

## 0 Introduction to Stochastic Processes

After introducing the basic elements we are going to work with, we can present the really interesting objects that are going to make use of the previously introduced building blocks: **stochastic process**. This chapter provides an introduction to stochastic processes and some of their properties.

### 0 Central Limit Theorem and Law of Large Numbers

Before actually defining stochastic processes, it is fundamental to introduce two key results in probability theory that are essential for understanding the behavior of stochastic processes: the **Central Limit Theorem** and the **Law of Large Numbers**. These theorems provide insights into how random variables behave when aggregated over time or across many trials.

#### 0 Sequences of Sums of Random Variables

Let us consider a sequence of random variables  $X_1, X_2, X_3, \dots$  and the **sum** of their first  $n$  elements. As we have seen in the previous chapter, this is again a random variable:

$$S_n = X_1 + X_2 + \dots + X_n$$

We should already be familiar with the following facts:

- if all random variables  $X_i$  have a **common mean**  $\mu = \mathbb{E}[X_i]$ , then the expected value of the sum is given by:  $\mathbb{E}[S_n] = n\mu$
- if all random variables  $X_i$  are **independent** and have a **common variance**  $\sigma^2 = \text{Var}\{X_i\}$ , then the variance of the sum is given by:  $\text{Var}\{S_n\} = n\sigma^2$

These two facts combined give us an idea that the distribution of these incremental sums  $S_n$  is normal (because it comes from the sum of normal random variables) with mean  $n\mu$  and variance  $n\sigma^2$ :  $S_n \sim \mathcal{N}(0, n)$ .

Additionally, for each  $n$ , we know that these variables are **not independent**, in fact,  $S_{n+1} = S_n + X_{n+1}$ , so we have that:

$$\mathbb{E}[S_{n+1} | S_n] = S_n, \text{Var}\{X_{n+1} | S_n\} = 1 \implies S_{n+1} | S_n \sim \mathcal{N}(S_n, 1)$$

The dependence between the variables creates a very different structure of the observed sequences (or **paths**) compared to what we would observe if the variables were independent. We can visualize this by simulating some realizations from the sequence of sums  $S_n$ ,  $n = 1, 2, \dots$  and comparing them with realizations of a sequence  $W_n \sim \mathcal{N}(0, n)$ .

To start off we try to produce a single realization of the first  $N = 1000$  variables in the initial i.i.d.  $X_n \sim \mathcal{N}(0, 1)$  sequence:

```
1 set.seed(42) # set random number generator seed
2 N = 1000 # number of random variables to simulate
3 X = rnorm(N) # simulate N i.i.d. standard normal r.v.'s
```

We can observe the results of this simulation by the command `head(X)` which provides the first few values of the simulated vector  $X$ . Alternatively it is possible to visualize the **scatter plot** of the simulated values. We prefer to visualize the results using a scatter plot to remind us of *individual realizations*, however, to represent the idea of a single random function  $X(n) = X_n$  changing value through time, we can also use a line plot `plot(x, type='l', xlab='n', ylab='expression(X[n])')`. This comparison is shown in Figure 0.1.

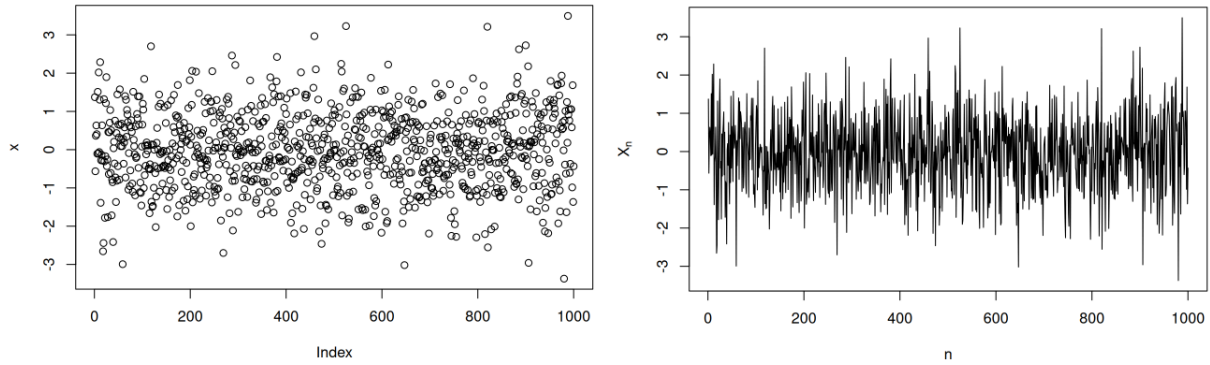


Figure 0.1: Left: Scatter plot of a single realization of  $X_n \sim \mathcal{N}(0, 1)$ . Right: Line plot of the same realization of  $X_n \sim \mathcal{N}(0, 1)$ . Both plots represent the same experiment run 1000 times.

