CHAPTER 3

Stochastic Processes: general theory

There are two ways to approach the definition of a stochastic process (S.P.)
The easier approach, which we follow in Section 3.1, is to view a S.P. as a collection of R.V.-s $\{X_t(\omega), t \in \mathcal{I}\}$, focusing on the finite dimensional distributions and ignoring sample path properties (such as continuity). Its advantages are:
• The index set $\mathcal{I}$ can be arbitrary;
• The measurability of $t \mapsto X_t(\omega)$ as a function from $\mathcal{I}$ to $\mathbb{R}$ (per fixed ω), is not part of the definition of the process;
• There is no need to master functional analysis techniques and results.

The main disadvantages of this approach are:
• It only works when $X_t(\omega)$ is well-defined for each fixed t;
• Using it, we cannot directly determine important properties such as continuity or monotonicity of the sample path, or the law of $\sup_t X_t$.

A more sophisticated approach to S.P., mostly when $\mathcal{I}$ is an interval on the real line, views the S.P. as a function-valued R.V: $\Omega \to$ [Space of functions] (examples of such spaces might be $C[0,1]$ or $L^2[0,1]$). The main advantages of this approach are the complements of the disadvantages of our approach and vice verse.

Defining and using characteristic functions we study in Section 3.2 the important class of Gaussian stochastic processes. We conclude this chapter by detailing in Section 3.3 sufficient conditions for the continuity of the sample path $t \mapsto X_t(\omega)$, for almost all outcomes ω.

3.1. Definition, distribution and versions

Our starting point is thus the following definition of what a stochastic process is.

Definition 3.1.1. *Given* $(\Omega, \mathcal{F}, \mathbf{P})$*, a* stochastic process *(S.P.)* $\{X_t\}$ *is a collection* $\{X_t : t \in \mathcal{I}\}$ *of R.V.-s where the index* t *belongs to the index set* $\mathcal{I}$*. Typically,* $\mathcal{I}$ *is an interval in* $\mathbb{R}$ *(in which case we say that* $\{X_t\}$ *is a* continuous time stochastic process*), or a subset of* $\{1, 2, \ldots, n, \ldots\}$ *(in which case we say that* $\{X_t\}$ *is a* discrete time stochastic process*. We also call* $t \mapsto X_t(\omega)$ *the* sample function *(or* sample path*) of the S.P.*

Recall our notation $\sigma(X_t)$ for the σ-field generated by X_t. The discrete time stochastic processes are merely countable collections of R.V.-s $X_1, X_2, X_3, \ldots$ defined on the same probability space. All relevant information about such a process during a finite time interval $\{1, 2, \ldots, n\}$ is conveyed by the σ-field $\sigma(X_1, X_2, \ldots, X_n)$, namely, the σ-field generated by the "rectangle" sets $\bigcap_{i=1}^n \{\omega : X_i(\omega) \leq \alpha_i\}$ for $\alpha_i \in \mathbb{R}$ (compare with Definition 1.2.8 of $\sigma(X)$). To deal with the full infinite time horizon we just take the σ-field $\sigma(X_1, X_2, \ldots)$ generated by the union of these sets

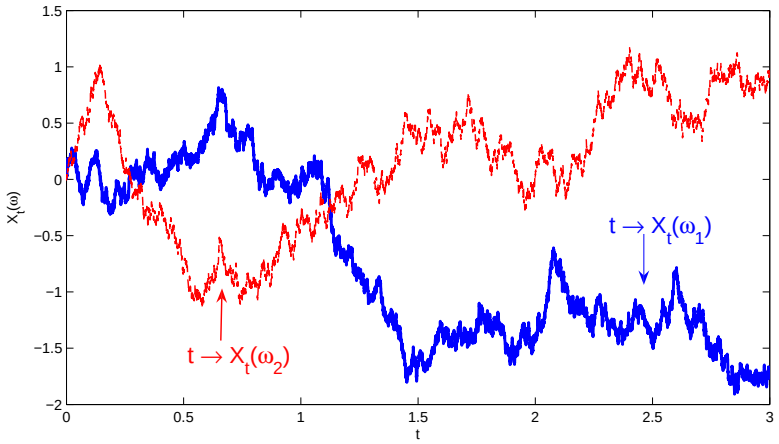

FIGURE 1. Two sample paths of a stochastic process, corresponding to two outcomes ω_1 and ω_2.

over $n = 1, 2, \ldots$ (this is exactly what we did to define $\mathcal{F}_c$ of the coin tosses Example 1.1.13). Though we do not do so here, it is not hard to verify that in this setting the σ-field $\sigma(X_1, X_2, \ldots)$ coincides with the smallest σ-field containing $\sigma(X_t)$ for all $t \in \mathcal{I}$, which we denote hereafter by $\mathcal{F}_{\mathbf{X}}$.

Perhaps the simplest example of a discrete time stochastic process is that of independent, identically distributed R.V.-s X_n (see Example 1.4.7 for one such construction). Though such a S.P. has little interesting properties worth study, it is the corner stone of the following, fundamental example, called the random walk.

Definition 3.1.2. *A* random walk *is the sequence* $S_n = \sum_{i=1}^n \xi_i$, *where* ξ_i *are independent and identically distributed real-valued R.V.-s defined on the same probability space* $(\Omega, \mathcal{F}, \mathbf{P})$. *When* ξ_i *are integer-valued we say that this is a random walk on the integers, and call the special case of* $\xi_i \in \{-1, 1\}$ *a simple random walk.*

When considering a continuous-time S.P. we deal with uncountable collections of R.V.. As we soon see, this causes many difficulties. For example, when we talk about the *distribution* (also called the *law*) of such a S.P. we usually think of the restriction of $\mathbf{P}$ to a certain σ-field. But which σ-field to choose? That is, which one carries all the information we have in mind? Clearly, we require at least $\mathcal{F}_{\mathbf{X}}$ so we can determine the law $\mathcal{P}_{X_t}$ of $X_t(\omega)$ per fixed t. But is this enough? For example, we might be interested in sets such as $H_f = \{\omega : X_t(\omega) \leq f(t), \forall 0 \leq t \leq 1\}$ for some functions $f : [0, 1] \mapsto \mathbb{R}$. Indeed, the value of $\sup(X_t : 0 \leq t \leq 1)$ is determined by such sets for $f(t) = \alpha$ independent of t. Unfortunately, such sets typically involve an uncountable number of set operations and thus are usually not in $\mathcal{F}_{\mathbf{X}}$, and hence $\sup(X_t : 0 \leq t \leq 1)$ might not even be measurable on $\mathcal{F}_{\mathbf{X}}$. So, maybe we should take instead $\sigma(\{H_f, f : [0, 1] \to \mathbb{R}\})$ which is quite different from $\mathcal{F}_{\mathbf{X}}$? We choose in these notes the smaller, i.e. simpler, σ-field $\mathcal{F}_{\mathbf{X}}$, but what is then the minimal information we need for specifying uniquely a probability measure on this space? Before tackling this issue, we provide some motivation to our interest in continuous-time S.P.

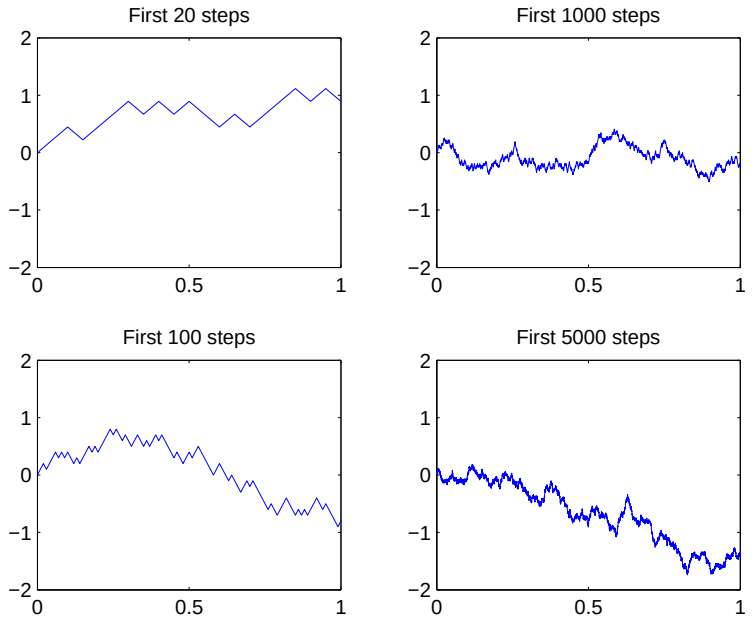

FIGURE 2. Illustration of scaled simple random walks for different values of n.

