Thomas Mikosch "Elementary Stochastic Calculus" (1998)
Published by World Scientific

Example 1.1.13 (Autocorrelations of a time series)
For a time series $X_0, X_1, X_2, \ldots$ the autocorrelation at lag h is defined by $\text{corr}(X_0, X_h)$, $h = 0, 1, \ldots$. A claim which can frequently be found in the literature is that financial time series (derived from stock indices, share prices, exchange rates, etc.) are nearly uncorrelated. This is supported by the sample autocorrelations of the daily log-returns X_t of the S&P index; see Figure 1.1.14. In contrast to this observation, the estimated autocorrelations of the absolute values $|X_t|$ are different from zero even for large lags h. This indicates that there is dependence in this time series. □

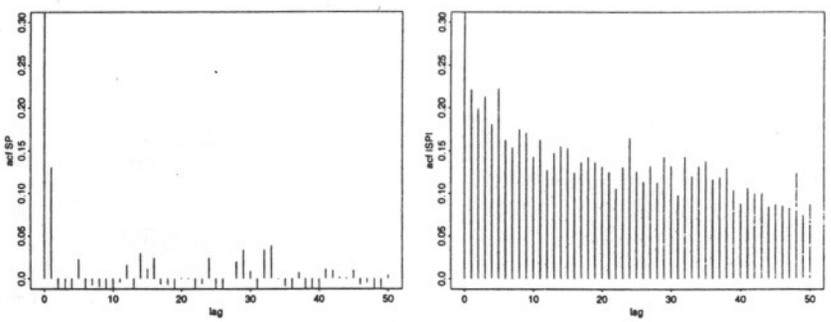

Figure 1.1.14 *The estimated autocorrelations of the S&P index* (left) *and of its absolute values* (right)*; see Example 1.1.13; cf. the comments in Figure 1.1.4.*

In what follows, we will often deal with infinite collections $(X_t, t \in T)$ of random variables X_t, i.e. T is an infinite index set. In this set-up, we may also introduce independence:

The collection of random variables $(X_t, t \in T)$ is *independent* if for every choice of distinct indices $t_1, \ldots, t_n \in T$ and $n \geq 1$ the random variables $X_{t_1}, \ldots, X_{t_n}$ are independent. This collection is *independent and identically distributed* (*iid*) if it is independent and all random variables X_t have the same distribution.

Notes and Comments

In this section we recalled some elementary probability theory which can be found in every textbook on the topic; see for instance Pitman (1993) for an elementary level and Gut (1995) for an intermediate course. Also many textbooks on statistics often begin with an introduction to probability theory; see for example Mendenhall, Wackerly and Scheaffer (1990).

Appendix 0

1.2 Stochastic Processes

We suppose that the exchange rate NZ$/US$ at every fixed instant t between 9 a.m. and 10 a.m. this morning is random. Therefore we can interpret it as a realization $X_t(\omega)$ of the random variable X_t, and so we observe $X_t(\omega)$, $9 \leq t \leq 10$. In order to make a guess at 10 a.m. about the exchange rate $X_{11}(\omega)$ at 11 a.m. it is reasonable to look at the whole evolution of $X_t(\omega)$ between 9 a.m. and 10 a.m. This is also a demand of the high standard technical devices which provide us with almost continuous information about the process considered. A mathematical model for describing such a phenomenon is called a stochastic process.

A *stochastic process* X is a collection of random variables

$$(X_t\,,\, t \in T) = (X_t(\omega)\,,\, t \in T\,,\, \omega \in \Omega)\,,$$

defined on some space Ω.

For our purposes, T is often an interval, for example $T = [a, b]$, $[a, b)$ or $[a, \infty)$ for $a < b$. Then we call X a *continuous-time* process in contrast to *discrete-time* processes. In the latter case, T is a finite or countably infinite set. For obvious reasons, the index t of the random variable X_t is frequently referred to as *time*, and we will follow this convention.

