

# **Chapter 1**

## **Stochastic Processes**

Compiled and adapted by Jesús Cid-Sueiro, Feb. 2023.



# Contents

<b>1</b>	<b>Stochastic Processes</b>	<b>1</b>
1.1	Introduction	2
1.1.1	Characterization of a stochastic process	3
1.2	Independent processes	3
1.3	Random walks	4
1.3.1	Properties of random walks	6
1.3.2	Stopping times	6
1.4	Markov Chains	8
1.4.1	Transition graph	9
1.4.2	State probabilities	10
1.4.3	Stationary distribution	10
1.4.4	Computing the stationary distribution	11
1.5	Stationary Processes	13
1.5.1	Strict-Sense Stationarity	13
1.5.2	First and Second-Order Statistics	14
1.5.3	Wide-Sense Stationarity	16
1.5.4	Linear systems with stochastic inputs	18
1.5.5	Gaussian processes	20
1.5.6	Power Spectral Density	21
1.5.7	Ergodicity	23
	References	26

***Note:** Secs 1.1-1.4 in this document are an adaptation of material taken from [1] (Creative Commons license BY-NC-SA 4.0). Sec. 1.5 is a translation, summary and adaptation of a chapter in [2] made by the author with permission of the co-authors. Secs. 1.1-1.4 is published under license BY-NC-SA 4.0. Sec. 1.5 cannot be published or redistributed in any form. It cannot be transformed or used partially without permission of the original authors*

This document introduces the concept of stochastic process and analyzes some important examples.

## 1.1 Introduction

A stochastic process (SP) is a collection of random variables  $\{X_n\}$  indexed by time,  $n$ . Depending on the nature of the time variable, we will consider different types of SP:

- **Discrete-time** SP: when the time variable is a real number ( $n \in \mathbb{R}$ )
- **Continuous-time** SP: when the time variable is integer ( $n \in \mathbb{Z}$ ).

This chapter discusses discrete-time processes only. Thus, under contrary stated, all SP will be assumed to be discrete.

Depending on the range of the time variable, two types of SP will be considered:

- **One-side** SP:  $\{X_n, n \geq 0\}$ . Note that a one-side SP is a sequence of random variables.
- **Two-side** SP:  $\{X_n, n \in \mathbb{Z}\}$ , the SP is defined for all integer times.

*Example 1* Consider the one-side stochastic process  $X_n$  such that, for all  $n$ ,  $X_n \in \{-n, n\}$  with  $P\{X_n = n\} = P\{X_n = -n\}$  and all variables in the collection are mutually independent.

A realization of all random variables in an SP defines a signal. For this reason, we can define stochastic processes in the following alternative but equivalent form:

**Definition 1 (Stochastic Process)** A discrete-time stochastic process is a probability measure over a space of (one-side or two-side) discrete-time signals.

*Example 2* Consider the one-side stochastic process  $X_n$  such that

$$X_n = n, \quad \text{for all } n \geq 0 \quad (1.1)$$

or

$$X_n = -n, \quad \text{for all } n \geq 0 \quad (1.2)$$

with equal probabilities. Note that, the Universe set of this SP is finite, containing only two signals.

### 1.1.1 Characterization of a stochastic process

Note that the random variables defining a SP can be continuous or discrete. In general, a SP is completely characterized when we can determine the joint distribution  $P(X_{n_0}, X_{n_1}, \dots, X_{n_{M-1}})$  for any subset of random variables from the process. However, for many SP, the computation of these distributions is unfeasible. We will pay attention to processes satisfying different properties that simplify the characterization.

## 1.2 Independent processes

**Definition 2 (Independent process)** A stochastic process  $X_n$  is statistically independent (or, simply, independent) if all its random variables are statistically independent, that is, for any  $M > 0$  and any subset of  $M$  random variables  $X_{n_0}, X_{n_1}, \dots, X_{n_{M-1}}$  with  $n_0 < n_1, \dots < n_{M-1}$  we have

$$P(X_{n_0}, X_{n_1}, \dots, X_{n_{M-1}}) = \prod_{k=0}^{M-1} P(X_{n_k}) \quad (1.3)$$

**Note:** A quick explanation on the mathematical notation. Eq. (1.3) should be interpreted as: the joint distribution of the set of  $M$  random variables  $X_{n_0}, X_{n_1}, \dots, X_{n_{M-1}}$  is the product of the marginal distribution of each variable. Thus, for instance, the cumulative distributions gets factorized:

$$F_{X_{n_0}, X_{n_1}, \dots, X_{n_{M-1}}}(x_{n_0}, x_{n_1}, \dots, x_{n_{M-1}}) = \prod_{k=0}^{M-1} F_{X_{n_k}}(x_{n_k}) \quad (1.4)$$

for any  $x_{n_0}, x_{n_1}, \dots, x_{n_{M-1}}$  in the respective domains of the variables. If all the random variables in the process are continuous, Eq. (1.4) is equivalent to the factorization of the density functions

$$P_{X_{n_0}, X_{n_1}, \dots, X_{n_{M-1}}}(x_{n_0}, x_{n_1}, \dots, x_{n_{M-1}}) = \prod_{k=0}^{M-1} P_{X_{n_k}}(x_{n_k}) \quad (1.5)$$

Also, if the random variables are discrete, this is equivalent to the factorization of the probability mass functions

$$P_{X_{n_0}, X_{n_1}, \dots, X_{n_{M-1}}}(x_{n_0}, x_{n_1}, \dots, x_{n_{M-1}}) = \prod_{k=0}^{M-1} P_{X_{n_k}}(x_{n_k}) \quad (1.6)$$

In general, we will use expressions like (1.3) instead of (1.4), (1.5), or (1.6) to state some relations between distributions, because it is simpler and encompasses all of them.

A particularly interesting case arises when all variables in the SP follow the same distribution.

### Definition 3 (IID process)

A stochastic process  $X_n$  is independent and identically distributed, or IID, if it is independent and all variables in the process follow the same distribution, i.e.  $P(X_n) = P(X_m)$

(which implies, for instance, that  $F_{X_n}(x) = F_{X_m}(x)$ , for all  $x$  and any  $n, m$  in the time domain of the process).

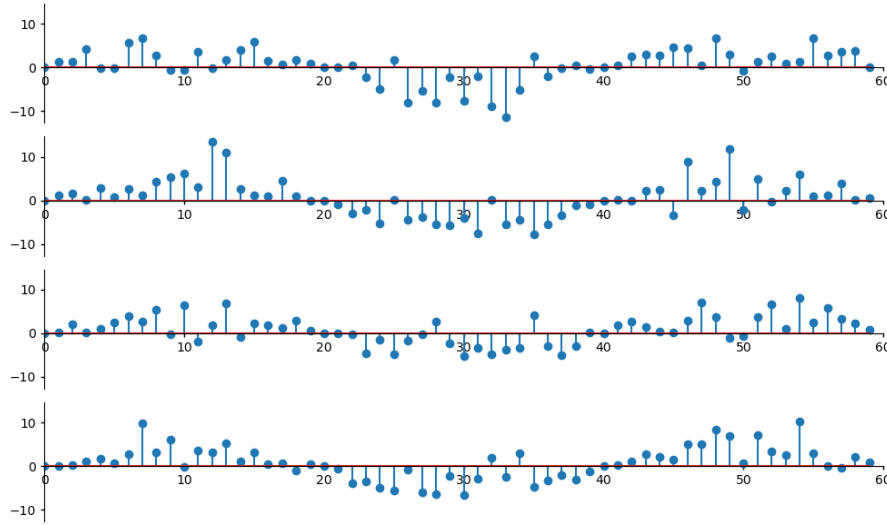
One of the simplest IID processes is the Bernoulli process.