It is important to remember that random variables are functions of the results of an experiment. The sequence we just produced can be seen as the initial 1000 outcomes of an infinite sequence corresponding to a single observed experiment result  $\omega \in \Omega$ . We may try to repeat the experiment multiple times, say 20, obtaining each time a different observed sequence, or path, of this process. This is done through the function `replicate` in R in the following code whose results are illustrated in Figure 0.2:

```

1 N = 1000
2 m = 20 # number of realizations
3 x_paths = replicate(m, rnorm(N))
4 plot(x_paths[,1], type='l', xlim=c(0,N), ylim=range(x_paths),
5      xlab='n', ylab=expression(X[n])) # plot the first realization
6
7 for (i in 2:m) # plot the remaining on the existent canvas
8   lines(x_paths[,i], col=i)

```

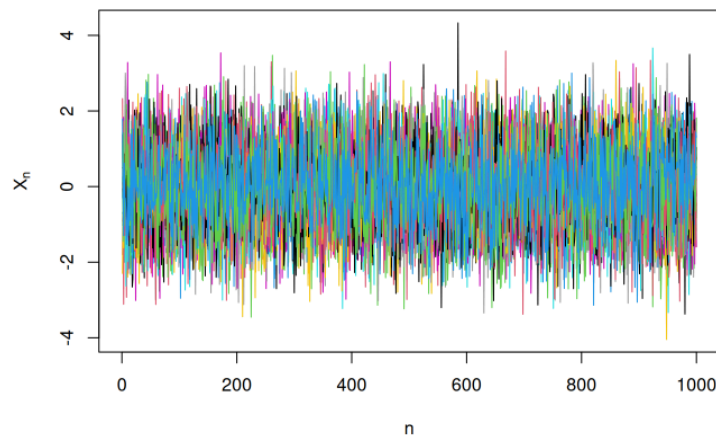


Figure 0.2: 20 realizations of the sequence  $X_n \sim \mathcal{N}(0, 1)$ .

From Figure 0.2 we can notice how all the paths exhibit a similar behavior. This is due to the independence of the individual  $X_n$  variables. These sequences of i.i.d. standard normal random variables are a mathematical representation of what is called **white noise** in *time series analysis*.

Now that we have defined the realizations of all the random variables we are going to work with, we can proceed to create the sums  $S_n$ . To start off, let's create a single realization of the sums  $S_n$  for  $n = 1, \dots, N$ :

```

1 set.seed(42)
2 N = 1000
3 S = rnorm(1)
4 for (n in 2:N)
5   S = c(S1, rnorm(1) + S[n-1])

```

```

1 set.seed(42)
2 N = 1000
3 S = cumsum(rnorm(N))

```

These codes are actually identical and produce the same resulting vector  $S$  which contains a single realization of the cumulative sum  $S_n = \sum_{i=1}^n X_i$  for  $n = 1, \dots, N$ . The only difference is that the code on the right uses much more efficient functions and should be preferred when working at large scale.

Following we illustrate the results of simulating a single realization of the sums  $S_n$  via the command `plot(S, type='l', xlim=c(0,N), ylim=max(abs(S))*c(-1,1))` in Figure 0.3.

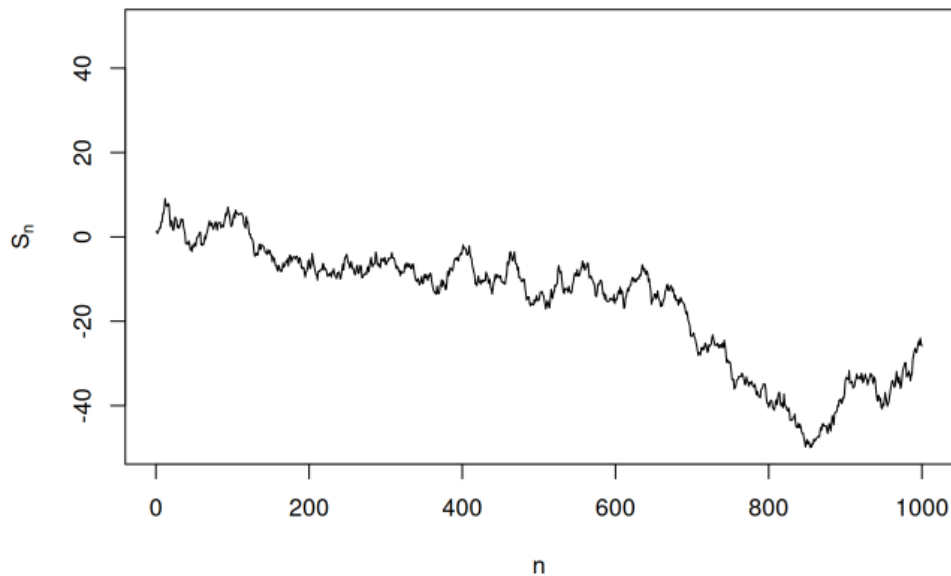


Figure 0.3: Single realization of the sum  $S_n = \sum_{i=1}^n X_i$  for  $n = 1, \dots, N$ .

In a very similar way as before, we can produce multiple realizations of the sums  $S_n$  by using the `replicate` function in R. This is shown in the following code:

```

1 set.seed(42)
2 N = 1000
3 m = 20 # number of realizations
4 S_paths = replicate(m, cumsum(rnorm(N)))
5
6 # plot the first realization to create the canvas
7 plot(S_paths[,1], type='l', xlim=c(0,N), ylim=range(S_paths),
8      xlab='n', ylab=expression(S[n]))
9
10 # plot the remaining realizations on the existent canvas
11 for (i in 2:m)
12   lines(S_paths[,i], col=i)

```

The results of this code are shown in Figure 0.4.

