) is an ensemble of random variables, there are different degrees of stationarity, depending on how many of its statistical moments are required to be time-invariant.

3.3.1 Strict-Sense Stationarity (SSS)

A random process X(t) is Strict-Sense Stationary (SSS) if all its joint probability distributions remain unchanged by a shift in time.

$$\text{Joint PDF of } (\mathbf{X}(t_1), \mathbf{X}(t_2), \ldots, \mathbf{X}(t_k)) = \text{Joint PDF of } (\mathbf{X}(t_1+\tau), \mathbf{X}(t_2+\tau), \ldots,$$

3.3.2 Wide-Sense Stationarity (WSS)

A random process X(t) is Wide-Sense Stationary (WSS) (also called Weak-Sense Stationary or Second-Order Stationary) if only its first and second statistical moments are time-invariant. This is the condition used for almost all-practical signal processing.
The two conditions for a process X(t) to be WSS are:

Condition 1: Constant Mean

The expected value (mean function) must be a constant, independent of time t.

$$\mu_{\mathbf{X}}(t) = E[\mathbf{X}(t)] = \mu_{\mathbf{X}} = \text{Constant}$$

Condition 2: Time-Lag Dependent Autocorrelation

The autocorrelation function must depend only on the **time difference** or **time lag**, $\tau = t_2 - t_1$, and not on the absolute time instants t_1 or t_2.

$$R_{\mathbf{X}}(t_1, t_2) = E[\mathbf{X}(t_1)\mathbf{X}(t_2)] = R_{\mathbf{X}}(\tau)$$

Relationship:

- **SSS ⇒WSS:** If a process is SSS and its second-order moments are finite, it is automatically WSS.
- **WSS ⇏ SSS (in general):** The converse is not generally true. A process can be WSS without being SSS.
- **WSS ⇔SSS for Gaussian Processes:** If the process is Gaussian (noise in most communication systems is modeled this way), then WSS implies SSS, making WSS a powerful and sufficient condition.

3.3.3 Statistical Properties of WSS Processes

For WSS processes, the key statistical descriptors are significantly simplified, becoming constants or functions of a single variable (τ).

3.3.3.1. Mean ($\boldsymbol{\mu_x}$)

The **Mean** is the constant DC component of the random signal.

$$\mu_X = E[X(t)]$$

3.3.3.2. Variance and Average Power

The variance, σ_X^2 , is a constant, as it is derived from moments that are time-invariant for WSS processes.

Measure	Formula	WSS Simplification	Interpretation
Average Total Power	$P_{Total}(t) = E[X^2(t)]$	$P_{Total} = R_X(0)$	The total average power (AC + DC) is constant. This is the value of the autocorrelation function at τ=0.
Variance (AC Power)	$\sigma_X^2(t) = E[(X(t) - \mu_X)^2]$	$\sigma_X^2 = R_X(0) - {\mu_X}^2$	The average power of the zero-mean component or the AC power. This is also constant.
Standard Deviation	$\sigma_X(t) = \sqrt{\sigma_X^2(t)}$	$\sigma_X = \sqrt{R_X(0) - {\mu_X}^2}$	The square root of the variance. It measures the typical deviation of the signal amplitude from the mean, in the same units as the signal.

3.4 Power Spectral Density (PSD)

The PSD, or Power Spectral Density, is the essential characteristic we seek to obtain through spectral analysis. It represents the distribution of the signal's power per unit of frequency.
The Power Spectral Density (PSD), denoted as $S_X(\omega)$ or $S_X(f)$, describes how the average power of a random signal is distributed over frequency.

3.4.1. The Wiener-Khinchine Theorem

For a Wide-Sense Stationary (WSS) process X(t), the PSD is defined as the Fourier Transform (FT) of its Autocorrelation Function (ACF), $R_X(\tau)$. This fundamental relationship is known as the **Wiener-Khinchine Theorem**.

$$S_X(\omega) = \mathcal{F}\{R_X(\tau)\} = \int_{-\infty}^{+\infty} R_X(\tau) e^{-j\omega\tau} d\tau$$

Conversely, the ACF can be recovered by taking the Inverse Fourier Transform (IFT) of the PSD:

$$R_X(\tau) = \mathcal{F}^{-1}\{S_X(\omega)\} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_X(\omega) e^{j\omega\tau} d\omega$$

Significance of the PSD

1. **Power Distribution:** It shows which frequencies contribute the most power to the signal.
2. **Filter Design:** It is essential for designing filters to pass desired signal frequencies and attenuate noise frequencies.
3. **Frequency Analysis:** It allows us to apply the tools of deterministic signal frequency analysis (e.g., system transfer functions) to random signals.

3.4.2. Properties of the Power Spectral Density

Because the ACF $R_X(\tau)$ has specific properties (as seen in the previous section), the PSD $S_X(\omega)$ must also have related properties:

- **Real and Non-Negative:** $S_X(\omega)$ is always a real and non-negative function for all ω. A negative power at any frequency is physically impossible.
- **Even Symmetry:** Since $R_X(\tau)$ is a real, even function, its Fourier Transform $S_X(\omega)$ must also be an even function.

$$S_X(\omega) = S_X(-\omega)$$

- **Total Average Power:** The total average power, P_{avg}, of the random signal is the area under its PSD curve (normalized by 2π):

$$P_{avg} = R_{\mathbf{X}}(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\mathbf{X}}(\omega)d\omega$$

Example: White Noise

The concept of PSD is best illustrated by **White Noise**, which is often modeled as having a constant power density.

Measure	Formula	Interpretation
ACF	$R_N(\tau) = \frac{N_0}{2}\delta(\tau)$	Samples are uncorrelated for any time lag $\tau \neq 0$.
PSD	$S_N(\omega) = F\{R_N(\tau)\} = \frac{N_0}{2}$	The power is uniformly distributed across all frequencies (flat spectrum).
Total Power	$P_{avg} = R_N(0) = \infty$	True white noise has infinite power, which is why it is an idealization. Real noise is "band-limited white noise."

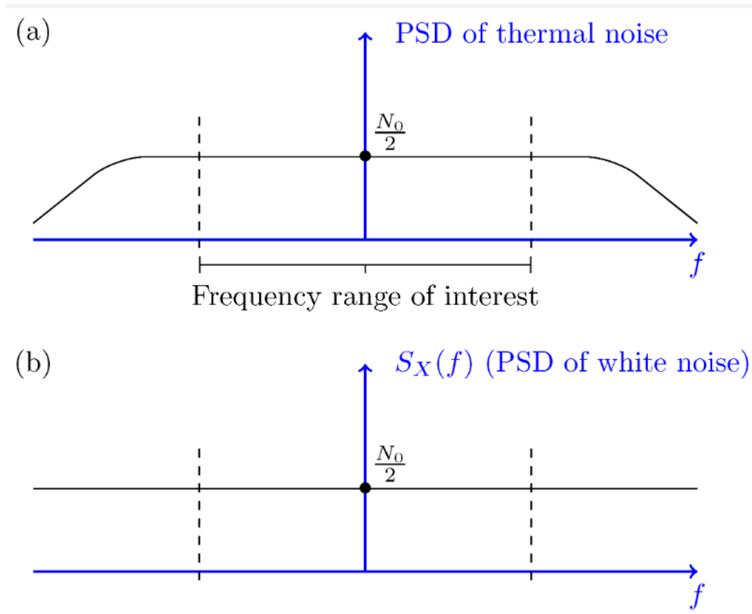


Figure 3.5: Types of Noise.

3.4.3. Cross Spectral Density

For two jointly WSS random processes X(t) and Y(t), we define the cross spectral density $S_{XY}(f)$ as the Fourier transform of the cross-correlation function $R_{XY}(\tau)$,

$$S_{XY}(f) = \mathcal{F}\{R_{XY}(\tau)\} = \int_{-\infty}^{+\infty} R_{XY}(\tau) e^{-2j\pi f\tau}\, d\tau$$

3.5. Sampling of Random Signals

The process of sampling a continuous-time random signal, X(t), yields a discrete-time random sequence, X[n], where $t = nTs$ and Ts is the sampling period.

$$X[\mathrm{n}] = X(nTs)$$

3.5.1. The Sampling Theorem (Nyquist-Shannon Theorem)

While the theorem was originally derived for deterministic, band-limited signals, it must be carefully applied to random signals, specifically to their **Power Spectral Density (PSD)**.

For a continuous-time random process X(t) whose PSD, $S_X(\omega)$, is band-limited (meaning $S_X(\omega) = \mathbf{0}$ for $|\omega| > \omega_B$):

1. **Nyquist Rate $\boldsymbol{\omega_N}$:** This is the minimum sampling frequency required to avoid aliasing.

$$\omega_N = 2\omega_B \text{ or } f_N = 2f_B$$

2. **Sampling Frequency (f_s):** The actual sampling frequency must be greater than or equal to the Nyquist rate.

$$f_s > f_N$$

If the signal is sampled at $f_s > f_N$, then the original continuous-time signal X(t) can theoretically be perfectly reconstructed from the discrete sequence X[n].

3.5.2. Aliasing in Random Signals

When the sampling frequency is too low ($f_s < f_N$), aliasing occurs.

- **Effect:** The high-frequency components of the signal's PSD are folded back into the lower-frequency range, distorting the spectrum.
- **Aliased PSD:** The PSD of the discrete-time sequence, $S_X^{DT}(\omega)$, is a superposition of shifted and scaled versions of the continuous-time PSD, , $S_X^{DT}(\omega)$:

$$S_X^{DT}(\omega) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} S_X^{CT}(\omega - k\omega_s)$$

where, $\omega_s = 2\pi/T_S$is the sampling frequency.
If $S_X^{DT}(\omega)$is not band-limited, the shifted copies overlap, causing the aliasing distortion in $S_X^{DT}(\omega)$. This is why anti-aliasing filters are mandatory before sampling to ensure the input signal is adequately band-limited.

3.6. Filtering of Random Signals-Adapted filter, Wiener filter

Filtering refers to passing a random process X(t) through a Linear Time-Invariant system, h(t), to produce an output random process Y(t).
When a WSS random process X(t) is input to an LTI system with impulse response h(t) and transfer function H(ω), the output process Y(t) is also WSS.

$$\boldsymbol{Y}(t) = \boldsymbol{X}(t) * h(t)$$

3.6.1. Frequency-Domain Relationship (Power Spectral Density)

The most useful and simplified relationship is in the frequency domain, linking the input and output PSDs. This is an extension of the LTI system frequency response for deterministic signals.

The output PSD, $S_Y(\omega)$, is found by multiplying the input PSD, $S_Y(\omega)$, by the squared magnitude of the system's frequency response (transfer function), H(ω):

$$S_Y(\omega) = |H(\omega)|^2 S_X(\omega)$$

$$|H(\omega)|^2 = H(\omega)H^*(\omega)$$

The filter's transfer function acts as a weighting function, shaping the spectral distribution of the power in the input signal.

Maximize the **peak Signal-to-Noise Ratio (SNR)** at the output:

3.6.2. Adapted Filter (Matched Filter)

The Adapted Filter or Matched Filter is a filter designed to maximize the peak instantaneous output power of a known deterministic signal s(t) in the presence of additive, wide-sense stationary noise n(t) at a specific sampling time t_0. This is crucial for detection systems in telecom (e.g., radar, digital receiver fronts).

$$SNR_{max} = \frac{|Y_s(t_0)|^2}{E[Y_n^2(t_0)]}$$

Where $Y_s(t_0)$ is the output signal component and $\boldsymbol{E}[Y_n^2(t)]$ is the average output noise power.

If the input noise N(t) is White Noise (i.e., SN(ω) = N_0/2, the optimal filter's impulse response h_{opt}(t) is:

$$h_{opt}(t) = s(t_0 - t)$$

The optimal filter's impulse response is a time-reversed and time-shifted version of the known signal s(t). It is "matched" to the signal.

Frequency Domain : The optimal transfer function $H_{opt}(\omega)$ is:

$$H_{opt}(\omega) = S^*(\omega)e^{-j\omega t_0}$$

Where *S(ω)* is the Fourier Transform of the signal s(t).

The maximum SNR achieved by the matched filter is proportional to the energy of the signal:

$$SNR_{max} = \frac{2E_s}{N_0}$$

Where E_s is the energy of the signal s(t) and $N_0/2$ is the noise PSD.

3.6.3. Wiener Filter

The Wiener Filter is the optimal LTI filter designed to minimize the mean-squared error (MSE) between the actual output Y(t) and some desired random signal D(t) (the estimation objective). This is the optimal filter for estimation and prediction.

The filter $H_W(\omega)$ is designed to minimize the cost function :

$$MSE = E[(D(t) - Y(t))^2]$$

The optimal transfer function $H_W(\omega)$ is given by :

$$H_W(\omega) = \frac{S_{DX}(\omega)}{S_X(\omega)}$$

Where:

- $S_{DX}(\omega)$: **Cross-PSD** between the desired signal $D(t)$ and the input signal $X(t)$.
- $S_X(\omega)$: **Auto-PSD** of the input signal $X(t)$. (Note: $X(t)$ is usually the sum of the signal and noise).

