Definition 3.1.4. If X and Y are both random variables in L^2, then the covariance of X and Y, written $\mathrm{Cov}(X,Y)$ is defined to be

$$\mathrm{Cov}(X,Y) := \mathbb{E}\left((X-\mu_X)(Y-\mu_Y)\right)$$

where $\mu_X := \mathbb{E}(X)$, $\mu_Y := \mathbb{E}(Y)$. Whenever the covariance of X and Y exists, we define the *correlation* of X and Y to be

$$\mathrm{Corr}(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \tag{$\dagger$}$$

where σ_X is the standard deviation of X, and σ_Y is the standard deviation of Y.

Remark. By *fiat*, $0/0 := 0$ in ($\dagger$). Although this is sinful in calculus, we advanced mathematicians understand that such a decree is permitted as long as we recognize that it is only a *convenience* which allows us to simplify the formula. We need not bother with the extra conditions about dividing by zero. (See Exercise 3.1.20.)

Definition 3.1.5. We say that X and Y are *uncorrelated* if $\mathrm{Cov}(X,Y) = 0$ (or, equivalently, if $\mathrm{Corr}(X,Y) = 0$).

Theorem 3.1.6 (Linearity of Expectation). *Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are (discrete or continuous) random variables with $X \in L^1$ and $Y \in L^1$. Suppose also that $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are both (piecewise) continuous and such that $f \circ X \in L^1$ and $g \circ Y \in L^1$. Then, $f \circ X + g \circ Y \in L^1$ and, furthermore,*

$$\mathbb{E}(f \circ X + g \circ Y) = \mathbb{E}(f \circ X) + \mathbb{E}(g \circ Y).$$

Exercise 3.1.7. Prove the above theorem separately for both the discrete case and the continuous case. Be sure to state any assumptions or theorems from elementary calculus that you use.

Fact. If $X \in L^2$ and $Y \in L^2$, then the following computational formulæ hold:

- $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$;
- $\mathrm{Var}(X) = \mathrm{Cov}(X,X) = \sigma^2$;
- $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Exercise 3.1.8. Verify the three computational formulæ above.

Exercise 3.1.9. Using the third computational formula, and the results of Exercise 1.3.4, quickly show that if $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathrm{Var}(X) = \sigma^2$. Together with Exercise 1.3.4, this is the reason that if $X \sim \mathcal{N}(\mu, \sigma^2)$, we say that X is normally distributed with mean μ and variance σ^2.

Definition 3.1.10. The random variables X and Y are said to be *independent* if $f(x,y)$, the joint density of (X,Y), can be expressed as

$$f(x,y) = f_X(x) \cdot f_Y(y)$$

where f_X is the density of X and f_Y is the density of Y.

Remark. Notice that we have combined the cases of a discrete and a continuous random variable into one definition. You can substitute the phrases *probability mass function* or *probability density function* as appropriate.

The following is an extremely deep, and important, result.

Theorem 3.1.11. *If X and Y are* independent *random variables with $X \in L^1$ and $Y \in L^1$, then*

- *the product XY is a random variable with $XY \in L^1$, and*
- $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$.

Exercise 3.1.12. Using this theorem, quickly prove that if X and Y are independent random variables, then they are necessarily uncorrelated. (As the next exercise shows, the converse, however, is not true: there do exist uncorrelated, dependent random variables.)

Exercise 3.1.13. Consider the random variable X defined by $P(X = -1) = 1/4$, $P(X = 0) = 1/2$, $P(X = 1) = 1/4$. Let the random variable Y be defined as $Y := X^2$. Hence, $P(Y = 0|X = 0) = 1$, $P(Y = 1|X = -1) = 1$, $P(Y = 1|X = 1) = 1$.

- Show that the density of Y is $P(Y = 0) = 1/2$, $P(Y = 1) = 1/2$.
- Find the joint density of (X, Y), and show that X and Y are not independent.
- Find the density of XY, compute $\mathbb{E}(XY)$, and show that X and Y are uncorrelated.

Exercise 3.1.14. Prove Theorem 3.1.11 in the case when both X and Y are continuous random variables.

