Chapter 1
Stochastic Processes

Contents

***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.*

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

$$X_n = n, \qquad \text{for all } n \geq 0 \tag{1.1}$$

or

$$X_n = -n, \qquad \text{for all } n \geq 0 \tag{1.2}$$

with equal probabilities. Note that, the Universe set of this SP if 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 $P(X_{n_0}, X_{n_1}, \ldots, X_{n_{M-1}})$ 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 $X_{n_0}, X_{n_1}, \ldots, X_{n_{M-1}}$ with $n_0 < n_1, \ldots < n_{M-1}$ we have

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

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 $X_{n_0}, X_{n_1}, \ldots, X_{n_{M-1}}$ is the product of the marginal distribution of each variable. Thus, for instance, the cumulative distributions gets factorized:

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

for any $x_{n_0}, x_{n_1}, \ldots, x_{n_{M-1}}$ 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

$$p_{X_{n_0}, X_{n_1}, \ldots, X_{n_{M-1}}}(x_{n_0}, x_{n_1}, \ldots, x_{n_{M-1}}) = \prod_{k=0}^{M-1} p_{X_{n_k}}(x_{n_k}) \tag{1.5}$$

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

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

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 $\mathscr{B}(p)$ is an IID process given by Bernoulli random variables:

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

(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 3 realizations of the independent stochastic process X_n, where each sample of the process en follow a exponential distribution with parameter $\lambda_n = 1/(n+1)$.

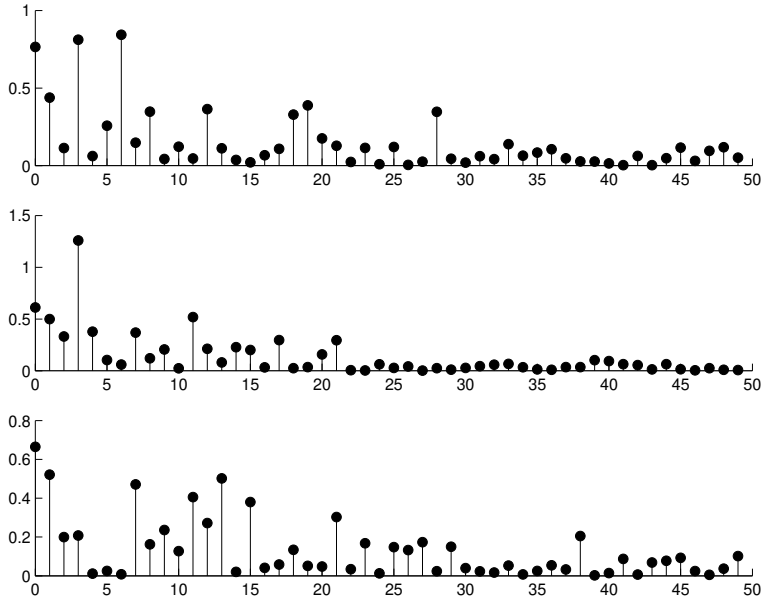

Fig. 1.1 3 realizations of process X_n described in example 4

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

$$X_n = \sum_{k=1}^{n} Y_k, \qquad n > 0 \tag{1.8}$$

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

$$X_n = X_{n-1} + Y_n, \qquad n > 0 \tag{1.9}$$

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.

$$-n \leq X_n \leq n \tag{1.10}$$

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

$$\mathbb{E}\{X_n\} = \sum_{k=0}^{n} \mathbb{E}\{Y_k\} = 0 \tag{1.11}$$

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

$$\text{var}\{X_n\} = \sum_{k=1}^{n} \text{var}\{Y_k\} = n \tag{1.12}$$

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 $Z_n = X_{n+h} - X_n$ 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:

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

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 $P\{X_N = U | X_0 = 0\}$. Thus, defining function

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

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

Applying the total probability theorem,

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

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, the 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 $X_0, X_1, \ldots, X_n$. 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 $\mathscr{S}$, which we call the **state space**. We say that the SP has the Markov property if

$$P\{X_{n+1} = i | X_n, X_{n-1}, \ldots, X_0\} = P\{X_{n+1} = i | X_n\} \tag{1.15}$$

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

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 $\mathscr{S}$ is a finite set. In this case, we usually take the elements of $\mathscr{S}$ as consecutive integers. For instance, for a state space with m elements, we will take $\mathscr{S} = \{0, 1, \ldots, m-1\}$ or $\mathscr{S} = \{1, 2, \ldots, m\}$ 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

