3 Random signals and probabilistic systems

We have defined stochastic signals in an earlier lecture. A stochastic signal \(X\) is a collection \((X_t)_{t \in T}\) of random variables indexed by a time variable \(t \in T\). Depending on the structure of the set \(T\), we speak of discrete- or continuous-time signals:

- If \(T\) is a subset of the integers \(\mathbb{Z}\), then \(X\) is a discrete-time stochastic signal.
- If \(T\) is a finite or an infinite interval of the real line \(\mathbb{R}\), then \(X\) is a continuous-time stochastic signal.

The set \(X\) of possible values each \(X_t\) can take is called the state space of the process. If the elements of \(X\) can be put in a one-to-one correspondence with a subset of \(\mathbb{Z}\), then we say that \(X\) is a discrete-state process; otherwise, if \(X\) is a finite or an infinite interval of \(\mathbb{R}\), we say that \(X\) is a continuous-state process. It is important to point out that, even though we are using the word “state,” it does not mean that \(X_t\) is a state variable in the sense that it contains all relevant information about the signal up to time \(t\).

Another way to view stochastic signals is as follows. We have a probability space \((\Omega, \mathcal{F}, P)\), such that each \(X_t\) is a random variable with values in \(X\), i.e., a function \(X_t : \Omega \rightarrow X\). This specifies a random variable for each \(t \in T\). On the other hand, fixing \(\omega \in \Omega\) specifies a function from \(T\) into \(X\) — this function simply assigns the value \(X_t(\omega)\) to each \(t \in T\). Thus, one can think of \(X\) as a random function of \(t\), where all the randomness is packed into \(\omega\).

This is best appreciated through examples, so let’s take a look at a few.

**Deterministic signals.** Deterministic signals are a special case of stochastic signals. That is, if \(x : T \rightarrow X\) is a deterministic signal with state space \(X\), we can just take \(X_t(\omega) = x(t)\) for all \(t \in T\) and all \(\omega \in \Omega\). Thus, the value of \(X_t\) at each time \(t\) is known and equal to \(x(t)\) — that is, \(X\) is a function of \(t\). Another way of stating that \(X\) is a deterministic signal is to note that, for any subset \(A\) of the state space \(X\), \(P[X_t \in A]\) takes only two values: 0 if \(x(t) \notin A\) and 1 if \(x(t) \in A\).

**i.i.d. processes.** Take \(T = \mathbb{Z}_+\) and let \(X_0, X_1, \ldots\) be a collection of independent and identically distributed (i.i.d.) random variables. Then \(X = (X_t)_{t \in \mathbb{Z}_+}\) is called an i.i.d. stochastic process, and the common probability distribution of the \(X_t\)’s is called the (one-time) marginal distribution of \(X\). For example, if each \(X_t\) is a Bernoulli\((p)\) random variable, then \(X = (X_t)\) records the outcomes of repeated tosses of a coin with bias \(p\). If we view it as a random function of \(t\), then the value of this function at any time \(t\) can be only 0 or 1. Moreover, because of independence, for any collection of times \(t_1, t_2, \ldots, t_n \in T\) and for any binary values \(x_1, \ldots, x_n \in \{0, 1\}\),

\[
P[X_{t_1} = x_1, \ldots, X_{t_n} = x_n] = \prod_{i=1}^{n} P[X_{t_i} = x_i] = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}.
\]  

\(\text{(3.1)}\)

\(^1\)Also known as stochasti processes – we will use both terms in this course.
Markov processes. We have already encountered Markov processes before. A Markov process with state space $X$ is a discrete time process $X = (X_t)_{t \in \mathbb{Z}}$ that can be realized in the form

$$X_{t+1} = f(X_t, U_t), \quad t = 0, 1, 2, \ldots$$

where $U = (U_t)_{t \in \mathbb{Z}}$ is an i.i.d. process, and the initial condition $X_0$ is independent of all the $U_t$’s. We have mostly focused on discrete-state Markov processes, but $X$ can take real values as well.