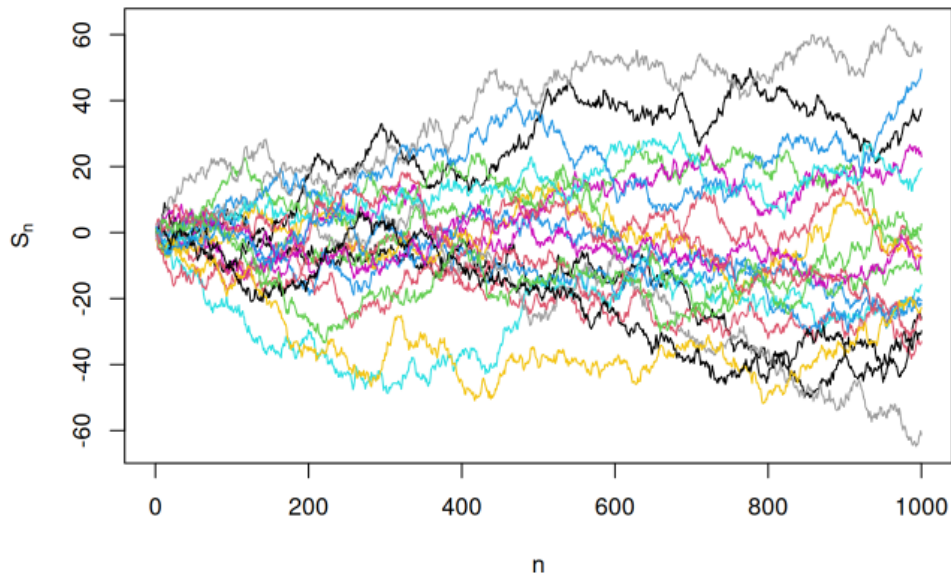


Figure 0.4: 20 realizations of the sum  $S_n = \sum_{i=1}^n X_i$  for  $n = 1, \dots, N$ .

We can notice that all these paths exhibit a similar “wandering” behavior, indeed each step only depends on the previous one plus a new independent standard normal random value, which will make the trajectory move up or down with equal probability. All values are centered around the same mean (0) but as the value of  $n$  increases we see that the paths tend to spread out more and more.

This means that we are actually able to describe the distribution of  $S_n$  for each fixed  $n$  characterizing it as a normal random variable with mean 0 and variance  $n$ :  $S_n \sim \mathcal{N}(0, n)$ . Even though we noticed that  $X_{n+1}$  is independent of  $S_n$ , the newly obtained sum  $S_{n+1}$  is **not independent** of  $S_n$ , in fact we have that  $S_{n+1} \mid S_n \sim \mathcal{N}(S_n, 1)$ . This may not look obvious at first, but it is a crucial point. To realize this, let’s have a look at what would happen if we tried to simulate  $N$  realizations of  $n$  independent random variables each with distribution  $\mathcal{N}(0, n)$ :

```
1 set.seed(42)
2 N=1000
3 W = rnorm(N, 0, sqrt(c(1:N)))
4 plot(W, type='l', xlab("n", ylab=expression(W[n])))
```

This code produces a single realization of  $N$  independent random variables  $W_n \sim \mathcal{N}(0, n)$  for  $n = 1, \dots, N$ . The results of this simulation are shown in Figure 0.5.

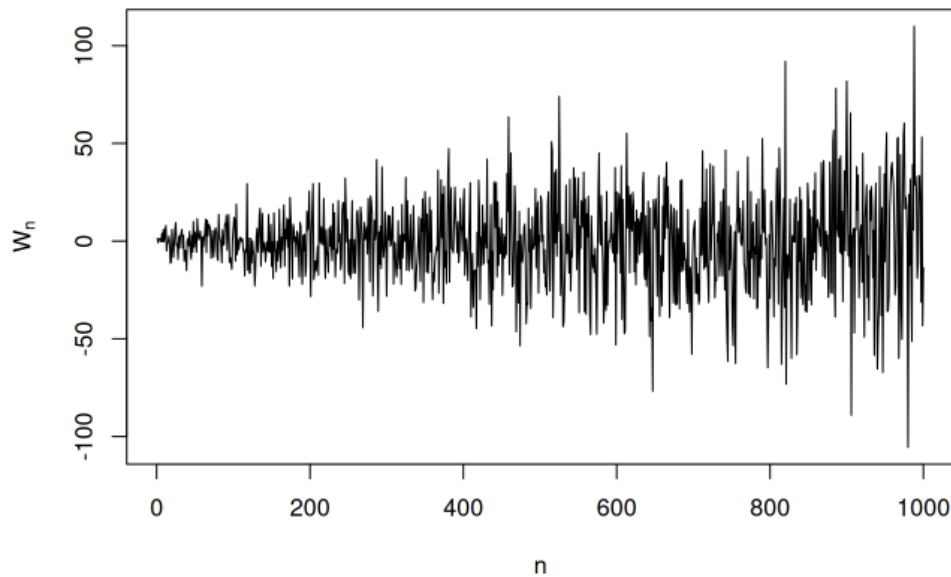


Figure 0.5: Realization of  $N$  independent random variables  $W_n \sim \mathcal{N}(0, n)$  for  $n = 1, \dots, N$ .

We can see how a single realization of this process produces a much more erratic behavior than any of the individual paths of dependent sums in Figure 0.4. Because the marginal distributions coincide, we can still see how the observed values  $w_n$  tend to get further away from the mean as  $n$  grows, but they are able to span the whole range of values in a single realization, while we needed  $m = 20$  realizations of the dependent processes to produce a similar coverage of the value range.

This shows the importance of the dependence between the  $S_n$  variable. Singularly, for each  $n$ , we have that  $S_n$  and  $W_n$  have the same marginal distribution, but within a single path the  $W_n$  are free to vary independently, while the  $S_n$  are constrained by their previous values.