Exercise 3.1.15. Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are independent, integrable, continuous random variables with densities f_X, f_Y, respectively. Let $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ be continuous functions such that $g \circ X \in L^1$ and $h \circ Y \in L^1$. Prove that $\mathbb{E}\left((g \circ X) \cdot (h \circ Y)\right) = \mathbb{E}(g \circ X)\,\mathbb{E}(h \circ Y)$.

As a consequence of the previous exercise, we have the following very important result.

Theorem 3.1.16 (Linearity of Variance when Independent). *Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are (discrete or continuous) random variables with $X \in L^2$ and $Y \in L^2$. If X and Y are* independent*, then*

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y).$$

It turns out that Theorem 3.1.11 is not quite true when X and Y are not independent. However, the following is a probabilistic form of the ubiquitous Cauchy-Schwarz inequality, and usually turns out to be good enough.

Theorem 3.1.17 (Cauchy-Schwarz Inequality). *Suppose that X and Y are both random variables with finite second moments. That is, $X \in L^2$, and $Y \in L^2$. It then follows that*

- *the product XY is a random variable with $XY \in L^1$, and*
- $(\mathbb{E}(XY))^2 \le \mathbb{E}(X^2)\,\mathbb{E}(Y^2)$, *and*
- $(\operatorname{Cov}(X, Y))^2 \le \operatorname{Var}(X)\ \operatorname{Var}(Y)$.

Exercise 3.1.18. Using the first part of the Cauchy-Schwarz inequality, show that if $X \in L^2$, then $X \in L^1$.

Exercise 3.1.19. Using the second part of the Cauchy-Schwarz inequality, show that if $X \in L^2$, then $X \in L^1$.

Exercise 3.1.20. Using the third part of the Cauchy-Schwarz inequality, you can now make sense of the Remark following Definition 3.1.4. Show that if X and Y are random variables with $\text{Var}(X) = \text{Var}(Y) = 0$, then $\text{Cov}(X, Y) = 0$.

The following facts are also worth mentioning.

Theorem 3.1.21. *If $a \in \mathbb{R}$ and $X \in L^2$, then $aX \in L^2$ and $\text{Var}(aX) = a^2 \text{Var}(X)$. In particular, $\text{Var}(-X) = \text{Var}(X)$.*

Theorem 3.1.22. *If $X_1, X_2, \ldots, X_n$ are L^2 random variables, then*

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}(X_i, X_j).$$

In particular, if $X_1, X_2, \ldots, X_n$ are uncorrelated *L^2 random variables, then*

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i).$$

3.2 Summarizing Stochastic Processes

On the other hand, for a stochastic process $\{X_t\}$, we have infinitely many random variables, and so we have infinitely many means, variances, covariances. These can be summarized with the help of two functions.

Definition 3.2.1. If $\{X_t\}$ is a stochastic process with $X_t \in L^1$ for each t, then the *mean function* (or *trend*) of $\{X_t\}$ is the non-random function $\mu(t) := \mathbb{E}(X_t)$.

Definition 3.2.2. If $\{X_t\}$ is a stochastic process with $X_t \in L^2$ for each t, then the *autocovariance function* of $\{X_t\}$ is the non-random function

$$\gamma(t, s) := \text{Cov}(X_t, X_s) = \mathbb{E}\left((X_t - \mu(t))(X_s - \mu(s))\right).$$

3.3 Weakly Stationary Stochastic Processes

Definition 3.3.1. We call the stochastic process $\{X_t\}$ *second-order (or weakly) stationary* if

- there is a constant μ such that $\mu(t) = \mu$ for all t, and
- $\gamma(t+h, t)$ only depends on h; that is, if $\gamma(t+h, t) = \gamma(h)$ for all t and for all h.