**Deterministic signals with stochastic parameters.** Let $A$ and $\Theta$ be two independent random variables, where $A \sim \text{Uniform}(0, 1)$ and $\Theta \in \text{Uniform}(0, 2\pi)$. Let $X_t = A \cos(2\pi ft + \Theta)$, $-\infty < t < \infty$. This is a sinusoidal signal with random amplitude and random phase, and is an example of a predictable stochastic signal: if we can observe $X_t$ over a full period (i.e., any interval of length $1/2\pi f$), then we can reconstruct the entire signal.

### 3.1 Describing stochastic signals

Given a deterministic signal $x : T \to \mathbb{R}$, we can compute its value at any time $t \in T$, at least in principle. With a stochastic signal, though, we can only deal in probabilities. Thus, if we have a discrete-state process, we can specify the probabilities $P[X_t = x]$ for all $t \in T$ and all $x \in X$; if we have a continuous-state process, we can specify the probabilities $P[a \leq X_t \leq b]$ for any finite or infinite subinterval $[a, b]$ of the state space $X$. It’s useful to think about the latter probability as the probability that the graph of the (random) function $t \mapsto X_t$ will “pass” through the “gate” $[a, b]$ at time $t$ – see Figure 1.

![Figure 1: The event \{a \leq X_t \leq b\}.

If we can do this for every $t \in T$, does this give us a complete description of the underlying stochastic signal? The answer is no. To see why, consider the following two discrete-time stochastic processes:

- $X = (X_t)_{t \in \mathbb{Z}}$ is an i.i.d. Bernoulli(1/2) process. That is, each $X_t$ encodes an independent toss of a fair coin.
We claim that these two processes have the same one-time marginal distributions, i.e., \( P(Y_n) \) for all \( n \). Later on, we will also be looking at two signals at a time, say, the autocorrelation function. We are interested in correlations about features of \( X \): the mean \( \mu_X = E[X] \), which is the average value \( \frac{}{} \) and the variance \( \sigma_X^2 = Var[X] \). We can also talk about correlations between the values of \( X \) at different times. This information is encoded in the autocovariance function

\[
R_X(s, t) := E[X_s X_t], \quad \forall s, t \in T
\]

and in the autocovariance function

\[
C_X(s, t) := E[(X_s - \mu_X(s))(X_t - \mu_X(t))] = E[X_s X_t] - m_X(s)m_X(t) = R_X(s, t) - m_X(s)m_X(t).
\]

Later on, we will also be looking at two signals at a time, say, \( X = (X_t) \) and \( Y = (Y_t) \), where \( X \) is the input and \( Y \) is the output of some system. Then we may want to look at the crosscorrelation

\[
R_{XY}(s, t) := E[X_s Y_t]
\]

and the crosscovariance

\[
C_{XY}(s, t) := E[(X_s - m_X(s))(Y_t - m_Y(t))] = R_{XY}(s, t) - m_X(s)m_Y(t).
\]
3.2 Random walk

We can spend all the time in the world on definitions and generalities, but the best way of appreciating key concepts is in the context of examples. This way, we will reinforce the already introduced definitions and motivate some new ones. Our first example is the random walk.

Formally, a random walk is a discrete-time stochastic signal $X = (X_t)_{t \in \mathbb{Z}_+}$ with the deterministic initial condition $X_0 = 0$ (this is imposed mostly for technical convenience) and with the update rule

$$X_{t+1} = X_t + U_t, \quad t = 0, 1, 2, \ldots \tag{3.2}$$

where $U = (U_t)_{t \in \mathbb{Z}_+}$ is some i.i.d. stochastic signal. In other words, $X$ describes a randomly moving particle, which at each time $t$ undergoes a displacement of random length and random direction (left or right) from its current position $X_t \in \mathbb{R}$. Thus, $X$ is a Markov process because it can be realized in the form

$$X_{t+1} = f(X_t, U_t)$$

with $f(x, u) = x + u$, such that the initial condition $X_0$ is independent of $U = (U_t)$.