## 0 Law of Large Numbers

We can notice that as we make  $n$  grow to infinity, both  $S_n$  and  $W_n$  will tend to diverge to infinity as well, because their variance grows without bound. Since the variance tends to grow linearly with  $n$  we can try to normalize these variables by dividing each sum by  $n$ . This produces a new sequence of random variables:

$$\bar{X}_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^n X_i \quad 0.1$$

Basically for each  $n$  we are computing the **sample mean** of the sample  $X_1, \dots, X_n$ . Let's look at the code to produce directly the same number of realizations of this new sequence of random variables as the previous example in Figure 0.4:

```
1 # reproduce same experimental setup
2 set.seed(42)
3 N = 1000
4 m = 20
5
6 sm_paths <- replicate(m, cumsum(rnorm(N)) / (1:N))
7
8 # plot the first realization to create the canvas
9 plot(sm_paths[,1], type='l', xlim=c(0,N), ylim=range(sm_paths),
10      xlab='n', ylab=expression(bar(X)[n]))
```

```

11
12 for (i in 2:m)
13   lines(sm_paths[,i], col=i)

```

Following we illustrate the results of this code in Figure 0.6.

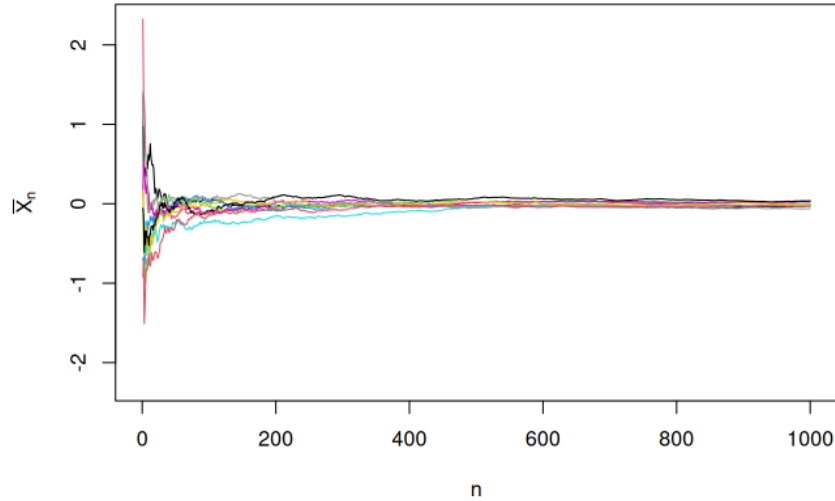


Figure 0.6: 20 realizations of the sample mean sequence  $\bar{X}_n = \frac{1}{n}S_n$

We can notice how now we have managed to control the variability of the sequence, but perhaps we did it too much. In fact all realizations seem to converge to have variance 0 quite fast. This is by no surprise since we have the following result

$$\text{Var}\{\bar{X}_n\} = \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0$$

Basically as  $n$  grows the mean remains fixed to 0, but the variance shrinks to 0 as well, making all realizations converge to the same constant value (the mean).

#### Theorem 1 (Law of Large Numbers)

Let  $X_1, X_2, \dots, X_n$  be a sequence of independent, identically distributed random variables with finite mean  $\mu = \mathbb{E}[X_i]$  and finite variance  $\sigma^2 = \text{Var}\{X_i\}$ . Then, the sample mean sequence defined in Equation 0.1 converges in probability to the true mean  $\mu$  as  $n$  approaches infinity:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{n \rightarrow \infty} \mu$$

#### 0 Central Limit Theorem

The Law of Large Number is a very important result that guarantees that the sample mean converges to the true mean as the sample size increases. However, it does not preserve any type of randomness in the limit. We would now like to control the variance of the sequence without making it vanish completely; in order to do so, we can try to divide the sums by a smaller factor than  $n$ , for instance  $\sqrt{n}$  (which is exactly the standard deviation of  $S_n$ ).

#### Remark

Dividing each sum by  $\sqrt{n}$  is equivalent to **standardizing** the sums:

$$Z_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}\{S_n\}}} = \frac{1}{\sqrt{n}} S_n$$

Following we illustrate the code to produce multiple realizations of this standardized sequence:

```

1  set.seed(42)
2  N = 1000
3  m = 20
4
5  z_paths -> replicate(m, cumsum(rnorm(N)) / sqrt(1:N))
6
7  # plots
8  plot(z_paths[,1], type='l', xlim=c(0,N), ylim=range(z_paths),
9       xlab='n', ylab=expression(Z[n]))
10 for (i in 2:m)
11   lines(z_paths[,i], col=i)

```

Figure 0.7 shows the results of this simulation, where we can notice how this time we have generated a sequence of identically distributed standard normal random variables.