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

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

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

Using (1.16), it is easy to see that

$$\sum_{j \in \mathscr{S}} p_{ij} = 1 \tag{1.18}$$

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

$$\mathbf{P}\mathbb{1}_m = \mathbb{1}_m \tag{1.19}$$

where $\mathbb{1}_m$ 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.

$$P\{X_{n+1} = j | X_n = i\} = P\{X_{m+1} = j | X_m = i\}, \qquad i \in \mathscr{S}, \quad j \in \mathscr{S} \tag{1.20}$$

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 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 $\mathscr{S} = \{0,1,2\}$ and transition probability matrix

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

is shown in Figure 1.3.

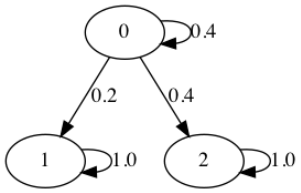

Fig. 1.3 The transition graph for the markov chain given by (1.21)

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

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

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

$$\mathbf{R}_n = \mathbf{R}_{n-1} \mathbf{P} \tag{1.23}$$

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

$$\mathbf{R}_n = \mathbf{P}^n \tag{1.24}$$

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

$$q_{n,i} = P\{X_n = i\}, \quad n \geq 0, \quad i \in \mathscr{S} \tag{1.25}$$

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

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

which, in vector form, can be computed as

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

1.4.3 Stationary distribution

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

$$\mathbf{P}^\intercal \boldsymbol{\pi} = \boldsymbol{\pi} \tag{1.28}$$

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

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

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

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

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

$$\lim_{n\to\infty} \mathbf{r}_i(n) = \boldsymbol{\pi}, \qquad \forall i, j \in \mathscr{S} \tag{1.29}$$

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 $\lim_{n\to\infty} \mathbf{r}_i(n)$ may depend on i.

Example 9 The MC in example in 7 has at least two limiting distributions: $\boldsymbol{\pi}_A = (0,1,0)^\intercal$ and $\boldsymbol{\pi}_B = (0,0,1)^\intercal$. 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

$$\lim_{n\to\infty} \mathbf{r}_i(n) = (0,0,1) \tag{1.30}$$

$$\lim_{n\to\infty} \mathbf{r}_i(n) = (0,1,0) \tag{1.31}$$

$$\lim_{n\to\infty} \mathbf{r}_i(n) = \left(0, \frac{1}{3}, \frac{2}{3}\right) \tag{1.32}$$

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 $\mathbb{1}^\intercal \mathbf{p} = 1$, that is

$$\begin{pmatrix} \mathbf{P}^\intercal - \mathbf{I} \\ \mathbb{1}^\intercal \end{pmatrix} \boldsymbol{\pi} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{1.33}$$

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

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

so that the rank of $\mathbf{P}^\intercal - \mathbf{I}$ 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

1.5.1 Characterization of stochastic processes

A discrete-time SP X_n, can be characterized uniquely by knowing, for any value of N and any arbitrary set of N time instants $\{n_0,\ldots,n_{N-1}\}$, the joint distributions

$$P(X_{n_0},X_{n_1},\ldots,X_{n_{N-1}}) \tag{1.35}$$

This characterization is not very useful because, in general, the knowledge of the statistical properties of the process may be indirect or incomplete, and determining and processing this information may be difficult, impractical or even impossible. Fortunately, in many cases, we can make some assumptions that simplify this characterization, or it is only necessary to carry out a partial characterization.

In this text we are interested, in particular, in the processes characterized by their mean and autocorrelation functions, which are defined below. We will assume, in general, stochastic processes which can take complex values:

Definition 9 (Mean of a stochastic process) The mean of a continuous stochastic process X_n is defined as

$$\mu_X[n] = \mathbb{E}\{X_n\} \tag{1.36}$$

Note that, in general, the mean depends on time.