A random walk has another important property: it has independent increments. To see what this means, fix two times $0 \leq s < t$ and write

$$X_t = X_t - X_0 = (X_t - X_s) + (X_s - X_0).$$

Here, $X_t - X_0$ is the net displacement of the particle at time $t$, and $X_t - X_s$ is the net displacement of the particle between times $s$ and $t$. Now, iterating (3.2), we can write

$$X_t = X_t - X_0 = \sum_{r=0}^{t-1} U_r$$

for every $t$. Therefore,

$$X_t - X_s = (X_t - X_0) - (X_s - X_0) = \sum_{r=0}^{t-1} U_r - \sum_{r=0}^{s-1} U_r = \sum_{r=s}^{t-1} U_r.$$

This tells us that $X_s - X_0$ is a function of $U_0, \ldots, U_{s-1}$, while $X_t - X_s$ is a function of $U_s, \ldots, U_{t-1}$. Since $U$ is an i.i.d. process, and independence is preserved by grouping, we see that the increments $X_t - X_s$ and $X_s - X_0$ are indeed independent. Thus, a random walk is a Markov process with independent increments.

We can easily compute first- and second-order moments of a random walk. Let $\mu = \mathbb{E}[U_0]$ and $\sigma^2 = \text{Var}[U_0]$ denote the common mean and variance of the $U_t$’s. Then, by linearity of expectation, we have

$$m_X(t) = \mathbb{E}[X_t] = \mathbb{E} \left[ \sum_{r=0}^{t-1} U_r \right] = \sum_{r=0}^{t-1} \mathbb{E}[U_r] = t \mu,$$

and
so the mean $m_X(t)$ increases linearly with $t$ if $\mu > 0$, stays at zero if $\mu = 0$, and decreases linearly with $t$ if $\mu < 0$. For the variance, we have

$$\sigma_X^2(t) = \text{Var}[X_t] = \text{Var} \left[ \sum_{r=0}^{t-1} U_r \right].$$

Since the variance of a sum of independent random variables is the sum of the variances, we can write

$$\sigma_X^2(t) = \sum_{r=0}^{t-1} \text{Var}[U_r] = t\sigma^2.$$

Thus, the variance of $X_t$ grows linearly with $t$. The mean $m_X(t)$ tells us where on average we expect to find the particle at time $t$, and the standard deviation $\sigma_X(t) = \sigma \sqrt{t}$ tells us the range of fluctuations of $X_t$. The same reasoning can be used to compute the means and the variances of the increments of $X$. First, let’s compute the mean: for $s < t$,

$$\mathbb{E}[X_t - X_s] = \mathbb{E}[X_t] - \mathbb{E}[X_s] = m_X(t) - m_X(s) = (t - s)\mu.$$

Now, to compute the variance we write

$$\text{Var}[X_t - X_s] = \text{Var} \left[ \sum_{r=s}^{t-1} U_r \right] = \sum_{r=s}^{t-1} \text{Var}[U_r] = (t-s)\sigma^2,$$

once again using the fact that the $U_r$’s are independent. We can also compute the autocorrelation function: Fix $s < t$. Then

$$R_X(s,t) = \mathbb{E}[X_s X_t]$$
$$= \mathbb{E}[(X_s - X_0)(X_t - X_0)]$$
$$= \mathbb{E}[(X_s - X_0)(X_t - X_s + X_s - X_0)]$$
$$= \mathbb{E}[(X_s - X_0)(X_t - X_s)] + \mathbb{E}[(X_s - X_0)^2].$$

Since $X$ has independent increments, we can write

$$\mathbb{E}[(X_s - X_0)(X_t - X_s)] = \mathbb{E}[X_s - X_0]\mathbb{E}[X_t - X_s] = s\mu \cdot (t-s)\mu = s(t-s)\mu^2.$$

Also,

$$\mathbb{E}[(X_s - X_0)^2] = \mathbb{E}[X_s^2] = \sigma_X^2(s) + m_X^2(s) = s\sigma^2 + s^2\mu^2.$$

Putting everything together, we write

$$R_X(s,t) = s(t-s)\mu^2 + s\sigma^2 + s^2\mu^2 = st\mu^2 + s\sigma^2.$$

On the other hand, if $t \leq s$, then we obtain

$$R_X(s,t) = st\mu^2 + t\sigma^2.$$
Expressed more succinctly, the autocorrelation function of our random walk $X$ is given by

$$R_X(s, t) = st\mu^2 + \sigma^2 \min(s, t).$$  \hfill (3.3)

The autocovariance is then given by

$$C_X(s, t) = R_X(s, t) - m_X(s)m_X(t) = \sigma^2 \min(s, t).$$  \hfill (3.4)

Next, we will use random walks as building blocks of two important types of continuous-time stochastic signals: the Wiener process and the Poisson process.