Assuming that $\mathbf{E}(\xi_1) = 0$ and $\mathbf{E}(\xi_1^2) = 1$ we have that $\mathbf{E}(S_n) = 0$ and

$$\mathbf{E}(S_n^2) = \mathbf{E}\Big[(\sum_{i=1}^n \xi_i)^2\Big] = \sum_{i,j=1}^n \mathbf{E}(\xi_i \xi_j) = n\,, \tag{3.1.1}$$

that is, $\mathbf{E}\Big[(n^{-1/2}S_n)^2\Big] = 1$. Further, in this case, by central limit theorem we have that

$$n^{-1/2}S_n \xrightarrow{\mathcal{L}} G = N(0,1)\,, \tag{3.1.2}$$

with $N(\mu, v)$ denoting a Gaussian R.V. of mean μ and variance v (c.f. Example 1.4.13). Replacing n by $[nt]$ (the integer part of nt), we get from (3.1.2) by rescaling that also $n^{-1/2}S_{[nt]} \xrightarrow{\mathcal{L}} N(0,t)$ for any fixed $0 \le t \le 1$. This leads us to state the functional C.L.T. where all values of $0 \le t \le 1$ are considered at once.

Theorem 3.1.3. *(see* [**Bre92**, Section 12.2]*) Consider the random walk S_n when $\mathbf{E}(\xi_1) = 0$ and $\mathbf{E}(\xi_1^2) = 1$. Take the linear interpolation of the sequence S_n, scale space by $n^{-\frac{1}{2}}$ and time by n^{-1} (see Figure 2). Taking $n \to \infty$ we arrive at a limiting object which we call the* Brownian motion *on $0 \le t \le 1$. The convergence here is weak convergence in the sense of Definition 1.4.20 with $\mathbb{S}$ the set of continuous functions on $[0,1]$, equipped with the topology induced by the supremum norm.*

Even though it is harder to define the Brownian motion (being a continuous time S.P.), computations for it typically involve relatively simple Partial Differential Equations and are often more explicit than those for the random walk. In addition,

unlike the random walk, the Brownian motion does not depend on the specific law of ξ_i (beyond having zero mean and unit variance). For a direct construction of the Brownian motion, to which we return in Section 5.1, see [**Bre92**, Section 12.7].

Remark. The condition $\mathbf{E}(\xi_1^2) < \infty$ is almost necessary for the $n^{-1/2}$ scaling of space and the Brownian limiting process. Indeed, if $\mathbf{E}(\xi_1^\alpha) = \infty$ for some $0 < \alpha < 2$, then both scaling of space and the limit S.P. are changed with respect to Theorem 3.1.3.

In the next example, which is mathematically equivalent to Theorem 3.1.3, we replace the sum of independent and identically distributed R.V.-s by the product of such non-negative R.V.-s.

Example 3.1.4. *Let $M_n = \prod_{i=1}^n Y_i$ where Y_i are positive, independent identically distributed random variables (for instance, the random daily return rates of a certain investment). Let $S_n = \log M_n$ and $\xi_i = \log Y_i$, noting that $S_n = \sum_{i=1}^n \xi_i$ is a random walk. Assuming $\mathbf{E}(\log Y_1) = 0$ and $\mathbf{E}[(\log Y_1)^2] = 1$ we know that $n^{-1/2} S_{[nt]} = n^{-1/2} \log M_{[nt]}$ converges to a Brownian motion W_t as $n \to \infty$. Hence, the limit behavior of M_n is related to that of the* Geometric Brownian motion e^{W_t}.

For continuous time stochastic processes we provide next few examples of events that are in the σ-field $\mathcal{F}_{\mathbf{X}}$ and few examples that in general may not belong to this σ-field (see [**Bre92**, Section 12.4] for more examples).

(1). $\{\omega : X_{t_1} \le \alpha\} \in \mathcal{F}_{\mathbf{X}}$.

(2). $\{\omega : X_{1/k} \le \alpha, k = 1, 2, \ldots, N\} \in \mathcal{F}_{\mathbf{X}}$.

(3). $\{\omega : X_{1/k} \le \alpha, \quad k = 1, 2, 3, \ldots\} = \{\omega : \sup_k X_{1/k}(\omega) \le \alpha\} \in \mathcal{F}_{\mathbf{X}}$.

(4). $\{\omega : \sup(X_t : 0 \le t \le 1) \le \alpha\}$ is not necessarily in $\mathcal{F}_{\mathbf{X}}$ since the supremum here is over an uncountable collection of R.V.-s. However, check that this event is in $\mathcal{F}_{\mathbf{X}}$ whenever all sample functions of the S.P. $\{X_t\}$ are right continuous.

(5). $\{\omega : X_t(\omega) : \mathcal{I} \mapsto \mathbb{R}$ is a measurable function$\}$ may also be outside $\mathcal{F}_{\mathbf{X}}$ (say for $\mathcal{I} = [0, 1]$).

We define next the stochastic process analog of the distribution function.

Definition 3.1.5. *Given $N < \infty$ and a collection $t_1, t_2, \ldots, t_N$ in $\mathcal{I}$, we denote the (joint) distribution of $(X_{t_1}, \ldots, X_{t_N})$ by $F_{t_1, t_2, \cdots, t_N}(\cdot)$, that is,*

$$F_{t_1, t_2, \cdots, t_N}(\alpha_1, \alpha_2, \cdots, \alpha_N) = \mathbf{P}(X_{t_1} \le \alpha_1, \cdots, X_{t_N} \le \alpha_N),$$

for all $\alpha_1, \alpha_2, \cdots, \alpha_N \in \mathbb{R}$. We call the collection of functions $F_{t_1, t_2, \cdots, t_N}(\cdot)$, the finite dimensional distributions *(f.d.d.) of the S.P.*

Having independent increments is one example of a property that is determined by the f.d.d.

Definition 3.1.6. *With $\mathcal{G}_t$ the smallest σ-field containing $\sigma(X_s)$ for any $0 \le s \le t$, we say that a S.P. $\{X_t\}$ has* independent increments *if $X_{t+h} - X_t$ is independent of $\mathcal{G}_t$ for any $h > 0$ and all $t \ge 0$. This property is determined by the f.d.d. That is, if the random variables $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are mutually independent, for all $n < \infty$ and $0 \le t_1 < t_2 < \cdots < t_n < \infty$ then the S.P. $\{X_t\}$ has independent increments.*

Remark. For example, both the random walk and the Brownian motion are processes with independent increments.

The next example shows that the f.d.d. do not determine some other important properties of the stochastic process, such as *continuity* of its sample path. That is, knowing all f.d.d. is not enough for computing $\mathbf{P}(\{\omega : t \mapsto X_t(\omega) \text{ is continuous on } [0,1]\})$.

Example 3.1.7. *Consider the probability space $\Omega = [0,1]$ with its Borel σ-field and the Uniform law on $[0,1]$ (that is, the probability of each interval equals its length, also known as Lebesgue measure restricted to $[0,1]$). Given $\omega \in \Omega$, we define two Stochastic Processes:*

$$Y_t(\omega) = 0, \quad \forall t, \omega \qquad\qquad X_t(\omega) = \begin{cases} 1, t = \omega \\ 0, \text{ otherwise} \end{cases}$$

Let $A_t = \{\omega : X_t \neq Y_t\} = \{t\}$. Since $\mathbf{P}(A_t) = 0$, we have that $\mathbf{P}(X_t = Y_t) = 1$ for each fixed t. Moreover, let $A^N = \bigcup_{i=1}^N A_{t_i}$, then $\mathbf{P}(A^N) = 0$ (a finite union of negligible sets is negligible). Since this applies for any choice of N and $t_1, \ldots, t_N$, we see that the f.d.d. of $\{X_t\}$ are the same as those of $\{Y_t\}$. Moreover, considering the set $A^\infty = \bigcup_{i=1}^\infty A_{t_i}$, involving a countable number of times, we see that $\mathbf{P}(A^\infty) = 0$, that is, almost surely, $X_t(\omega)$ agrees with $Y_t(\omega)$ at any fixed, countable, collection of times. But note that some global sample-path properties do not agree. For example,

$$\mathbf{P}(\{\omega : (\sup\{X_t(\omega) : 0 \leq t \leq 1\}) \neq 0\}) = 1,$$

$$\mathbf{P}(\{\omega : (\sup\{Y_t(\omega) : 0 \leq t \leq 1\}) \neq 0\}) = 0.$$

Also,

$$\mathbf{P}(\{\omega : t \mapsto X_t(\omega) \text{ is continuous}\}) = 0,$$

$$\mathbf{P}(\{\omega : t \mapsto Y_t(\omega) \text{ is continuous}\}) = 1.$$

While the maximal value and continuity of sample path are different for the two S.P. of Example 3.1.7, we should typically consider such a pair to be the same S.P., motivating our next two definitions.