*Example 3* A Bernoulli process  $X_n$  with parameter  $p$ , denoted as  $\mathcal{B}(p)$  is an IID process given by Bernoulli random variables:

$$P_{X_n}(x) = p^x(1-p)^{1-x}, \quad x \in \{0, 1\} \quad (1.7)$$

(Note that Eq. (1.7) is equivalent to claim that  $P\{X_n = 1\} = p$  and  $P\{X_n = 0\} = 1 - p$ ). Bernoulli and other IID processes are commonly used as building blocks of other, more complex, processes.

*Example 4* Figure 1.1 shows 4 realizations of the independent stochastic process  $X_n$ , where each sample of the process follows a Gaussian distribution with mean and standard deviation equal to  $4 \sin(0.05\pi n)$ .



**Fig. 1.1** 4 realizations of process  $X_n$  described in example 4

### 1.3 Random walks

**Definition 4 (Simple random walk (SRW))** Let  $Y_n \in \{-1, 1\}$  be a one-sided IID process with  $P\{Y_n = 1\} = \frac{1}{2}$ . A simple random walk is defined by  $X_0 = 0$  and

$$X_n = \sum_{k=1}^n Y_k, \quad n > 0 \quad (1.8)$$

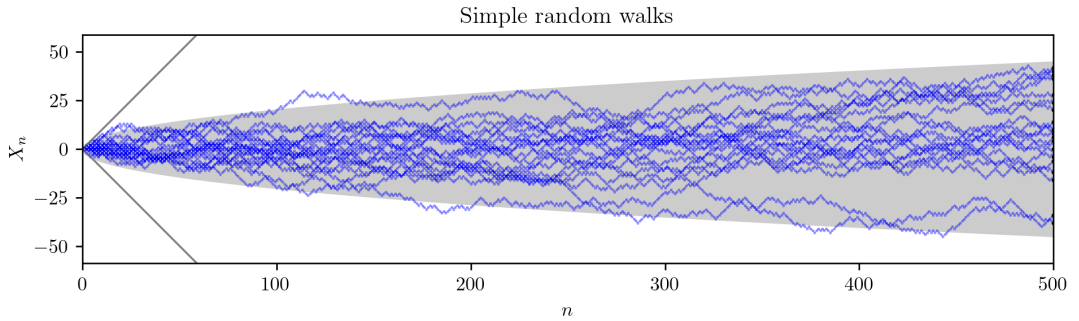
Note that we can express the random variables in a SRW recursively as

$$X_n = X_{n-1} + Y_n, \quad n > 0 \quad (1.9)$$

Thus, a SRW is not an independent process. It is not an identically distributed process, either, since the range of the random variables in the process grows with  $n$ .

$$-n \leq X_n \leq n \quad (1.10)$$

Fig. 1.2 shows several realizations of the SRW. Note that the realizations are quite far from the bounds. This is because the probability that  $X_n$  is close to  $n$  or  $-n$  vanishes for large  $n$ .



**Fig. 1.2** 20 realizations of a simple random walk. The straight lines represent the boundaries of the region containing all possible trajectories. The shaded area covers two standard deviations: all trajectories tend to be inside the gray area around 95% of the time.

The mean of the process is given by

$$\mathbb{E}\{X_n\} = \sum_{k=0}^n \mathbb{E}\{Y_k\} = 0 \quad (1.11)$$

and, since  $Y_n$  is IID, the variance is given by

$$\text{var}\{X_n\} = \sum_{k=1}^n \text{var}\{Y_k\} = n \quad (1.12)$$

Therefore, the standard deviation is  $\sqrt{n}$  and we can expect from the process to lie between the limits  $-\sqrt{n}$  and  $\sqrt{n}$  most of the time. This can be expressed a bit more precisely, by noting that, since  $X_n$  is the sum of  $n$  independent random variable with zero-mean and bounded variance, and as a consequence of the central limit theorem,  $X_n$  converges in distribution to a zero-mean Gaussian random variable with variance  $n$ . This implies that, after large  $n$ , we can expect that  $|X_n| < 2\sqrt{n}$  approximately 95% of time.

### 1.3.1 Properties of random walks

- **Independent increment** For all  $0 \leq n_0 \leq n_1 \leq n_2 \leq n_3$ , the random variables  $X_{n_1} - X_{n_0}$  and  $X_{n_3} - X_{n_2}$  are mutually independent
- **Time invariance** For all  $h \geq 1$ , the SP  $Z_n = X_{n+h} - X_h$  is a simple random walk.

### 1.3.2 Stopping times

The SRW can be used as a model of the total winning and losses of a fair game. Assume that we play a game based on repeated trials of flipping a coin. Every time the result is 'heads', we win 1 point while we loose 1 point otherwise. By taking  $Y_n = 1$  if the  $n$ -th toss is 'heads', process  $X_n$  accounts for the total score after  $n$  trials.

Since the SRW is zero mean, we cannot expect a positive gain after an arbitrary number of steps. One may wonder if there is a way to ensure a positive score by choosing the appropriate stopping time. Let's assume we set an upper and lower limits,  $U > 0$  and  $-L < 0$ , respectively, such that, when  $X_n = U$  or  $X_n = -L$ , we stop betting. By choosing  $L$ , we set the maximum loss we can tolerate, by choosing  $U$ , we set the gains that will make us stop playing.

We can call stopping time,  $N$  to the first time where  $X_N$  reaches  $U$  or  $-L$ . Note that  $N$  is a random variable, because it is a function of the random variables in the SP:

$$N = \min\{n \mid X_n = U \text{ or } X_n = -L\} \quad (1.13)$$

Is there a way to select  $U$  and  $L$  in order to expect a positive gain? To answer this question, let us compute the probability of stopping at level  $U$ , that is  $P\{X_N = U \mid X_0 = 0\}$ . Thus, defining function

$$f(i) = P\{X_N = U \mid X_0 = i\}$$

our goal is to compute  $f(0)$ .

Applying the total probability theorem,

$$\begin{aligned} f(i) &= P\{X_N = U \mid X_0 = i\} \\ &= P\{X_1 = i+1 \mid X_0 = i\}P\{X_N = U \mid X_0 = i, X_1 = i+1\} \\ &\quad + P\{X_1 = i-1 \mid X_0 = i\}P\{X_N = U \mid X_0 = i, X_1 = i-1\} \\ &= \frac{1}{2}P\{X_N = U \mid X_1 = i+1\} + \frac{1}{2}P\{X_N = U \mid X_1 = i-1\} \\ &= \frac{1}{2}f(i+1) + \frac{1}{2}f(i-1) \end{aligned} \quad (1.14)$$

Noting, also, that

$$f(U) = 1,$$

$$f(-L) = 0,$$

if we let  $f(-L+1) = \alpha$  then it follows that  $f(1-L+r) = \alpha r$  for all  $r \leq U+L$ . Therefore,  $\alpha = U+L$ , and it follows that



$$f(0) = P\{X_N = U | X_0 = 0\} = \frac{L}{U+L}$$

Thus, the gambler can increase the probability of scoring  $U$  points by increasing  $L$ . However, this implies a higher risk to suffer large temporal losses. The expected gain becomes:

$$\mathbb{E}\{X_N\} = Uf(0) - L(1 - f(0)) = 0$$

Thus, there is no way to set the bounds  $U$  and  $L$  in order to expect a positive gain.