Definition 10 (Autocorrelation) The autocorrelation of a stochastic process X_n is defined as

$$R_X[n_1,n_2] = \mathbb{E}\{X_{n_1}X_{n_2}^*\} \tag{1.37}$$

Definition 11 (Cross correlation) The cross-correlation between the stochastic processes X_n and Y_n is

$$R_{XY}[n_1,n_2] = \mathbb{E}\{X_{n_1}Y_{n_2}^*\} \tag{1.38}$$

Definition 12 (Cross covariance) The cross-covariance between the stochastic processes X_n and Y_n of means $\mu_X[n]$ and $\mu_Y[n]$, respectively, is

$$C_{XY}[n_1,n_2] \doteq \mathbb{E}\{(X_{n_1}-\mu_X[n_1])(Y_{n_2}-\mu_Y[n_2])^*\} \tag{1.39}$$

Definition 13 (Autocovariance) The autocovariance (or simply covariance) of the process X_n with mean $\mu_X[n]$ is

$$C_X[n_1,n_2] \doteq \mathbb{E}\{(X_{n_1}-\mu_X[n_1])(X_{n_2}-\mu_X[n_2])^*\} \tag{1.40}$$

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

$$C_X[n_1,n_2] = R_X[n_1,n_2] - \mu_X[n_1]\mu_X^*[n_2] \tag{1.41}$$

Example 10 (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

$$P_{X_n}(k) = \delta[k-2n] \tag{1.42}$$

and, in general,

$$P_{X_{n_0},\ldots,X_{n_{N-1}}}(k_0,\ldots,k_{N-1}) = \prod_{i=0}^{N-1} \delta\left[k_i - 2n_i\right] \tag{1.43}$$

Likewise, it is immediate to verify that its mean is $\mu_X[n] = 2n$, its autocorrelation is $R_X[n_1,n_2] = 4n_1n_2$ and its autocovariance is zero.

Example 11 Consider the stochastic process X_n given by

$$X_n = A \cdot (1+n^2) \tag{1.44}$$

where $u[n]$ is the step function and A is a real Gaussian random variable with mean 0 and unit variance.

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

$$\mu_X[n] = \mathbb{E}\{A\} \cdot (1+n^2) = 0 \tag{1.45}$$

$$R_X(n,m) = \mathbb{E}\{X_nX_m\} = \mathbb{E}\{A^2\} \cdot (1+n^2) \cdot (1+m^2) = (1+n^2) \cdot (1+m^2) \tag{1.46}$$

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

$$\sigma_X^2[n] = R_n(n,n) = \left(1+n^2\right)^2 \tag{1.47}$$

Note that, at any time, n, X_n is the product of a Gaussian random variable by a deterministic factor $(1+n^2)$. Thus, X_n is Gaussian. Despite X_n is not a deterministic process, the random variables in the process are strongly dependent: knowing the value of the process at some time $m > 0$, all values of X_n become completely determined. For instance, under the event $X_m = x_m$, for some $x_m \in \mathbb{R}$, we have $A \cdot (1+m^2) = x_m$, thus $A = \frac{x_m}{(1+m^2)}$ and therefore $X_n = \frac{nx_m}{1+n^2}$, for all n, with probability 1. This can be more formally expressed in terms of conditional distributions:

$$p_{X_n|X_m}(x_n|x_m) = \delta\left[x_n - \frac{nx_m}{(1+m^2)}\right] \tag{1.48}$$

From this expression we can compute the joint distribution

$$P_{X_n,X_m}(x_n,x_m) = P_{X_n|X_m}(x_n|x_m)P_{X_m}(x_m) \tag{1.49}$$

$$= \delta\left[x_n - \frac{nx_m}{(1+m^2)}\right] \frac{1}{m\sqrt{2\pi}} e^{-\frac{x_m^2}{2m^2}}, \tag{1.50}$$

Example 12 (Bernoulli process)

A Bernoulli process, $\mathscr{B}(p)$ given by (1.7) is independent and, thus, its joint probability mass function will have the form

$$P_{X_{n_0},\ldots,X_{n_{N-1}}}(x_0,\ldots,x_{N-1}) = \prod_{i=0}^{N-1} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=0}^{N-1} x_i}(1-p)^{N-\sum_{i=0}^{N-1} x_i} \tag{1.51}$$

The mean, the autocorrelation and the autocovariance are

$$\mu_X[n] = \mathbb{E}\{X_n\} = p, \tag{1.52}$$

$$R_X[n_1,n_2] = \mathbb{E}\{X_{n_1}X_{n_2}\} = \begin{bmatrix} p & \text{if } n_1 = n_2 \\ p^2 & \text{if } n_1 \neq n_2 \end{bmatrix} = p^2 + p(1-p)\delta[n_1 - n_2] \tag{1.53}$$

and

$$C_X[n_1,n_2] = p(1-p)\delta[n_1 - n_2] \tag{1.54}$$

respectively.