### 3.3 The Wiener process

Recall our example of the symmetric random walk on the integers: at time $t = 0$ a particle starts at $x = 0$ and, at each time $t = 1, 2, \ldots$ it hops either one step to the left or one step to the right with equal probability. This is a random walk whose increments are sums of independent Rademacher random variables (remember that a Rademacher random variable takes values $\pm 1$ with probability $\frac{1}{2}$). Now suppose that the particle hops every $\tau$ seconds, and each hop is $\pm h$ meters. This gives us a continuous-time stochastic process $X = (X_t)_{t \in \mathbb{R}_+}$ that can be described in terms of an i.i.d. Rademacher sequence $U = (U_k)_{k \in \mathbb{Z}_+}$ as follows:

$$X_t = 0, \quad \text{for } 0 \leq t < \tau$$  \hfill (3.5)

and

$$X_t = h(U_0 + \ldots + U_{n-1}), \quad \text{if } n\tau \leq t < (n+1)\tau \text{ for } n = 1, 2, \ldots.$$  \hfill (3.6)

A typical path of $X$ is shown in Figure 2.

![Figure 2: A continuous-time interpolation of a random walk.](image)

We will now consider the case when both the times between hops and the sizes of hops become infinitesimally small, i.e., when $\tau \to 0$ and $h \to 0$. The continuous-time stochastic signal that results in the limit is known as the Wiener process or as the Brownian motion.

The Wiener process has a rich history. In 1828, the English botanist Robert Brown noticed a curious thing: when he looked through his microscope at some particles of pollen suspended in
water, he noticed that the particles were constantly and irregularly moving. At first, he attributed this to the pollen somehow being alive, but he repeated the same experiment with inanimate particles (such as sawdust) and observed the same result. No one could explain this phenomenon until the early 1900’s, when Albert Einstein in 1905 and, independently of him, Marian Smoluchowski in 1906 came up with an explanation based on the atomic hypothesis and gave a mathematical description of this Brownian motion of particles in water. The explanation was as follows: Imagine a vessel containing water at room temperature. If the atomic hypothesis is false, then water is a smooth continuous medium, and, by symmetry, we should not expect the pollen particles to move around. On the other hand, if the atomic hypothesis is true, then water is not a continuous fluid but consists of discrete molecules, and these molecules will be undergoing constant random agitation due to heat. If we suspend pollen in the water, then its particles will be kicked around by the randomly moving water molecules. This is what Robert Brown saw when he looked in his microscope. Both Einstein and Smoluchowski assumed that each such kick is delivered by a different molecule, and that the random motions of different molecules are independent. Now, a particle of pollen is immense compared to a molecule of water, so each kick will be very weak; moreover, being so tiny compared to pollen particles, the water molecules will move very fast, so the time between different kicks will be very short. Our stochastic process $X$ defined in (3.5) and (3.6) is a simplified model of Brownian motion of a single particle, where we assume that the motion takes place in one dimension. Einstein’s model made some concrete quantitative predictions, which were verified experimentally by Jean Perrin. For this work, Perrin was awarded the Nobel Prize in physics in 1926. After Perrin’s experiments, the atomic hypothesis, first advanced by John Dalton in the early 1800’s, was finally settled.

It should be pointed out that neither Einstein nor Smoluchowski were mathematically rigorous in their derivations (and that wasn’t the point). The first rigorous analysis of Brownian motion was published by Norbert Wiener in 1921, hence the name “Wiener process.” These days, the Wiener process is ubiquitous in stochastic modeling. Apart from the Brownian motion, it is used in physics to model random diffusion phenomena, in engineering to describe thermal noise in electric conductors, and in finance to model fluctuations of stock prices. In fact, the same mathematical model was used by Louis Bachelier in 1900 (before Einstein and Smoluchowski) to model the fluctuations in prices of stock options on the Paris exchange. We will come back to some of these uses of the Wiener process later in the course.