## 1.4 Markov Chains

One important property of the simple random walk is that the effect of the past on the future is summarized only by the current state, rather than the whole history. In other words, the distribution of  $X_{n+1}$  depends only on the value of  $X_n$  and not on the whole set of values of  $X_0, X_1, \dots, X_n$ . A stochastic process with such property is called a **Markov chain**.

**Definition 5 (Markov property)** Let  $X_n$  be a discrete-time SP that, for each  $n$ , takes value in some discrete set  $\mathcal{S}$ , which we call the **state space**. We say that the SP has the Markov property if

$$P\{X_{n+1} = i | X_n, X_{n-1}, \dots, X_0\} = P\{X_{n+1} = i | X_n\} \quad (1.15)$$

for all  $n$  in the time-domain of the process and for all  $i \in \mathcal{S}$ .

**Definition 6 (Markov chain)**

Let  $X_n$  be a discrete-time SP. We say that the SP is a Markov chain (MC) if it satisfies the Markov property.

We will say that an MC is **finite** if  $\mathcal{S}$  is a finite set. In this case, we usually take the elements of  $\mathcal{S}$  as consecutive integers. For instance, for a state space with  $m$  elements, we will take  $\mathcal{S} = \{0, 1, \dots, m-1\}$  or  $\mathcal{S} = \{1, 2, \dots, m\}$  under our convenience.

Note that the SRW is an MC, but it is not finite.

Any finite MC can be described in terms of the transition probabilities

$$p_{ij} = P\{X_{n+1} = j | X_n = i\} \quad (1.16)$$

All the elements of a MC model can be encoded in a transition probability matrix

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} & \dots & p_{0,m-1} \\ p_{10} & p_{11} & \dots & p_{1,m-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m-1,0} & p_{m-1,1} & \dots & p_{m-1,m-1} \end{pmatrix} \quad (1.17)$$

Using (1.16), it is easy to see that

$$\sum_{j \in \mathcal{S}} p_{ij} = 1 \quad (1.18)$$

thus, all rows in  $\mathbf{P}$  sum up to one. For this reason, we say that  $\mathbf{P}$  is a right-stochastic matrix. In matrix form, this property can be expressed as

$$\mathbf{P}\mathbf{1}_m = \mathbf{1}_m \quad (1.19)$$

where  $\mathbf{1}_m$  is an all-ones vector with dimension  $m$ .

Note that, in general, the transition probabilities  $p_{ij}$  (and, thus, the transition matrix) may change with time. However, the case of constant probabilities is particularly interesting.

**Definition 7 (Homogeneous Markov Chain)** A markov chain  $X_n$  is homogeneous if the transition probabilities do not depend on time, that is.

$$P\{X_{n+1} = j | X_n = i\} = P\{X_{m+1} = j | X_m = i\}, \quad i \in \mathcal{S}, \quad j \in \mathcal{S} \quad (1.20)$$

for any  $m$  and  $n$  in the time domain of the process.

*Example 5 (Two-state machine)* A machine can be either working or broken on a given day. If it is working, it will break down in the next day with probability 0.01, and will continue working with probability 0.99. If it breaks down on a given day, it will be repaired and be working in the next day with probability 0.8, and will continue to be broken down with probability 0.2. We can model this machine by a Markov chain with two states: working, and broken down. The transition probability matrix is given by

$$\begin{pmatrix} 0.99 & 0.8 \\ 0.01 & 0.2 \end{pmatrix}$$

*Example 6 (SRW)* A simple random walk is an example of an MC. However, there is no transition probability matrix associated with the simple random walk since the sample space is infinite.

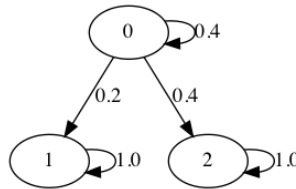
### 1.4.1 Transition graph

The transition probability matrix of a given homogeneous MC can be represented by a directed graph, whose nodes are the states, and edges from node  $i$  to node  $j$  represents a transition from state  $i$  to state  $j$  with nonzero probability. The weight of the edge is the transition probability.

*Example 7* The transition graph of an MC given by state space  $\mathcal{S} = \{0, 1, 2\}$  and transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 0.4 & 0.2 & 0.4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.21)$$

is shown in Figure 1.3.



**Fig. 1.3** The transition graph for the markov chain given by (1.21)

### 1.4.2 State probabilities

The transition probability matrix is useful to compute any probabilities about the state of the process at any given time. For instance, the  $n$ -th step transition probabilities, defined as  $r_{ij}(n) = P\{X_n = j | X_0 = i\}$  can be computed recursively, by applying the total probability theorem, as

$$r_{ij}(n) = \sum_{k \in \mathcal{S}} P\{X_n = j | X_{n-1} = k, X_0 = i\} P\{X_{n-1} = k | X_0 = i\} = \sum_{k \in \mathcal{S}} r_{ik}(n-1) p_{kj}, \quad n > 1 \quad (1.22)$$

Note that  $r_{ij}(1) = p_{ij}$ .

Defining the  $n$ -step transition probability matrix  $\mathbf{R}_n$  as the matrix with components  $r_{ij}(n)$ , the recurrent relation (1.22) can be written in matrix form as

$$\mathbf{R}_n = \mathbf{R}_{n-1} \mathbf{P} \quad (1.23)$$

and, noting that  $\mathbf{R}_1 = \mathbf{P}$ , we get

$$\mathbf{R}_n = \mathbf{P}^n \quad (1.24)$$

Using these relations we can compute the state probabilities at any given time, given the initial state or the initial state probabilities.

For instance, let  $\mathbf{q}_n$  be the state probability vector at time  $n$ , that is

$$q_{n,i} = P\{X_n = i\}, \quad n \geq 0, \quad i \in \mathcal{S} \quad (1.25)$$

then we can express the state probability vector at time  $n$  as a function of the initial state probabilities using (1.24) as

$$\begin{aligned} q_{n,j} &= P\{X_n = j\} = \sum_{i \in \mathcal{S}} P\{X_n = j | X_0 = i\} P\{X_0 = i\} \\ &= \sum_{i \in \mathcal{S}} r_{ij}(n) P\{X_0 = i\} \end{aligned} \quad (1.26)$$

which, in vector form, can be computed as

$$\mathbf{q}_n = \mathbf{R}_n^\top \mathbf{q}_0 = (\mathbf{P}^\top)^n \mathbf{q}_0 \quad (1.27)$$

### 1.4.3 Stationary distribution

**Definition 8 (Stationary distribution)** A stationary distribution of an MC is a probability distribution over the state space  $\mathcal{S}$  (where  $P\{X_0 = j\} = \pi_j$ ) such that

$$\mathbf{P}^\top \boldsymbol{\pi} = \boldsymbol{\pi} \quad (1.28)$$

*Example 8* Let  $\mathcal{S} = \mathbb{Z}_m$  and  $X_0 = 0$ . Consider the MC  $X_n$  such that  $X_{n+1} = (X_n + 1) \bmod m$  with probability  $\frac{1}{2}$  and  $X_{n+1} = (X_n - 1) \bmod m$  with probability  $\frac{1}{2}$ .