Definition 3.1.8. *Two S.P. $\{X_t\}$ and $\{Y_t\}$ are called* versions *of one another if they have the same finite-dimensional distributions.*

Definition 3.1.9. *A S.P. $\{Y_t\}$ is called a* modification *of another S.P. $\{X_t\}$ if $\mathbf{P}(Y_t = X_t) = 1$ for all $t \in \mathcal{I}$.*

We consider next the relation between the concepts of modification and version, starting with:

Exercise 3.1.10. *Show that if $\{Y_t\}$ is a modification of $\{X_t\}$, then $\{Y_t\}$ is also a version of $\{X_t\}$.*

Note that a modification has to be defined on the same probability space as the original S.P. while this is not required of versions.

The processes in Example 3.1.7 are modifications of one another. In contrast, our next example is of two versions on the same probability space which are not modifications of each other.

Example 3.1.11. *Consider* $\Omega_2 = \{HH, TT, HT, TH\}$ *with the uniform probability measure* $\mathbf{P}$*, corresponding to two independent fair coin tosses, whose outcome is* $\omega = (\omega_1, \omega_2)$*. Define on* $(\Omega_2, 2^{\Omega_2}, \mathbf{P})$ *the S.P*

$$X_t(\omega) = \mathbf{1}_{[0,1)}(t) I_H(\omega_1) + \mathbf{1}_{[1,2)}(t) I_H(\omega_2)\,, \qquad 0 \le t < 2\,,$$

and $Y_t(\omega) = 1 - X_t(\omega)$ *for* $0 \le t < 2$*.*

Exercise 3.1.12. *To practice your understanding you should at this point check that the processes* X_t *and* Y_t *of Example 3.1.11 are versions of each other but are not modifications of each other.*

Definition 3.1.13. *We say that a collection of finite dimensional distributions is* consistent *if*

$$\lim_{\alpha_k \uparrow \infty} F_{t_1,\cdots,t_N}(\alpha_1, \cdots, \alpha_N) = F_{t_1,\cdots,t_{k-1},t_{k+1},t_N}(\alpha_1, \cdots, \alpha_{k-1}, \alpha_{k+1}, \cdots, \alpha_N)\,,$$

for any $1 \le k \le N$*,* $t_1 < t_2 < \cdots < t_N \in \mathcal{I}$ *and* $\alpha_i \in \mathbb{R}$*,* $i = 1, \ldots, N$*.*

Convince yourself that the f.d.d. of any S.P. must be consistent. Conversely,

Proposition 3.1.14. *For any consistent collection of finite dimensional distributions, there exists a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *and a stochastic process* $\omega \mapsto X_t(\omega)$ *on it, whose f.d.d. are in agreement with the given collection (c.f.* [**Bre92**, Theorem 12.14]*, or* [**GS01**, Theorem 8.6.3]*). Further, the restriction of the probability measure* $\mathbf{P}$ *to the* σ*-field* $\mathcal{F}_{\mathbf{X}}$ *is uniquely determined by the given collection of f.d.d.*

We note in passing that the construction of Proposition 3.1.14 builds on the easier case of discrete time stochastic processes (which is treated for example in [**Bre92**, Section 2.4]).

We can construct a σ-field that is the image of $\mathcal{F}_{\mathbf{X}}$ on the range of $\omega \mapsto \{X_t(\omega) : t \in \mathcal{I}\} \subseteq \mathbb{R}^{\mathcal{I}}$.

Definition 3.1.15. *For an interval* $\mathcal{I} \subseteq \mathbb{R}$*, let* $\mathbb{R}^{\mathcal{I}}$ *denote the set of all functions* $x : \mathcal{I} \to \mathbb{R}$*. A* finite dimensional rectangle *in* $\mathbb{R}^{\mathcal{I}}$ *is any set of the form* $\{x : x(t_i) \in J_i, i = 1, \ldots, n\}$ *for a non-negative integer* n*, intervals* $J_i \subseteq \mathbb{R}$ *and times* $t_i \in \mathcal{I}$*,* $i = 1, \ldots, n$*. The* cylindrical σ-field $\mathcal{B}^{\mathcal{I}}$ *is the* σ*-field generated by the collection of finite dimensional rectangles.*

• *What follows may be omitted at first reading.*

While f.d.d. do not determine important properties of the sample path $t \mapsto X_t(\omega)$ of the S.P. (see Example 3.1.7), they uniquely determine the probabilities of events in $\mathcal{F}_{\mathbf{X}}$ (hence each property of the sample path that can be expressed as an element of the cylindrical σ-field $\mathcal{B}^{\mathcal{I}}$).

Proposition 3.1.16. *For an interval* $\mathcal{I} \subseteq \mathbb{R}$ *and any S.P.* $\{X_t\}$ *on* $t \in \mathcal{I}$*, the* σ*-field* $\mathcal{F}_{\mathbf{X}}$ *consists of the events* $\{\omega : X.(\omega) \in \Gamma\}$ *for* $\Gamma \in \mathcal{B}^{\mathcal{I}}$*. Further, if a S.P.* $\{Y_t\}$ *is a version of* $\{X_t\}$*, then* $\mathbf{P}(X. \in \Gamma) = \mathbf{P}(Y. \in \Gamma)$ *for all such* Γ *(see* [**Bre92**, Corollary 12.9 and Proposition 12.12] *for proofs).*

3.2. Characteristic functions, Gaussian variables and processes

Subsection 3.2.1 is about the fundamental concept of characteristic function and its properties. Using it, we study in Subsection 3.2.2 the Gaussian random vectors and stochastic processes. Subsection 3.2.3 deals with *stationarity*, an important concept in the general theory of stochastic processes which is simpler to check for the Gaussian processes.

3.2.1. Characteristic function. We start with the definition of the characteristic function of a random vector. In doing so we adopt the convention that a complex valued random variable Z is a function from Ω to $\mathbb{C}$ such that both the real and imaginary parts of Z are Borel measurable, and if $Z = X + iY$ with $X, Y \in \mathbb{R}$ integrable random variables (and $i = \sqrt{-1}$), then $\mathbf{E}(Z) = \mathbf{E}(X) + i\mathbf{E}(Y) \in \mathbb{C}$.

Definition 3.2.1. *A* random vector $\underline{X} = (X_1, X_2, \cdots, X_n)$ *with values in* $\mathbb{R}^n$ *has the* characteristic function

$$\Phi_{\underline{X}}(\underline{\theta}) = \mathbf{E}[e^{i\sum_{k=1}^n \theta_k X_k}]\,,$$

where $\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_n) \in \mathbb{R}^n$ *and* $i = \sqrt{-1}$.

Remark. The characteristic function $\Phi_{\underline{X}} : \mathbb{R}^n \to \mathbb{C}$ exists for any $\underline{X}$ since

$$e^{i\sum_{k=1}^n \theta_k X_k} = \cos(\sum_{k=1}^n \theta_k X_k) + i\sin(\sum_{k=1}^n \theta_k X_k)\,, \tag{3.2.1}$$

with both real and imaginary parts being bounded (hence integrable) random variables. Actually, we see from (3.2.1) that $\Phi_{\underline{X}}(\underline{0}) = 1$ and $|\Phi_{\underline{X}}(\underline{\theta})| \le 1$ for all $\underline{\theta} \in \mathbb{R}^n$ (see [**Bre92**, Proposition 8.27] or [**GS01**, Section 5.7] for other properties of the characteristic function).