Example 3.3.2 (White Noise). Suppose that $\{X_t\}$ is collection of *uncorrelated* random variables, each with mean 0 and variance σ^2. We say that $\{X_t\}$ is *White Noise.* (The reason for this luminous name will become clear when we discuss the spectrum of a stationary process.) We now verify that $\{X_t\}$ is second-order stationary. First, it is obvious that $\mu(t) = 0$. Second, if $s \neq t$, then the assumption that the collection is uncorrelated implies that $\gamma(t, s) = 0$, $s \neq t$. On the other hand, if $s = t$, then $\gamma(t, t) = \text{Var}(X_t) = \sigma^2$. Thus, $\mu(t) = 0$, and

$$\gamma(t+h, t) = \begin{cases} \sigma^2, & h = 0, \\ 0, & h \neq 0, \end{cases}$$

so that $\{X_t\}$ is, in fact, second-order stationary. We write $\{X_t\} \sim WN(0, \sigma^2)$ to indicate that $\{X_t\}$ is white noise with $\text{Var}(X_t) = \sigma^2$, each t.

Example 3.3.3 (iid Noise). Suppose instead that $\{X_t\}$ is collection of *independent* random variables, each with mean 0 and variance σ^2. We say that $\{X_t\}$ is *iid Noise.* As with white noise, we easily see that iid noise is stationary with trend $\mu(t) = 0$ and

$$\gamma(t+h,t) = \begin{cases} \sigma^2, & h = 0, \\ 0, & h \neq 0. \end{cases}$$

We write $\{X_t\} \sim IID(0, \sigma^2)$ to indicate that $\{X_t\}$ is iid noise with $\text{Var}(X_t) = \sigma^2$, each t.

Remark. With these two examples, we see that two different processes may both have the same trend and autocovariance function. Thus, $\mu(t)$ and $\gamma(t+h, t)$ are NOT always enough to distinguish stationary processes. (However, for stationary *Gaussian* processes they are enough.) For a more substantial example of two different stationary processes having the same second order properties, see Example 5.2.3.

Remark. Our textbook [2] carefully distinguishes between white noise and iid noise. In fact, if $\{X_t\} \sim IID(0, \sigma^2)$, then $\{X_t\} \sim WN(0, \sigma^2)$. However, the converse is not true; see Problem 1.8 in [2].

As a consequence of the second condition for second-order stationarity, we make the following definition.

Definition 3.3.4. Suppose that $\{X_t\}$ is a second-order stationary process. The *autocovariance function (ACVF) at lag h* of $\{X_t\}$ is

$$\gamma(h) := \text{Cov}(X_{t+h}, X_t).$$

The *autocorrelation function (ACF) at lag h* of $\{X_t\}$ is

$$\rho(h) := \frac{\gamma(h)}{\gamma(0)}.$$

Exercise 3.3.5. For a second-order stationary process, show that $\text{Var}(X_t) = \gamma(0)$ for each t.

Note that $\text{Cov}(X_{t+h}, X_t) = \gamma(h)$, and by the exercise above, $\text{Var}(X_{t+h}) = \text{Var}(X_t) = \sigma^2 = \gamma(0)$ so that $\sigma_{t+h} = \sigma_t = \sigma$. Thus, by definition, we find that

$$\text{Corr}(X_{t+h}, X_t) = \frac{\text{Cov}(X_{t+h}, X_t)}{\sigma_{t+h}\sigma_t} = \frac{\gamma(h)}{\sigma\sigma} = \frac{\gamma(h)}{\gamma(0)} = \rho(h).$$

This verifies the second equality in the definition of the ACF in Definition 1.4.3 on page 16 of [2].

3.4 Other Important Examples of Stochastic Processes

In this section, we give some further examples of stochastic processes, and practice computing autocovariance functions. Although these stochastic processes turn out not to be stationary, they are of the upmost importance in the theory of probability and statistics.