The stationary distribution of this MC is  $\pi_i = \frac{1}{m}$ .

Note that the vector  $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_m)^\top$  is an eigenvector of  $\mathbf{P}^\top$  with eigenvalue 1. Hence the following theorem can be deduced from the Perron-Frobenius theorem.

**Theorem 1** *If  $p_{ij} > 0$  for all  $i, j \in \mathcal{S}$ , then there exists a unique stationary distribution of the system.*

*Moreover, if  $\mathbf{r}_i(n)$  is the state conditional probability vector with components  $r_{ij}(n)$ ,  $j \in \mathcal{S}$  (i.e., the values in the  $i$ -th row of  $\mathbf{R}_n$ ),*

$$\lim_{n \rightarrow \infty} \mathbf{r}_i(n) = \boldsymbol{\pi}, \quad \forall i, j \in \mathcal{S} \quad (1.29)$$

A corresponding theorem is not true if we consider infinite state spaces.

Finite Markov chains can be shown to have at least one stationary distribution. However, if at least one of the transition probabilities is zero, the behavior of the process may depend on the topology of the transition graph:

- The stationary distribution may be not unique.
- The limiting distributions  $\lim_{n \rightarrow \infty} \mathbf{r}_i(n)$  may depend on  $i$ .

*Example 9* The MC in example in 7 has at least two limiting distributions:  $\boldsymbol{\pi}_A = (0, 1, 0)^\top$  and  $\boldsymbol{\pi}_B = (0, 0, 1)^\top$ . Also, any convex combination of them  $\boldsymbol{\pi} = \alpha \boldsymbol{\pi}_A + (1 - \alpha) \boldsymbol{\pi}_B$ , with  $0 \leq \alpha \leq 1$  is a stationary distribution.

The limiting distributions depend on the initial state: is is easy to verify that

$$\lim_{n \rightarrow \infty} \mathbf{r}_1(n) = (0, 0, 1) \quad (1.30)$$

$$\lim_{n \rightarrow \infty} \mathbf{r}_2(n) = (0, 1, 0) \quad (1.31)$$

$$\lim_{n \rightarrow \infty} \mathbf{r}_3(n) = \left(0, \frac{1}{3}, \frac{2}{3}\right) \quad (1.32)$$

#### 1.4.4 Computing the stationary distribution

In general, the stationary distribution are all the solution of the set of equations (1.28) and the probabilistic constraint  $\mathbb{1}^\top \mathbf{p} = 1$ , that is

$$\begin{pmatrix} \mathbf{P}^\top - \mathbf{I} \\ \mathbb{1}^\top \end{pmatrix} \boldsymbol{\pi} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \quad (1.33)$$

The system of linear equations given by (1.33) is consistent (it has at least one solution) but overdetermined, because the left matrix has  $m + 1$  rows but only  $m$  columns. In fact, using (1.19), it is easy to see that

$$\mathbb{1}_m^\top (\mathbf{P}^\top - \mathbf{I}) = \mathbf{0} \quad (1.34)$$

so that the rank of  $\mathbf{P}^\top - \mathbf{I}$  is at most  $m - 1$ . Thus, we can remove one of the rows in the matrix (any one of them) to solve the system.

## 1.5 Stationary Processes

*[Note: This section is copyrighted. It cannot be published or redistributed in any form. It cannot be transformed or used partially without permission of the original authors]*

Stationarity refers to the time invariance of some statistical properties of a stochastic process. We have already seen some forms of time invariance in the previous sections:

- The homogeneity of Markov chain,  $X_n$ , which implies the time invariance of the conditional probabilities:  $P(X_n|X_m) = P(X_{n+k}|X_{m+k})$ , for any  $k \geq 0, n \geq m \geq 0$ .
- The time invariance of IID processes, which implies the invariance of the single-variable marginal distributions,  $P(X_{n+k}) = P(X_n)$  for any  $h > 0$
- The time invariance property of a SRW, which implies that  $P(X_{n+h} - X_h) = P(X_n)$ .

In this section we introduce two forms of stationarity which are particularly interesting in statistical signal processing applications. In its more restrictive form, stationarity implies the invariance of all joint distributions.

### 1.5.1 Strict-Sense Stationarity

**Definition 9 (Strict-Sense Stationarity)** A stochastic process is said to be Strict-Sense Stationary (or simply SSS) if all joint distributions are invariant to a time shift, that is, for any  $N > 0$ , any sampling times  $n_0, \dots, n_{N-1}$  (in the time domain of the process) and any time shift  $m > 0$ ,

$$P(X_{n_0}, X_{n_1}, \dots, X_{n_{N-1}}) = P(X_{n_0+m}, X_{n_1+m}, \dots, X_{n_{N-1}+m}) \quad (1.35)$$

Thus, if a stochastic process  $X_n$  is SSS, the process  $Y_n = X_{n+m}$  defines the same probability measure over the signal space.

*Example 10 (Bernoulli process)*

A Bernoulli process,  $\mathcal{B}(p)$  given by (1.7) is IID, thus, its joint probability mass function has the form

$$P_{X_{n_0}, \dots, X_{n_{N-1}}}(x_0, \dots, x_{N-1}) = \prod_{i=0}^{N-1} p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=0}^{N-1} x_i} (1-p)^{N - \sum_{i=0}^{N-1} x_i} \quad (1.36)$$

which does not depend on  $n_0, \dots, n_{N-1}$ . Thus, the Bernoulli process is SSS.

In general, all IID processes are SSS. On the contrary, independent but not identically distributed processes are not SSS. This is the case of non-constant deterministic process, as the following example shows

*Example 11 (Deterministic process)*

Consider the deterministic process given by  $X_n = 2n$  (with probability 1), that is, for each time  $n$ ,  $X_n$  is a discrete random variable with PMF

$$P_{X_n}(k) = \delta[k - 2n] \quad (1.37)$$

Since  $P_{X_n}(k)$  depends on  $n$ ,  $X_n$  is not SSS.

In general, strict stationarity is difficult to check and overly restrictive, and weaker forms of stationarity are more interesting. In the following, we will focus in the invariance of first and second order statistics<sup>1</sup>

### 1.5.2 First and Second-Order Statistics

In order to define wide-sense stationarity, we will first define some first and second-order statistics of stochastic processes. We will assume, in general, stochastic processes which can take complex values:

**Definition 10 (First and second order statistics)** Given stochastic process  $X_n$  we define the following statistics:

- **Mean:**

$$\mu_X[n] = \mathbb{E}\{X_n\} \quad (1.38)$$

- **Autocorrelation:**

$$r_X[n_1, n_2] = \mathbb{E}\{X_{n_1} X_{n_2}^*\} \quad (1.39)$$

- **Cross correlation**

$$r_{XY}[n_1, n_2] = \mathbb{E}\{X_{n_1} Y_{n_2}^*\} \quad (1.40)$$

Given stochastic processes  $X_n$  and  $Y_n$  with means  $\mu_X[n]$  and  $\mu_Y[n]$ , we define the following joint statistics:

- **Cross covariance** between  $X_n$  and  $Y_n$

$$c_{XY}[n_1, n_2] = \mathbb{E}\{(X_{n_1} - \mu_X[n_1])(Y_{n_2} - \mu_Y[n_2])^*\} \quad (1.41)$$

- **Autocovariance**

$$c_X[n_1, n_2] = \mathbb{E}\{(X_{n_1} - \mu_X[n_1])(X_{n_2} - \mu_X[n_2])^*\} \quad (1.42)$$

Note that  $c_X[n, n]$  is the variance of  $X_n$ . It is easy to verify that the autocovariance and the autocorrelation are related by

$$c_X[n_1, n_2] = r_X[n_1, n_2] - \mu_X[n_1] \mu_X^*[n_2] \quad (1.43)$$

---

<sup>1</sup> In this document, first and second order statistics refer to the expected value of polynomial functions of a random variable up to degree 2, that is, the mean, the correlation and other related statistics, like the covariance. This should not be confused with the so called "order statistics".