Example 13 (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 $(Y_1, \ldots Y_n)$. We have seen that the process is zero-mean with variance n. The autocorrelation (equal to the autocovariance) can be computed as

$$\begin{aligned} R_X[n_1,n_2] = \mathbb{E}\{X_{n_1}X_{n_2}\} &= \sum_{k=1}^{n_1}\sum_{\ell=0}^{n_2} \mathbb{E}\{Y_k Y_\ell\} = \sum_{k=1}^{n_1}\sum_{\ell=0}^{n_2} \delta[k-\ell] \\ &= \min(n_1,n_2) \end{aligned} \tag{1.55}$$

1.5.2 Stationarity

Stationarity is a very important property in the analysis of many stochastic processes. It refers to the time invariance of some statistical properties of the process.

Definition 14 (Strict-Sense Stationarity) A two-sided stationary process is said to be Strict-Sense Stationary (or simply SSS) if all its statistics are invariant to a time shift, that is, for any $N > 0$, any sampling instants $n_0, \ldots, x_{n_{N-1}}$ and any time shift $m \in \mathscr{Z}$

$$P(X_{n_0}, X_{n_1}, \ldots, X_{n_{N-1}}) = P(X_{n_0+m}, X_{n_1+m}, \ldots, X_{n_{N-1}+m}) \tag{1.56}$$

Thus, if a stochastic process X_n is SSS, the process $Y_n = X_{n+m}$ has the same statistical properties as X_n. In practice, strict stationarity is difficult to check and overly restrictive. Very frequently, the invariance of some statistics is enough for the analysis, particularly the invariance of first and second order statistics (mean and autocorrelation). In such cases, the process is said to be *wide-sense stationary*.

Definition 15 (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 $m \in \mathbb{Z}$,

$$\mu_X[n] = \mu_X[n+m] \tag{1.57}$$

and

$$R_X[n_1,n_2] = R_X[n_1+m, n_2+m] \tag{1.58}$$

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

$$\mu_X = \mu_X[n] \tag{1.59}$$

$$R_X[n] = R_X[m+n,m] \tag{1.60}$$

Also, combining (1.59) and (1.60) with (1.41), it is straighforward to see that the covariance function is also invariant to a time shift, and we can write

$$C_X[n] = C_X[m+n,m] \tag{1.61}$$

Example 14 Let's go back to the examples from the previous section.

- For the deterministic process $X_n = 2n$ in Example 10, we have $\mu_X[n] = 2n$, which depends on n. Thus, it is not WSS.
- The process X_n in Example 11 has zero mean (thus, the mean is invariant to a time shift), but the autocorrelation $R_X[n_1,n_2]$ (1.46) cannot be expressed as function of the time difference $n_1 - n_2$ only. Thus, the process is not WSS.
- The Bernoulli process X_n from Example 12 is SSS, since their joint probability function, (1.51), does not depend on the time variables. Consequently, it is also WSS, as can be seen by observing that the average in (1.52) is constant and the autocorrelation in (1.53) depends on the time difference only.
- The random walk from Example 13 is not stationary since its autocorrelation (1.55) is not invariant to a time shift: for any $m > 0$

$$\begin{aligned} R_X[n_1+m,n_2+m] &= \min(n_1+m,n_2+m) = \min(n_1,n_2)+m = R_X(n_1,n_2)+m \\ &\neq R_X[n_1,n_2] \end{aligned} \tag{1.62}$$

1.5.2.1 Properties of stationary processes

The autocorrelation of a stationary stochastic process has three properties of interest:

1. The autocorrelation is a Hermitian function: $R_X[k] = R_X^*[-k]$. This property is an immediate consequence of the definition.
2. $R_X[0] = \mathbb{E}\{|X_n|^2\}$.
3. R_X has an absolute maximum at the origin: for all n, $|R_X[n]| \leq R_X[0]$.