Example 3.4.1 (Simple Symmetric Random Walk). Suppose that $\{X_j\}$, $j = 0, 1, \ldots$, is a collection of iid random variables (called the *jumps*) with $X_0 = 0$ and $P(X_j = 1) = P(X_j = -1) = 1/2$, $j \in \mathbb{N}$ (called the *jump density*). Set $S_0 := X_0$, and for $n \in \mathbb{N}$, let S_n be defined recursively by

$$S_n := S_{n-1} + X_n.$$

Then the stochastic process $\{S_n\}$, $n = 0, 1, \ldots$, is called a *simple symmetric random walk.* In order to analyze the second order structure of $\{S_n\}$, we first observe that

$$S_n = \sum_{j=0}^{n} X_j.$$

Since $\mathbb{E}(X_j) = 0$ for all j, using the linearity of expectations, we find that $\mu(n) = \mathbb{E}(S_n) = 0$. As the $\{X_j\}$ are an iid collection of random variables, the linearity of variance can be applied so that

$$\gamma(n,n) = \mathrm{Var}(S_n) = \sum_{j=0}^{n} \mathrm{Var}(X_j) = \sum_{j=0}^{n} \mathbb{E}(X_j^2) = 0 + \sum_{j=1}^{n} 1 = n.$$

since $\mathbb{E}(X_0^2) = 0$ and $\mathbb{E}(X_j^2) = 1$ for all $j \in \mathbb{N}$. We immediately see that it is impossible for $\{S_n\}$ to be stationary since $\gamma(n,n)$ *depends* on n. However, to finish the computation of the autocovariance function, note that for $h \in \mathbb{N}$,

$$\begin{aligned}
\mathrm{Cov}(S_{n+h}, S_n) &= \mathbb{E}\left(S_{n+h}S_n\right) - \mathbb{E}\left(S_{n+h}\right)\mathbb{E}\left(S_n\right) \\
&= \mathbb{E}\left(S_{n+h}S_n\right) \\
&= \mathbb{E}\left(\sum_{j=0}^{n+h} X_j \sum_{k=0}^{n} X_k\right) \\
&= \mathbb{E}\left(\sum_{j=0}^{n} X_j \sum_{k=0}^{n} X_k\right) + \mathbb{E}\left(\sum_{j=n+1}^{n+h} X_j \sum_{k=0}^{n} X_k\right)
\end{aligned}$$

Since $\{X_j\}$ is an independent collection, we have $\mathbb{E}(X_j X_k) = 0$ for $j \neq k$. From this, we see that

$$\mathbb{E}\left(\sum_{j=n+1}^{n+h} X_j \sum_{j=0}^{n} X_j\right) = 0$$

because each possible product contains *different* random variables. Similarly,

$$\mathbb{E}\left(\sum_{j=0}^{n} X_j \sum_{k=0}^{n} X_k\right) = \mathbb{E}\left(\sum_{j=0}^{n} X_j^2 + \sum_{j\neq k,\, 0\leq j,k\leq n}\sum X_j X_k\right) = \mathbb{E}\left(\sum_{j=0}^{n} X_j^2\right) = n.$$

Thus, we find that $\gamma(n+h, n) = n$ for all $h = 0, 1, 2, \ldots$.

Example 3.4.2 (Generalized Random Walk). Here are some generalizations of the simple symmetric random walk.

- The adjective *simple* refers to a random walk whose possible jumps are ± 1 only.
- The adjective *symmetric* refers to a random walk whose jump density is symmetric.
- If $X_0 = 0$ and $P(X_j = 1) = 1 - P(X_j = -1) = p$, $j \in \mathbb{N}$, for some $p \in (0, 1/2) \cup (1/2, 1)$, then S_n is a *simple asymmetric random walk.*
- Suppose that $X_0 = 0$ and $P(X_j = 1) = 1/3$, $P(X_j = -1) = 1/3$, $P(X_j = 0) = 1/3$, $j \in \mathbb{N}$. Then S_n is a *non-simple symmetric random walk.*

- Similarly, if $X_0 = 0$ and $P(X_j = 2) = 1/2$, $P(X_j = -2) = 1/2$, $j \in \mathbb{N}$, then S_n is a *non-simple symmetric random walk.*
- Suppose that $X_0 = 0$ and $P(X_j = 2) = 1/4$, $P(X_j = 1) = 1/4$, $P(X_j = -2) = 1/2$, $j \in \mathbb{N}$. Then S_n is a *non-simple asymmetric random walk.*