*Example 12 (Bernoulli process)*

For a Bernoulli process,  $\mathcal{B}(p)$ , using (1.7) we get

$$\mu_X[n] = \mathbb{E}\{X_n\} = p, \quad (1.44)$$

also

$$\begin{aligned} r_X[n_1, n_2] &= \mathbb{E}\{X_{n_1}X_{n_2}\} = \begin{bmatrix} \mathbb{E}\{X_{n_1}^2\} & \text{if } n_1 = n_2 \\ \mathbb{E}\{X_{n_1}\}\mathbb{E}\{X_{n_2}\} & \text{if } n_1 \neq n_2 \end{bmatrix} = \begin{bmatrix} p & \text{if } n_1 = n_2 \\ p^2 & \text{if } n_1 \neq n_2 \end{bmatrix} \\ &= p^2 + p(1-p)\delta[n_1 - n_2] \end{aligned} \quad (1.45)$$

and  $c_X[n_1, n_2] = r_X[n_1, n_2]$ .

*Example 13 (Deterministic process)*

Consider the deterministic process in Example 11, we have

$$\mu_X[n] = \mathbb{E}\{X_n\} = \mathbb{E}\{2n\} = 2n, \quad (1.46)$$

$$r_X[n_1, n_2] = \mathbb{E}\{X_{n_1}X_{n_2}\} = \mathbb{E}\{4n_1n_2\} = 4n_1n_2 \quad (1.47)$$

and, using (1.43)

$$c_X[n_1, n_2] = 0 \quad (1.48)$$

*Example 14* Consider the stochastic process  $X_n$  given by

$$X_n = (1 + n^2)Y \quad (1.49)$$

where  $Y$  is a real Gaussian random variable with zero mean and unit variance.

The mean and the autocorrelation of the process can be easily calculated as

$$\mu_X[n] = \mathbb{E}\{Y\} \cdot (1 + n^2) = 0 \quad (1.50)$$

$$r_X(n, m) = \mathbb{E}\{X_nX_m\} = (1 + n^2) \cdot (1 + m^2) \cdot \mathbb{E}\{Y^2\} = (1 + n^2) \cdot (1 + m^2) \quad (1.51)$$

Since  $X_m$  has zero mean, the autocovariance is equal to the autocorrelation, and the variance is

$$\sigma_X^2[n] = r_X[n, n] = (1 + n^2)^2 \quad (1.52)$$

*Example 15 (Simple Random Walk)*

For the SRW  $X_n$  given by (1.8), the value of the process at any time  $n$  depends on the value of a set of  $n$  random variables  $(Y_1, \dots, Y_n)$ . We have seen that the process is zero-mean with variance  $n$ . The autocorrelation (equal to the autocovariance) can be computed as

$$\begin{aligned} r_X[n_1, n_2] &= \mathbb{E}\{X_{n_1}X_{n_2}\} = \sum_{k=1}^{n_1} \sum_{\ell=0}^{n_2} \mathbb{E}\{Y_kY_\ell\} = \sum_{k=1}^{n_1} \sum_{\ell=0}^{n_2} \delta[k - \ell] \\ &= \min(n_1, n_2) \end{aligned} \quad (1.53)$$

### 1.5.3 Wide-Sense Stationarity

We are ready to define a weaker form of stationarity:

**Definition 11 (Wide-Sense Stationarity)** A two-sided stationary process is Wide-Sense Stationary (or simply WSS) if its mean and autocorrelation functions are invariant to a time shift, that is, for any integers  $n, n_1, n_2$  and any time shift  $m \in \mathbb{Z}$ ,

$$\mu_X[n] = \mu_X[n+m] \quad (1.54)$$

and

$$r_X[n_1, n_2] = r_X[n_1 + m, n_2 + m] \quad (1.55)$$

Thus, if a process is WSS, its mean is constant and its autocorrelation only depends on the difference between  $n_1$  and  $n_2$ . For this reason, we will typically use the notation

$$\mu_X = \mu_X[n] \quad (1.56)$$

$$r_X[n] = r_X[m+n, m] \quad (1.57)$$

Also, combining (1.56) and (1.57) with (1.43), it is straightforward to see that the covariance function is also invariant to a time shift, and we can write

$$c_X[n] = c_X[m+n, m] \quad (1.58)$$

*Example 16* Let's go back to the examples from the previous section.

- **Bernoulli process:** The Bernoulli process  $X_n$  from Example 10 is SSS, since their joint probability function, (1.36), does not depend on the time variables. Consequently, it is also WSS, as can be seen by observing that the average in (1.44) is constant and the autocorrelation in (1.45) depends on the time difference only.
- **Deterministic process:** For the deterministic process  $X_n = 2n$  in Example 11, we have  $\mu_X[n] = 2n$ , which depends on  $n$ . Thus, it is not WSS.
- The process  $X_n$  in Example 14 has zero mean (thus, the mean is invariant to a time shift), but the autocorrelation  $r_X[n_1, n_2]$  (1.51) cannot be expressed as function of the time difference  $n_1 - n_2$  only. Thus, the process is not WSS.
- **Simple Random Walk:** The SRW from Example 15 is not stationary since its autocorrelation (1.53) is not invariant to a time shift: for any  $m > 0$

$$\begin{aligned} r_X[n_1 + m, n_2 + m] &= \min(n_1 + m, n_2 + m) = \min(n_1, n_2) + m = r_X(n_1, n_2) + m \\ &\neq r_X[n_1, n_2] \end{aligned} \quad (1.59)$$

#### 1.5.3.1 Properties of stationary processes

**Theorem 2** The autocorrelation WSS process  $X_n$  satisfies the following properties:

1. **Hermiticity:**  $r_X[n] = r_X^*[-n]$ .
2.  $r_X[0] = \mathbb{E}\{|X_n|^2\}$ .

3. It is **positive-definite**: for any integer  $N > 0$  and any coefficients  $c_n \in \mathbb{C}, n = 0, \dots, N-1$ ,

$$\sum_{m=0}^{N-1} \sum_{n=0}^{N-1} c_m c_n^* r_X[n-m] \geq 0 \quad (1.60)$$

4. **Maximality at the origin**:  $|r_X[n]| \leq r_X[0]$ .