We can prove the third property by knowing that, for any value of $a \in \mathbb{C}$,

$$\mathbb{E}\{|X_n - aX_0|^2\} \geq 0 \tag{1.63}$$

Expanding the modulus of the difference, we get

$$2\mathscr{R}e\{a^* R_X[n]\} \leq R_X[0] + |a|^2 R_X[0] \tag{1.64}$$

Since this is true for any $a \in \mathbb{C}$, we can take

$$a = \frac{R_X[n]}{|R_X[n]|} \tag{1.65}$$

so that $|a| = 1$ and, substituting in (1.64)

$$|R_X[n]| \leq R_X[0] \tag{1.66}$$

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

$$\mathbb{E}\{Y_n\} = (-1)^n \mathbb{E}\{X_n\} = 0 \tag{1.67}$$

and

$$R_Y[n_1, n_2] = (-1)^{n_1+n_2} R_X[n_2 - n_1] \tag{1.68}$$

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

$$R_Y[n_1, n_2] = (-1)^{n_2-n_1} R_X[n_2 - n_1] \tag{1.69}$$

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

$$\mathbb{E}\{Y_n^3\} = (-1)^n \mathbb{E}\{X_n^3\} \tag{1.70}$$

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

$$p_X[k] = \frac{3}{4}\delta[k-1] + \frac{1}{4}\delta[k+3] \tag{1.71}$$

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

$$\mathbb{E}\{Y_n^3\} = -\frac{3}{2}(-1)^n \tag{1.72}$$

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

1.5.2.2 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 $n_0, \ldots, n_{N-1}$ and $n'_0, \ldots, n'_{N'-1}$,

$$P(X_{n_0}, \ldots, X_{n_{N-1}}, Y_{n'_0}, \ldots, Y_{n'_{N'-1}}) = P(X_{n_0+m}, \ldots, X_{n_{N-1}+m}, Y_{n'_0+m}, \ldots, Y_{n'_{N'-1}+m}) \tag{1.73}$$

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

$$R_{XY}[n_1, n_2] = R_{XY}[n_1 + m, n_2 + m] \tag{1.74}$$

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

$$R_{XY}[n] \doteq R_{XY}[m + n, m] \tag{1.75}$$

1.5.2.3 White processes.

A particular case of WSS processes of special interest are those whose autocorrelation at any pair of (different) times is zero. Thus, it we say that a stationary stochastic process X_n is white if it is zero-mean and its autocorrelation function has the form

$$R_X[n] = \sigma_X^2 \delta[n] \tag{1.76}$$

where $\sigma_X^2 \geq 0$ is the variance.

1.5.3 Ergodicity

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

$$\mathbb{E}\{X\} \approx \frac{1}{K} \sum_{k=0}^{K-1} x_k \tag{1.77}$$

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 16 (Mean-ergodicity)
A WSS process X_n with mean μ is mean-ergodic if the time average

$$S_N = \frac{1}{2N+1} \sum_{n=-N}^{N} X_n \tag{1.78}$$

converges in squared mean to the mean, μ, that is

$$\lim_{N \to \infty} |S_N - \mu|^2 = 0 \tag{1.79}$$

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

$$\mathbb{E}\{S_N\} = \frac{1}{2N+1}\sum_{n=-N}^{N}\mathbb{E}\{X_n\} = \mu_X \tag{1.80}$$

However, its variance (which can be interpreted as the mean square error of estimation of the mean) is,

$$\begin{aligned}\mathbb{E}\{|S_N-\mu_X|^2\} =& E\left\{\left|\frac{1}{2N+1}\sum_{n=-N}^{N}(X_n-\mu_X)\right|^2\right\} = \\ =& \frac{1}{(2N+1)^2}\sum_{n=-N}^{N}\sum_{m=-N}^{N}\mathbb{E}\{(X_n-\mu_X)(X_m^*-\mu_X^*)\} \\ =& \frac{1}{(2N+1)^2}\sum_{n=-N}^{N}\sum_{m=-N}^{N}C_X[n-m]\end{aligned} \tag{1.81}$$

If the process is independent, $C_X[n-m] = \sigma_X^2\delta[n-m]$ and we can write

$$\mathbb{E}\{|S_N-\mu_X|^2\} = \frac{1}{2N+1}\sigma_X^2 \tag{1.82}$$

Thus S_N is a random variable whose variance asymptotically reduces to zero as N increases. This indicates that even having a single realization of the process, averaging a good number of observations, it is possible to obtain a good estimate of its mean. This proves the following

Theorem 2 (Ergodicity of an independent WSS process)
If X_n is an independent WSS process with finite variance, then it is mean-ergodic. In particular, IID processes with finite variance are mean-ergodic.