Our next proposition justifies naming $\Phi_{\underline{X}}$ the characteristic function of (the law of) $\underline{X}$.

Proposition 3.2.2. *The characteristic function determines the* law *of a random vector. That is, if* $\Phi_{\underline{X}}(\underline{\theta}) = \Phi_{\underline{Y}}(\underline{\theta})$ *for all* $\underline{\theta}$ *then* $\underline{X}$ *has the same law (= probability measure on* $\mathbb{R}^n$*) as* $\underline{Y}$ *(for proof see* [**Bre92**, Theorems 11.4 and 8.24] *or* [**GS01**, Corollary 5.9.3]*).*

Remark. The law of a non-negative random variable X is also determined by its moment generating function $M_X(s) = \mathbf{E}[e^{sX}]$ at $s < 0$ (see [**Bre92**, Proposition 8.51] for a proof). While the real-valued function $M_X(s)$ is a simpler object, it is unfortunately useless for the many random variables X which are neither non-negative nor non-positive and for which $M_X(s) = \infty$ for all $s \neq 0$.

The characteristic function is very useful in connection with convergence in law. Indeed,

Exercise 3.2.3. *Show that if* $X_n \xrightarrow{\mathcal{L}} X$ *then* $\Phi_{X_n}(\theta) \to \Phi_X(\theta)$ *for any* $\theta \in \mathbb{R}$.

Remark. Though much harder to prove, the converse of Exercise 3.2.3 is also true, namely if $\Phi_{X_n}(\theta) \to \Phi_X(\theta)$ for each $\theta \in \mathbb{R}$ then $X_n \xrightarrow{\mathcal{L}} X$.

We continue with a few explicit computations of the characteristic function.

Example 3.2.4. *Consider X a* Bernoulli*(p) random variable, that is, $\mathbf{P}(X=1) = p$ and $\mathbf{P}(X=0) = 1-p$. Its characteristic function is by definition*

$$\Phi_X(\theta) = \mathbf{E}[e^{i\theta X}] = pe^{i\theta} + (1-p)e^{i0\theta} = pe^{i\theta} + 1 - p\,.$$

The same type of explicit formula applies to any R.V. $X \in$ SF. Moreover, such formulas apply for any discrete valued R.V. For example, if $X \sim$ Poisson(λ) then

$$\Phi_X(\theta) = \mathbf{E}[e^{i\theta X}] = \sum_{k=0}^{\infty} \frac{(\lambda e^{i\theta})^k}{k!} e^{-\lambda} = e^{\lambda(e^{i\theta}-1)}\,. \tag{3.2.2}$$

The characteristic function has an explicit form also when the R.V. X has a probability density function f_X as in Definition 1.2.23. Indeed, then by Proposition 1.2.29 we have that

$$\Phi_X(\theta) = \int_{-\infty}^{\infty} e^{i\theta x} f_X(x)dx\,, \tag{3.2.3}$$

which is merely the Fourier transform of the density f_X. For example, applying this formula we see that the Uniform random variable U of Example 1.1.11 has characteristic function $\Phi_U(\theta) = (e^{i\theta}-1)/(i\theta)$. Assuming that the density f_X is bounded and continuous, we also have the explicit inversion formula

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\theta x}\Phi_X(\theta)d\theta\,, \tag{3.2.4}$$

as a way to explain Proposition 3.2.2 ([**Bre92**, Theorem 8.39] shows that this inversion formula is valid whenever $\int |\Phi_X(\theta)|d\theta < \infty$, see also [**GS01**, Theorem 5.9.1]).

We next recall the extension of the notion of density as in Definition 1.2.23 to a random vector (as done already in (2.4.1)).

Definition 3.2.5. *We say that a random vector $\underline{X} = (X_1, \ldots, X_n)$ has a* probability density function *$f_{\underline{X}}$ if*

$$\mathbf{P}(\{\omega : a_i \le X_i(\omega) \le b_i, i = 1, \ldots, n\}) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_{\underline{X}}(x_1, \ldots, x_n)dx_n \cdots dx_1\,,$$

for every $a_i < b_i$, $i = 1, \ldots, n$. Such density $f_{\underline{X}}$ must be a non-negative Borel measurable function with $\int_{\mathbb{R}^n} f_{\underline{X}}(\underline{x})d\underline{x} = 1$ ($f_{\underline{X}}$ is sometimes called the joint density of $X_1, \ldots, X_n$ as in [**GS01**, Definition 4.5.2]*).*

Adopting the notation $(\underline{\theta}, \underline{x}) = \sum_{k=1}^n \theta_k x_k$ we have the following extension of the Fourier transform formula (3.2.3) to random vectors $\underline{X}$ with density,

$$\Phi_{\underline{X}}(\underline{\theta}) = \int_{\mathbb{R}^n} e^{i(\underline{\theta},\underline{x})} f_{\underline{X}}(\underline{x})d\underline{x}$$

(this is merely a special case of the extension of Proposition 1.2.29 to $h : \mathbb{R}^n \to \mathbb{R}$). Though we shall not do so, we can similarly extend the explicit inversion formula of (3.2.4) to $\underline{X}$ having bounded continuous density, or alternatively, having an absolutely integrable characteristic function.

The computation of the characteristic function is much simplified in the presence of independence, as shown by the following alternative of Proposition 1.4.40.

Proposition 3.2.6. *If $\underline{X} = (X_1, X_2, \ldots, X_n)$ with X_i mutually* independent *R.V., then clearly,*

$$\Phi_{\underline{X}}(\underline{\theta}) = \mathbf{E}[\prod_{k=1}^{n} e^{i\theta_k X_k}] = \prod_{k=1}^{n} \Phi_{X_k}(\theta_k) \qquad \forall\ \underline{\theta} \in \mathbb{R}^n \tag{3.2.5}$$

Conversely, if (3.2.5) holds then the random variables X_i, $i = 1, \ldots, n$ are mutually independent of each other.

3.2.2. Gaussian variables, vectors and processes. We start by recalling some linear algebra concepts we soon need.

Definition 3.2.7. *An $n \times n$ matrix $\mathbf{A}$ with entries A_{jk} is called* non-negative definite *(or positive semidefinite) if $A_{jk} = A_{kj}$ for all j, k, and for any $\underline{\theta} \in \mathbb{R}^n$*

$$(\underline{\theta}, \mathbf{A}\underline{\theta}) = \sum_{j=1}^{n}\sum_{k=1}^{n} \theta_j A_{jk} \theta_k \geq 0.$$

We next define the Gaussian random vectors via their characteristic functions.

Definition 3.2.8. *We say that a* random vector *$\underline{X} = (X_1, X_2, \cdots, X_n)$ has a* Gaussian *(or multivariate Normal) distribution if*

$$\Phi_{\underline{X}}(\underline{\theta}) = e^{-\frac{1}{2}(\underline{\theta}, \Sigma\underline{\theta})} e^{i(\underline{\theta}, \underline{\mu})},$$

for some non-negative definite $n \times n$ matrix Σ, some $\underline{\mu} = (\mu_1, \mu_2, \cdots, \mu_n) \in \mathbb{R}^n$ and all $\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_n) \in \mathbb{R}^n$.

Remark. In the special case of $n = 1$, we say that a random variable X is Gaussian if for some $\mu \in \mathbb{R}$, some $\sigma^2 \geq 0$ and all $\theta \in \mathbb{R}$,

$$\mathbf{E}[e^{i\theta X}] = e^{-\frac{1}{2}\theta^2\sigma^2 + i\theta\mu}.$$

As we see next, the classical definition of Gaussian distribution via its density amounts to a strict subset of the distributions we consider in Definition 3.2.8.