In a common scenario where the input X(t) is composed of the desired signal S(t) plus additive noise N(t), and the signal and noise are uncorrelated (i.e., X(t) = S(t) + N(t) and $S_{SN}(\omega) = 0$), the formula simplifies to:

$$H_W(\omega) = \frac{S_S(\omega)}{S_S(\omega) + S_N(\omega)}$$

The Wiener filter acts as a non-causal spectral smoother. If the signal PSD $S_S(\omega)$ is much larger than the noise PSD $S_N(\omega)$ at a certain frequency, $H_W(\omega) \approx 1$. If the noise dominates, $H_W(\omega) \approx 0$. It essentially performs an optimal trade-off between smoothing the noise and preserving the signal.

3.7. Statistical estimation and spectral estimation

3.7.1. Statistical Estimation (Parametric Method)

In practice, we never have access to the infinite ensemble of a random signal, X(t). We only observe a single, finite-length sequence of data, x[n], for n=0,1………… N-1.

Statistical Estimation is the process of using this finite-length data segment to find approximate values for the signal's underlying statistical parameters (like mean, variance, and, most importantly, the PSD).

- **Bias**: Whether the estimator's expected value equals the true value.
- **Variance**: How much estimates vary from sample to sample.
- **Consistency**: Estimator improves with more data.

Common methods: Maximum Likelihood, Least Squares, Method of Moments.

3.7.2. Spectral Estimation (Non-Parametric Methods)

Spectral estimators are generally grouped into two classes, parametric and nonparametric. A parametric estimator assumes a certain model for the random process with several unknown parameters and then attempts to estimate the parameters. Given the model parameters, the PSD is then computed analytically from the model.

On the other hand, a nonparametric estimator makes no assumptions about the nature of the random process and estimates the PSD directly. Since parametric estimators take advantage of some prior knowledge of the nature of the process, it would be expected that these estimators are more accurate. However, in some cases, prior knowledge may not be available, in which case a non parametric estimator may be more appropriate. We start with a description of some basic techniques for nonparametric spectral estimation.

Problem: We don't have infinite data or true Rxx(τ), so we must estimate $S_x(f)$ from N samples.

1. **Nonparametric**: Doesn't assume a specific signal model (e.g., Periodogram-based methods).
2. **Parametric**: Assumes a model (AR, MA, ARMA) and estimates model parameters first.

3.8. Periodogram, correlogram, averaged periodogram, smoothed periodogram

These estimators are based on Fourier analysis, leveraging the Wiener-Khinchine theorem PSD = F\{ACF\}.

3.8.1 The Periodogram (Schuster, 1898)

The **Periodogram**, $\hat{S}_X^{PER}(\omega)$, is the most straightforward but often problematic estimator. It is calculated by taking the Fourier Transform of the data sequence itself (or its estimate). Given N samples x[0],x[1],…,x[N−1]x[0],x[1],…,x[N−1]:

$$\hat{S}_x^{PER}(\omega) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]e^{-j\omega n}\right|^2 = \frac{1}{N}|X_N(\omega)|^2$$

It is the normalized magnitude-squared of the Discrete Fourier Transform (DFT) of the data. The Periodogram is an **inconsistent estimator**. Its variance does not decrease as the data record length N increases. Its estimate is highly noisy, making it a poor choice on its own.

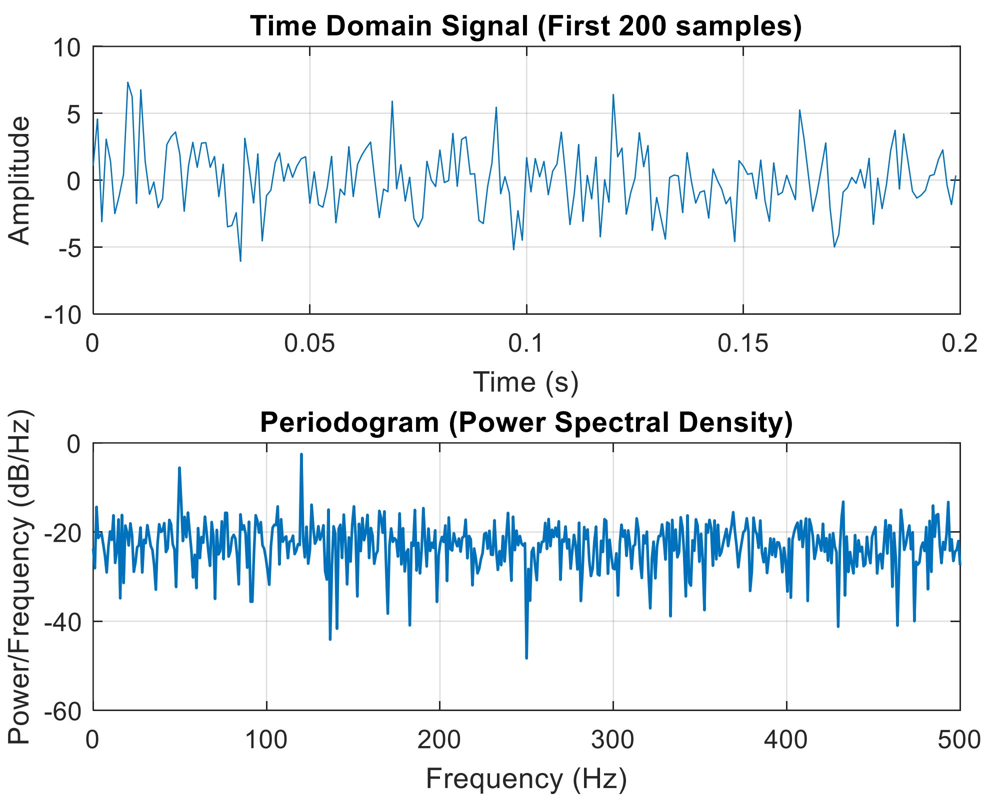


Figure 3.6: Periodogram PSD Estimate.

3.8.2 Correlogram (Blackman-Tukey method)

The **Correlogram** uses the Wiener-Khinchine theorem directly, by first calculating the ACF estimate $\hat{\boldsymbol{R}}_{\boldsymbol{X}}^{\boldsymbol{b}}[\boldsymbol{k}]$, and then taking its Fourier Transform.

$$\hat{S}_X^{COR}(\omega) = \sum_{k=-M}^{M} \hat{R}_X^b[k]e^{-j\omega k}$$

Windowing: A data window must be applied to the ACF estimate (hence the index M, where M < N) to ensure the resulting PSD is smooth and physically meaningful.

This estimator is often biased but can achieve lower variance than the periodogram by controlling the length of the window M, i.e, smaller M → more smoothing (lower variance, higher bias in frequency resolution).

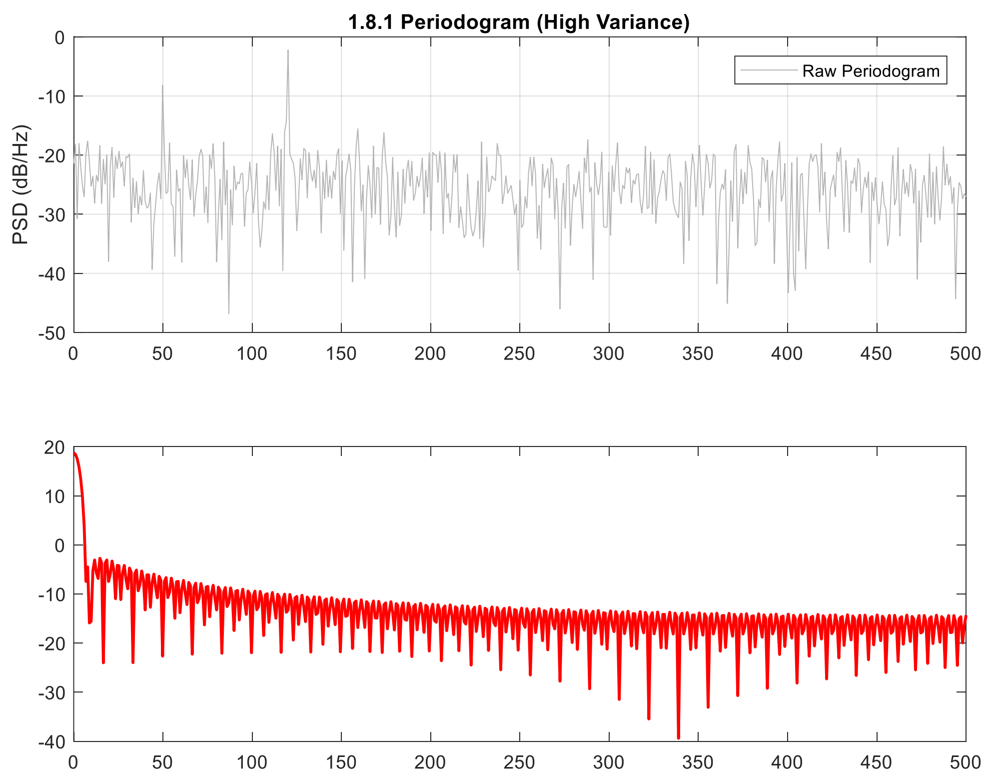


Figure 3.7: Periodogram vs. Correlogram (Blackman-Tukey).

3.8.3 Averaged Periodogram (Welch's Method)

To overcome the inconsistency of the standard Periodogram, the data is typically broken into shorter, overlapping segments, and the periodogram of each segment is calculated and then averaged.

$$\hat{S}_X^{WELCH}(\omega) = \frac{1}{L}\sum_{i=0}^{L-1} \hat{S}_{X,i}^{PER}[\omega]$$

Averaging the periodograms significantly reduces the variance of the estimate, resulting in a much smoother and more reliable PSD estimate, at the expense of slight frequency resolution loss.

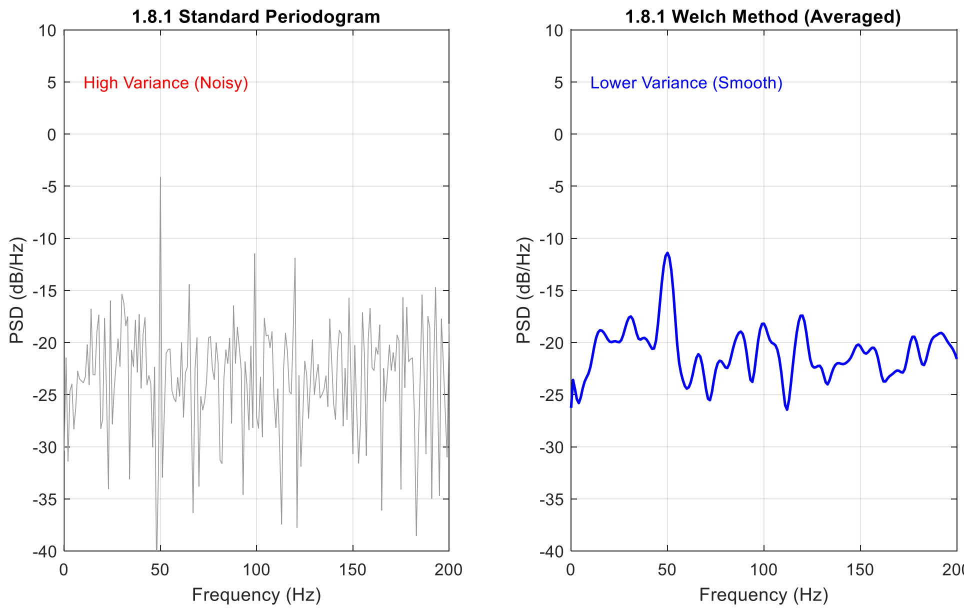


Figure 3.8: Comparing Periodogram vs. Welch's Method

3.8.4 Smoothed Periodogram

The **Smoothed Periodogram** (or often just a specific implementation of the Welch method) refers to applying a smoothing filter (convolution) to the Periodogram in the frequency domain, which is equivalent to multiplying the ACF estimate by a window function in the time domain.

Method:

- Split data into overlapping segments.
- Apply a data window (e.g., Hamming) to each segment before periodogram.
- Average the modified periodograms.

Reduces variance at expense of resolution (similar to averaging), but overlapping segments can increase number of segments for same length L.