In fact, the jumps $\{X_j\}$, $j = 0, 1, \ldots$, can have almost ANY distribution. Suppose that $\{X_j\} \sim IID(0, \sigma^2)$. Let $S_0 = 0$, and for $n \in \mathbb{N}$, define S_n to be

$$S_n := S_{n-1} + X_n = \sum_{j=1}^{n} X_j.$$

Then $\{S_n\}$ is said to be a *random walk* with jump distribution $\{X_j\}$. Furthermore, we find that $\mu(n) = 0$ for all n, and by modifying Example 3.4.1 slightly, $\gamma(n+h, n) = n\sigma^2$. This shows that *no* (non-trivial) random walk is stationary.

Random walks are among the most important, and fundamental, stochastic processes. In fact, they are the building blocks for, and the discrete time analogue of, *Brownian motion*, the *most* important of all stochastic processes.

Example 3.4.3 (Brownian Motion). Consider a collection of random variables $\{B_t\}$, $0 \le t < \infty$, having the following properties:

- $B_0 = 0$,
- for $0 \le s < t \le \infty$, $B_t - B_s \sim \mathcal{N}(0, t - s)$,
- for $0 \le s < t \le \infty$, $B_t - B_s$ is independent of B_s,
- the trajectories $t \mapsto B_t$ are continuous.

The stochastic process $\{B_t\}$, $0 \le t < \infty$, is called *Brownian motion.* It is actually a very deep result that there *exists* a stochastic process having these properties (continuous trajectories is the tough part). One way to prove the existence of Brownian motion is to take an appropriate limit of appropriately scaled simple symmetric random walks. This concept of a *scaling limit* is of fundamental importance in modern probability research.

Remark. The history of Brownian motion is fascinating. In the summer of 1827, the Scottish botanist Robert Brown observed that microscopic pollen grains suspended in water move in an erratic, highly irregular, zigzag pattern. Following Brown's initial report, other scientists verified the strange phenomenon. Brownian motion was apparent whenever very small particles were suspended in a fluid medium, for example smoke particles in air. It was eventually determined that finer particles move more rapidly, that their motion is stimulated by heat, and that the movement is more active when the fluid viscosity is reduced.

However, it was only in 1905 that Albert Einstein, using a probabilistic model, could provide a satisfactory explanation of the Brownian motion. He asserted that the Brownian motion originates in the continual bombardment of the pollen grains by the molecules of the surrounding water, with successive molecular impacts coming from different directions and contributing different impulses to the particles. As a result of the continual collisions, the particles themselves had the same average kinetic energy as the molecules. Thus, he showed that Brownian motion provided a solution (in a certain sense) to the famous partial differential equation $u_t = u_{xx}$, the so-called *heat equation.*

Note that in 1905, belief in atoms and molecules was far from universal. In fact, Einstein's "proof" of Brownian motion helped provide convincing evidence of atomic existence. Einstein had a busy 1905, also publishing seminal papers on the special theory of relativity and the photoelectric effect. In fact, his work on the photoelectric effect won him a Nobel prize. Curiously, though, history has shown that the photoelectric effect is the *least* monumental of his three 1905 triumphs. The world at that time simply could not accept special relativity!

Since Brownian motion described the physical trajectories of pollen grains suspended in water, Brownian paths must be continuous. But they were seen to be so irregular that the French physicist Jean Perrin believed them to be non-differentiable. (The German mathematician Karl Weierstrass had recently discovered such pathological functions do exist. Indeed the continuous function

$$g(x) = \sum_{n=1}^{\infty} b^n \cos(a^n \pi x)$$

where a is odd, $b \in (0, 1)$, and $ab > 1 + 3\pi/2$ is nowhere differentiable.) Perrin himself worked to show that colliding particles obey the gas laws, calculated Avogadro's number, and won the 1926 Nobel prize.

Finally, in 1923, the mathematician Norbert Wiener established the mathematical existence of Brownian motion by verifying the existence of a stochastic process with the given properties.