Definition 3.2.9. *We say that $\underline{X}$ has a* non-degenerate *Gaussian distribution if the matrix Σ is invertible, or alternatively, when Σ is (strictly) positive definite matrix, that is $(\underline{\theta}, \Sigma\underline{\theta}) > 0$ whenever $\underline{\theta}$ is a non-zero vector (for an equivalent definition see* [**GS01**, Section 4.9]*).*

Proposition 3.2.10. *A random vector $\underline{X}$ with a non-degenerate* Gaussian distribution *has the* density

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2}(det\Sigma)^{1/2}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu}, \Sigma^{-1}(\underline{x}-\underline{\mu}))}$$

(see also [**GS01**, Definition 4.9.4]*). In particular, if $\sigma^2 > 0$, then a Gaussian random variable X has the density*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

(for example, see [**GS01**, Example 4.4.4]*).*

Our next proposition links the vector $\underline{\mu}$ and the matrix Σ to the first two moments of the Gaussian distribution.

Proposition 3.2.11. *The* parameters *of the* Gaussian distribution *are* $\mu_j = \mathbf{E}(X_j)$ *and* $\Sigma_{jk} = \mathbf{E}[(X_j - \mu_j)(X_k - \mu_k)]$, $j,k = 1,\ldots,n$ *(c.f.* [**GS01**, Theorem 4.9.5]*). Thus* $\underline{\mu}$ *is the mean vector and* Σ *is the covariance matrix of* $\underline{X}$.

As we now demonstrate, there is more to a Gaussian random vector than just having coordinates that are Gaussian random variables.

Exercise 3.2.12. *Let* X *be a Gaussian R.V. independent of* S, *with* $\mathbf{E}(X) = 0$ *and* $\mathbf{P}(S = 1) = \mathbf{P}(S = -1) = 1/2$.

(a) *Check that* SX *is Gaussian.*
(b) *Give an example of uncorrelated, zero-mean, Gaussian R.V.* X_1 *and* X_2 *such that the vector* $\underline{X} = (X_1, X_2)$ *is not Gaussian and where* X_1 *and* X_2 *are not independent.*

Exercise 3.2.13. *Suppose* (X,Y) *has a bi-variate Normal distribution (per Definition 3.2.8), with mean vector* $\underline{\mu} = (\mu_X, \mu_Y)$ *and the covariance matrix* $\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$, *with* σ_X, $\sigma_Y > 0$ *and* $|\rho| \le 1$.

(a) *Show that* (X,Y) *has the same law as* $(\mu_X + \sigma_X\rho U + \sigma_X\sqrt{1-\rho^2}V, \mu_Y + \sigma_Y U)$, *where* U *and* V *are independent Normal R.V.-s of mean zero and variance one. Explain why this implies that* $Z = X - (\rho\sigma_X/\sigma_Y)Y$ *is independent of* Y.
(b) *Explain why such* X *and* Y *are independent whenever they are uncorrelated (hence also whenever* $\mathbf{E}(X|Y) = \mathbf{E}X$*).*
(c) *Verify that* $\mathbf{E}(X|Y) = \mu_X + \frac{\rho\sigma_X}{\sigma_Y}(Y - \mu_Y)$.

Part (b) of Exercise 3.2.13 extends to any Gaussian random vector. That is,

Proposition 3.2.14. *If a Gaussian random vector* $\underline{X} = (X_1,\ldots,X_n)$ *has uncorrelated coordinates, then its coordinates are also mutually independent.*

PROOF. Since the coordinates X_k are uncorrelated, the corresponding matrix Σ has zero entries except at the main-diagonal $j = k$ (see Proposition 3.2.11). Hence, by Definition 3.2.8, the characteristic function $\Phi_{\underline{X}}(\underline{\theta})$ is of the form of $\prod_{k=1}^n \Phi_{X_k}(\theta_k)$. This in turn implies that the coordinates X_k of the random vector $\underline{X}$ are mutually independent (see Proposition 3.2.6). ∎

Definition 3.2.8 allows for Σ that is non-invertible, so for example the random variable $X = \mu$ a.s. is considered a Gaussian variable though it obviously does not have a density (hence does not fit Definition 3.2.9). The reason we make this choice is to have the collection of Gaussian distributions closed with respect to convergence in 2-mean, as we prove below to be the case.

Proposition 3.2.15. *Suppose a sequence of* n*-dimensional Gaussian random vectors* $\underline{X}^{(k)}$, $k = 1,2,\ldots$ *converges in* 2*-mean to an* n*-dimensional random vector* $\underline{X}$, *that is,* $\mathbf{E}[(X_i - X_i^{(k)})^2] \to 0$ *as* $k \to \infty$, *for* $i = 1,2,\ldots,n$. *Then,* $\underline{X}$ *is a Gaussian random vector, whose parameters* $\underline{\mu}$ *and* Σ *are the limits of the corresponding parameters* $\underline{\mu}^{(k)}$ *and* $\Sigma^{(k)}$ *of* $\underline{X}^{(k)}$.

PROOF. We start by verifying the convergence of the parameters of $\underline{X}^{(k)}$ to those of $\underline{X}$. To this end, fixing $1 \le i,j \le n$ and applying the inequality

$$|(a+x)(b+y) - ab| \le |ay| + |bx| + |xy|$$

for $a = X_i$, $b = X_j$, $x = X_i^{(k)} - X_i$ and $y = X_j^{(k)} - X_j$, we get by monotonicity of the expectation that

$$\begin{aligned}\mathbf{E}|X_i^{(k)}X_j^{(k)} - X_iX_j| &\leq \mathbf{E}|X_i(X_j^{(k)} - X_j)| + \mathbf{E}|X_j(X_i^{(k)} - X_i)| \\ &+ \mathbf{E}|(X_i^{(k)} - X_i)(X_j^{(k)} - X_j)|\,.\end{aligned}$$

Thus, by the Schwarz inequality, c.f. Proposition 1.2.41, we see that

$$\begin{aligned}\mathbf{E}|X_i^{(k)}X_j^{(k)} - X_iX_j| &\leq \|X_i\|_2\|X_j^{(k)} - X_j\|_2 + \|X_j\|_2\|X_i^{(k)} - X_i\|_2 \\ &+ \|X_i^{(k)} - X_i\|_2\|X_j^{(k)} - X_j\|_2\,.\end{aligned}$$

So, the assumed convergence in 2-mean of $\underline{X}^{(k)}$ to $\underline{X}$ implies the convergence in 1-mean of $X_i^{(k)}X_j^{(k)}$ to X_iX_j as $k \to \infty$. This in turn implies that $\mathbf{E}X_i^{(k)}X_j^{(k)} \to \mathbf{E}X_iX_j$ (c.f. Exercise 1.3.21). Further, the assumed convergence in 2-mean of $X_l^{(k)}$ to X_l (as $k \to \infty$) implies the convergence of $\mu_l^{(k)} = \mathbf{E}X_l^{(k)}$ to $\mu_l = \mathbf{E}X_l$, for $l = 1, 2, \ldots, n$, and hence also that

$$\Sigma_{ij}^{(k)} = \mathbf{E}X_i^{(k)}X_j^{(k)} - \mu_i^{(k)}\mu_j^{(k)} \to \mathbf{E}X_iX_j - \mu_i\mu_j = \Sigma_{ij}\,.$$

In conclusion, we established the convergence of the mean vectors $\underline{\mu}^{(k)}$ and the covariance matrices $\Sigma^{(k)}$ to the mean vector $\underline{\mu}$ and the covariance matrix Σ of $\underline{X}$, respectively.

Fixing $\underline{\theta} \in \mathbb{R}^n$ the assumed convergence in 2-mean of $\underline{X}^{(k)}$ to $\underline{X}$ also implies the convergence in 2-mean, and hence in probability, of $(\underline{\theta}, \underline{X}^{(k)})$ to $(\underline{\theta}, \underline{X})$. Hence, by bounded convergence, $\Phi_{\underline{X}^{(k)}}(\underline{\theta}) \to \Phi_{\underline{X}}(\underline{\theta})$ for each fixed $\underline{\theta} \in \mathbb{R}^n$ (see Corollary 1.4.28). Since $\Phi_{\underline{X}^{(k)}}(\underline{\theta}) = e^{-\frac{1}{2}(\underline{\theta},\Sigma^{(k)}\underline{\theta})}e^{i(\underline{\theta},\underline{\mu}^{(k)})}$ for each k, the convergence of the parameters $(\underline{\mu}^{(k)}, \Sigma^{(k)})$ implies that the function $\Phi_{\underline{X}}(\underline{\theta})$ must also be of such form. That is, necessarily $\underline{X}$ has a Gaussian distribution, whose parameters are the limits of the corresponding parameters of $\underline{X}^{(k)}$, as claimed. ∎

The next proposition provides an alternative to Definition 3.2.8.