A stochastic process X is a function of two variables.
For a fixed instant of time t, it is a random variable:

$$X_t = X_t(\omega)\,, \quad \omega \in \Omega\,.$$

For a fixed random outcome $\omega \in \Omega$, it is a function of time:

$$X_t = X_t(\omega)\,, \quad t \in T\,.$$

This function is called a *realization*, a *trajectory* or a *sample path* of the process X.

These two aspects of a stochastic process are illustrated in Figure 1.2.1.

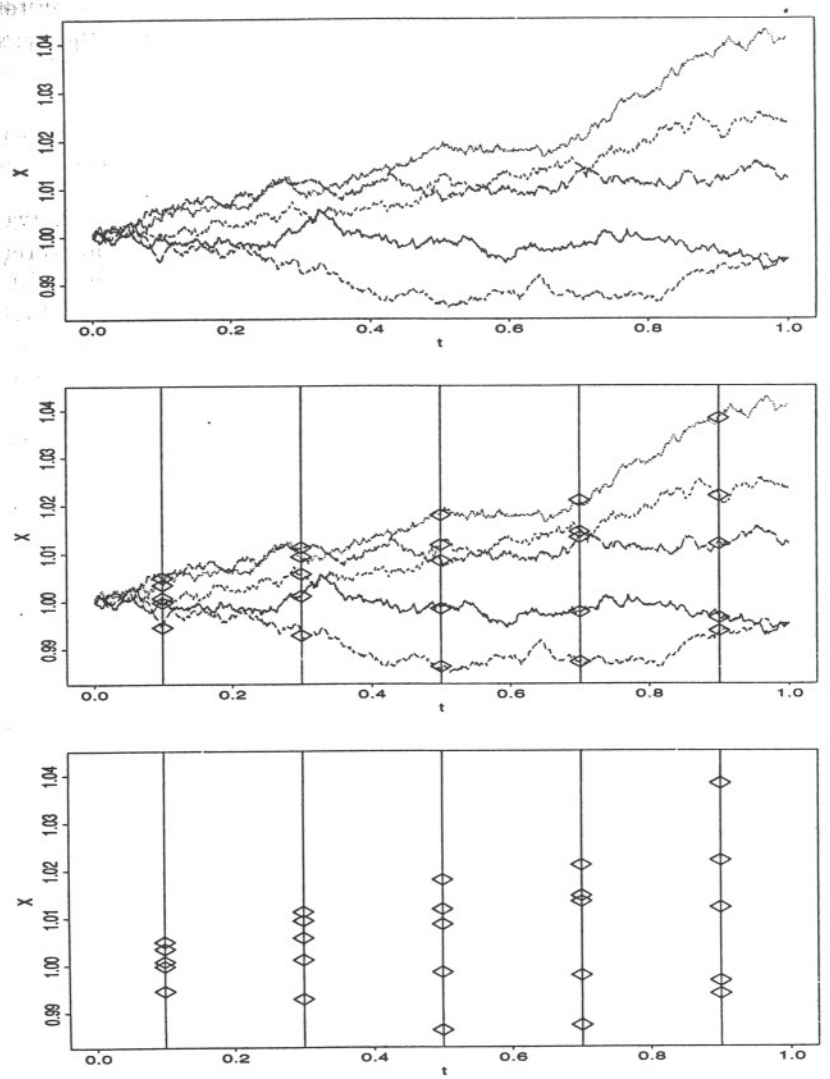

Figure 1.2.1 5 *sample paths of a stochastic process* $(X_t, t \in [0,1])$. Top: *every path corresponds to a different* $\omega \in \Omega$. Middle and bottom: *the values on the vertical lines at* $t = 0.1, \ldots, 0.9$ *visualize the random variables* $X_{0.1}, \ldots, X_{0.9}$*; they occur as the projections of the sample paths on the vertical lines.*