Now let us return to analyzing the process $X$. As already mentioned, we will pass to the limit as $h, \tau \to 0$. However, we cannot treat these time and space parameters as independent. Let’s see why. From (3.6), we can compute the variance of $X_t$: if $n\tau \leq t < (n+1)\tau$, then

$$\text{Var}[X_t] = \text{Var}[h(U_0 + \ldots + U_{n-1})]$$
$$= h^2 \text{Var}[U_0 + \ldots + U_{n-1}]$$
$$= nh^2,$$

where we have used the fact that the Rademacher random variables $U_k$ are independent, and each has unit variance. Now, if $\tau$ is very small, then $n = \lfloor t/\tau \rfloor \approx t/\tau$, so we can approximate $\text{Var}[X_t] \approx$ 

---

3Today, every schoolchild knows that matter consists of atoms. The atomic hypothesis was still a matter of controversy in the early XX century.
When $P\tau$ whenever this probability is nonzero. Now, we are interested in the limit when the particle hopped left and when it hopped right, as long as the end result is the same. For will keep things simple and just assume that (3.7), we can write

in this limit.

$Z$ and is referred to as the binomial random variable

heads

probability

in

example, the path in Figure 2 was generated by the following pattern of hops: rllrrrrlrrll. There are $\binom{n}{k}$ patterns of signs for $U_0,\ldots,U_{n-1}$, so that exactly $k$ are positive and $n-k$ are negative, and each sign is equally likely. Therefore, we arrive at the following:

$$
\mathbb{P}[X_{n\tau} = mh] = \begin{cases} 
\frac{1}{2^n} \binom{n}{\frac{m+n}{2}}, & \text{if } m = -n, 2-n, \ldots, n-2, n \\
0, & \text{otherwise}
\end{cases}
$$

(3.7)

Since we are concerned with the net displacement of the particle after $n$ hops, it does not matter when the particle hopped left and when it hopped right, as long as the end result is the same. For example, the path in Figure 2 was generated by the following pattern of hops: rllrrrrlrrll. There are $\binom{n}{k}$ patterns of signs for $U_0,\ldots,U_{n-1}$, so that exactly $k$ are positive and $n-k$ are negative, and each sign is equally likely. Therefore, we arrive at the following:

$$
\mathbb{P}[X_{n\tau} = mh] = \begin{cases} 
\frac{1}{2^n} \binom{n}{\frac{m+n}{2}}, & \text{if } m = -n, 2-n, \ldots, n-2, n \\
0, & \text{otherwise}
\end{cases}
$$

When $\mathbb{P}[X_{n\tau} = mh]$ is nonzero, it can be expressed in terms of the binomial distribution, which arises in the following way. Suppose we have a coin with bias $p$, and we toss it $n$ times. The outcome of the $i$th toss is described by an independent Bernoulli$(p)$ random variable $B_i$, taking the value 1 with probability $p$ and 0 with probability $1-p$. Thus, $Z = \sum_{i=1}^{n} B_i$ is the number of times the coin comes up heads in $n$ tosses. It can take values in the set $\{0, 1, \ldots, n\}$ with probabilities

$$
\mathbb{P}[Z = k] = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n
$$

and is referred to as the binomial random variable with parameters $n$ and $p$. Symbolically, we write $Z \sim \text{Binom}(n, p)$ to denote this fact. In particular, examining our expression for $\mathbb{P}[X_{n\tau} = mh]$ in (3.7), we can write

$$
\mathbb{P}[X_{n\tau} = mh] = \mathbb{P}\left[\text{Binom}(n, 1/2) = \frac{m+n}{2}\right]
$$