The independence is not a necessary condition for mean-ergodicity. Noting that (1.81) is a sum in which the value $C_X[k]$, with $k \geq 0$, occurs $2N+1-k$ times (from $n = k-N$, $m = -N$ to $n = N$, $m = N-k$) and the values $C[k]$ with $k < 0$ appear $2N+1+k$ times, we can write (1.81) as

$$\begin{aligned}\mathbb{E}\{|S_N-\mu_X|^2\} =& \frac{1}{(2N+1)^2}\sum_{k=-2N+1}^{2N-1}(2N+1-|k|)C_X[k] \\ =& \frac{1}{(2N+1)^2}\sum_{k=-2N+1}^{2N-1}\left(1-\frac{|k|}{2N+1}\right)C_X[k]\end{aligned} \tag{1.83}$$

Thus a WSS process is mean ergodic if (1.83) converges to 0 as N tends to infinity. For instance, it is not difficult to prove the all processes satisfying

$$\sum_{k=-\infty}^{\infty}|C_X[k]| < \infty \tag{1.84}$$

satisfy the above condition, and are, thus, mean-ergodic.

Example 16 Consider the stationary process given by $X_n = Y_n + 0.8Y_{n-1}$, where Y_n is a Gaussian IID process of unit variance. It is easy to see that X_n is a stationary process with zero mean and

$$\begin{aligned} C_X[k] =& \mathbb{E}\{(Y_n + 0.8Y_{n-1})(Y_{n+k} + 0.8Y_{n+k-1})\} \\ =& 1.64\delta[k] + 0.8\delta[k-1] + 0.8\delta[k+1] \end{aligned} \tag{1.85}$$

Since the autocovariance vanishes for $|k| > 1$,the process is ergodic in the mean. Figure 1.4 represents the first samples (from $n = 0$ to $n = 50$) of 4 realizations. The average of 50 realizations is represented in the lower part. The 50 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.

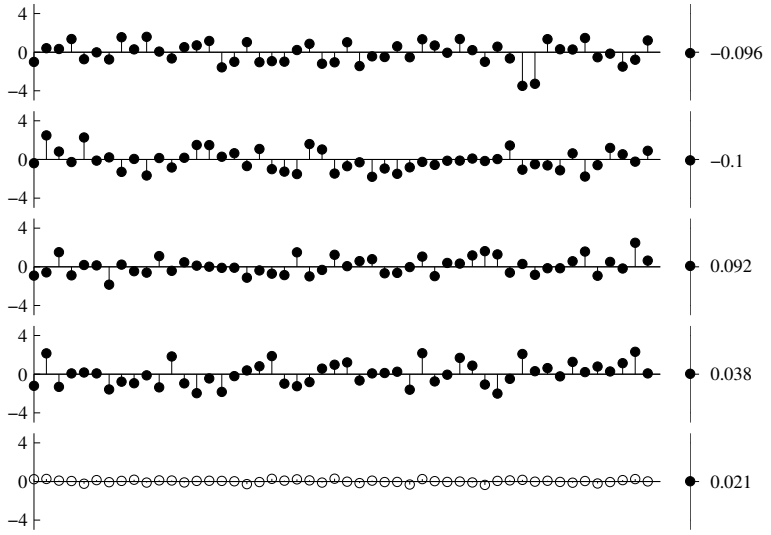

Fig. 1.4 4 realizations of the stochastic process of the example 16 (top), the average of 50 realizations (bottom) and the averages of the represented samples (right). Because of the stationarity, the average signal approaches a constant of value equal to the mean of the process. Because of the ergodicity, time averages also approximate the mean.

1.5.3.1 Ergodicity in the autocorrelation

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

$$R_N = \frac{1}{2N+1} \sum_{m=-N}^{N} X_{m+n} X_m^* \tag{1.86}$$

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.