Proposition 3.2.16. *A random vector $\underline{X}$ has the Gaussian distribution if and only if $(\sum_{i=1}^n a_{ji}X_i, j = 1, \ldots, m)$ is a Gaussian random vector for any non-random coefficients $a_{11}, a_{12}, \ldots, a_{mn} \in \mathbb{R}$ (c.f.* [**GS01**, Definition 4.9.7]*).*

It is usually much easier to check Definition 3.2.8 than to check the conclusion of Proposition 3.2.16. However, it is often very convenient to use the latter en-route to the derivation of some other property of $\underline{X}$.

We are finally ready to define the class of Gaussian stochastic processes.

Definition 3.2.17. *A stochastic process (S.P.) $\{X_t,\ t \in \mathcal{I}\}$ is* Gaussian *if for all $n < \infty$ and all $t_1, t_2, \cdots, t_n \in \mathcal{I}$, the random vector $(X_{t_1}, X_{t_2}, \cdots, X_{t_n})$ has a Gaussian distribution, that is, all finite dimensional distributions of the process are Gaussian.*

To see that you understood well the definitions of Gaussian vectors and processes, convince yourself that the following corollary holds.

Corollary 3.2.18. *All distributional properties of Gaussian processes are determined by* the mean $\mu(t) = \mathbf{E}(X_t)$ *of the process and its* auto-covariance function $\rho(t, s) = \mathbf{E}[(X_t - \mu(t))\ (X_s - \mu(s))]$.

Applying Proposition 3.2.14 for the Gaussian vector $\underline{X} = (Y_{t_2} - Y_{t_1}, \ldots, Y_{t_n} - Y_{t_{n-1}})$ of increments of a Gaussian stochastic process $\{Y_t\}$ (with arbitrary finite n and $0 \le t_1 < t_2 < \cdots < t_n$), we conclude from Definition 3.1.6 that

Proposition 3.2.19. *If* $\mathrm{Cov}(Y_{t+h} - Y_t, Y_s) = 0$ *for a Gaussian stochastic process* $\{Y_t\}$*, all* $t \ge s$ *and* $h > 0$*, then the S.P.* $\{Y_t\}$ *has uncorrelated hence* independent increments *(which is thus also equivalent to* $\mathbf{E}(Y_{t+h} - Y_t | \sigma(Y_s, s \le t)) = \mathbf{E}(Y_{t+h} - Y_t)$ *for any* $t \ge 0$ *and* $h > 0$*).*

The special class of Gaussian processes plays a key role in our construction of the Brownian motion. When doing so, we shall use the following extension of Proposition 3.2.15.

Proposition 3.2.20. *If the S.P.* $\{X_t,\ t \in \mathcal{I}\}$ *and the Gaussian S.P.* $\{X_t^{(k)},\ t \in \mathcal{I}\}$ *are such that* $\mathbf{E}[(X_t - X_t^{(k)})^2] \to 0$ *as* $k \to \infty$*, for each fixed* $t \in \mathcal{I}$*, then* X_t *is a Gaussian S.P. with mean and auto-covariance functions that are the pointwise limits of those for* $X_t^{(k)}$*.*

PROOF. Fixing $n < \infty$ and $t_1, t_2, \cdots, t_n \in \mathcal{I}$, apply Proposition 3.2.15 for the sequence of Gaussian random vectors $(X_{t_1}^{(k)}, X_{t_2}^{(k)}, \cdots, X_{t_n}^{(k)})$ to see that the distribution of $(X_{t_1}, X_{t_2}, \cdots, X_{t_n})$ is Gaussian. Since this applies for all finite dimensional distributions of the S.P. $\{X_t,\ t \in \mathcal{I}\}$ we are done (see Definition 3.2.17). ∎

Here is the derivation of the C.L.T. statement of Example 1.4.13 and its extension towards a plausible construction of the Brownian motion.

Exercise 3.2.21. *Consider the random variables* $\widehat{S}_k$ *of Example 1.4.13.*

(a) *Applying Proposition 3.2.6 verify that the corresponding characteristic functions are*

$$\Phi_{\widehat{S}_k}(\theta) = [\cos(\theta/\sqrt{k})]^k \,.$$

(b) *Recalling that* $\delta^{-2}\log(\cos\delta) \to -0.5$ *as* $\delta \to 0$*, find the limit of* $\Phi_{\widehat{S}_k}(\theta)$ *as* $k \to \infty$ *while* $\theta \in \mathbb{R}$ *is fixed.*

(c) *Suppose random vectors* $\underline{X}^{(k)}$ *and* $\underline{X}$ *in* $\mathbb{R}^n$ *are such that* $\Phi_{\underline{X}^{(k)}}(\underline{\theta}) \to \Phi_{\underline{X}}(\underline{\theta})$ *as* $k \to \infty$*, for any fixed* $\underline{\theta}$*. It can be shown that then the laws of* $\underline{X}^{(k)}$*, as probability measures on* $\mathbb{R}^n$*, must converge weakly in the sense of Definition 1.4.20 to the law of* $\underline{X}$*. Explain how this fact allows you to verify the C.L.T. statement* $\widehat{S}_n \xrightarrow{\mathcal{L}} G$ *of Example 1.4.13.*

Exercise 3.2.22. *Consider the random vectors* $\underline{X}^{(k)} = (\frac{1}{\sqrt{k}} S_{k/2}, \frac{1}{\sqrt{k}} S_k)$ *in* $\mathbb{R}^2$*, where* $k = 2, 4, 6, \ldots$ *is even, and* S_k *is the* simple random walk *of Definition 3.1.2, with* $\mathbf{P}(\xi_1 = -1) = \mathbf{P}(\xi_1 = 1) = 0.5$*.*

(a) *Verify that*

$$\Phi_{\underline{X}^{(k)}}(\underline{\theta}) = [\cos((\theta_1 + \theta_2)/\sqrt{k})]^{k/2} [\cos(\theta_2/\sqrt{k})]^{k/2} \,,$$

where $\underline{\theta} = (\theta_1, \theta_2)$*.*

(b) *Find the mean vector* $\underline{\mu}$ *and the covariance matrix* Σ *of a Gaussian random vector* $\underline{X}$ *for which* $\Phi_{\underline{X}^{(k)}}(\underline{\theta})$ *converges to* $\Phi_{\underline{X}}(\underline{\theta})$ *as* $k \to \infty$*.*

(c) *Upon appropriately generalizing what you did in part (b), I claim that the* Brownian motion *of Theorem 3.1.3 must be a Gaussian stochastic process. Explain why, and guess what is the mean $\mu(t)$ and auto-covariance function $\rho(t,s)$ of this process (if needed take a look at Chapter 5).*

We conclude with a concrete example of a Gaussian stochastic process.

Exercise 3.2.23. *Let $Y_n = \sum_{k=1}^n \xi_k V_k$ for i.i.d. random variables $\{\xi_k\}$ such that $p = \mathbf{P}(\xi_k = 1) = 1 - \mathbf{P}(\xi_k = -1)$ and i.i.d. Gaussian random variables $\{V_k\}$ of zero mean and variance one that are independent of the collection $\{\xi_k\}$.*

(a) *Compute the mean $\mu(n)$ and auto-covariance function $\rho(\ell, n)$ for the discrete time stochastic process $\{Y_n\}$.*
(b) *Find the law of $\xi_1 V_1$, explain why $\{Y_n\}$ is a Gaussian process and provide the joint density $f_{Y_n, Y_{2n}}(x,y)$ of Y_n and Y_{2n}.*

3.2.3. Stationary processes. We conclude this section with a brief discussion of the important concept of stationarity, that is, invariance of the law of the process to translation of time.