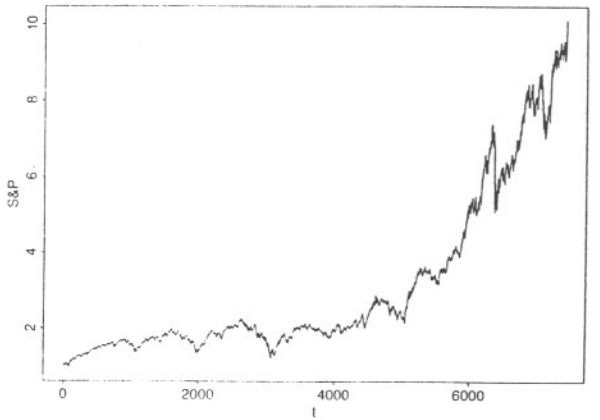

Figure 1.2.2 *The (scaled) daily values of the S&P index over a period of* 7,422 *days. The graph suggests that we consider the S&P time series as the sample path of a continuous-time process. If there are many values in a time series such that the instants of time* $t \in T$ *are "dense" in an interval, then one may want to interpret this discrete-time process as a continuous-time process. The sample paths of a real-life continuous-time process are always reported at discrete instants of time. Depending on the situation, one has to make a decision which model (discrete- or continuous-time) is more appropriate.*

Example 1.2.3 *A time series*

$$X_t\,, \quad t = 0, \pm 1, \pm 2, \ldots,$$

is a discrete-time process with $T = \mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$. Time series constitute an important class of stochastic processes. They are relevant models in many applications, where one is interested in the evolution of a real-life process. Such series represent, for example, the daily body temperature of a patient in a hospital, the daily returns of a price or the monthly number of air traffic passengers in the US. The most popular theoretical time series models are the *ARMA* (**A**uto**R**egressive **M**oving **A**verage) processes. They are given by certain difference equations in which an iid sequence (Z_t) (see p. 22), the so-called *noise*, is involved. For example, a moving average of order $q \geq 1$ is defined as

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}\,, \quad t \in \mathbb{Z}\,,$$

and an autoregressive process of order 1 is given by

$$X_t = \phi X_{t-1} + Z_t\,, \quad t \in \mathbb{Z}\,.$$

Here $\theta_1, \ldots, \theta_q$ and ϕ are given real parameters. Time series models can be understood as discretizations of stochastic differential equations. We will see this for the autoregressive process on p. 141.

Figure 1.2.4 shows two examples. □

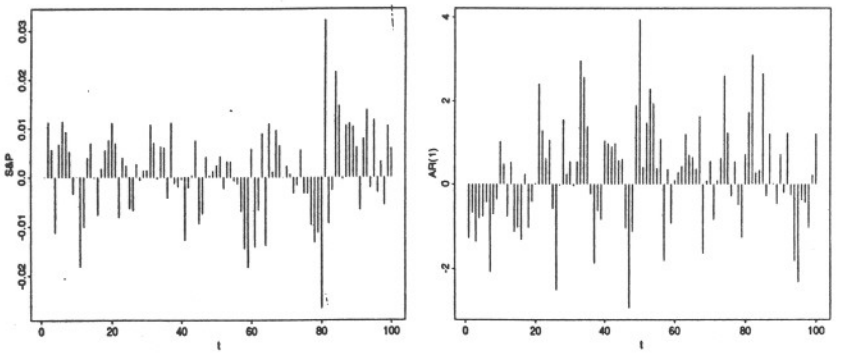

Figure 1.2.4 *Two time series X_t, $t = 1, \ldots, 100$.* Left: 100 *successive daily log-returns of the S&P index; see Figure 1.1.4.* Right: *a simulated sample path of the autoregressive process $X_t = 0.5X_{t-1} + Z_t$, where Z_t are iid $N(0,1)$ random variables; see Example 1.2.3.*

We see that the concepts of a random variable X and of a stochastic process $(X_t, t \in T)$ are not so much different. Both have random realizations, but the realization $X(\omega)$ of a random variable is a number, whereas the realization $X_t(\omega), t \in T$, of a stochastic process is a function on T. So we are completely correct if we understand a stochastic process to be a "random element" taking functions as values. Moreover, we can interpret a random variable and a random vector as special stochastic processes with a finite index set T.