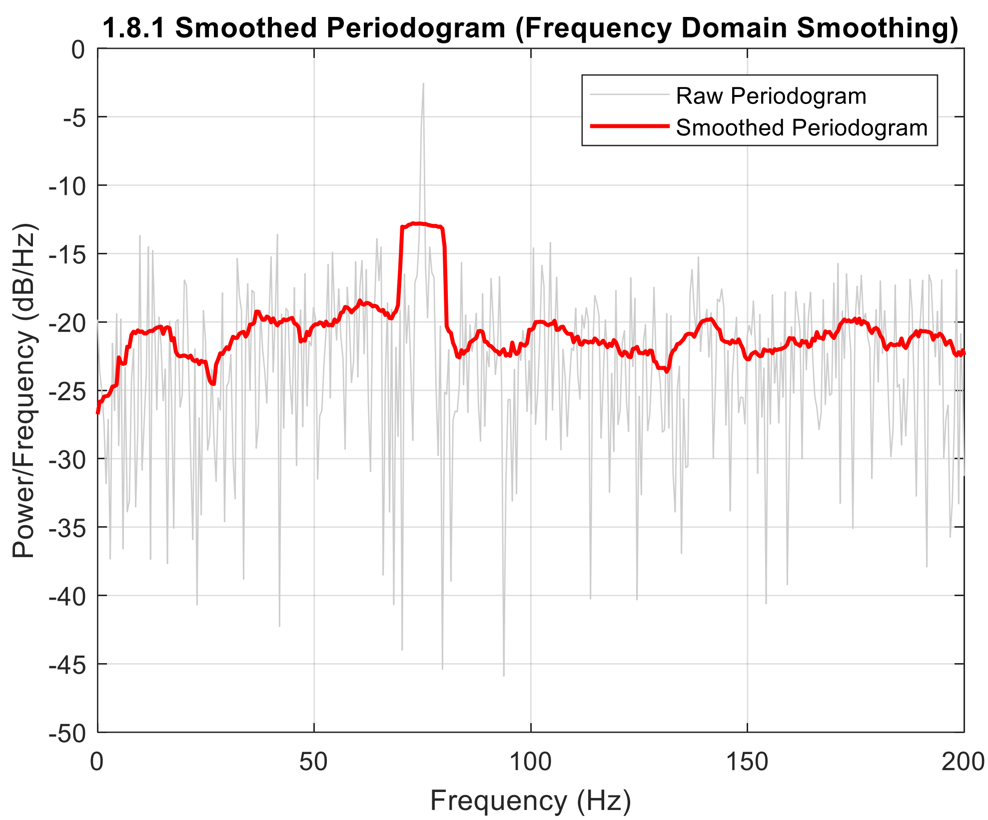


Figure 3.9: Comparing Smoothed Periodogram vs. Raw Periodogram.

3.9. AR, MA and ARMA models

These are modern techniques that assume the signal X[n] is the output of a filter driven by white noise W[n]. The estimation task is reduced to finding the optimal filter coefficients, which is often easier and more accurate than directly estimating the PSD non-parametrically.

3.9.1. AutoRegressive (AR) Model

The current sample X[n] is a linear combination of previous samples of X[n] plus a white noise excitation W[n].

$$\mathbf{X}[n] = -a_1\mathbf{X}[n-1] - a_2\mathbf{X}[n-2] - \cdots - a_p\mathbf{X}[n-p] + \mathbf{W}[n]$$

Transfer Function: The process is modeled as an All-Pole filter (like the output of a system described by difference equations).

Estimation: The coefficients a_i are typically found using the Yule-Walker equations.

It is used for modeling sharp spectral peaks (formants in speech processing).

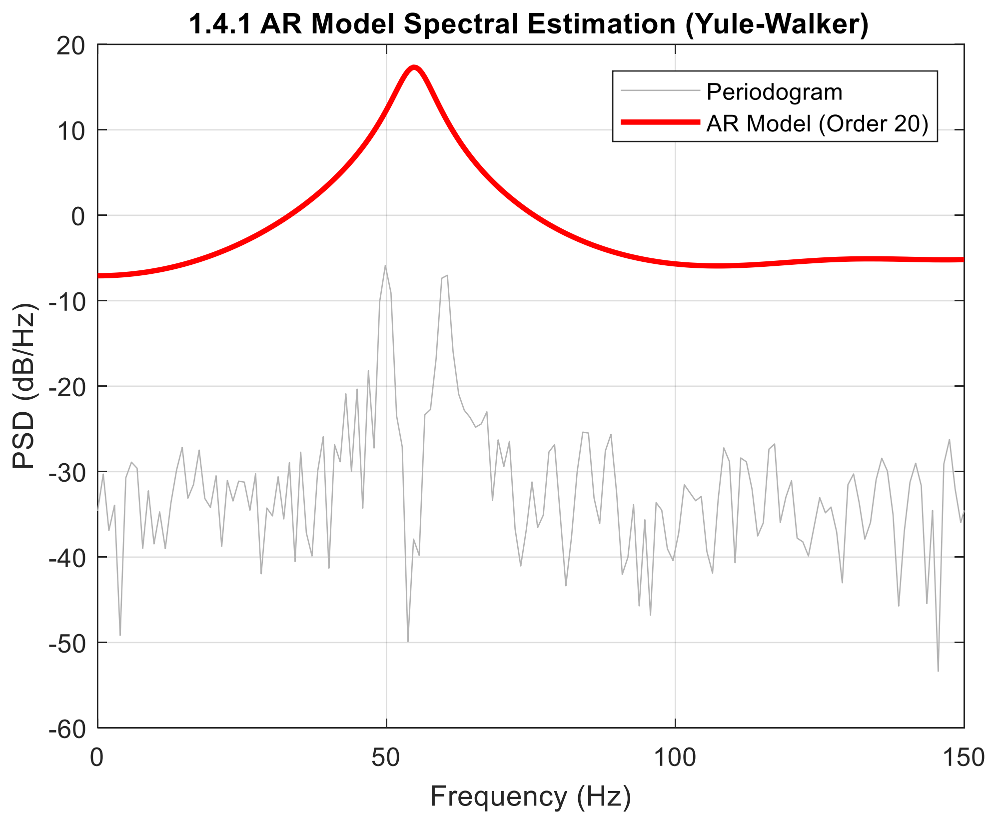


Figure 3.10: Periodogram vs. AR model.

3.9.2. Moving Average (MA) Model

The current sample X[n] is a linear combination of current and previous white noise samples.

$$\mathbf{X}[n] = b_0\mathbf{W}[n] + b_1\mathbf{W}[n-1] + \cdots + b_q\mathbf{W}[n-q]$$

Transfer Function: The process is modeled as an All-Zero filter (Finite Impulse Response, or FIR, filter).

It is used for modeling sharp spectral nulls (antiformants).

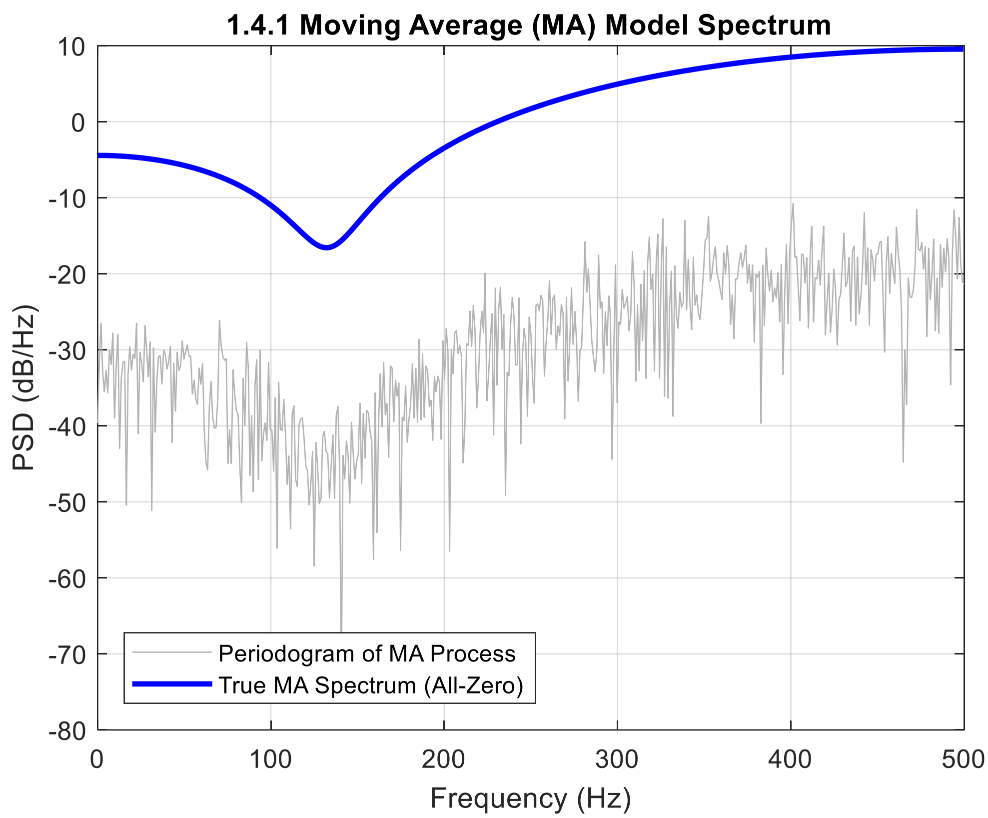


Figure 3.11: Moving Average (MA) Model.

3.9.3. AutoRegressive Moving Average (ARMA) Model

The **ARMA** model combines both AR and MA components, making it the most general (and complex) of the three.

$$\mathbf{X}[n] = \sum_{i=1}^{p} a_i \mathbf{X}[n-i] + \sum_{j=0}^{q} b_j \mathbf{W}[n-j]$$

Transfer Function: Modeled as a **Pole-Zero** filter.

Can often model a complex process with fewer parameters than pure AR or MA models.

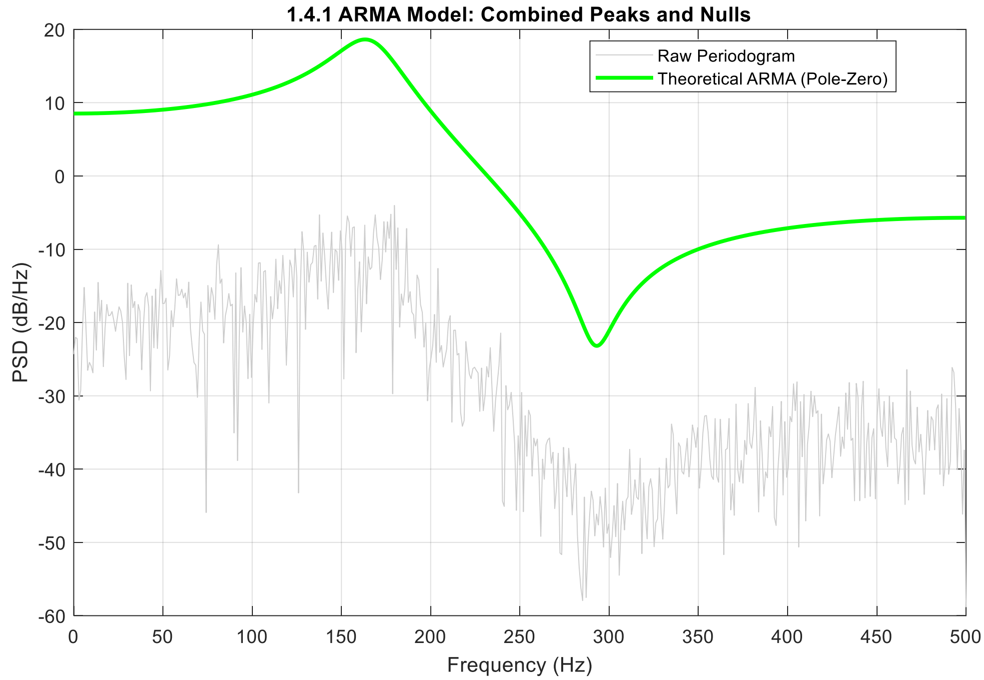


Figure 3.12: Moving Average (MA) Model.

EXERCISES

Exercise 3.1:

A Wide-Sense Stationary (WSS) random process, X(t), has an Autocorrelation Function (ACF) defined as:

$$R_X(\tau) = 5e^{-2|\tau|} + 9$$

1. Calculate the mean and the total average power P_{Total} of the process X(t).
2. Determine the Power Spectral Density (PSD), $S_X(\omega)$, of the process X(t).

Exercise 3.2:

Consider a WSS random process X t) with :

$$R_X(\tau) = e^{-a|\tau|}$$

Where *a* is a positive real number. Find the PSD of X(t)

Exercise 3.3: Matched Filter (Adapted Filter)

A known deterministic signal, s(t), is given by:

$$s(t) = \begin{cases} 4 & 0 \leq t \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

This signal is corrupted by White Noise with a PSD of $S_X(\omega)$, = $N_0/2 = 0.5$. The goal is to design a matched filter to maximize the output SNR at time $t_0 = 2$ seconds.

1. Determine the optimal impulse response of the matched filter.
2. Calculate the maximum achievable output Signal-to-Noise Ratio (SNR)

Exercise 3.4: Parametric Model Identification and PSD

A discrete-time WSS random process, X[n], is generated by passing white noise, W[n], through a filter. The process X[n] satisfies the following difference equation:

$$\mathbf{X}[n] = 0.5\mathbf{X}[n-1] + \mathbf{W}[n]$$

The white noise input, W[n], has a variance of $W^2 = 1$.

Exercise 3.5: Create a signal in MATLAB that consists of:

1. Two closely spaced sine waves at **150 Hz** and **155 Hz**.
2. A "spectral null" (a dip) at **300 Hz**.
3. Additive White Gaussian Noise (AWGN).
4. Use a sampling frequency Fs = 1000 Hz and a short duration of 0.5 seconds.

Exercise 3.6: Apply the following methods to your signal and answer the observation questions:

1. **The Raw Periodogram:** Use periodogram. Can you clearly see two distinct peaks at 150 and 155 Hz, or do they look like one messy "blob"?
2. **Welch’s Method:** Use pwelch with a segment length of 128. Does the noise look smoother compared to the Periodogram?
3. **AR Model (Yule-Walker):** Use pyulear with an order of 20. Does this model resolve the two closely spaced peaks better than the Periodogram?
4. **ARMA Model:** Estimate the spectrum using the known coefficients or an ARMA estimator. Does this model show the dip at 300 Hz more clearly than the AR-only model?