Definition 3.2.24. *A stochastic process $\{X_t\}$ indexed by $t \in \mathbb{R}$ is called* (strong sense) stationary *if its f.d.d. satisfy*

$$\begin{aligned} F_{t_1,t_2,\ldots,t_N}(\alpha_1,\alpha_2,\ldots,\alpha_N) &= \mathbf{P}(X_{t_1} \le \alpha_1, \ldots, X_{t_N} \le \alpha_N) \\ &= \mathbf{P}(X_{t_1+\tau} \le \alpha_1, \ldots, X_{t_N+\tau} \le \alpha_N) \\ &= F_{t_1+\tau,t_2+\tau,\ldots,t_N+\tau}(\alpha_1,\alpha_2,\ldots,\alpha_N)\,, \end{aligned}$$

for all $\tau \in \mathbb{R}$, $N < \infty$, $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, N$ and any monotone $t_1 < t_2 \cdots < t_N \in \mathbb{R}$. A similar definition applies to discrete time S.P. indexed by t on the integers, just then t_i and τ take only integer values.

It is particularly easy to verify the stationarity of Gaussian S.P. since

Proposition 3.2.25. *A Gaussian S.P. is stationary if and only if $\mu(t) = \mu$ (a constant) and $\rho(t,s) = r(|t-s|)$, where $r : \mathbb{R} \to \mathbb{R}$ is a function of the time difference $|t-s|$. (A stochastic process whose mean and auto-covariance function satisfy these two properties is called* weak sense (or covariance) stationary*. In general, a weak sense stationary process is not a strong sense stationary process (for example, see* [**GS01**, Example 8.2.5]*). However, as the current Proposition shows, the two notions of stationarity are equivalent in the Gaussian case.)*

Convince yourself that Proposition 3.2.25 is an immediate consequence of Corollary 3.2.18 (alternatively, use directly Proposition 3.2.11). For more on stationary Gaussian S.P. solve the following exercise and see [**GS01**, Section 9.6] (or [**Bre92**, Section 11.5] for the case of discrete time).

Exercise 3.2.26. *Suppose $\{X_t\}$ is a zero-mean, (weak sense) stationary process with auto-covariance function $r(t)$.*

(a) *Show that $|r(h)| \le r(0)$ for all $h > 0$.*
(b) *Show that if $r(h) = r(0)$ for some $h > 0$ then $X_{t+h} \overset{a.s.}{=} X_t$ for each t.*
(c) *Explain why part (c) of Exercise 3.2.13 implies that if $\{X_t\}$ is a zero-mean, stationary, Gaussian process with auto-covariance function $r(t)$ such that $r(0) > 0$, then $\mathbf{E}(X_{t+h}|X_t) = \frac{r(h)}{r(0)} X_t$ for any t and $h \ge 0$.*

(d) *Conclude that there is no zero-mean, stationary, Gaussian process of independent increments other than the trivial process* $X_t \equiv X_0$.

Definition 3.2.27. *We say that a process* $\{X_t, t \geq 0\}$ *has* stationary increments *if* $X_{t+h} - X_t$ *and* $X_{s+h} - X_s$ *have the same law for all* $s, t, h, \geq 0$. *The same definition applies to discrete time S.P. indexed by* t *on the integers, just with* t, s *and* h *taking only integer values (and if the S.P. is indexed by a non-negative integer time, then so are the values of* s, t *and* h*).*

Example 3.2.28. *Clearly, a sequence of independent and identically distributed random variables* $\ldots, X_{-2}, X_{-1}, X_0, X_1, \ldots$ *is a discrete time stationary process. However, many processes are not stationary. For example, the random walk* $S_n = \sum_{i=1}^n X_i$ *of Definition 3.1.2 is a non-stationary S.P. when* $\mathbf{E}X_1 = 0$ *and* $\mathbf{E}X_1^2 = 1$. *Indeed, if* $\{S_n\}$ *was a stationary process then the law of* S_n*, and in particular its second moment, would not depend on* n *– in contradiction with (3.1.1). Convince yourself that every stationary process has stationary increments, but note that the random walk* S_n *has stationary increments, thus demonstrating that stationary increments are not enough for stationarity.*

For more on stationary discrete time S.P. see [**Bre92**, Section 6.1], or see [**GS01**, Chapter 9] for the general case.

3.3. Sample path continuity

As we have seen in Section 3.1, the distribution of the S.P. does not specify uniquely the probability of events outside the rather restricted σ-field $\mathcal{F}_{\mathbf{X}}$ and in particular provides insufficient information about the behavior of its supremum, as well as about the continuity of its sample path.

Our goal in this section is thus to find relatively easy to check sufficient conditions for the existence of a modification of the S.P. that has a somewhat "nice" sample paths. The following definition of sample path continuity is the first step in this direction.

Definition 3.3.1. *We say that* $\{X_t\}$ *has* continuous sample path *w.p.1 if* $\mathbf{P}(\{\omega : t \mapsto X_t(\omega)$ *is continuous*$\}) = 1$. *Similarly, we use the term* continuous modification *to denote a modification* $\{\widetilde{X}_t\}$ *of a given S.P.* $\{X_t\}$ *such that* $\{\widetilde{X}_t\}$ *has continuous sample path w.p.1.*

The next definition of *Hölder continuity* provides a quantitative refinement of this notion of continuity, by specifying the maximal possible smoothness of the sample path of X_t.

Definition 3.3.2. *A S.P.* Y_t *is* locally Hölder continuous *with exponent* γ *if for some* $c < \infty$ *and a R.V.* $h(\omega) > 0$,

$$\mathbf{P}(\{\omega : \sup_{0 \leq s,t \leq T,\ |t-s| \leq h(\omega)} \frac{|Y_t(\omega) - Y_s(\omega)|}{|t-s|^\gamma} \leq c\}) = 1.$$

Remark. The word "locally" in the above definition refers to the R.V. $h(\omega)$. When it holds for unrestricted $t, s \in [0, T]$ we say that Y_t is *globally (or uniformly) Hölder continuous* with exponent γ. A particular important special case is that of $\gamma = 1$, corresponding to *Lipschitz continuous* functions.

Equipped with Definition 3.3.2 our next theorem gives a very useful criterion for the existence of a continuous modification (and even yields a further "degree of smoothness" in terms of the Hölder continuity of the sample path).

Theorem 3.3.3 (Kolmogorov's continuity theorem)**.** *Given a S.P.* $\{X_t, t \in [0,T]\}$*, suppose there exist* $\alpha, \beta, c, h_0 > 0$ *such that*

$$\mathbf{E}(|X_{t+h} - X_t|^\alpha) \le ch^{1+\beta}, \qquad \text{for all} \quad 0 \le t,\, t+h \le T,\;\; 0 < h < h_0\,. \tag{3.3.1}$$

Then, there exists a continuous modification Y_t *of* X_t *such that* Y_t *is also locally Hölder continuous with exponent* γ *for any* $0 < \gamma < \beta/\alpha$*.*

Remark. In case you have wondered why exponent γ near $(1+\beta)/\alpha$ does not work in Theorem 3.3.3, read its proof, and in particular, the derivation of [**KS97**, inequality (2.9), page 54]. Or, see [**Oks03**, Theorem 2.6, page 10], for a somewhat weaker result.

It is important to note that condition (3.3.1) of Theorem 3.3.3 involves only the joint distribution of (X_t, X_{t+h}) and as such is verifiable based on the f.d.d. of the process. In particular, either all versions of the given S.P. satisfy (3.3.1) or none of them does.

The following example demonstrates that we must have $\beta > 0$ in (3.3.1) to deduce the existence of a continuous modification.

Example 3.3.4. *Consider the stochastic process* $X_t(\omega) = I_{\{U \le t\}}(\omega)$*,* $t \in [0,1]$*, where* U *is a Uniform*$[0,1]$ *random variable (that is,* $U(\omega) = \omega$ *on the probability space* $(U, \mathbb{R}, \mathcal{B})$ *of Example 1.1.11). Note that* $|X_{t+h} - X_t| = 1$ *if* $0 \le t < U \le t+h$*, and* $|X_{t+h} - X_t| = 0$ *otherwise. So,* $\mathbf{E}(|X_{t+h} - X_t|^\alpha) = U((t, t+h]) \le h$ *for any* $h > 0$ *and* $t \ge 0$*. That is,* $\{X_t, t \ge 0\}$ *satisfies (3.3.1) with* $c = 1$*,* $\beta = 0$ *any* α *and* h_0*. However, clearly the sample path of* $X_t(\omega)$ *is discontinuous at* $t = U(\omega)$ *whenever* $\omega \ne 0$*. That is, almost surely* $\{X_t\}$ *has discontinuous sample paths (and it is further possible to show that this S.P. has no continuous modification).*

The following application of Kolmogorov's continuity theorem demonstrates the importance of choosing wisely the free parameter α in this theorem.