Distribution

In analogy to random variables and random vectors we want to introduce non-random characteristics of a stochastic process such as its distribution, expectation, etc. and describe its dependence structure. This is a task much more complicated than the description of a random vector. Indeed, a non-trivial stochastic process $X = (X_t, t \in T)$ with infinite index set T is an infinite-dimensional object; it can be understood as the infinite collection of the random variables X_t, $t \in T$. Since the values of X are functions on T, the *distribution of* X should be defined on subsets of a certain "function space", i.e.

$$P(X \in A)\,, \quad A \in \mathcal{F}\,, \tag{1.8}$$

where $\mathcal{F}$ is a collection of suitable subsets of this space of functions. This approach is possible, but requires advanced mathematics, and so we try to find some simpler means.

The key observation is that a stochastic process can be interpreted as a collection of random vectors.

> The *finite-dimensional distributions* (*fidis*) of the stochastic process X are the distributions of the finite-dimensional vectors
>
> $$(X_{t_1}, \ldots, X_{t_n})\,, \quad t_1, \ldots, t_n \in T\,,$$
>
> for all possible choices of times $t_1, \ldots, t_n \in T$ and every $n \geq 1$.

We can imagine the fidis much easier than the complicated distribution (1.8) of a stochastic process. It can be shown that the fidis determine the distribution of X. In this sense, we refer to the collection of the fidis as the *distribution of the stochastic process*.

Stochastic processes can be classified according to different criteria. One of them is the kind of fidis.

Example 1.2.5 (Gaussian process)
Recall from (1.6) the definition of an n-dimensional Gaussian density. A stochastic process is called *Gaussian* if all its fidis are multivariate Gaussian. We learnt in Example 1.1.9 that the parameters μ and Σ of a Gaussian vector are its expectation and covariance matrix, respectively. Hence the distribution of a Gaussian stochastic process is determined only by the collection of the expectations and covariance matrices of the fidis.

A simple Gaussian process on $T = [0,1]$ consists of iid $N(0,1)$ random variables. In this case the fidis are characterized by the distribution functions

$$\begin{aligned} & P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n) \\ &= P(X_{t_1} \leq x_1) \cdots P(X_{t_n} \leq x_n) \\ &= \Phi(x_1) \cdots \Phi(x_n) \end{aligned}$$

$$0 \le t_1 \le \cdots \le t_n \le 1\,, \quad (x_1, \ldots, x_n) \in \mathbb{R}^n\,.$$

The sample paths of this process are very irregular. See Figure 1.2.6 for an illustration. □

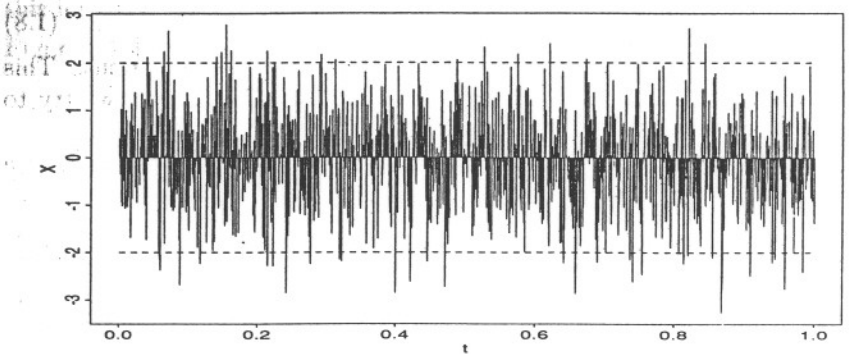

Figure 1.2.6 *A sample path of the Gaussian process $(X_t, t \in [0,1])$, where the X_ts are iid $N(0,1)$; see Example 1.2.5. The expectation function is $\mu_X(t) = 0$ and the dashed lines indicate the curves $\pm 2\sigma_X(t) = \pm 2$; see Example 1.2.7.*

Expectation and Covariance Function

For a random vector $\mathbf{X} = (X_1, \ldots, X_n)$ we defined the expectation $\mu_{\mathbf{X}} = (EX_1, \ldots, EX_n)$ and the covariance matrix $\Sigma_{\mathbf{X}} = (\mathrm{cov}(X_i, X_j)\,, i, j = 1, \ldots, n)$. A stochastic process $X = (X_t, t \in T)$ can be considered as the collection of the random vectors $(X_{t_1}, \ldots, X_{t_n})$ for $t_1, \ldots, t_n \in T$ and $n \ge 1$. For each of them we can determine the expectation and covariance matrix. Alternatively, we can consider these quantities as functions of $t \in T$:

The *expectation function* of X is given by

$$\mu_X(t) = \mu_{X_t} = EX_t\,, \quad t \in T\,.$$

The *covariance function* of X is given by

$$c_X(t,s) = \mathrm{cov}(X_t, X_s) = E[(X_t - \mu_X(t))\,(X_s - \mu_X(s))]\,, \quad t, s \in T\,.$$

The *variance function* of X is given by

$$\sigma_X^2(t) = c_X(t,t) = \mathrm{var}(X_t)\,, \quad t \in T\,.$$

We learnt in Example 1.2.5 that Gaussian processes are determined only via their expectation and covariance functions. This is not correct for a non-Gaussian process.

As for a random vector, the expectation function $\mu_X(t)$ is a deterministic quantity around which the sample paths of X are concentrated. The covariance function $c_X(t,s)$ is a measure of dependence in the process X. The variance function $\sigma_X^2(t)$ can be considered as a measure of spread of the sample paths of X around $\mu_X(t)$. In contrast to the one-dimensional case, a statement like "95% of all sample paths lie between the graphs of $\mu_X(t) - 2\sigma_X(t)$ and $\mu_X(t) + 2\sigma_X(t)$" is very difficult to show (even for Gaussian processes), and is in general not correct. We will sometimes consider computer graphs with paths of certain stochastic processes and also indicate the curves $\mu_X(t)$ and $\mu_X(t) \pm 2\sigma_X(t)$, $t \in T$. The latter have to be interpreted for every fixed t, i.e. for every individual random variable X_t. Only in a heuristic sense, do they give bounds for the paths of the process X. See Figure 1.2.6 for an illustration.

Example 1.2.7 (Continuation of Example 1.2.5)
Consider the Gaussian process $(X_t, t \in [0,1])$ of iid $N(0,1)$ random variables X_t. Its expectation and covariance functions are given by

$$\mu_X(t) = 0 \quad \text{and} \quad c_X(t,s) = \begin{cases} 1 & \text{if} \quad t = s\,, \\ 0 & \text{if} \quad t \neq s\,. \end{cases}$$ □

Dependence Structure

We have already introduced Gaussian processes by specifying their fidis as multivariate Gaussian. Another way of classifying stochastic processes consists of imposing a special dependence structure.

The process $X = (X_t, t \in T)$, $T \subset \mathbb{R}$, is *strictly stationary* if the fidis are invariant under shifts of the index t:

$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_n+h}) \tag{1.9}$$

for all possible choices of indices $t_1, \ldots, t_n \in T$, $n \ge 1$ and h such that $t_1 + h, \ldots, t_n + h \in T$. Here $\stackrel{d}{=}$ stands for the identity of the distributions; see p. 211 for the definition. For the random vectors in (1.9) this means that their distribution functions are identical.

Example 1.2.8 (Stationary Gaussian processes)
Consider a process $X = (X_t, t \in T)$ with $T = [0, \infty)$ or $T = \mathbb{Z}$. A trivial example of a strictly stationary process is a sequence of iid random variables X_t, $t \in \mathbb{Z}$. Since a Gaussian process X is determined by its expectation and covariance functions, condition (1.9) reduces to

$$\mu_X(t+h) = \mu_X(t) \quad \text{and} \quad c_X(t,s) = c_X(t+h, s+h)$$

for all $s, t \in T$ such that $s+h, t+h \in T$. But this means that $\mu_X(t) = \mu_X(0)$ for all t, whereas $c_X(t,s) = \widetilde{c}_X(|t-s|)$ for some function $\widetilde{c}_X$ of one variable. Hence, for a Gaussian process, strict stationarity means that the expectation function is constant and the covariance function only depends on the distance $|t-s|$. More generally, if a (possibly non-Gaussian) process X has the two aforementioned properties, it is called a *stationary (in the wide sense)* or *(second-order) stationary* process. □

If we describe a real-life process by a (strictly or in the wide sense) stationary stochastic process, then we believe that the characteristic properties of this process do not change when time goes by. The dependence structure described by the fidis or the covariance function is invariant under shifts of time. This is a relatively strong restriction on the underlying process. However, it is a standard assumption in many probability related fields such as statistics and time series analysis.