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 ouput proces Y_n is given by

$$Y_n = X_n * h[n] \tag{1.87}$$

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

$$\begin{aligned}\mu_Y[n] =& \mathbb{E}\{Y_n\} = \mathbb{E}\{X_n * h[n]\} = E\left\{\sum_{k=-\infty}^{\infty} h[k] * X_{n-k}\right\} \\ =& \sum_{k=-\infty}^{\infty} h[k] * \mathbb{E}\{X_{n-k}\} = h[n] * \mathbb{E}\{X_n\} \\ =& h[n] * \mu_X[n]\end{aligned} \tag{1.88}$$

The cross-correlation of the input and output is

$$\begin{aligned}R_{YX}[n_1,n_2] =& \mathbb{E}\{Y_{n_1}X^*_{n_2}\} = E\left\{X^*_{n_2}\sum_{k=-\infty}^{\infty} X_k h[n_1-k]\right\} \\ =& \sum_{k=-\infty}^{\infty} h[n_1-k]R_X[k,n_2] \\ =& h[n_1] * R_X[n_1,n_2]\end{aligned} \tag{1.89}$$

Finally, the autocorrelation of the output is

$$\begin{aligned}R_Y[n_1,n_2] =& \mathbb{E}\{Y_{n_1}Y^*_{n_2}\} = E\left\{Y_{n_1}\sum_{k=-\infty}^{\infty} h^*[k]X^*_{n_2-k}\right\} \\ =& \sum_{k=-\infty}^{\infty} h^*[k]R_{YX}[n_1,n_2-k] \\ =& R_{YX}[n_1,n_2] * h^*[n_2]\end{aligned} \tag{1.90}$$

Combining (1.89) and (1.90), we get

$$R_Y[n_1,n_2] = h[n_1] * R_X[n_1,n_2] * h^*[n_2] \tag{1.91}$$

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 (1.88) and (1.91) 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.

If X_n is stationary, $\mu_X[n] = \mu_X$, and

$$\mu_Y = \mu_X \sum_{n=-\infty}^{\infty} h[n] \tag{1.92}$$

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

$$R_Y[n_1+m, n_2+m] = h[n_1] * R_X[n_1+m, n_2+m] * h^*[n_2] \tag{1.93}$$

If X_n is stationary $R_X[n_1+m, n_2+m] = R_X[n_1, n_2]$, then

$$R_Y[n_1+m, n_2+m] = R_Y[n_1, n_2] \tag{1.94}$$

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

$$\begin{aligned} R_{YX}[n_1, n_2] =& h[n_1] * R_X[n_1 - n_2] = \sum_{k=-\infty}^{\infty} h[n_1 - k] R_X[k - n_2] \\ =& \sum_{k=-\infty}^{\infty} h[n_2 + n - k] R_X[k - n_2] = \sum_{k=-\infty}^{\infty} h[n-k] R_X[k] \\ =& h[n] * R_X[n] \end{aligned} \tag{1.95}$$

Finally, the autocorrelation of the output is

$$\begin{aligned} R_Y[n] =& R_Y[n_1, n_2] = R_{YX}[n_1, n_2] * h^*[n_2] \\ =& \sum_{k=-\infty}^{\infty} h^*[k] R_{YX}[n_1 - n_2 + k] = \sum_{k=-\infty}^{\infty} h^*[k] \sum_{\ell=-\infty}^{\infty} h[n+k-\ell] R_X[\ell] \\ =& \sum_{\ell=-\infty}^{\infty} R_X[\ell] \sum_{k=-\infty}^{\infty} h^*[k] h[n+k-\ell] = R_X[n] * h[n] * h^*[-n] \\ =& R_X[n] * r_h[n] \end{aligned} \tag{1.96}$$

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

Example 17 Let X_n be a unit variance white noise process. The stationary process Y_n resulting from passing X_n through a linear and invariant and causal filter, given by the difference equation

$$Y_n = 0.6 \cdot Y_{n-1} + X_n \tag{1.97}$$

it is stationary. According to (1.88), since the impulse response of this system is $h[n] = 0.6^n u[n]$, the process mean will be,

$$\mu_Y[n] = h[n] * \mu_X[n] = 0 \tag{1.98}$$

and its autocorrelation

$$R_Y[n] = R_X[n] * h[n] * h[-n] = \frac{0.6^{|n|}}{0.64} \tag{1.99}$$

1.5.5 Gaussian processes

Definition 17 (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 $\{n_0, \ldots, n_{N-1}\}$, the pdf of order N,

$$p_{X_{n_0},\ldots,X_{n_{N-1}}}(x_0, \ldots, x_{N-1}) \tag{1.100}$$

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. (1.88) and (1.116).