3.10. Conclusion

This chapter successfully transitioned the focus from deterministic signal analysis to the statistical treatment required for real-world telecommunications. The key was establishing Wide-Sense Stationarity (WSS), which allows complex random signals to be analyzed using simple statistical moments (mean, variance) and the Autocorrelation Function $R_X(\tau)$. This led directly to the core analytical tool, the Power Spectral Density $S_X(\omega)$, linked via the Wiener-Khinchine Theorem. Applying this structure to LTI systems provided the essential relationship $S_Y(\omega) = \lceil \mathrm{H}(\omega) \rceil^2 \, S_X(\omega)$, enabling the design of optimal filters.

We specifically examined the Matched Filter for maximizing signal detection and the Wiener Filter for minimizing estimation error. Finally, recognizing that statistics must be estimated from finite data, the chapter introduced practical Spectral Estimation methods, ranging from non-parametric techniques (like the Averaged Periodogram) to advanced parametric models (AR, MA, ARMA), thus providing the complete theoretical and applied toolkit necessary for processing random signals in modern telecom systems.

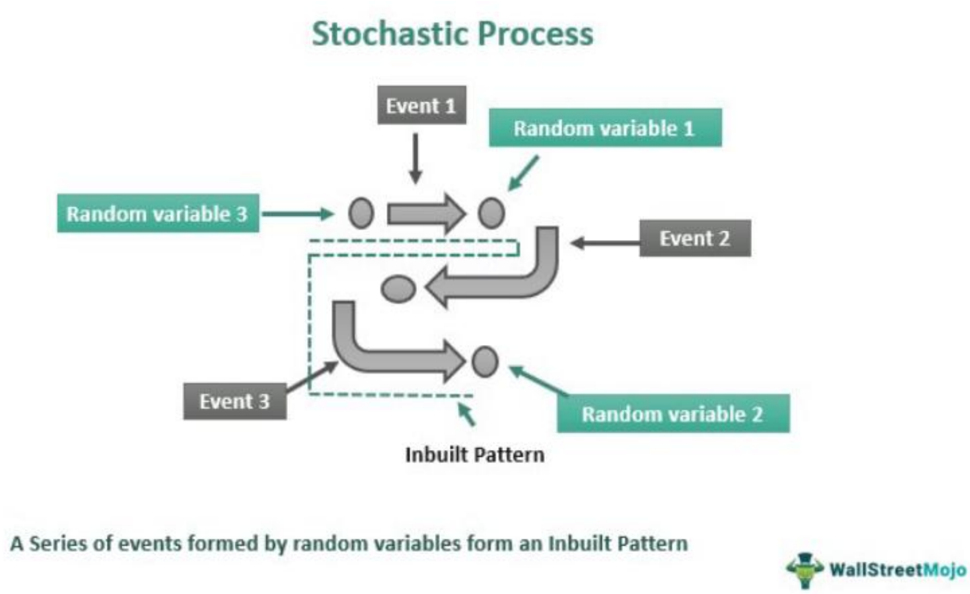
Chapter 4: Stochastic Processes

Chapter 4: Stochastic Processes

Course objective

At the end of this chapter, the student will be able to:

1. **Define** fundamental notions of stochastic processes and distinguish between stationarity in the strict and broad sense, as well as the concept of Ergodicity.
2. **Characterize** the behavior of linear systems driven by stochastic inputs and analyze their output properties.
3. **Describe** and **classify** key stochastic processes, including Poisson (point processes), Gaussian, and Markovian processes, by their defining properties.
4. **Explain** the limitations of second-order statistics for non-Gaussian processes and utilize higher-order statistics (moments, cumulants, polyspectra) for their analysis.
5. **Understand** the principle of particle filtering as a sequential Monte Carlo method for state estimation in non-linear and non-Gaussian dynamic systems.



4.1. Introduction

Much of your background in signals and systems is assumed to have focused on the effect of LTI systems on deterministic signals, developing tools for analyzing this class of signals and systems, and using what you learned in order to understand applications in communication (e.g., AM and FM modulation), control (e.g., stability of feedback systems), and signal processing (e.g., filtering). It is important to develop a comparable understanding and associated tools for treating the effect of LTI systems on signals modeled as the outcome of probabilistic experiments, i.e., a class of signals referred to as random signals (alternatively referred to as random processes or stochastic processes). Such signals play a central role in signal and system design and analysis, and throughout the remainder of this text. In this chapter we define random processes via the associated ensemble of signals, and begin to explore their properties. In successive chapters we use random processes as models for random or uncertain signals that arise in communication, control and signal processing applications.

4.2. Notions of stochastic processes

We defined a random variable X as a function that maps each outcome of a probabilistic experiment to a real number. In a similar manner, a real-valued CT or DT random process, X(t) or X[n] respectively, is a function that maps each outcome of a probabilistic experiment to a real CT or DT signal respectively, termed the realization of the random process in that experiment. For any fixed time instant $t = t_0$ or $n = n_0$, the quantities $X(t_0)$ and $X[n_0]$ are just random variables. The collection of signals that can be produced by the random process is referred to as the ensemble of signals in the random process.

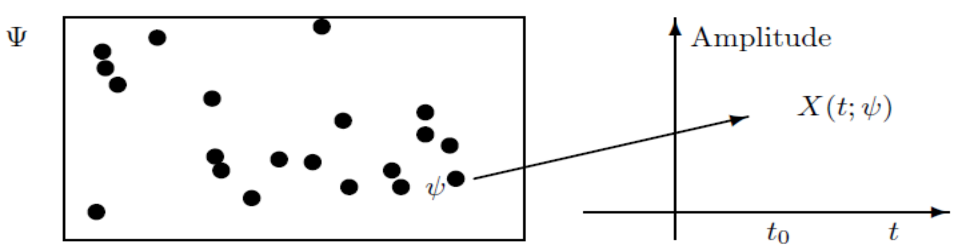


Figure 4.1: A random process.

Definition: A random process is a function of the elements of a sample space, S, as well as another independent variable, t. Given an experiment, E, with sample space, S, the random process, X(t), maps each possible outcome, $\zeta \in S$, to a function of t, $x(t, \zeta)$, as specified by some rule.

Example : Suppose an experiment consists of flipping a coin. If the outcome is heads, $\zeta = H$, the random process takes on the functional form $x_H(t) = \sin(\omega_o t)$; whereas, if the outcome is tails, $\zeta = T$, the realization $x_T(t) = \sin(2\omega_o t)$ occurs, where ω_o is some fixed frequency. The two realizations of this random process are illustrated in Figure 4.2.

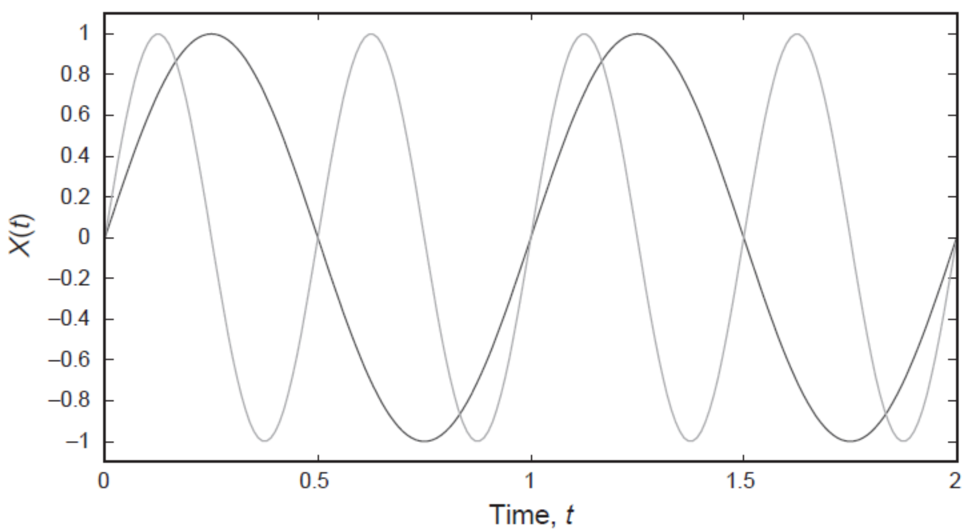


Figure 4.2: Member functions for the random process.

4.3. Strict and Wild Sense Stationarities, Ergodicity

4.3.1. STRICT-SENSE STATIONARITY

Definition: A continuous time random process X(t) is strict sense stationary if the statistics of the process are invariant to a time shift. Specifically, for any time shift τ and any integer $n \geq 1$.

In general, we would expect that the joint PDFs associated with the random variables obtained by sampling a random process at an arbitrary number k of arbitrary times will be time-dependent, i.e., the joint PDF :

$$f_{X(t_1),\ldots\ldots,X(t_k)}X(x_1, \ldots\ldots\ldots, x_k)$$

will depend on the specific values of $t_1,\ldots,t_k$. If all the joint PDFs stay the same under arbitrary time shifts, i.e., if

$$f_{X(t_1),\ldots\ldots,X(t_k)}X(x_1,\ldots\ldots\ldots,x_k) = f_{X(t_1+\tau),\ldots\ldots,X(t_k+\tau)}X(x_1,\ldots\ldots\ldots,x_k)$$

for arbitrary τ , then the random process is said to be strict-sense stationary (SSS). Said another way, for a strict-sense stationary process, the statistics depend only on the relative times at which the samples are taken, not on the absolute times.

4.3.2. WIDE-SENSE STATIONARITY

Definition: A random process is wide sense stationary (WSS) if the mean function and autocorrelation function are invariant to a time shift.

Of particular use to us is a less restricted type of stationarity. Specifically, if the mean value $\mu_X(t_i)$ is independent of time and the autocorrelation $R_{XX}(t_i,t_j)$ or equivalently the autocovariance $C_{XX}(ti,tj)$ is dependent only on the time difference $(t_i - t_j)$, then the process is said to be wide-sense stationary (WSS). Clearly a process that is SSS is also WSS. For a WSS random process X(t), therefore, we have

$$\mu_{X(t)} = \mu_X$$

$$R_{XX}(t_1,t_2) = RXX(t_1 + \alpha, t_2 + \alpha)\ for\ every\ \alpha = R_{XX}(t_1 - t_2, 0)$$

(Note that for a Gaussian process (i.e., a process whose samples are always jointly Gaussian) WSS implies SSS, because jointly Gaussian variables are entirely determined by the their joint first and second moments.)

Two random processes X(t) and Y(t) are jointly WSS if their first and second moments (including the cross-covariance) are stationary. In this case we use the notation $R_{XY}(\tau)$ to denote $E[X(t + \tau)Y(t)]$.

4.3.3. ERGODICITY

The concept of ergodicity is sophisticated and subtle, but the essential idea is described here. We typically observe the outcome of a random process (e.g., we record a noise waveform) and want to characterize the statistics of the random process by measurements on one ensemble member. For instance, we could consider the time-average of the waveform to represent the mean value of the process (assuming this mean is constant for all time).

Definition: A WSS random process is ergodic if ensemble averages involving the process can be calculated using time averages of any realization of the process. Two limited forms of ergodicity are:

1. Ergodic in the mean: $x(t) = E[X(t)]$;
2. Ergodic in the autocorrelation: $x(t + \tau) = E[X(t)X(t + \tau)]$.

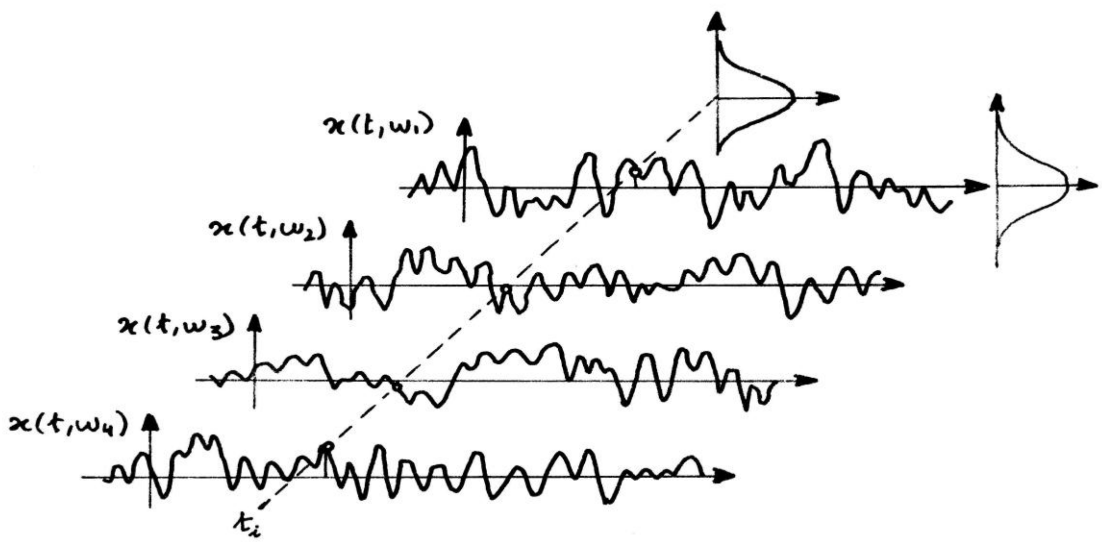


Figure 4.3: The concept of Ergodicity.

Example: Now consider the sinusoid with random phase $X(t) = a\,sin(\omega_o t + \theta)$, where θ is uniform over [0, 2π). This process is WSS. But is it ergodic?