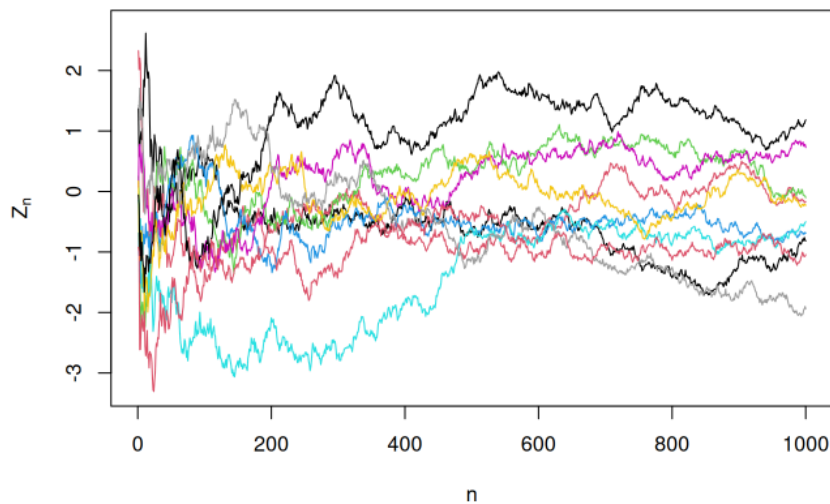


Figure 0.7: 20 realizations of the standardized sequence  $Z_n = \frac{1}{\sqrt{n}} S_n$

Even though all variables  $Z_n$  are identically distributed as standard normal random variables, they are **not independent** as we can notice from the regularity of the individual paths. We can actually visualize the marginal distribution of some of the  $Z_n$  variables using a histogram, for different values of  $n$ :  $n = 1, n = \frac{N}{2}, n = N$ . To effectively visualize the histograms we are going to simulate a larger number of realizations, say  $m = 1000$ . This is shown in Figure 0.8.

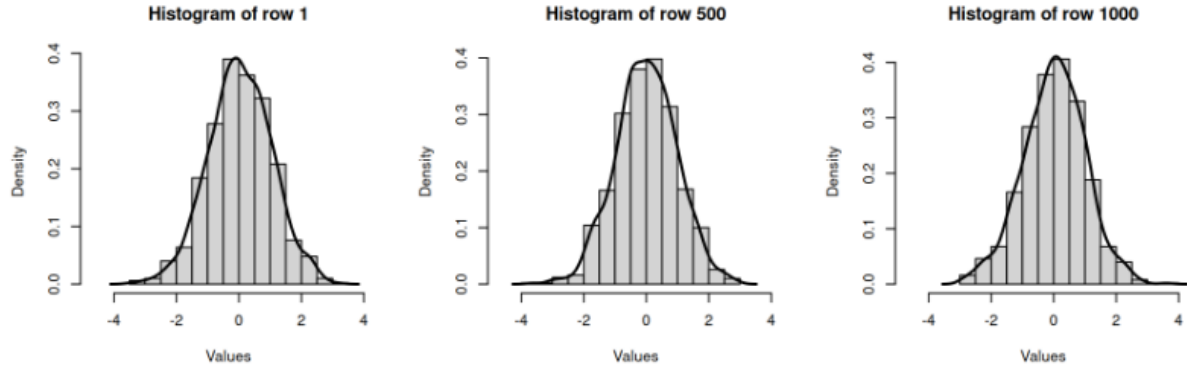


Figure 0.8: Histograms of the marginal distributions of  $Z_n$  for different values of  $n$ . From left to right:  $n = 1$ ,  $n = \frac{N}{2}$ ,  $n = N$ .

### Theorem 2 (Central Limit Theorem)

Let  $X_1, X_2, \dots, X_n$  be a sequence of independent, identically distributed random variables with finite mean  $\mu = \mathbb{E}[X_i]$  and finite variance  $\sigma^2 = \text{Var}\{X_i\}$  and let

$$S_n = \sum_{i=1}^n X_i = X_1 + X_2 + \dots + X_n$$

As  $n \rightarrow \infty$ , the **standardized sum** defined as

$$Z_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}\{S_n\}}} = \frac{S_n - n\mu}{\sigma} \sqrt{n}$$

converges in distribution to a **standard normal random variable**, that is:

$$F_{Z_n}(z) = \mathbb{P}\left[\frac{S_n - n\mu}{\sigma} \sqrt{n} \leq z\right] \rightarrow \Phi(z) \quad \forall z \quad 0.2$$

This is quite straightforward to understand in this case, since we have started from standard normal random variables  $X_i$  and summed them up to obtain normal random variables  $S_n$ . However, the **central limit theorem** holds in much more general settings, in fact it holds even when the original distribution of the sample  $X_n$  is not normal, but anything else. In such case the distribution of  $Z_n$  will no longer be exactly normal but will **converge** to it. This holds also in case of the Law of Large Numbers: regardless of the original distribution of the  $X_n$  variables, the sample mean will always converge to the true mean if the  $X_n$  sample has a common mean  $\mu$ .

## 0 Introduction to Stochastic Processes

After providing two of the most important results in probability theory, it's finally time to introduce the long awaited **stochastic processes**.

### Definition 1 (Stochastic Process)

A **stochastic process** is a random variable that also depends on **time**. It is therefore a function of *two arguments*  $X(t, \omega)$ , where:

- $t \in \mathcal{T}$  is time, with  $\mathcal{T}$  being a set of possible times, usually  $[0, \infty)$ ,  $(-\infty, \infty)$ ,  $\{0, 1, 2, 3, \dots\}$  or  $\{\dots, -2, -1, 0, 1, 2, \dots\}$
- $\omega \in \Omega$  is the outcome of a random experiment, with  $\Omega$  being the whole sample space

In this context, the values of  $X(t, \omega)$  are called the **states** of the process.

The one we have just introduced in Definition 1 is actually one of the two possible ways we have at our disposal for defining a stochastic process. At any fixed time  $t$  we have a random variable  $X_{t(\omega)}$ , a function of a random outcome; on the other hand, if we fix the outcome  $\omega$ , we obtain a function of time  $X_{\omega(t)}$ . This function is called a **realization** or **sample path** or a **trajectory** of the process  $X = \{X(t) : t \in \mathcal{T}\}$ .