**Proof** The hermiticity and property 2 are a direct consequence of the definition. To prove the positive definiteness, note that, for any coefficients  $c_n$ , we have

$$0 \leq \mathbb{E} \left\{ \left| \sum_{n=0}^{N-1} c_n X_n \right|^2 \right\} = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} c_m c_n^* r_X[n-m] \quad (1.61)$$

The maximality at the origin is a consequence of the positive definiteness: since (1.60) is true for any complex coefficients, we can take  $c_0 = 1, c_1 = c_2 = \dots = c_{N-1} = 0$  to get

$$(1 + |c_n X_n|^2) r_X[0] + c_n r_X[n] + c_n^* r_X^*[n] \geq 0 \quad (1.62)$$

and, taking  $c_n = -\frac{r_X^*[n]}{|r_X[n]|}$  (which implies  $|c_n|^2 = 1$ ), we get

$$2r_X[0] - 2|r_X[n]| \geq 0 \quad (1.63)$$

which completes the proof.  $\square$

### 1.5.3.2 WSS and SSS processes

Note that stationarity in the strict sense implies the wide sense stationarity, but the opposite is not generally true: there are SSS processes that are not WSS, as the following example shows:

*Example 17* Let  $X_n$  be a zero mean stationary stochastic process and  $Y_n = (-1)^n X_n$ . It is easy to check that

$$\mathbb{E}\{Y_n\} = (-1)^n \mathbb{E}\{X_n\} = 0 \quad (1.64)$$

and

$$r_Y[n_1, n_2] = (-1)^{n_1+n_2} r_X[n_2 - n_1] \quad (1.65)$$

Since  $(-1)^{n_1+n_2} = (-1)^{n_2-n_1}$ , it follows

$$r_Y[n_1, n_2] = (-1)^{n_2-n_1} r_X[n_2 - n_1] \quad (1.66)$$

which only depends on  $n_2 - n_1$ . Therefore,  $Y_n$  is WSS. However, we can check that, for example,

$$\mathbb{E}\{Y_n^3\} = (-1)^n \mathbb{E}\{X_n^3\} \quad (1.67)$$

which, in general, depends on  $n$ . For example, if, for all  $n$ ,  $X_n$  follows the discrete distribution

$$p_X[k] = \frac{3}{4}\delta[k-1] + \frac{1}{4}\delta[k+3] \quad (1.68)$$

we have  $\mathbb{E}\{X_n\} = 0$  and  $\mathbb{E}\{X_n^3\} = -3/2$ , then

$$\mathbb{E}\{Y_n^3\} = -\frac{3}{2}(-1)^n \quad (1.69)$$

which depends on  $n$ . Therefore, some statistics from  $X_n$  are not invariant to a time shift and, thus,  $X_n$  is not SSS.

Also, note that all IID processes are SSS, with autocovariance

$$c_X[n] = \mathbb{E}\{(X_n - \mu_X)(X_0 - \mu_X)^*\} = \sigma_X^2 \delta[n] \quad (1.70)$$

If the process is zero mean, we have

$$r_X[n] = \sigma_X^2 \delta[n] \quad (1.71)$$

In general, a SP whose autocovariance is a delta function is said to be white. All zero-mean IID processes are white, but the opposite is not true in general: some white processes may be non IID.

### 1.5.3.3 Jointly stationary processes.

The two-side stationary processes  $X_n$  and  $Y_n$  are said to be jointly stationary *in the strict sense* if their joint statistical properties do not vary with a displacement in time, that is, if, for any values of  $N$  and  $N'$ , any value of  $m$  and any instants  $n_0, \dots, n_{N-1}$  and  $n'_0, \dots, n'_{N'-1}$ ,

$$P(X_{n_0}, \dots, X_{n_{N-1}}, Y_{n'_0}, \dots, Y_{n'_{N'-1}}) = P(X_{n_0+m}, \dots, X_{n_{N-1}+m}, Y_{n'_0+m}, \dots, Y_{n'_{N'-1}+m}) \quad (1.72)$$

Likewise, we will say that  $X_n$  and  $Y_n$  are jointly wide-sense stationary if both are WSS and for any integer values of  $n_1, n_2$  and  $m$  we have

$$r_{XY}[n_1, n_2] = r_{XY}[n_1 + m, n_2 + m] \quad (1.73)$$

which allows us to describe the cross-correlation function using a single variable

$$r_{XY}[n] = r_{XY}[m + n, m] \quad (1.74)$$

### 1.5.4 Linear systems with stochastic inputs

Let  $X_n$  be a stochastic process that is the input to a linear time-invariant system with impulse response  $h[n]$ . The output process  $Y_n$  is given by

$$Y_n = X_n * h[n] \quad (1.75)$$

In general, the statistical characterization of  $Y_n$  is not easy, except in particular cases (such as the Gaussian), but we can find some useful expressions for some first and second order moments.

For example, the mean of the output process is

$$\begin{aligned} \mu_Y[n] &= \mathbb{E}\{Y_n\} = \mathbb{E}\{h[n] * X_n\} = h[n] * \mathbb{E}\{X_n\} \\ &= h[n] * \mu_X[n], \end{aligned} \quad (1.76)$$

the cross-correlation of the input and output is

$$\begin{aligned} r_{YX}[n_1, n_2] &= \mathbb{E}\{Y_{n_1} X_{n_2}^*\} = E \left\{ X_{n_2}^* \sum_{k=-\infty}^{\infty} X_k h[n_1 - k] \right\} \\ &= \sum_{k=-\infty}^{\infty} h[n_1 - k] r_X[k, n_2] \\ &= h[n_1] * r_X[n_1, n_2] \end{aligned} \quad (1.77)$$

and the autocorrelation of the output is

$$\begin{aligned} r_Y[n_1, n_2] &= \mathbb{E}\{Y_{n_1} Y_{n_2}^*\} = E \left\{ Y_{n_1} \sum_{k=-\infty}^{\infty} h^*[k] X_{n_2-k}^* \right\} \\ &= \sum_{k=-\infty}^{\infty} h^*[k] r_{YX}[n_1, n_2 - k] \\ &= r_{YX}[n_1, n_2] * h^*[n_2] \end{aligned} \quad (1.78)$$

Combining (1.77) and (1.78), we get

$$r_Y[n_1, n_2] = h[n_1] * r_X[n_1, n_2] * h^*[n_2] \quad (1.79)$$

In the above equation, there are two convolution operations. We must perform the first on the variable  $n_1$  and the second on the variable  $n_2$ . Equations (1.76) and (1.79) show that both the mean and the autocorrelation of the output depend exclusively on the mean and the autocorrelation of the input, respectively, as well as on the impulse response of the system.