Given any realization $x(t) = a\,sin(\omega_o t + \theta)$, the time average is :

$$\langle x(t)\rangle = \langle a\,sin(\omega o t + \theta)\rangle = 0$$

That is, the average value of any sinusoid is zero. So this process is ergodic in the mean since the ensemble average of this process was also zero. Next, consider the sample autocorrelation function:

$$\langle x(t)x(t+\tau)\rangle = a^2\langle \sin(\omega_0 t+\theta)\sin(\omega_0 t+\omega_0\tau+\theta)\rangle$$

$$= \frac{a^2}{2}\langle \cos(\omega_0\tau)\rangle - \frac{a^2}{2}\langle \cos(2\omega_0 t+\omega_0\tau+2\theta)\rangle = \frac{a^2}{2}\cos(\omega_0\tau).$$

This also is exactly the same expression obtained for the ensemble averaged autocorrelation function. Hence, this process is also Ergodic in the autocorrelation.

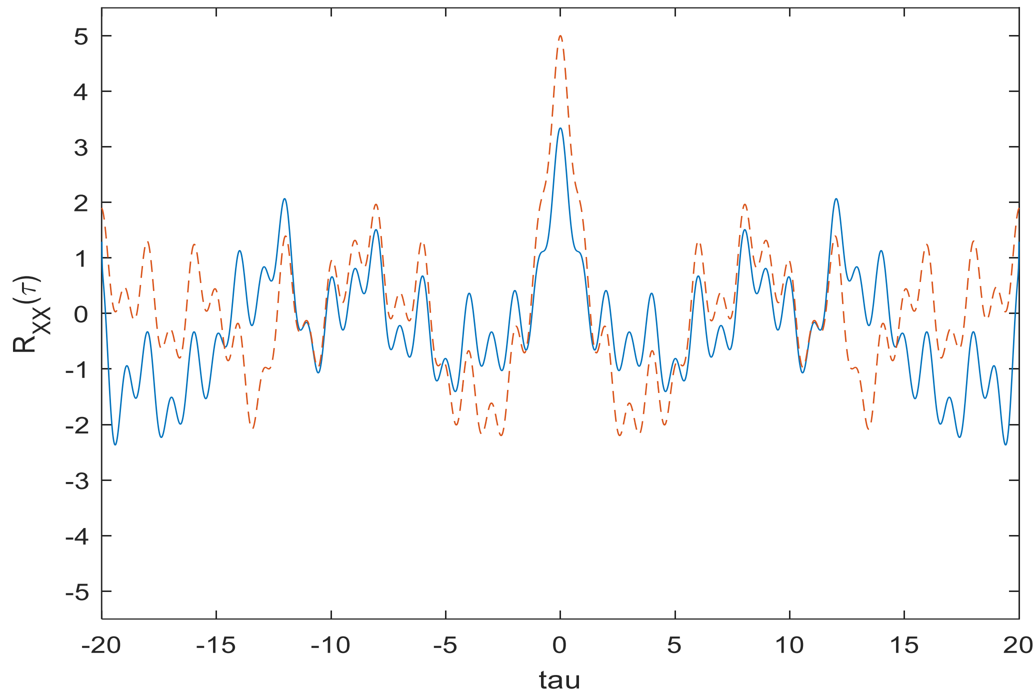


Figure 4.4 : Comparison of the time-average autocorrelation and the ensemble-average autocorrelation for the sum of sinusoids process of Example.

The solid line is the time average autocorrelation, while the dashed line is the ensemble-average autocorrelation.

4.4. Stochastic input systems

A common class of problems in a variety of aspects of communication, control and signal processing involves the estimation of one random process from observations of another, or estimating (predicting) future values from the observation of past values.

For example, it is common in communication systems that the signal at the receiver is a corrupted (e.g., noisy) version of the transmitted signal, and we would like to estimate the transmitted signal from the received signal. Other examples lie in predicting weather and financial data from past observations. We will be treating this general topic in much more detail in later chapters, but a first look at it here can be beneficial in understanding random processes.

As a simple illustration of linear prediction, consider a discrete-time process x[n]. Knowing the value at time n_0 we may wish to predict what the value will be m samples into the future, i.e. at time $n_0 + m$. We limit the prediction strategy to a linear one, i.e., with $\hat{x}[n_0 + m]$ denoting the predicted value, we restrict $\hat{x}[n_0 + m]$ to be of the form :

$$\hat{x}[n_0 + m] = ax[n_0] + b$$

and choose the prediction parameters a and b to minimize the expected value of the square of the error, i.e., choose a and b to minimize :

$$\epsilon = E\{(x[n_0 + m] - ax[n_0] - b)^2\}.$$

If we assume that the process is WSS so that $R_{xx}[n_0+m, n_0] = R_{xx}[m]$, $R_{xx}[n_0, n_0] = R_{xx}[0]$, and also assume that it is zero mean, ($\mu_x = 0$), then equations reduce to :

$$a = R_{xx}[m]/R_{xx}[0]$$
$$b = 0$$

So that,

$$\widehat{x}[n_0 + m] = \frac{R_{xx}[m]}{R_{xx}[0]} x[n_0].$$

If the process is not zero mean, then it is easy to see that,

$$\widehat{x}[n_0 + m] = \mu_x + \frac{C_{xx}[m]}{C_{xx}[0]}(x[n_0] - \mu_x) .$$

4.5. Examples of stochastic processes (Poisson, Gaussian and Markovian processes)

As we call before, a stochastic process is a collection of random variables indexed by time or space, representing the evolution of some random phenomenon. These three examples represent different types of randomness. The three types listed are fundamental models used across science, engineering, and finance.

4.5.1. Gaussian Random Processes

One of the most important classes of random processes is the Gaussian random process, which is defined as follows.

Definition : A random process, $X(t)$, for which any n samples, $X_1 = X(t_1), X_2 = X(t_2), \ldots, X_n = X(t_n)$, taken at arbitrary points in time $t_1, t_2, \ldots, t_n$, form a set of jointly Gaussian random variables for any $n = 1, 2, 3, \ldots$ is a Gaussian random process.

In vector notation, the vector of n samples, $\boldsymbol{X} = [X1, X2, \ldots, Xn]^T$, will have a joint PDF given by :

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det(C_{XX})}} \exp\left(-\frac{1}{2}(x - \mu_X)^T C_{XX}^{-1}(x - \mu_X)\right).$$

As with any joint Gaussian PDF, all that is needed to specify the PDF is the mean vector and the covariance matrix. When the vector of random variables consists of samples of a random process, to specify the mean vector, all that is needed is the mean function of the random process, $\mu_{X(t)}$, since that will give the mean for any sample time. Similarly, all that is needed to specify the elements of the covariance matrix, $C_{i,j} = \text{Cov}(X(t_i),X(t_j))$, would be the autocovariance function of the random process, $C_{XX}(t_1, t_2)$ or, equivalently, the autocorrelation function,

$R_{XX}(t_1, t_2)$, together with the mean function. Hence, the mean and autocorrelation functions provide sufficient information to specify the joint PDF for any number of samples of a Gaussian random process. Note that since any nth order PDF is completely specified by $\mu_{X(t)}$ and $R_{XX}(t_1, t_2)$, if a Gaussian random process is WSS, then the mean and autocorrelation functions will be invariant to a time shift and hence any PDF will be invariant to a time shift. Therefore, any WSS Gaussian random process is also stationary in the strict sense.

4.5.2. Poisson Processes

Definition : A Poisson process is a counting process that models the occurrence of events randomly and independently over time. It is defined by a single parameter, the rate (λ) , which represents the average number of events per unit of time.

Consider a process X(t) that counts the number of occurrences of some event in the time interval [0, t). The event might be the telephone calls arriving at a certain switch in a public telephone network, customers entering a certain store, or the birth of a certain species of animal under study.

Since the random process is discrete (in amplitude), we will describe it in terms of a probability mass function, $P_X(i;\ t) = P_r(X(t) = i)$.

Each occurrence of the event being counted is referred to as an arrival, or a point. These types of processes are referred to as counting processes, or birth processes. Suppose this random process has the following general properties:

- Independent Increments : The number of arrivals in two nonoverlapping intervals are independent. That is, for two intervals $[t_1, t_2)$ and $[t_3, t_4)$ such that $t_1 \le t_2 \le t_3 \le t_4$, the number of arrivals in $[t_1, t_2)$ is statistically independent of the number of arrivals in $[t_3, t_4)$.
- Stationary Increments : The number of arrivals in an interval $[t, t+\tau)$ depends only on the length of the interval τ and not on where the interval occurs, t.

Consider the PMF of the counting process at time *t*+*t*. In particular, consider finding the probability of the event $\{X(t + t) = 0\}$.

$$\begin{aligned} P_X(0; t + \Delta t) &= \Pr(\text{no arrivals in } [0, t + \Delta t)) \\ &= \Pr(\text{no arrivals in } [0, t)) \Pr(\text{no arrivals in } [t, t + \Delta t)) \\ &= P_X(0; t)[1 - \lambda \Delta t + o(\Delta t)] \end{aligned}$$

Subtracting $P_X(0; t)$ from both sides and dividing by *t* results in :

$$\frac{P_X(0; t + \Delta t) - P_X(0; t)}{\Delta t} = -\lambda P_X(0; t) + \frac{o(\Delta t)}{\Delta t} P_X(0; t).$$

Passing to the limit as $t \to 0$ gives the first order differential equation

$$\frac{d}{dt} P_X(0; t) = -\lambda P_X(0; t).$$

The solution to this equation is of the general form :

$$P_X(0; t) = c \exp(-\lambda t) u(t)$$

for some constant *c*. The constant *c* is found to be equal to unity by using the fact that at time zero, the number of arrivals must be zero; that is, $P_X(0; 0) = 1$. Hence,

$$P_X(0; t) = \exp(-\lambda t) u(t).$$

The general solution to the family of differential equations is :

$$P_X(i; t) = \frac{(\lambda t)^i}{i!} e^{-\lambda t} u(t).$$

Starting with the properties made about the nature of this counting process at the start of this section, we have demonstrated that X(t) follows a Poisson distribution, hence this process is referred to as a Poisson counting process.

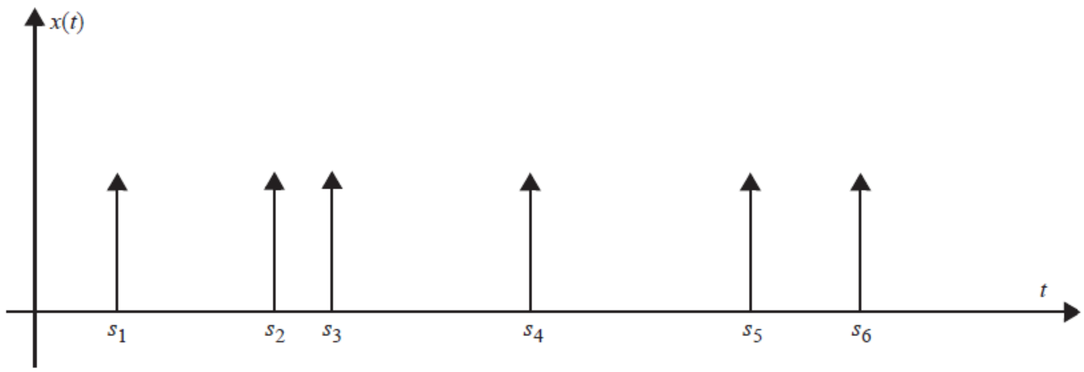


Figure 4.5 : A sample realization of the Poisson impulse process.

4.5.3. Markovian Process (Markov Chain/Process)

Definition: A Markovian process is a stochastic process that satisfies the Markov Property, also known as the "memoryless" property. This property states that the probability of transitioning to any future state depends only on the current state and not on the sequence of events that preceded it.

Markov Property: The future is independent of the past, given the present.

A sequence of random variables $\{X_0, X_1, X_2, ...\}$ with **discrete state space** $S = \{s_1, s_2, ..., s_n\}$ satisfying:

$$P(X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \ldots) = P(X_{n+1} = x_{n+1} | X_n = x_n)$$

Transition Matrix: For a discrete-time Markov chain, the dynamics are defined by a transition probability matrix P, where P_{ij} is the probability of moving from state i to state j in one step.

Real-World Examples :

Domain	States Being Modeled	Application
Search Engines	A web page in the Internet graph.	Google's PageRank Algorithm: Calculating the importance of a webpage based on the probability of a user randomly clicking links.
Finance & Economics	Bull, Bear, or Stagnant market conditions.	Modeling the probability of a market switching between different regimes.
Weather Forecasting	Sunny, Cloudy, or Rainy weather on a given day.	Predicting tomorrow's weather based only on today's weather and a transition probability matrix.
Queueing Theory	The number of customers in a service system (e.g., a phone queue).	Designing efficient service systems by modeling the transitions in queue length.
Genetics/Biology	Nucleotides (A, T, C, G) in a DNA sequence.	Modeling the evolution and substitution of bases in genetic code.

4.6. Higher order statistics (Moments and cumulants, Polyspectra, non-Gaussian processes, non-linear treatments)

Higher-Order Statistics (HOS) are statistical measures that go beyond the traditional first-order (mean) and second-order (variance, autocorrelation, power spectrum) statistics. They involve moments and cumulants of order three or higher, and are essential tools for analyzing non-Gaussian and non-linear systems.