### Definition 2 (Stochastic Process - Alternative Definition)

For a given sample space  $\mathcal{S}$  of some experiment, a **random process** is any rule that associates a time-dependent function with each outcome in  $\mathcal{S}$ . Any such function that may result is a **sample function** of the random process. The collection of all possible sample function is called the **ensemble** of the random process.

A stochastic process may be classified according to the nature, either of the outcomes of the experiment or of the time parameter in the following way:

- $X(t, \omega)$  is **discrete-state** if the variable  $X_{t(\omega)}$  is **discrete** for each time  $t$ , and it is **continuous-state** if  $X_{t(\omega)}$  is **continuous** for each time  $t$ .
- $X(t, \omega)$  is a **discrete-time** process if the set of times  $\mathcal{T}$  is discrete, that is, it consists of separate isolated points. It is a **continuous-time** process if the set of times  $\mathcal{T}$  is a connected, possibly unbounded, interval of the real line.

### Example: Example of Stochastic Process

A communication system uses phase-shift keying to transmit information. A quaternary phase-shift keying system can transmit four distinct symbols (often used to encode two bits at a time). The four symbols are distinguished by varying the phase at which they are transmitted; specifically for  $k \in \{1, 2, 3, 4\}$ , the  $k$ -th symbol is transmitted for  $T$  seconds with the following wave form:

$$x_{k(t)} = \cos\left(2\pi f_0 t + \frac{\pi}{4} + k\frac{\pi}{2}\right), \quad 0 \leq t \leq T$$

for some determined frequency  $f_0$ . Considering the transmission of a single randomly selected symbol, and letting  $X(t)$  denoting the corresponding transmitted wave we obtain the graph in Figure 0.9.

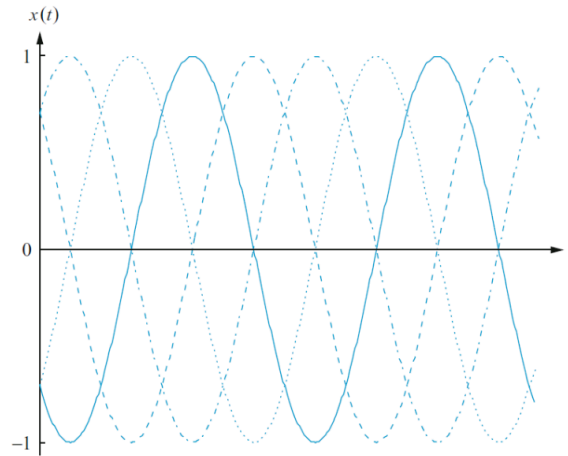


Figure 0.9: Ensemble of a continuous-time, continuous-state stochastic process

Notice how this scenario is also quite strange: whenever we fix a time, the support of the random variable  $X(t)$  may change, indeed for some times  $t$  the possible values of  $X(t)$  are only two, while for other times they are four. This is because at some times the different waveforms may overlap.

At any fixed time point  $t_0$ , the ensemble of a random process  $X(t)$  forms a probability distribution:  $X(t_0)$  is a random variable with support determined by the ensemble. Just like when we discussed vectors of random variables, we did not limit ourselves to describing only *marginal distribution*, they would not be enough: to fully characterize the behavior of a stochastic process we need to consider also the **joint distributions** of  $X(t_1), \dots, X(t_r)$  for all finite sets of time points  $t_1 < \dots < t_r$ . The collection of all such joint distributions constitutes the **finite-dimensional distributions** of the process.

For the purpose of this course, we are going to deal mainly with continuous-time stochastic processes.

### 0 Mean and Variance Functions

In this section we are going to introduce some of the most important functions that can be used to describe the behavior of a stochastic process. Suppose we have a random process  $X = \{X(t) : t \in \mathcal{T}\}$ . For each fixed  $t \in \mathcal{T}$ ,  $X(t)$  is a random variable, therefore we can define its expected value and variance.

If we keep into account all the random variables we can obtain by varying the fixed time  $t$ , we can then define some functions that describe how the mean and the variance of the random process change over time.

### Definition 3 (Mean and Variance Functions)

The **mean function** of a random process  $X(t)$  is given by

$$\mu_X(t) = \mathbb{E}[X(t)] \quad 0.3$$

where  $\mathbb{E}[X(t)]$  is the expected value of the random variable  $X(t)$  for the fixed time point  $t$ . Similarly we can define the **variance** and **standard deviation functions** of the process as:

$$\begin{aligned} \sigma_{X(t)}^2 &= \text{Var}\{X(t)\} = \mathbb{E}[(X(t) - \mu_X(t))^2] \\ &= \mathbb{E}[X^2(t)] - [\mu_X(t)]^2 \end{aligned} \quad 0.4$$

The **standard deviation function** is simply given by  $\sigma_X(t) = \sqrt{\text{Var}\{X(t)\}}$ .

### Remark

The mean, variance and standard deviation functions are **deterministic** (non random) functions of time  $t$ . This is because normal mean and variances are not random variables, they are simply plain numbers.