Exercise 3.3.5. *Suppose the stochastic process* X_t *is such that* $\mathbf{E}(X_t) = 0$ *and* $\mathbf{E}(X_t^2) = 1$ *for all* $t \in [0,T]$*.*

(a) *Show that* $|\mathbf{E}(X_t X_{t+h})| \le 1$ *for any* $h > 0$ *and* $t \in [0, T-h]$*.*

(b) *Suppose that for some* $\lambda < \infty$*,* $p > 1$ *and* $h_0 > 0$*,*

$$\mathbf{E}(X_t X_{t+h}) \ge 1 - \lambda h^p \qquad \text{for all } \; 0 < h \le h_0\,. \tag{3.3.2}$$

Using Kolmogorov's continuity theorem show that then X_t *has a continuous modification.*

(c) *Suppose* X_t *is a Gaussian stochastic process such that* $\mathbf{E}(X_t) = 0$ *and* $\mathbf{E}(X_t^2) = 1$ *for all* $t \in [0,T]$*. Show that if* X_t *satisfies the inequality (3.3.2) for some* $\lambda < \infty$*,* $p > 0$ *and* $h_0 > 0$*, then for any* $0 < \gamma < p/2$*, the process* X_t *has a modification which is locally Hölder continuous with exponent* γ*.*

Hint: *see Section 5.1 for the moments of Gaussian random variable.*

As we show next, there exist non-Gaussian S.P.-s satisfying (3.3.2) with $p = 1$ for which there is no continuous modification.

Example 3.3.6. *One such example is the* "random telegraph signal" R_t *which is defined as follows. Let* τ_i, $i = 1, 2, \ldots$ *be independent random times, each having the* Exponential(1) distribution*, that is,* $\mathbf{P}(\tau_i \leq x) = 1 - e^{-x}$ *for all* i *and* $x \geq 0$*. Starting at* $R_0 \in \{-1, 1\}$ *such that* $\mathbf{P}(R_0 = 1) = 1/2$*, the S.P.* R_t *alternately jumps between* -1 *and* $+1$ *at the random times* $s_k = \sum_{i=1}^{k} \tau_i$ *for* $k = 1, 2, 3, \ldots$*, so* $R_t \in \{-1, 1\}$ *keeps the same value in each of the intervals* (s_k, s_{k+1})*. Since almost surely* $s_1 < \infty$*, this S.P. does not have a continuous modification. However,*

$$\mathbf{E}(R_t R_{t+\varepsilon}) = \mathbf{P}(R_t = R_{t+\varepsilon}) - \mathbf{P}(R_t \neq R_{t+\varepsilon}) = 1 - 2\mathbf{P}(R_t \neq R_{t+\varepsilon})\,,$$

and since for any $t \geq 0$ *and* $\varepsilon > 0$*,*

$$\varepsilon^{-1}\mathbf{P}(R_t \neq R_{t+\varepsilon}) \leq \varepsilon^{-1}\mathbf{P}(\tau_i \leq \varepsilon) = \varepsilon^{-1}(1 - e^{-\varepsilon}) \leq 1\,,$$

we see that R_t *indeed satisfies (3.3.2) with* $p = 1$ *and* $\lambda = 2$*.*

The stochastic process R_t of Example 3.3.6 is a special instance of the continuous-time *Markov jump processes*, which we study in Section 6.3. Though the sample path of this process is almost never continuous, it has the right-continuity property of the following definition, as is the case for all continuous-time Markov jump processes of Section 6.3.

Definition 3.3.7. *We say that a S.P.* X_t *has* right-continuous with left limits *(in short, RCLL) sample path, if for a.e.* $\omega \in \Omega$*, the sample path* $X_t(\omega)$ *is right-continuous and of left-limits at any* $t \geq 0$ *(that is, for* $h \downarrow 0$ *both* $X_{t+h}(\omega) \to X_t(\omega)$ *and the limit of* $X_{t-h}(\omega)$ *exists). Similarly, a modification having RCLL sample path with probability one is called* RCLL modification *of the S.P.*

Remark. To practice your understanding, check that any S.P. having continuous sample path also has RCLL sample path (in particular, the Brownian motion of Section 5.1 is such). The latter property plays a major role in continuous-time martingale theory, as we shall see in Sections 4.2 and 4.3.2. For more on RCLL sample path see [**Bre92**, Section 14.2].

Perhaps you expect any two S.P.-s that are modifications of each other to have (a.s.) *indistinguishable* sample path, i.e. $\mathbf{P}(X_t = Y_t \text{ for all } t \in \mathcal{I}) = 1$. This is indeed what happens for discrete time, but in case of continuous time, in general such property may fail, though it holds when both S.P.-s have right-continuous sample paths (a.s.).

Exercise 3.3.8.

(a) *Let* $\{X_n\}$*,* $\{Y_n\}$ *be discrete time S.P.-s that are modifications of each other. Show that* $\mathbf{P}(X_n = Y_n \text{ for all } n \geq 0) = 1$*.*

(b) *Let* $\{X_t\}$*,* $\{Y_t\}$ *be continuous time S.P.-s that are modifications of each other. Suppose that both processes have right-continuous sample paths a.s. Show that* $\mathbf{P}(X_t = Y_t \text{ for all } t \geq 0) = 1$*.*

(c) *Provide an example of two S.P.-s which are modifications of one another but which are not indistinguishable.*

We conclude with a hierarchy of the sample path properties which have been considered here.

Proposition 3.3.9. *The following implications apply for the sample path of any stochastic process:*

$$\text{Hölder continuity} \;\Rightarrow\; \text{Continuous w.p.1} \;\Rightarrow\; \text{RCLL}\,.$$

PROOF. The stated relations between local Hölder continuity, continuity and right-continuity with left limits hold for any function $f : [0, \infty) \to \mathbb{R}$. Considering $f(t) = X_t(\omega)$ for a fixed $\omega \in \Omega$ leads to the corresponding relation for the sample path of the S.P. ∎

All the S.P. of interest to us shall have at least a RCLL modification, hence with all properties implied by it, and unless explicitly stated otherwise, hereafter we always assume that we are studying the RCLL modification of the S.P. in question.

The next theorem shows that objects like $Y_t = \int_0^t X_s ds$ are well defined under mild conditions and is crucial for the successful rigorous development of stochastic calculus.

Theorem 3.3.10 (Fubini's theorem)**.** *If X_t has RCLL sample path and for some interval I and σ-field $\mathcal{H}$, almost surely $\int_I \mathbf{E}[|X_t|\,|\mathcal{H}]dt$ is finite then almost surely $\int_I X_t dt$ is finite and*

$$\int_I \mathbf{E}[X_t|\mathcal{H}]\ dt = \mathbf{E}\ [\int_I X_t\ dt|\mathcal{H}]\,.$$

Remark. Taking $\mathcal{H} = \{\emptyset, \Omega\}$ we get as a special case of Fubini's theorem that $\int_I \mathbf{E}X_t\ dt = \mathbf{E}\ (\int_I X_t\ dt)$ whenever $\int_I \mathbf{E}|X_t|dt$ is finite.

Here is a converse of Fubini's theorem, where the differentiability of the sample path $t \mapsto X_t$ implies the differentiability of the mean $t \mapsto \mathbf{E}X_t$ of a S.P.

Exercise 3.3.11. *Let $\{X_t, t \geq 0\}$ be a stochastic process such that for each $\omega \in \Omega$ the sample path $t \mapsto X_t(\omega)$ is differentiable at any $t \geq 0$.*

(a) *Verify that $\frac{\partial}{\partial t}X_t$ is a random variable for each fixed $t \geq 0$.*

(b) *Suppose further that there is an integrable random variable Y such that $|X_t - X_s| \leq |t - s|Y$ for almost every $\omega \in \Omega$ and all $t, s \geq 0$. Using the dominated convergence theorem, show that $t \mapsto \mathbf{E}X_t$ is then differentiable with a finite derivative such that for all $t \geq 0$,*

$$\frac{d}{dt}\mathbf{E}(X_t) = \mathbf{E}\Big(\frac{\partial}{\partial t}X_t\Big).$$