The core motivation for using HOS stems from the property that Gaussian processes are completely characterized by their first- and second-order statistics. Thus, any information contained in the higher orders points directly to non-Gaussianity or non-linearity.

4.6.1. Moments and Cumulants

Both moments and cumulants describe the shape of a probability distribution function (PDF), but they have distinct mathematical and practical advantages.

4.6.1.1. Moments (μ_K)

The *k*-th order moment of a random variable X is the expected value of X^k.

Definition (Central Moment M_k): The moment about the mean $\boldsymbol{\mu_K} = \mathbf{E[X]}$ is $\boldsymbol{M_K} = \mathbf{E[(X - \mu)^k]}$.

Order (k)	Name	Formula (for zero-mean)	Description
1st	Mean (μ)	$E[X]$	Center of the distribution.
2nd	Variance (σ^2)	$E[X^2]$	Spread or power of the distribution.
3rd	Skewness	$E[X^3]$	Measure of asymmetry.
4th	Kurtosis	$E[X^4]$	Measure of "tailedness" (peakedness).

4.6.1.2. Cumulants ($\boldsymbol{k_k}$)

Cumulants are related to moments but have a critical property of independence that makes them superior for signal processing.

Definition: Cumulants are the coefficients in the Taylor series expansion of the cumulant-generating function, which is the natural logarithm of the characteristic function $\Phi(\omega) = E[e^{j\omega X}]$.

$$\Psi(\omega) = \ln(\Phi(\omega)) = \sum_{k=1}^{\infty} \frac{(j\omega)^k}{k!} \kappa_k$$

Order (k)	Name	Relationship to Central Moments (M_k)	Key Property
1st	k_1	$k_1 = M_1 = \mu$ (Mean)	Same as the mean.
2nd	k_2	$k_2 = M_2 = \sigma^2$ (Variance)	Same as the variance.
3rd	k_3	$k_3 = M_3$ (Skewness)	Measures asymmetry.
4th	k_4	$k_4 = M_4 - 3M_2^2$ (Excess Kurtosis)	Measures deviation from Gaussianity.

4.6.1.3. Why Cumulants are Preferred

The primary advantage of cumulants in signal processing is their additivity property and Gaussian blind spot:

1. **Gaussian Annihilation:** The cumulants of order $k \geq 3$ for a Gaussian process are identically zero ($k_k = 0$). This means that methods based on HOS cumulants (e.g., k_3, k_4) are *blind* to any additive Gaussian noise in the signal, effectively suppressing it.
2. **Independence:** The joint cumulant of a set of random variables vanishes if any subset of the variables is statistically independent of the remaining variables. This is not true for moments.
3. **Phase Preservation:** Cumulants, and their Fourier transforms (Polyspectra), preserve both the **amplitude and phase information** of the signal. Second-order statistics (like the Power Spectrum) only retain amplitude information.

4.6.2. Polyspectra (Higher-Order Spectra)

Just as the Power Spectral Density (PSD) is the Fourier Transform of the second-order statistic (autocorrelation), Polyspectra are the Fourier Transform of the higher-order cumulants (autocumulants).

Polyspectra are also known as Higher-Order Spectra (HOS).

4.6.2.1. Bispectrum (Third-Order Spectrum)

The Bispectrum $B(\omega_1, \omega_2)$ is the 3rd-order Polyspectrum, defined as the two-dimensional Fourier Transform of the 3rd-order cumulant (τ_1, τ_2) :

$$B(\omega_1, \omega_2) = \mathcal{F}_{\tau_1,\tau_2}\{C_3(\tau_1, \tau_2)\}$$

Application: Detecting quadratic phase coupling (non-linear interaction between two frequency components ω_1 and ω_2 producing energy at $\omega_1 + \omega_2$). It is non-zero only for non-Gaussian processes.

HOS is the necessary framework when dealing with signals and systems that violate the Gaussian or Linear assumptions:

4.7. Non-Gaussian Processes

Most real-world signals are non-Gaussian (e.g., radar clutter, atmospheric noise, seismic data, finance data). HOS are used to:

- **Detect and Quantify Non-Gaussianity:** The magnitude of k_k for $k \geq 3$ directly measures the deviation of the process from Gaussianity.

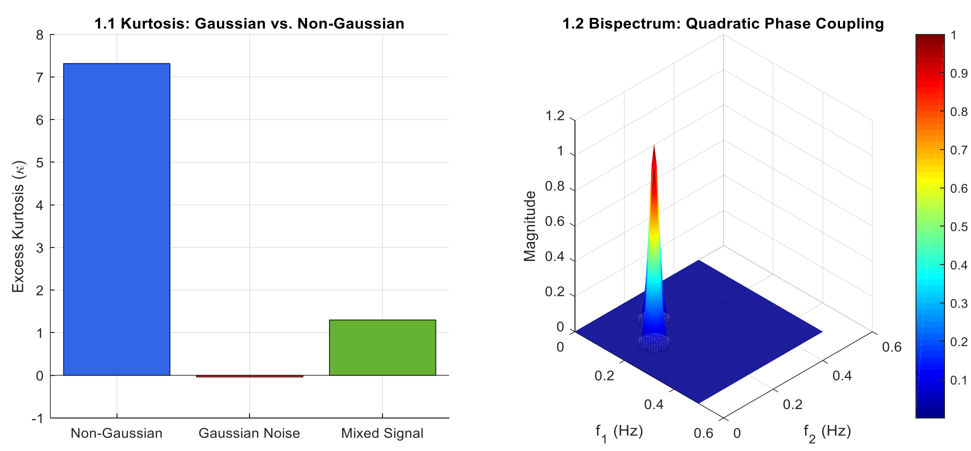
- **Suppress Gaussian Noise:** Because $k_k = 0$ for $k \geq \mathbf{3}$ in WGN, cumulant-based processing automatically filters out additive Gaussian noise, even if the noise is colored (i.e., not white).
- **Blind Deconvolution/Equalization:** HOS can estimate the phase of a transfer function (which second-order methods cannot do) and therefore restore a non-minimum phase signal from its output measurements alone.

4.8. Non-Linear Treatments

HOS provides a direct method for detecting and characterizing non-linear interactions within a system.

- **Non-Linear System Identification:** If a system is linear, the phase of its polyspectra should be zero. A non-zero polyspectrum phase indicates non-linear signal generation (e.g., a process generated by $y(t) = a\,x^2(t)$).
- **Coupling Detection:** The non-zero magnitude of the Bispectrum at specific frequency triplets $(\omega_1,\ \omega_2,\ \omega_1 + \omega_2)$ is proof of quadratic non-linear coupling within the underlying system.

The use of Higher-Order Statistics fundamentally changes the approach to signal analysis, moving from the simplistic assumptions of linearity and Gaussianity to a richer, more descriptive model of complex real-world phenomena.



(a) (b)

Figure 4.6: (a) Higher-Order Statistic Analysis: Excess Kurtosis as a Measure of Non-Gaussianity, (b) Bispectral Magnitude Surface: Identification of Quadratic Phase Coupling at (f_1, f_2).

Figure (a): This bar chart compares the Excess Kurtosis (k) of three signal types. It demonstrates that while the Non-Gaussian signal has a high statistical signature, the Gaussian Noise remains near zero. This confirms that Higher-Order Statistics (HOS) can effectively suppress additive Gaussian noise, allowing for the detection of non-Gaussian signals even in high-noise environments.

Figure (b): This 2D/3D surface plot represents the Bispectrum Magnitude, which identifies phase relationships between frequency components. The distinct peaks at the coordinates (f_1, f_2) provide mathematical proof of Quadratic Phase Coupling (non-linear interaction), a feature that remains hidden in a standard 2nd-order Power Spectral Density (PSD) plot.

4.9. Introduction to Particle Filtering (Sequential Monte Carlo Methods)

Definition: Particle Filtering (PF), also known as Sequential Monte Carlo (SMC) methods, is a powerful numerical technique used for solving the Bayesian filtering problem in complex dynamic systems. Fundamentally, a simulation-based approach approximates the system's state using a large number of random samples, or particles.

Particle filtering is particularly crucial when dealing with systems that violate the strict assumptions required by the classic Kalman Filter.

4.9.1. The Core Problem: Bayesian Filtering

The goal of state estimation (filtering) is to recursively determine the posterior probability density function (PDF) of the system's unobserved state (x_k) given all available noisy observations up to the current time ($z_{1:k}$):

$$\text{Filtering PDF} : P(x_k |\, z_{1:k}).$$

This posterior PDF represents all the knowledge we have about the system's state.

The filtering process involves two recursive steps derived from Bayes' theorem:

1. **Prediction (Time Update):** Propagating the previous posterior PDF $P(x_{k-1}|\, z_{1:k-1})$ forward in time using the system dynamics to get the prior PDF $P(x_k|\, z_{1:k-1})$.
2. **Update (Measurement Update):** Correcting the prior PDF $P(x_{k-1}|\, z_{1:k-1})$ using the new measurement z_k and the measurement model (likelihood) to get the final posterior PDF $P(x_k|\, z_{1:k})$.

The challenge is that for most non-linear or non-Gaussian systems, these probability distributions (and the integrals required to compute them) do not have a simple analytical solution.

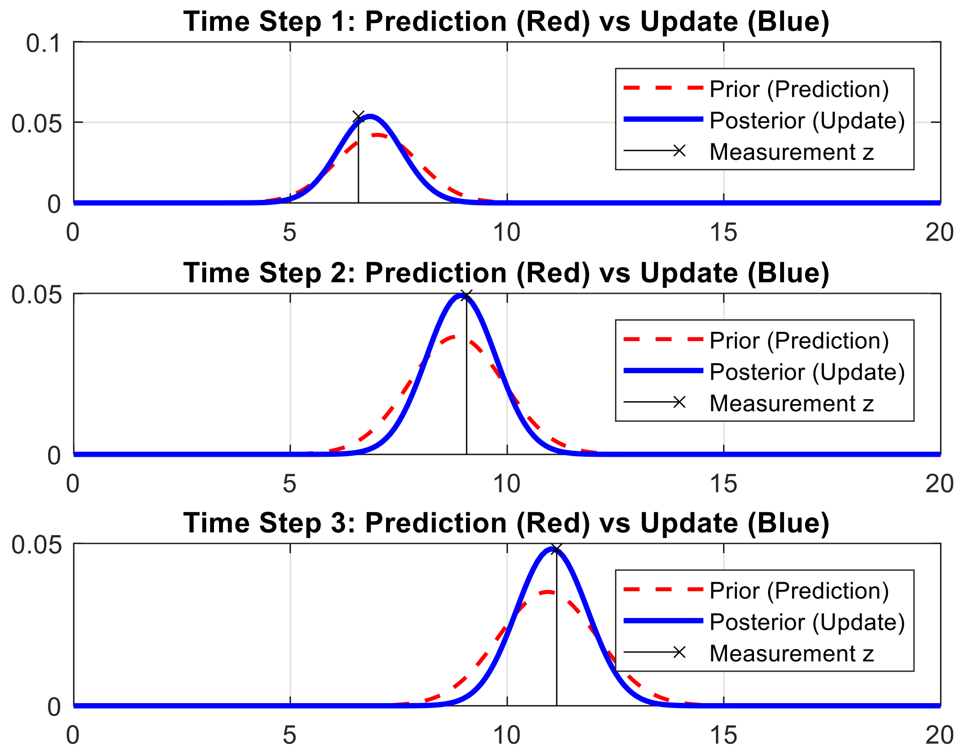


Figure 4.7: Recursive Bayesian Estimation: Prediction vs. Update.

This figure illustrates the recursive cycle of the **Bayesian Filter** in a 1D tracking scenario. It visualizes how the probability density function (PDF) evolves as the system incorporates both motion models and sensor measurements.

- **Prior PDF (Red Dashed Line):** Represents the **Prediction (Time Update)** step. It shows where the system expects the object to be after moving, according to the physical model. Note that the curve is "flatter" and "wider" than the previous step, representing the increase in uncertainty due to process noise.
- **Posterior PDF (Blue Solid Line):** Represents the **Update (Measurement Update)** step. This is the final result of the "Core Problem." By multiplying the

Prior with the sensor's Likelihood, the distribution "sharpens" around the most likely position.

- **Measurement (z) (Black 'X'):** Indicates the actual noisy data received from the sensor at that specific time step.

4.9.2. The Particle Filter Solution: Monte Carlo Approximation

Instead of trying to calculate the complex analytical PDF, the Particle Filter approximates it using a set of *N* weighted random samples (particles):

$$p(x_k|z_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(x_k - x_k^{(i)})$$

Where:

- $x_k^{(i)}$ is the state of the *i-th* particle (a hypothetical sample of the true state).
- $\omega_k^{(i)}$ is the weight of the *i-th* particle (a measure of how likely that particle is to represent the true state, given the measurements).
- $\delta(.)$ is the Dirac delta function.

As the number of particles N ncreases, this approximation converges to the true posterior PDF.