### Example: Mean and Variance Functions of a Random Process

An ideal signal has the form  $v_0 + a \cos(\omega_0 t + \theta_0)$  but amplitude variations may occur. We can model this situation with the random process:

$$X(t) = v_0 + A \cos(\omega_0 t + \theta_0)$$

where  $A$  is a random variable whose distribution describes the amplitude variation. This is often modeled as a Rayleigh random variable with parameter  $\sigma_A$  which has mean and variance:

$$\mathbb{E}[A] = \sigma \sqrt{\frac{\pi}{2}} \quad \text{Var}\{A\} = \frac{4 - \pi}{2} \sigma^2$$

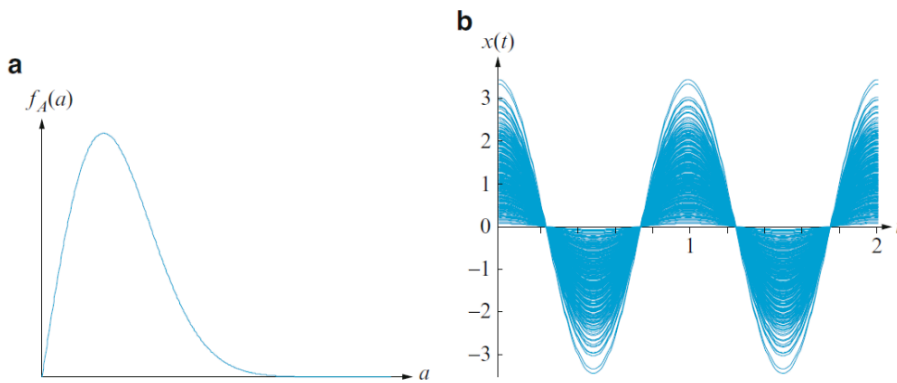


Figure 0.10: a): Rayleigh p.d.f. for  $\sigma_A = 1$ ; b): Ensemble of the random process  $X(t)$  for the parameters  $v_0 = 0$ ,  $\omega_0 = 2\pi$ ,  $\theta_0 = 0$ .

Both the p.d.f. of the Rayleigh random variable  $A$  and the ensemble of the random process  $X(t)$  are shown in Figure 0.10. If we fixed  $t$  we can apply the properties of expected value and variance to find the mean and variance functions of the process  $X$ .

Since, given any fixed  $t$ , we have that  $\cos(\omega_0 t + \theta_0)$  is a **constant**, we can write:

$$\begin{aligned}\mu_X(t) &= \mathbb{E}[X(t)] = \mathbb{E}[v_0 + A \cos(\omega_0 t + \theta_0)] = v_0 + \mathbb{E}[A] \cos(\omega_0 t + \theta_0) \\ \sigma_X^2(t) &= \text{Var}\{X(t)\} = \text{Var}\{A \cos(\omega_0 t + \theta_0)\} = \text{Var}\{A\} \cos^2(\omega_0 t + \theta_0)\end{aligned}$$

We can notice that, once the constants are fixed, both the mean and variance functions are **deterministic functions** that oscillate over time.

### 0 Auto-covariance Function

The mean and variance functions provide information about the behavior of the ensemble at each single point in time. However, they do not provide any information about the **dependence** between the random variables at different time points. Not surprisingly if we pick two distinct time point  $t, s$  we'll have that the random variables  $X(t)$  and  $X(s)$  are generally **related**. To capture this dependence we can use the **auto-covariance function** of the process.

#### Definition 4 (Auto-covariance Function)

The **auto-covariance function** of a random process  $X(t)$  is defined as:

$$C_{XX}(t, s) = \text{Cov}\{X(t), X(s)\} = \mathbb{E}[(X(t) - \mu_X(t))(X(s) - \mu_X(s))] \quad 0.5$$

The auto-covariance function is typically denoted  $\sigma_X(t, s)$  and, when  $t = s$ , we recover the variance function:  $\sigma_X(t, t) = \sigma_X^2(t)$ . Following we outline some properties of the auto-covariance function:

- $C_{XX}(t, s) = C_{XX}(s, t)$ , that is the auto-covariance function is **symmetric**
- $C_{XX}(t, s) = \mathbb{E}[X(t)X(s)] - \mu_X(t)\mu_X(s)$
- $\sigma_X^2(t) = \text{Var}\{X(t)\} = \text{Cov}\{X(t), X(t)\} = C_{XX}(t, t) = \mathbb{E}[X^2(t)] - \mu_X^2(t)$

This function has the same interpretation of the covariance between two random variables. The problem is that covariance does not only contain information about the strength of the dependence between two variables, but also about their variance. To solve this problem we can introduce the **auto-correlation function** of the process.

#### Remark

If we take as a reference Figure 0.10 (b), we can, according to the selection of time points  $t$  and  $s$ , observe different auto-covariance functions:

- if we pick  $t = 0.5, s = 1.5$ , we are going to have a positive auto-covariance since the values of the process at those times tend to vary in the same direction.
- if we pick  $t = 0.5, s = 1.0$ , we are going to have a negative auto-covariance since the values of the process at those times tend to vary in opposite directions.
- if we pick  $t, s$  such that all the ensemble paths cross the  $x$  axis at those times, we are going to have an auto-covariance close (if not equal) to zero. But this **does not mean** that  $X(t) \perp X(s)$

### 0 Auto-correlation Function

For the purpose of this course, we are going to refer to auto-correlation in a very similar way as we did for random variables, just we consider a function of two time points instead of two random variables.

#### Definition 5 (Auto-correlation Function)

Given a random process  $X(t)$ , and two time points  $t, s \in \mathcal{T}$ , the **auto-correlation function** of the process is given by the following equation:

$$\rho_X(t, s) = \sigma_X \frac{t, s}{\sigma_X(t)\sigma_X(s)} \quad 0.6$$

and its interpretation is analogous to that of the correlation between random variables. In particular  $-1 \leq \rho_X(t, s) \leq 1$  indicates the magnitude and the direction of the association between  $X(t) \wedge X(s)$ .