Exercise 3.4.4. Deduce from the definition of Brownian motion that for each t, the random variable B_t is normally distributed with mean 0 and variance t. Why does this implies that $\mathbb{E}(B_t^2) = t$?

Exercise 3.4.5 (Trend and ACVF of Brownian motion). Use the result of the previous exercise to prove that if $\{B_t\}$ is a Brownian motion, then $\mu(t) = 0$ for all t, and $\gamma(t, s) = s$ if $0 \le s < t$. This shows that Brownian motion is not a stationary process. (*Hint*: To compute $\gamma(t, s)$, write $B_s B_t = (B_s B_t - B_s^2) + B_s^2$, take expectations, and then use the third part of the definition of Brownian motion and the previous exercise.)

Exercise 3.4.6. Deduce from the definition of Brownian motion that for $0 \le s < t \le \infty$, the distribution of the random variable $B_t - B_s$ is the *same* as the distribution of the random variable B_{t-s}.

Remark. Although Brownian motion is not a stationary process, it does have what are called *stationary increments*; that is, the distribution of the *increment* $B_t - B_s$ only depends on $|t - s|$.

To extend the above exercises, an alternative way to define Brownian motion is via its second order properties.

Definition 3.4.7. A stochastic process $\{X_t\}$ is called a *Gaussian process* if for any $m \in \mathbb{N}$ and for any times $t_1, \ldots, t_m$, the joint distribution of $(X_{t_1}, X_{t_2}, \ldots, X_{t_m})$ is multivariate normal.

Example 3.4.8. Brownian motion $\{B_t\}$ is a Gaussian process with

- $\mu(t) = \mathbb{E}(B_t) = 0$ for each t,
- $\gamma(t, s) = \text{Cov}(B_t, B_s) = \min(s, t)$, and

- trajectories, $t \mapsto B_t$, which are continuous.

We end this section with two remarks which briefly indicate some of the connections between Brownian motion and time series.

Remark. Brownian motion was also an early model of stock prices. However, it was quickly realized that Brownian motion was an inadequate model since the trend of Brownian motion is 0, while the trend of most stocks is increasing (at least for the *good* ones). To compensate, the processes $X_t := \sigma B_t + \mu t$ (called *Brownian motion with drift*) and $Y_t := Y_0 \, e^{X_t} = Y_0 \exp(\sigma B_t + \mu t)$ (called *exponential Brownian motion*) were introduced. The merits and limitations of these processes are well-known. A further process, called *fractional Brownian motion*, attempts to address some of these limitations.

Exercise 3.4.9. Here is a quick review of ordinary differential equations. Suppose that the (deterministic) function $y(t)$ is defined by the differential equation

$$dy(t) := x(t)dt \qquad (\dagger)$$

for some given function $x(t)$. You will recall from elementary calculus that equation (†) has no meaning. Instead, it is simply *shorthand* for

$$\frac{dy(t)}{dt} = x(t).$$

Provided things make sense, we can *solve* the ODE (†) for $y(t)$, namely

$$y(t) = \int x(t)dt.$$

Find all possible functions $y(t)$ for which

$$dy(t) = \frac{1}{1+t^2}\, dt,$$

i.e., $x(t) = (1+t^2)^{-1}$.

Remark. You might be tempted to think that Brownian motion provides an example of white noise (as in Example 3.3.2). Since the random variables B_s and B_t, $s \neq t$ are not uncorrelated, Brownian motion is not white noise. However, there is a strange relationship between Brownian motion and white noise. Although we have noted that Brownian motion is non-differentiable in the traditional (elementary calculus) sense, it is possible to make sense of $\frac{dB_t}{dt}$ as a *generalized function.* In the *statistical communication theory* literature, a slightly different definition of white noise is made, namely as the generalized derivative of Brownian motion; i.e, if $\{X_t\}$ is white noise, then $dB_t := X_t dt$. The theory of *stochastic integration* gives meaning to such a *stochastic differential equation.*

4 Representations and Decompositions

It is often useful in mathematical analysis to have a representation of a complicated object in terms of other, more simple objects.

You saw a representation result in elementary calculus!!!