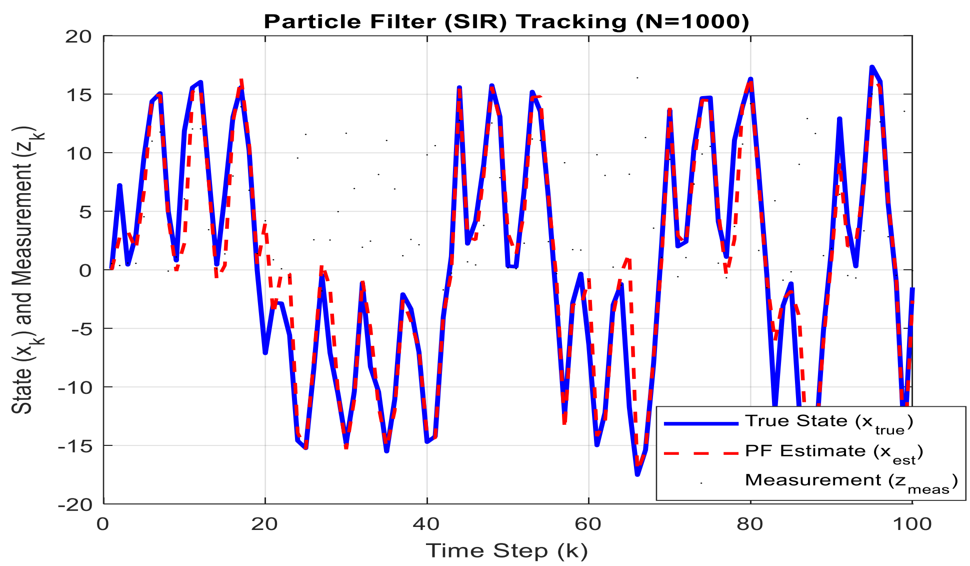


Figure 4.8: Particle Filter approximation.

EXERCISES

Exercise 4.1: You are measuring the temperature of a room every hour for a year.

1. If the average temperature in July is 25C but the average in January is 5C, is the process Wide-Sense Stationary (WSS)? Why or why not?
2. If you take the average temperature of 100 different rooms at exactly 12:00 PM (Ensemble Average) and it equals the average temperature of a single room measured over 24 hours (Time Average), what property does this process satisfy?

Exercise 4.2: Consider the process $X(t) = A\,cos(\,\omega t + \emptyset)$ where A and ω are constants and $\emptyset$ is a random variable uniformly distributed on $[0, 2\emptyset]$.

1. Calculate the Mean E[X(t)].
2. Calculate the Autocorrelation $R_{XX}(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{\tau})$.
3. Prove that this process is WSS.

Exercise 4.3: Ergodicity vs. Stationarity

A random process is defined as X(t) = A, where A is a random variable that takes the value +1 with probability 0.5 and -1 with probability 0.5. Once A is chosen for a specific realization, it remains constant for all t.

1. **Calculate the Ensemble Mean:** E[X(t)].
2. **Calculate the Time Mean:**$\langle \boldsymbol{X(t)} = \lim_{T\to\infty} \int_{-T}^{+T} \boldsymbol{X(t)dt} \rangle$.
3. **Conclusion:** Is this process WSS? Is this process **Ergodic**? Explain the physical meaning of your answer.

Exercise 4.4: A stochastic process X[n] is a First-Order Markov Chain if:

$$P(X[n]|X[n-1], X[n-2], \ldots, X[0]) = P(X[n]|X[n-1])$$

1. In plain English, what does this equation imply about the "memory" of the system?
2. If a system has two states (On/Off) and the probability of staying "On" is 0.9 while the probability of staying "Off" is 0.8, write the Transition Probability Matrix P.
3. How would you theoretically find the "Long-term" probability of the system being "On"?

4.10. Conclusion

This chapter has established the formal basis for analyzing stochastic processes, extending signal theory from deterministic functions to random functions of time or space. By adopting a probabilistic viewpoint, we have developed the tools necessary to characterize systems where signals are not entirely predictable yet possess an underlying statistical structure.

We have moved from the foundational definitions of Stationarity and Ergodicity, which bridge the gap between theory and measurement to the study of canonical models such as Poisson, Gaussian, and Markovian processes. Furthermore, we addressed the limitations of traditional second-order analysis by introducing Higher-Order Statistics (HOS) for non-linear systems and Particle Filtering for sequential state estimation.

They remain vital as we proceed to design optimal filters and evaluate the performance of modern engineering systems operating under uncertainty.

Bibliographic References

[1] A. V. Oppenheim and A. S. Willsky, "Signals and Systems," Second Edition, Prentice Hall, 1996.

[2] A. Papoulis and S. U. Pillai, "Probability, Random Variables, and Stochastic Processes," Fourth Edition, McGraw-Hill, 2002.

[3] P. Stoica and R. L. Moses, "Spectral Analysis of Signals," Prentice Hall, 2005.

[4] M. H. Hayes, "Statistical Digital Signal Processing and Modeling," John Wiley & Sons, 1996.

[5] C. L. Nikias and A. P. Petropulu, "Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework," Prentice Hall, 1993.

[6] B. Ristic, S. Arulampalam, and N. Gordon, "Beyond the Kalman Filter: Particle Filters for Tracking Applications," Artech House, 2004.

[7] J. G. Proakis and D. G. Manolakis, "Digital Signal Processing: Principles, Algorithms, and Applications," Fourth Edition, Pearson, 2007.

[8] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," IEEE Transactions on Signal Processing, vol. 50, no. 2, pp. 174-188, 2002.

[9] S. Haykin, "Adaptive Filter Theory," Fifth Edition, Pearson Education, 2013.

[10] H. L. Van Trees, "Detection, Estimation, and Modulation Theory, Part I," John Wiley & Sons, 2001.

[11] M. Kunt, "Traitement numérique des signaux," Volume 20, PPUR presses polytechniques, 1980.

[12] G. Blanchet and M. Charbit, "Digital Signal and Image Processing using MATLAB," ISTE Ltd, 2006.

[13] B. Picinbono, "Random Signals and Systems," Prentice Hall, 1993.

[14] P. Brémaud, "Introduction aux Probabilités: Modélisation des Phénomènes Aléatoires," Springer-Verlag, 1984.

[15] D. J. Daley and D. Vere-Jones, "An Introduction to the Theory of Point Processes," Springer, 2003.

[16] A. Doucet, N. de Freitas, and N. Gordon, "Sequential Monte Carlo Methods in Practice," Springer-Verlag, 2001.

[17] F. Auger, "Traitement du signal et statistiques," Cours et exercices corrigés, Edition Ellipses, 2013.

Web reference

[18] H. Pishro-Nik, "Introduction to Probability, Statistics, and Random Processes: Stationary Processes," Available at: probabilitycourse.com, Chapter 10.1.4. [https://www.probabilitycourse.com/chapter10/10_1_4_stationary_processes.php#:~:text=We%20can%20classify%20random%20processes,do%20not%20change%20by%20time.]

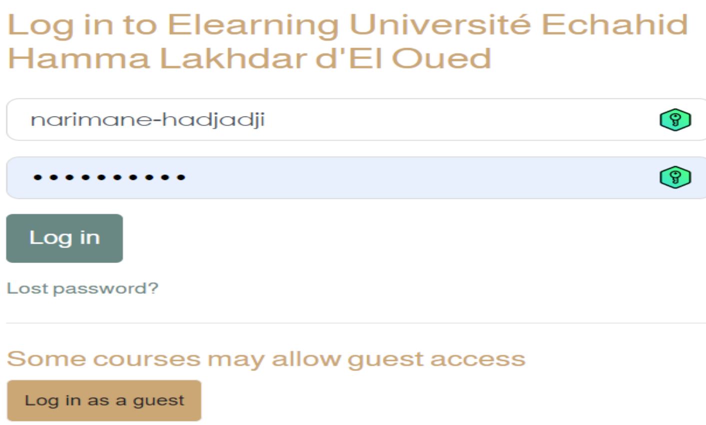
[19] AskFilo Educational Resources, "Explanation of Strict-Sense Stationary (SSS) Processes," [https://askfilo.com/user-question-answers-smart-solutions/a-explain-in-brief-strict-sense-stationary-sss-process-and-3336313730363630].

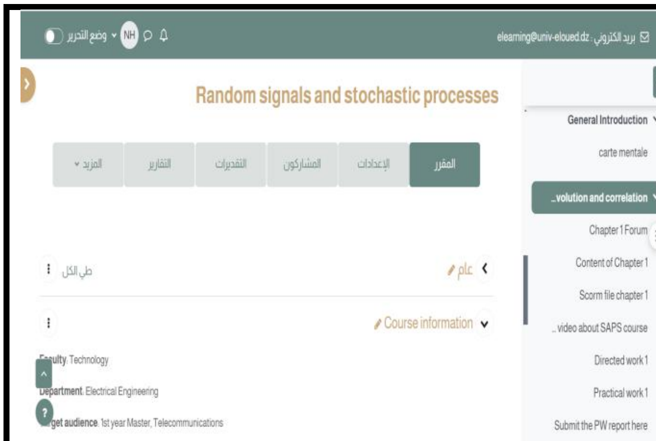
[20] iStock Illustrations, "Acoustic Wave Patterns and Signal Visualization," Vector Graphics Repository, 2024.[https://www.istockphoto.com/vector/acoustic-wave-gm1496878330-519415849]

[21] Grainger Operations, "Earthquake Preparedness and Seismic Monitoring Equipment," [https://www.grainger.com/know-how/operations/preparedness/kh-earthquake-supplies].

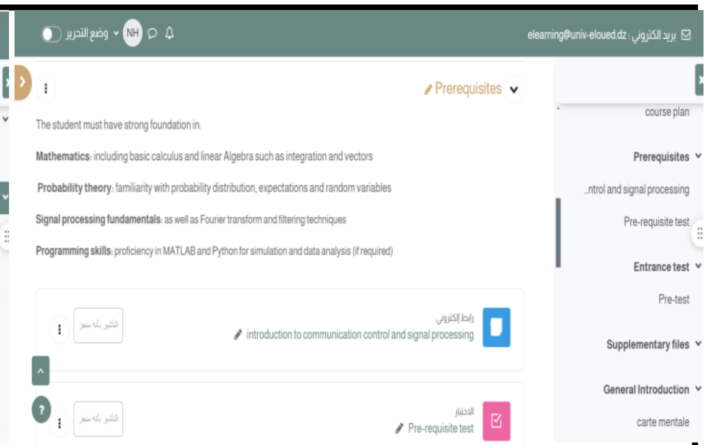
ANNEXES

- Uploading the courses on the platform "E-learning" of Echahid Hama Lakhdar University of El Oued;
Uploading: https://e-learning.univ-eloued.dz/course/view.php?id=1710