(3.8)

whenever this probability is nonzero. Now, we are interested in the limit $\tau \to 0$, which corresponds to letting $n = t/\tau$ approach infinity. So, we need to see what happens to the right-hand side of (3.8) in this limit.
To that end, we will use a remarkable asymptotic expression for the binomial distribution, called the DeMoivre–Laplace theorem. Imagine repeatedly tossing a coin with bias \( p \) and recording \( Z \), the number of heads in \( n \) tosses. We know that \( Z \sim \text{Binom}(n, p) \), and thus \( \mathbb{E}[Z] = np \) and \( \text{Var}[Z] = np(1 - p) \). That is, on average we expect the proportion of heads in \( n \) tosses to be somewhere around \( p \), and the standard deviation \( \sqrt{np(1 - p)} \) gives us an idea of the fluctuations of \( Z \) around its mean. Suppose we repeat such an experiment a large number of times (say, \( N \)) and let \( Z_j \) be the number of heads out of \( n \) tosses in the \( j \)th experiment. Then, if for each \( k \in \{0, 1, \ldots, n\} \) we plot the number of times \( j \in \{1, \ldots, N\} \) when \( Z_j \) took the value \( k \), we would end up with a nice bell-shaped plot with the peak around \( np \) and with rapid decay of the height of the plot for \( |k - np| > \sqrt{np(1 - p)} \). The DeMoivre–Laplace theorem expresses this quantitatively and says the following: for all sufficiently large values of \( n \) and for all \( |k - np| \leq \sqrt{np(1 - p)} \),

\[
\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right). \tag{3.9}
\]

In our case, \( p = 1/2 \) and \( k = \frac{m+n}{2} \), so, for \( n \) large and for \( |m| \leq \sqrt{n} \), Eq. (3.9) gives

\[
\mathbb{P}[X_{nt} = mh] \approx \frac{1}{\sqrt{\pi n}} \exp\left(-\frac{m^2}{2n}\right). \tag{3.10}
\]

Now, in the limit \( \tau \to 0 \), we will obtain a continuous-time stochastic signal \( W = (W_t)_{t \in \mathbb{R}_+} \), which is our Wiener process. Unlike \( X \), whose state space is discrete and consists of integer multiples of \( h \), the state space of \( W \) will be the entire real line. Hence, as we pass to the limit of \( \tau \to 0 \), the probability mass function \( \mathbb{P}[X_t = mh] \) will become a probability density function \( f_t(x) \) of \( W_t \) — that is, by definition, \( f_t(x) \, dx \approx \mathbb{P}[x \leq W_t < x + dx] \), where \( dx \) denotes an infinitesimally small change in \( x \). More formally,

\[
f_t(x) = \lim_{\delta \to 0} \frac{\mathbb{P}[x \leq W_t < x + \delta]}{\delta}.
\]

Now, for a fixed value of \( \tau \), \( \mathbb{P}[X_{nt} = mh] = \mathbb{P}[mh \leq X_{nt} < (m + 2)h] \) — since \( m \) can take the values \( \{0, -n, -n + 2, \ldots, n - 2, n\} \), any two possible values of \( X_{nt} \) are at least \( 2h \) apart. So, if we fix \( t \) and \( x \), let \( \tau = t/n \), \( h = x = \sqrt{\frac{\Delta t}{n}} \), and \( \delta = 2h \), and then take the limit as \( n \to \infty \), then we will obtain

\[
f_t(x) = \lim_{n \to \infty} \frac{\mathbb{P}[mh \leq X_{nt} < (m + 2)h]}{2h} = \lim_{n \to \infty} \frac{1}{2h\sqrt{\pi n/2}} \exp\left(-\frac{m^2}{2n}\right) = \lim_{n \to \infty} \frac{1}{\sqrt{2\pi nh^2}} \exp\left(-\frac{m^2}{2n}\right) = \frac{1}{\sqrt{2\pi D}} \exp\left(-\frac{x^2}{2Dt}\right),
\]

where we have used (3.10). This shows that the marginal distribution of $W_t$ is Gaussian with mean 0 and variance $Dt$.