#### 1.5.4.1 Stationary processes.

We can use the general expressions for the output mean and the output autocorrelation to derive specific expression for WSS processes. If  $X_n$  is stationary,  $\mu_X[n] = \mu_X$ , and

$$\mu_Y = \mu_X \sum_{n=-\infty}^{\infty} h[n] \quad (1.80)$$

Also, by the properties of the convolution operator, for any  $m \in \mathbb{Z}$  we can write

$$\begin{aligned}
r_Y[n_1 + m, n_2 + m] &= h[n_1] * r_X[n_1 + m, n_2 + m] * h^*[n_2] = h[n_1] * r_X[n_1, n_2] * h^*[n_2] \\
&= r_Y[n_1, n_2]
\end{aligned} \tag{1.81}$$

and, therefore  $Y_n$  is also stationary. In such case, we can express the autocorrelation as a function of a single variable. By calling  $n = n_1 - n_2$ , the cross-correlation of the input and output is

$$\begin{aligned}
r_{YX}[n_1, n_2] &= h[n_1] * r_X[n_1 - n_2] = \sum_{k=-\infty}^{\infty} h[n_1 - k] r_X[k - n_2] \\
&= \sum_{k=-\infty}^{\infty} h[n_2 + n - k] r_X[k - n_2] = \sum_{k=-\infty}^{\infty} h[n - k] r_X[k] \\
&= h[n] * r_X[n]
\end{aligned} \tag{1.82}$$

Finally, the autocorrelation of the output is

$$\begin{aligned}
r_Y[n] &= r_Y[n_1, n_2] = r_{YX}[n_1, n_2] * h^*[n_2] \\
&= \sum_{k=-\infty}^{\infty} h^*[k] r_{YX}[n_1 - n_2 + k] = \sum_{k=-\infty}^{\infty} h^*[k] \sum_{\ell=-\infty}^{\infty} h[n + k - \ell] r_X[\ell] \\
&= \sum_{\ell=-\infty}^{\infty} r_X[\ell] \sum_{k=-\infty}^{\infty} h^*[k] h[n + k - \ell] = r_X[n] * h[n] * h^*[-n] \\
&= r_X[n] * r_h[n]
\end{aligned} \tag{1.83}$$

where  $r_h[n] = h[n] * h^*[-n]$ .

*Example 18* Let  $X_n$  be a unit variance white process. The stationary process  $Y_n$  resulting from passing  $X_n$  through a linear and invariant and causal filter, given by the difference equation

$$Y_n = 0.6 \cdot Y_{n-1} + X_n \tag{1.84}$$

is stationary. According to (1.76), since the impulse response of this system is  $h[n] = 0.6^n u[n]$ , the process mean will be,

$$\mu_Y[n] = h[n] * \mu_X[n] = 0 \tag{1.85}$$

and its autocorrelation

$$r_Y[n] = r_X[n] * h[n] * h[-n] = \frac{0.6^{|n|}}{0.64} \tag{1.86}$$

### 1.5.5 Gaussian processes

**Definition 12 (Gaussian Process)** A real SP  $X_n$  is a Gaussian Process (GP) if, for any value of  $N$  and any arbitrary set of  $N$  time instants  $\{n_0, \dots, n_{N-1}\}$ , the pdf of order  $N$ ,

$$p_{X_{n_0}, \dots, X_{n_{N-1}}}(x_0, \dots, x_{N-1}) \tag{1.87}$$

is Gaussian.

Gaussian processes have several interesting properties:

1. Since the parameters of the (multidimensional) Gaussian distribution are the mean and the covariance matrix, any GP is completely characterized by its mean and its autocorrelation function.
2. Thus, if a GP is WSS, it is SSS.
3. Since any linear combination of Gaussian random variables is Gaussian, if a GP is the input to a linear time-invariant system, the output is also GP characterized by the output mean and the output autocorrelation driven by Eqs. (1.76) and (1.97).

### 1.5.6 Power Spectral Density

The Power Spectral Density is the basic tool to extend the Fourier Analysis of signals to stochastic processes.

Since the realizations of a stochastic process are signals, it is tempting to define the Fourier Transform of a process stochastic  $X_n$  as a new process whose realizations are the Fourier transforms of the realizations of  $X_n$ ,  $x_n$ . However, there is an insurmountable difficulty in this definition: the realizations of the stationary stochastic processes are signals of finite and non-zero average power and infinite energy and, therefore, they do not have a Fourier Transform.

However, in general, the Fourier Transform of a truncated process (limited in time) can be computed. Let  $X_n$  be a stochastic process of mean  $\mu_X[n]$  and autocorrelation  $r_X[n_1, n_2]$ , and consider the truncated process  $X_{N,n}$  given by

$$X_{N,n} = \begin{cases} X_n, & \text{if } |n| \leq N \\ 0, & \text{if } |n| > N \end{cases} \quad (1.88)$$

The Fourier Transform of  $X_{N,n}$ , given by

$$X_N(e^{j\omega}) = \sum_{n=-\infty}^{\infty} X_{N,n} e^{-j\omega n} = \sum_{n=-N}^N X_n e^{-j\omega n} \quad (1.89)$$

is, in turn, a stochastic process of mean

$$\mathbb{E}\{X_N(e^{j\omega})\} = \sum_{n=-N}^N \mu_X[n] e^{-j\omega n} \quad (1.90)$$

and squared mean

$$\mathbb{E}\{|X_N(e^{j\omega})|^2\} = E \left\{ \left| \sum_{n=-N}^N X_n e^{-j\omega n} \right|^2 \right\} \quad (1.91)$$

If the process is stationary, the mean quadratic value grows indefinitely with  $N$  until it becomes infinity, but we can avoid this effect by normalizing (1.91) with respect to the

length of the integration interval,  $2N$ . We define the *power spectral density* or PSD of a stochastic process to the limit of the expression above.

**Definition 13 (Power Spectral Density)** The power spectral density (PSD) of a stochastic process  $X_n$  is defined as

$$S_X(e^{j\omega}) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \mathbb{E}\{|X_N(e^{j\omega})|^2\} \quad (1.92)$$

Note that, by definition, the PSD is always real and not negative: for all  $\omega$ ,

$$S_X(e^{j\omega}) \geq 0 \quad (1.93)$$

Also, by the properties of the Fourier Transform, the PSD of a real stochastic process is an even function.

The expression in (1.92) is not practical, but we can find an alternative expression for stationary processes, that allows to compute the PSD using the autocorrelation, using a classical result from Wiener and Khinchine that we state without proof.

**Theorem 3 (Wiener-Khinchine theorem)**

*If  $X_n$  is WSS and  $r_X[n]$  is absolutely summable,*

$$S_X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} r_X[n] e^{-j\omega n} \quad (1.94)$$

Therefore, the power spectral density of a stationary process  $X_n$  is equal to the Fourier Transform of its autocorrelation function and, conversely

$$r_X[n] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(e^{j\omega}) e^{j\omega n} d\omega \quad (1.95)$$

*Example 19 White processes*

If  $X_n$  is a white WSS process with autocorrelation  $r_X[k] = \sigma_X^2 \delta[k]$ , its power spectral density is constant at all frequencies.

$$S_X(e^{j\omega}) = \sigma_X^2 \quad (1.96)$$

### 1.5.6.1 Power spectral density at the output of linear systems

It is interesting to analyze the power spectral density of stochastic processes that pass through linear filters. Starting from the expression for the output autocorrelation obtained in (1.79), we can write

$$r_Y[n] = r_X[n] * h[n] * h^*[-n] \quad (1.97)$$

and applying the convolution property to (1.97), we get

$$S_Y(e^{j\omega}) = S_X(e^{j\omega}) |H(e^{j\omega})|^2 \quad (1.98)$$



### 1.5.7 Ergodicity

If the expected value of a random variable  $X$  is unknown, it can be estimated as the sample average of  $K$  independent realizations  $x_0, \dots, x_{K-1}$

$$\mathbb{E}\{X\} \approx \frac{1}{K} \sum_{k=0}^{K-1} x_k \quad (1.99)$$

This estimation is supported by the (weak and strong) laws of large numbers, that guarantee, under quite general conditions, the convergence of the sample average to the mean as the number of samples goes to infinity.

We can estimate the mean of a stochastic process in the same way, by averaging multiple realizations. However, in many practical applications, this is not possible because only a single sample of the process is available.