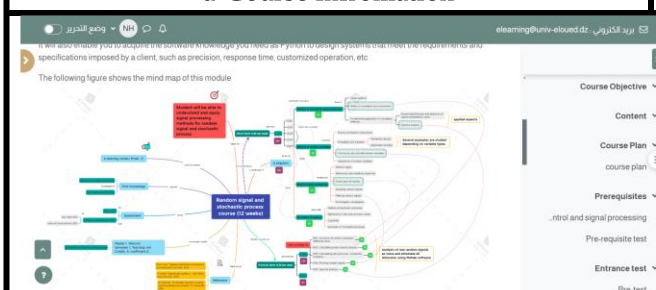




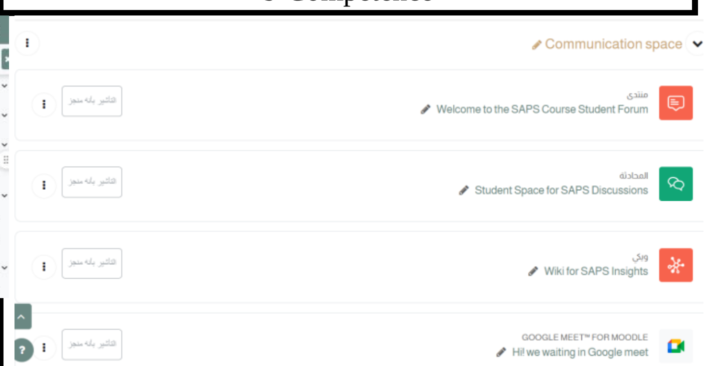
a-Course infromation



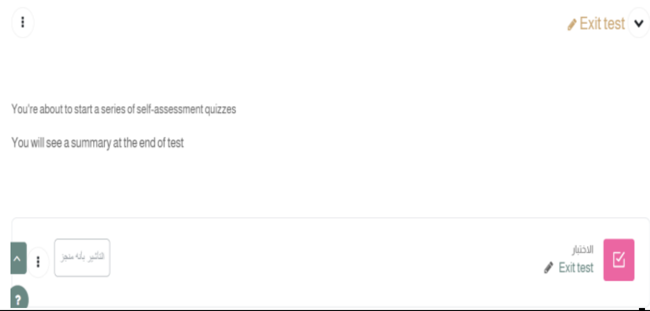
b-Competence



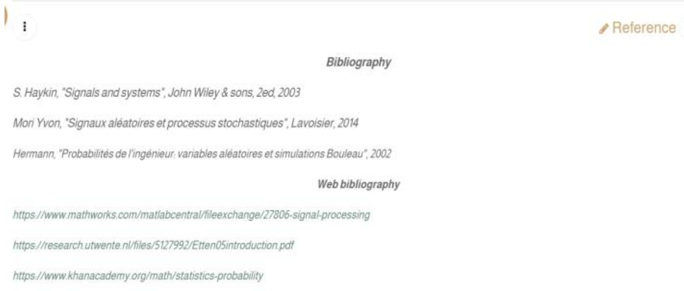
c- Mind carte



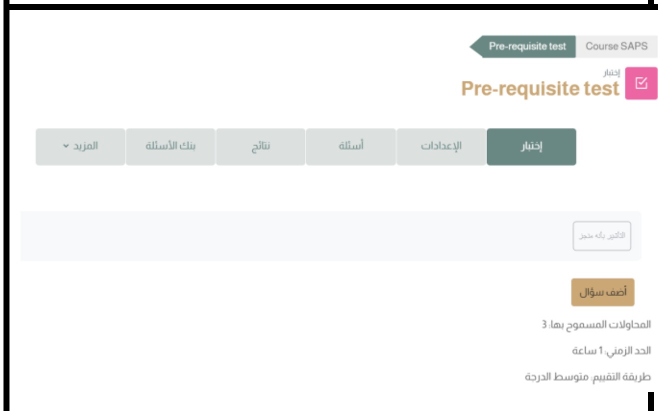
d- communication space



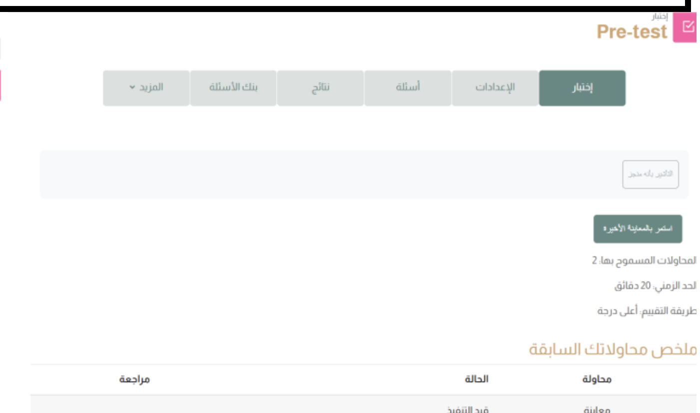
e- Exit Test



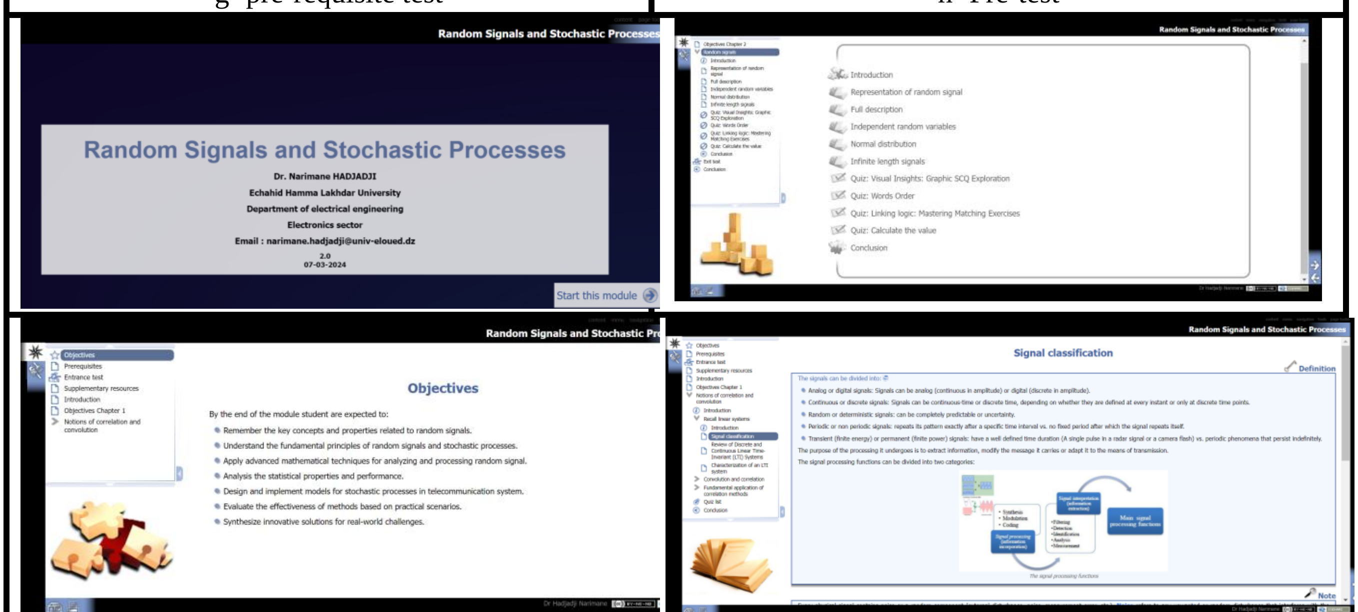
f-Reference



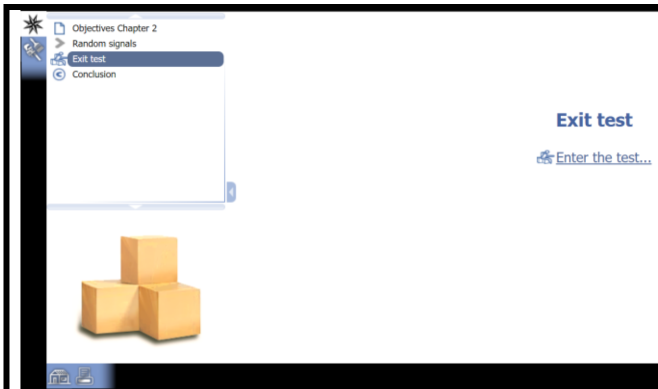
g- pre-requisite test

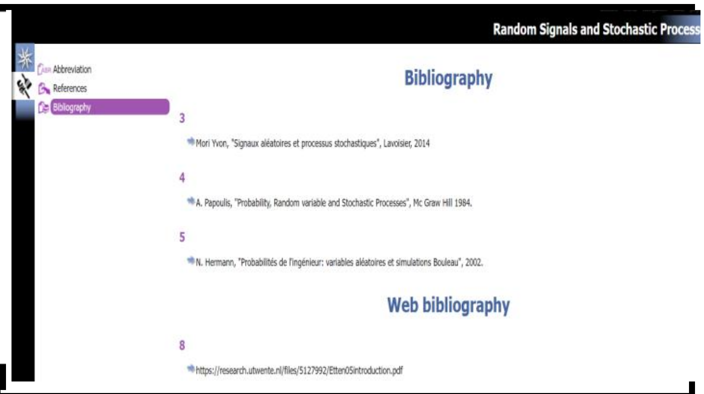


h- Pre-test



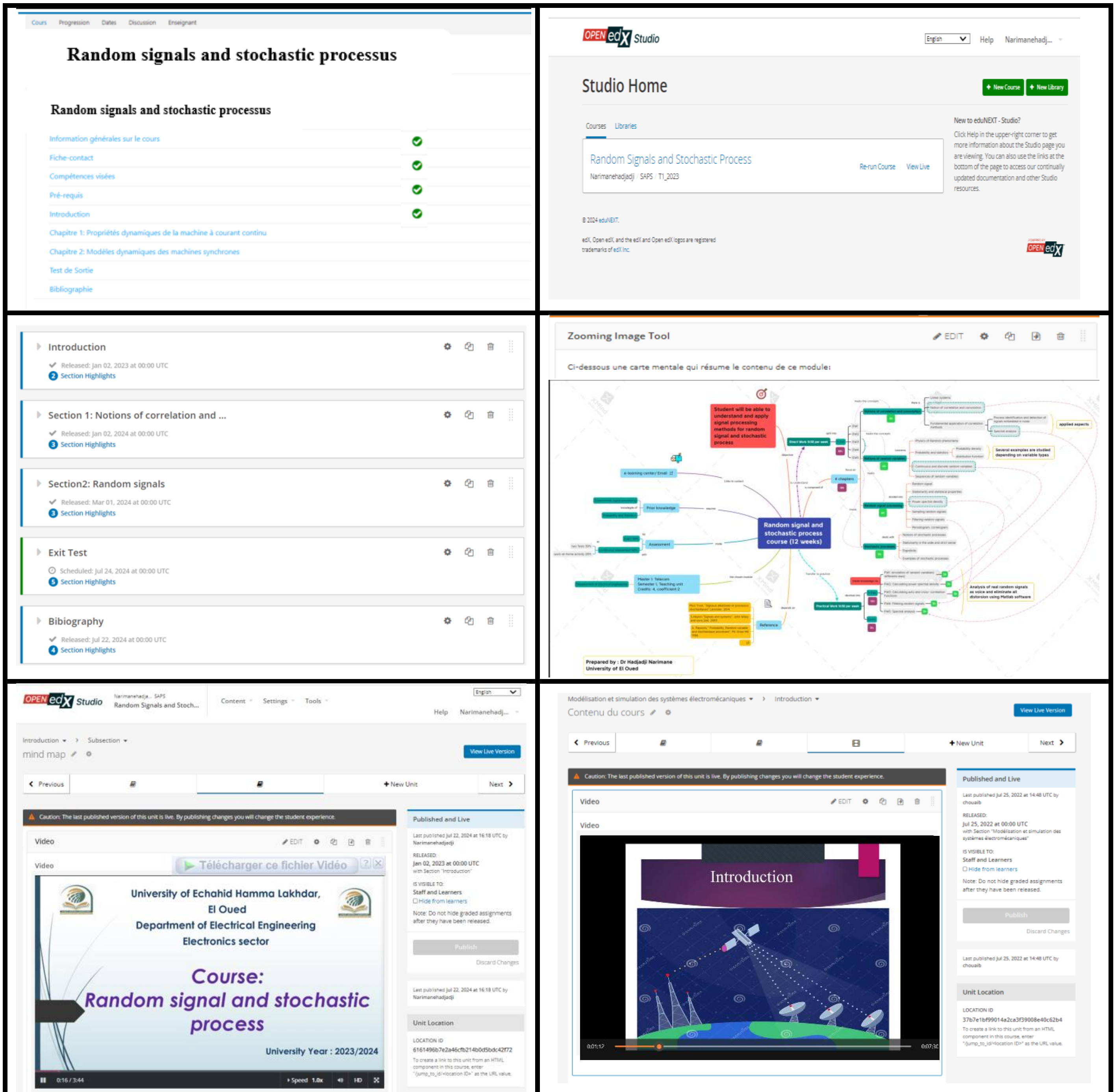
i- SCORM Formats in Moodle.





j- Exit Test	j- References
Online course deployment on the Echahid Hama Lakhdar University of El-Oued platform.	

- Educational Support Materials – To facilitate deeper understanding of the lessons, students are invited to consult my Edx space, which provides pedagogical videos with comprehensive audio-visual explanations of the course content.



Pedagogical course formats within the LMS on the EdX platform.

Abstract

This module provides a comprehensive framework for the analysis, modeling, and processing of both deterministic and stochastic signals within linear and non-linear systems. Beginning with the fundamental principles of Linear Time-Invariant (LTI) systems, the course explores the essential roles of convolution and correlation in signal identification and noise detection. It transitions into the statistical treatment of Random Variables, establishing rigorous definitions for convergence and probability distributions.

A significant portion of the module is dedicated to the processing of Random Signals, where students master Power Spectral Density (PSD) estimation through both non-parametric techniques (Periodogram, Welch) and parametric modeling (AR, MA, and ARMA). The curriculum extends into advanced Stochastic Processes, covering stationarity, ergodicity, and the unique advantages of Higher-Order Statistics (HOS) for characterizing non-Gaussian and non-linear phenomena. Finally, the module introduces modern state-space estimation via Bayesian and Particle Filtering, providing the tools necessary for tracking dynamic systems.

Keywords: LTI Systems, Stochastic Processes, Spectral Estimation, Probability, Higher-Order Statistics, Bayesian Filtering, Particle Filters, Signal Processing.

Résumé

Ce module fournit un cadre complet pour l'analyse, la modélisation et le traitement des signaux déterministes et stochastiques au sein de systèmes linéaires et non linéaires. Commençant par les principes fondamentaux des systèmes Linéaires Invariants dans le Temps (LTI), le cours explore les rôles essentiels de la convolution et de la corrélation dans l'identification des signaux et la détection du bruit. Il passe au traitement statistique des variables aléatoires, établissant des définitions rigoureuses pour la convergence et les distributions de probabilité.

Une part importante du module est consacrée au traitement des Signaux Aléatoires, où les étudiants maîtrisent l'estimation de la Densité Spectrale de Puissance (PSD) à travers des techniques non paramétriques (Périodogramme, Welch) et la modélisation paramétrique (AR, MA et ARMA).

Le programme s'étend aux processus stochastiques avancés, couvrant la stationnarité, l'ergodicité et les avantages uniques des statistiques d'ordre supérieur (HOS) pour caractériser les phénomènes non gaussiens et non linéaires. Enfin, le module introduit l'estimation moderne de l'état via le filtrage bayésien et les filtres particulaires, fournissant les outils nécessaires pour suivre les systèmes dynamiques.

Mots-clés : Systèmes LTI, Processus Stochastiques, Estimation Spectrale, Probabilité, Statistiques d'Ordre Supérieur, Filtrage Bayésien, Filtres de Particules, Traitement du Signal.

"It is my sincere hope that this module serves as a clear and valuable resource for your academic journey. In the spirit of continuous improvement, I welcome all comments, suggestions, or corrections from students and colleagues alike to enhance the quality of this work."