Moreover, let us again return to our process $X = (X_t)_{t \in \mathbb{R}_+}$ for fixed finite values of $h$ and $r$. Then, using (3.6), for any three times $r \leq s \leq t$ we can write

$$X_r = h\left(U_0 + \ldots + U_{[r/r]} - 1\right)$$

and

$$X_t - X_s = h\left(U_0 + \ldots + U_{[t/r]} - 1\right) - h\left(U_0 + \ldots + U_{[s/r]} - 1\right) = h\left(U_{[s/r]} + \ldots + U_{[t/r]} - 1\right).$$

Since $[r/r] - 1 < [s/r]$, we see that $X_r$ and the increment $X_t - X_s$ are functions of two disjoint subsets of the $U_k$'s, which are independent. Therefore, $X_t - X_s$ is independent of all $\{X_r\}_{r \in \mathbb{R}_+}$, so $X$ has independent increments. Passing to the limit as $\tau \to 0$, we see that the Wiener process $W$ also has independent increments: for any three times $r \leq s \leq t$, the increments $W_t - W_s$ and $W_s - W_r$ are independent. Another immediate observation that follows from (3.6) is that, for $s \leq t$, $X_t - X_s$ has the same distribution as $X_{t-s}$. Therefore, the same holds in the limit for the Wiener process: The increment $W_t - W_s$ has the same distribution as $W_{t-s}$, which, as we have already seen, is Gaussian with mean 0 and variance $D(t-s)$. We then say that $W$ has stationary increments: the distribution of $W_t - W_s$ depends only on $t-s$, and therefore the increment $W_{t+r} - W_{s+r}$ has the same distribution as $W_t - W_s$ for any $r \in \mathbb{R}$, such that $s+r \geq 0$.

Before moving on, let us summarize the three key properties of the Wiener process $W = (W(t))_{t \geq 0}$:

1. $W_0 = 0$ (zero initial condition)
2. For any three times $0 \leq r \leq s \leq t$, $W_t - W_s$ is independent of $W_r$ (independent increments)
3. For any $0 \leq s \leq t$, $W_t - W_s \sim N(0, D(t-s))$ (stationary Gaussian increments)

Another key property of the Wiener process (which we will not prove) is that we can always perform the passage to the limit in such a way that its sample paths are continuous functions of $t$. Then it can also be proved that they are nowhere differentiable. A typical sample path of a Wiener process is shown in Figure 1. Finally, the Wiener process is Markov in the sense that, for any $t \geq 0$, the future of the process, i.e., $(W_s)_{s \geq t}$, is determined by $W_t$ only, and not by the entire past $(W_s)_{s \leq t}$. You can easily convince yourselves of this by looking at the discrete-time random-walk approximations (for finite values of $\tau$ and $h$) — and we know that these processes are Markov. In general, the definition of a continuous-time Markov process is a bit technical, but we can state it as follows: Consider a continuous-time stochastic process $X = (X_t)_{t \in T}$. For any increasing sequence of “sampling times” $t_0 < t_1 < \ldots$ in $T$, we can define a discrete-time process $Y = (Y_k)_{k \in \mathbb{Z}_+}$ by taking $Y_k = X_{t_k}$. Then we say that $X$ is Markov if the sampled process $Y$ is Markov for any choice of the sampling instants.

Let us apply this criterion to the Wiener process. For any sequence of sampling instants $t_0 < t_1 < \ldots$, we can write

$$Y_{k+1} = Y_k + (Y_{k+1} - Y_k) = W_{t_k} - W_0 + W_{t_{k+1}} - W_{t_k}.$$
Since $W$ has independent increments, $Y_k = W_{t_k} - W_0$ and $W_{t_{k+1}} - W_{t_k}$ are independent. In particular, $Y_0 = W_{t_0} - W_0$ is independent of $Y_1 - Y_0$, $Y_2 - Y_1$, etc. Moreover, $W_{t_{k+1}} - W_{t_k} \sim N(0, D(t_{k+1} - t_k))$. Now, for each $k$ define a function
\[ f_k(y, u) := y + \sqrt{D(t_{k+1} - t_k)}u, \]
and let $U = (U_k)_{k \in \mathbb{Z}}$ be a sequence of i.i.d. $N(0, 1)$ random variables, which are also independent of $Y_0 = X_{t_0}$. Then we can realize $Y$ using the imperative model
\[ Y_{k+1} = f_k(Y_k, U_k). \]
Thus, $Y$ is a time-inhomogeneous Markov process. Since this can be done for any sequence of sampling instants, we see that the Wiener process is, indeed, Markov.