It is worth to mention that this is not the only possible interpretation of auto-correlation. Another common interpretation comes from signal processing. The alternative definition is given by  $R_{XX}(t, s) = \mathbb{E}[X(t)X(s)]$  and it is equivalent to  $\rho_X(t, s)$  only when the mean and variance function are respectively equal to 0 and 1. To see why this is true, let  $X, Y$  be two random variables; the correlation between  $X$  and  $Y$  is given by:

$$\rho_{X,Y} = \frac{\text{Cov}\{X, Y\}}{\sqrt{\text{Var}\{X\} \text{Var}\{Y\}}} = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sqrt{\text{Var}\{X\} \text{Var}\{Y\}}}$$

This can easily be retrieved via [ciaoneeq\\_2\\_covariance](#) and [ciaoneeq:correlation\\_coefficient](#). Let's try to understand when it is the case that  $\rho_{X,Y} = \mathbb{E}[XY]$ : first of all it is necessary that either one of  $\mathbb{E}[X]$  or  $\mathbb{E}[Y]$  is equal to zero; secondly the product of the variances in the denominator must be equal to 1.

Let now  $\tilde{X}, \tilde{Y}$  be two random variables such as  $\mu_{\tilde{X}} = \mathbb{E}[\tilde{X}]$ ,  $\sigma_{\tilde{X}} = \text{std}(\tilde{X})$  and  $\mu_{\tilde{Y}} = \mathbb{E}[\tilde{Y}]$ ,  $\sigma_{\tilde{Y}} = \text{std}(\tilde{Y})$ .

If we put  $X = \frac{\tilde{X} - \mu_{\tilde{X}}}{\sigma_{\tilde{X}}}$  and  $Y = \frac{\tilde{Y} - \mu_{\tilde{Y}}}{\sigma_{\tilde{Y}}}$ , that is, we consider the standardized versions of  $\tilde{X}$  and  $\tilde{Y}$ , we have that the two conditions mentioned before are satisfied, indeed we have  $\mathbb{E}[X] = \mathbb{E}[Y] = 1$  and  $\text{Var}\{X\} = \text{Var}\{Y\} = 1$ ; therefore  $\rho_{X,Y} = \mathbb{E}[XY]$ .

## 0 Stationarity and Weak Stationarity

After defining some of the most important functions that characterize the behavior of a stochastic process, we now introduce a concept which tells us about the “**stability**” of the random process itself. We are referring to **stationarity** and **weak stationarity**.

### Definition 6 (Stationary Stochastic Process)

A random process  $X(t)$  is said to be **strictly stationary** if all of its statistical properties are **invariant** with respect to time. More precisely,  $X(t)$  is stationary if, for any time points  $t_1, \dots, t_r$  and any value  $\tau$ , the joint distribution of  $X(t_1), \dots, X(t_r)$  is the same as the joint distribution of  $X(t_1 + \tau), \dots, X(t_r + \tau)$ .

To verify strict stationarity it is therefore necessary to check that **all** finite-dimensional distributions of the process are invariant w.r.t. time shifts. This definition implies that all statistical properties of the process are also unchanged, since we require all joint distributions to be invariant over time.

Stationarity is very important: granting that the statistical properties of a process will not change over time allows us to make inferences about the future behavior of the process based on its past behavior. The problem of this definition is that it requires the invariance for any time frame

considered, even infinitely large ones. This is often too restrictive for practical applications, where we usually deal with **finite time frames**.

Strict stationarity is also very hard to verify in practice, since it requires the knowledge of all finite-dimensional distributions of the process, along with some knowledge about the characterizing functions (mean, variance, auto-covariance, etc.). To overcome this, we introduce a weaker form of stationarity.

#### Definition 7 (Wide Sense Stationarity)

A random process  $X(t)$  is said to be **weakly stationary** (or **stationary in the wide sense**) if the following two conditions hold:

- The mean function of  $X(t)$ ,  $\mu_X(t)$  is a constant, i.e. there exists a constant  $\mu$  such that  $\mu_X(t) = \mu$  for all  $t \in \mathcal{T}$ .
- The auto-covariance function of  $X(t)$ ,  $C_{XX}(t, s)$  depends only on the time difference  $s - t$

While the first condition of Definition 7 is quite straightforward to understand, the second one may require some further explanation. The idea is that we require that the degree of association between  $X(t)$  and  $X(s)$ , measured by the auto-covariance, depends only on the distance between the times  $s, t$  but not on the position of those times on an absolute scale.

## 0 Markov Processes

In this course we are going to focus on a very specific kind of stochastic process called **Markov process**. In this section we are going to introduce the basic definitions and properties of these processes.

#### Definition 8 (Markov Process)

A stochastic process  $X(t)$  is **Markov** if for any  $t_1 < \dots < t_n < t$  and any sets  $A; A_1, \dots, A_n$  we have that:

$$\begin{aligned} \mathbb{P}[X(t) \in A \mid X(t_1) \in A_1, \dots, X(t_n) \in A_n] \\ = \mathbb{P}[X(t) \in A \mid X(t_n) \in A_n] \end{aligned} \quad 0.7$$

To put this in practical terms, the Markov property states that, knowing the present, there is no additional information from the past that can be used to better predict the future:

$$\mathbb{P}[\text{future} \mid \text{past}, \text{present}] = \mathbb{P}[\text{future} \mid \text{present}]$$

For the future development of a Markov process, only its present state is important, and it does not matter how the process arrived to this state.