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

$$X_{N,n} = \begin{cases} X_n, & \text{if } |n| \leq N \\ 0, & \text{if } |n| > N \end{cases} \tag{1.101}$$

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

$$X_N\left(e^{j\omega}\right) = \sum_{n=-\infty}^{\infty} X_{N,n} e^{-j\omega n} = \sum_{n=-N}^{N} X_n e^{-j\omega n} \tag{1.102}$$

is, in turn, a stochastic process of mean

$$\mathbb{E}\{X_N\left(e^{j\omega}\right)\} = \sum_{n=-N}^{N} \mu_X[n]e^{-j\omega n} \tag{1.103}$$

and squared mean

$$\mathbb{E}\left\{\left|X_N\left(e^{j\omega}\right)\right|^2\right\} = E\left\{\left|\sum_{n=-N}^{N} X_n e^{-j\omega n}\right|^2\right\} \tag{1.104}$$

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 (1.104) 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 18 (Power Spectral Density) The power spectal density (PSD) of a stochastic process X_n is defined as

$$S_X(e^{j\omega}) = \lim_{N\to\infty} \frac{1}{2N+1}\mathbb{E}\{|X_N(e^{j\omega})|^2\} \tag{1.105}$$

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

$$S_X\left(e^{j\omega}\right) \geq 0 \tag{1.106}$$

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

Our goal now is to determine what relationship exists between the PSD as defined and the measures we have used to characterize stochastic processes, and more specifically, the relationship of the PSD with the autocorrelation function. Developing the expression (1.104)

$$\begin{aligned}\mathbb{E}\{|X_T(e^{j\omega})|^2\} =& E\left\{\sum_{n_1=-N}^{N} X_{n_1}e^{-j\omega n_1}\sum_{n_2=-N}^{N} X^*_{n_2}e^{j\omega n_2}\right\}\\ =& \sum_{n_1=-N}^{N}\sum_{n_2=-N}^{N}\mathbb{E}\{X_{n_1}X^*_{n_2}\}e^{-j\omega(n_1-n_2)}\\ =& \sum_{n_1=-N}^{N}\sum_{n_2=-N}^{N} R_X[n_1,n_2]e^{-j\omega(n_1-n_2)}\end{aligned} \tag{1.107}$$

thus

$$S_X\left(e^{j\omega}\right) = \lim_{N\to\infty}\frac{1}{2N+1}\sum_{n_1=-N}^{N}\sum_{n_2=-N}^{N} R_X[n_1,n_2]e^{-j\omega(n_1-n_2)} \tag{1.108}$$

and, making the change of variable $n = n_1 - n_2$, we get

$$S_X\left(e^{j\omega}\right) = \lim_{N\to\infty}\frac{1}{2N+1}\sum_{n_2=-N}^{N}\sum_{n=-N-n_2}^{N-n_2} R_X[n_2+n,n_2]e^{-j\omega n} \tag{1.109}$$

Under certain restrictions on the autocorrelation function, we can simplify this expression until we arrive at

$$S_X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} \overline{R}_X[n]e^{-j\omega n} \tag{1.110}$$

where

$$\overline{R}_X[n] = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} R_X[k+n,k] \tag{1.111}$$

This result is known as the Wiener-Khinchine Theorem.

We will now analyze this result for the stationary processes.

1.5.6.1 Stationary processes.

If the process X_n is WSS, $R_X[k+n,k] = R_X[n]$ and, therefore, $\overline{R}_X[n] = R_X[n]$, so that

$$S_X\left(e^{j\omega}\right) = \sum_{n=-\infty}^{\infty} R_X[n]e^{-j\omega n} \tag{1.112}$$

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

According to this definition, note that

$$R_X[n] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(e^{j\omega})e^{j\omega n} d\omega \tag{1.113}$$

so that

$$\mathbb{E}\{|X_n|^2\} = R_X[0] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X\left(e^{j\omega}\right) d\omega \tag{1.114}$$

Example 18 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.

$$S_X(e^{j\omega}) = \sigma_X^2 \tag{1.115}$$

1.5.6.2 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 (1.91), we can write

$$R_Y[n] = R_X[n] * h[n] * h^*[-n] \tag{1.116}$$

and applying the convolution property to (1.116), we get

$$S_Y\left(e^{j\omega}\right) = S_X\left(e^{j\omega}\right) \left|H\left(e^{j\omega}\right)\right|^2 \tag{1.117}$$

References

1. Kempthorne, Peter, Choongbum Lee, Vasily Strela, and Jake Xia. "Topics in mathematics with applications in finance." Massachusetts Institute of Technology: MIT OpenCouseWare), 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`).