If a stochastic process  $X_n$  is WSS, the mean is constant, and we can try to estimate it as the average of all samples from a single realization. Thus, we may wonder if time averages of the SP converge to the mean. A process satisfying this property is called mean-ergodic.

**Definition 14 (Mean-ergodicity)**

A WSS process  $X_n$  with mean  $\mu_X$  is mean-ergodic if the time average

$$S_N = \frac{1}{2N+1} \sum_{n=-N}^N X_n \quad (1.100)$$

converges in squared mean to the mean,  $\mu_X$ , that is

$$\lim_{N \rightarrow \infty} \mathbb{E} \left\{ |S_N - \mu_X|^2 \right\} = 0 \quad (1.101)$$

Note that  $S_N$  is itself a random variable with the same mean as the process

$$\mathbb{E}\{S_N\} = \frac{1}{2N+1} \sum_{n=-N}^N \mathbb{E}\{X_n\} = \mu_X \quad (1.102)$$

and the expectation in (1.101) is the variance. Expanding the expression of this variance, it is straightforward to obtain the following condition for ergodicity:

**Theorem 4** A WSS process  $X_n$  with mean  $\mu$  is mean-ergodic iff

$$\lim_{N \rightarrow \infty} \frac{1}{(2N+1)^2} \sum_{n=-N}^N \sum_{m=-N}^N c_X[n-m] = 0 \quad (1.103)$$

**Proof** Noting that

$$\begin{aligned}
\mathbb{E}\{|S_N - \mu_X|^2\} &= E \left\{ \left| \frac{1}{2N+1} \sum_{n=-N}^N (X_n - \mu_X) \right|^2 \right\} = \\
&= \frac{1}{(2N+1)^2} \sum_{n=-N}^N \sum_{m=-N}^N \mathbb{E}\{(X_n - \mu_X)(X_m^* - \mu_X^*)\} \\
&= \frac{1}{(2N+1)^2} \sum_{n=-N}^N \sum_{m=-N}^N c_X[n-m] \tag{1.104}
\end{aligned}$$

the proof is completed.  $\square$

We can use the above theorem to state a sufficient condition for ergodicity, that is satisfied by many WSS processes of practical interest:

**Theorem 5** *If the covariance function of a WSS process  $X_n$  is absolutely summable, that is*

$$\sum_{n=-\infty}^{\infty} |c_X[n]| < \infty \tag{1.105}$$

*the process is mean-ergodic.*

**Proof** Since  $c_X[n]$  is absolutely summable,  $B = \sum_{n=-\infty}^{\infty} |c_X[n]|$  is finite, and we can use  $B$  to upper bound the variance in (1.106) as

$$|\mathbb{E}\{|S_N - \mu_X|^2\}| \leq \frac{1}{(2N+1)^2} \sum_{n=-N}^N \sum_{m=-N}^N |c_X[n-m]| = \frac{1}{(2N+1)^2} \sum_{n=-N}^N B = \frac{B}{2N+1} \tag{1.106}$$

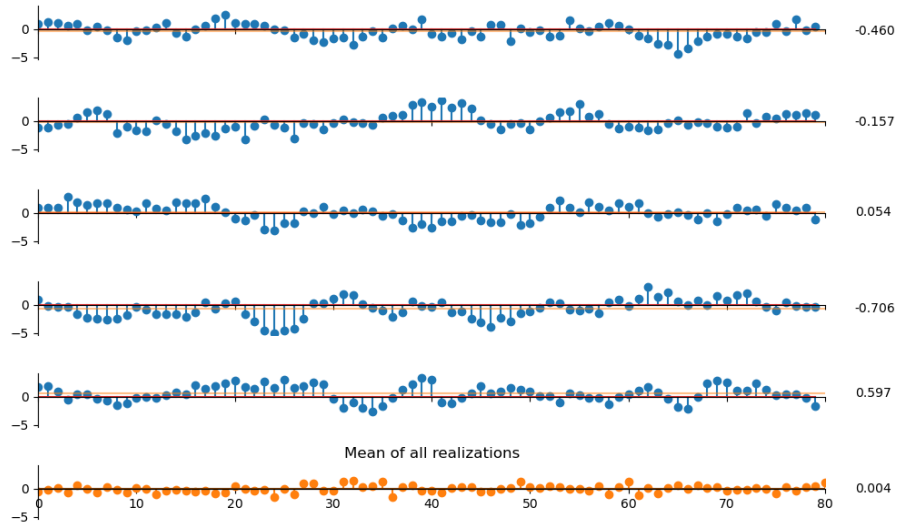
Since the upper bound converges to zero for large  $N$ , the process is mean-ergodic.  $\square$

Note that all white processes (and, in particular, IID processes) satisfy (1.105). Thus, they are mean-ergodic.

**Example 20** Consider the stationary process given by the causal recursive filter  $X_n = 0.8X_{n-1} + Z_n$ , where the input,  $Z_n$ , is a Gaussian IID process of unit variance. It is easy to see that the impulse response of the filter is  $h[n] = 0.8^n \cdot u[n]$ , where  $u[n]$  is the step function. Thus,  $X_n$  is a zero-mean stationary process with autocovariance given by (1.83), which is equal to

$$c_X[n] = \frac{0.8^{|n|}}{0.36} \tag{1.107}$$

Since the autocovariance vanishes for  $|k| > 1$ , the process is mean-ergodic. Figure 1.4 represents the first samples (from  $n = 0$  to  $n = 79$ ) of 5 realizations. The average of 1000 realizations is represented in the lower part. The 80 sample average of each realization is shown on the right margin. As a consequence of stationarity, the average signal in the bottom approaches a constant signal around the mean, which is zero. As a consequence of the ergodicity, the averages of each signal also approaches the mean.



**Fig. 1.4** 5 realizations of the stochastic process of the example 20 (top), the average of 1000 realizations (bottom) and the averages of the represented samples (right). Because of the stationarity, the average signal (at the bottom) approaches a zero signal, which is mean of the process. Because of the ergodicity, time averages also approximate the mean.

### 1.5.7.1 Ergodicity in the autocorrelation

In a similar way, ergodicity is defined in autocorrelation. A WSS process  $X_n$  is autocorrelation-ergodic if

$$r_N = \frac{1}{2N+1} \sum_{m=-N}^N X_{m+n} X_m^* \quad (1.108)$$

converges to  $r_X[n]$  in squared mean. In general,  $X_n$  is autocorrelation-ergodic if, for every value of  $m$ , the process  $Z_n[m] = X_{m+n} X_m^*$  is mean-ergodic.

## References

1. Kempthorne, Peter, Choongbum Lee, Vasily Strela, and Jake Xia. "Topics in mathematics with applications in finance." Massachusetts Institute of Technology: MIT OpenCourseWare), <https://ocw.mit.edu/courses/18-s096-topics-in-mathematics-with-applications-in-finance-fall-2013/> (Accessed Feb, 2023). License: Creative Commons BY-NC-SA.
2. Artés-Rodríguez, A., Pérez-González, F., Cid-Sueiro, J., López-Valcarce, R., Mosquera-Nartallo, C., Pérez-Cruz, F. (2012). Comunicaciones digitales ([http://www.tsc.uc3m.es/~antonio/libro\\_comunicaciones](http://www.tsc.uc3m.es/~antonio/libro_comunicaciones)).