Stationarity can also be imposed on the increments of a process. The process itself is then not necessarily stationary.

> Let $X = (X_t, t \in T)$ be a stochastic process and $T \subset \mathbb{R}$ be an interval.
> X is said to have *stationary increments* if
>
> $$X_t - X_s \stackrel{d}{=} X_{t+h} - X_{s+h} \text{ for all } t, s \in T \text{ and } h \text{ with } t+h, s+h \in T.$$
>
> X is said to have *independent increments* if for every choice of $t_i \in T$ with $t_1 < \cdots < t_n$ and $n \geq 1$,
>
> $$X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$$
>
> are independent random variables.

One of the prime examples of processes with independent, stationary increments is the homogeneous Poisson process. Homogeneity is here another wording for stationarity of the increments.

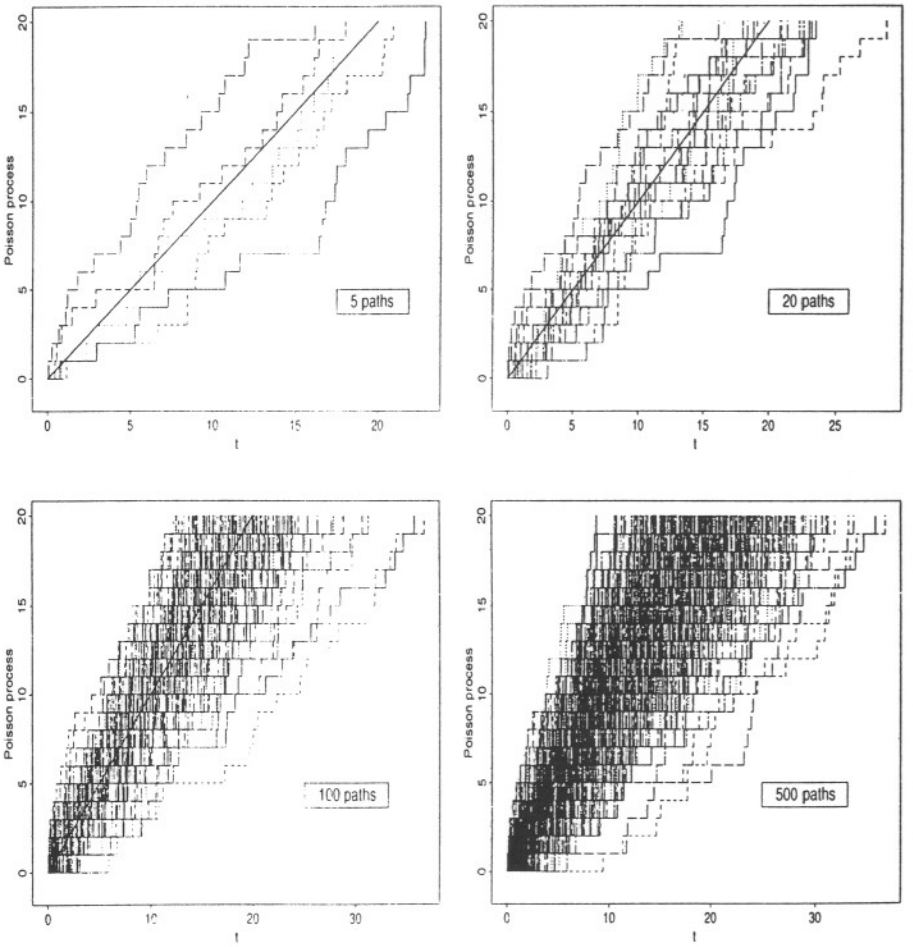

Figure 1.2.9 *Sample paths of a homogeneous Poisson process $(X_t, t \in [0, \infty))$ with intensity $\lambda = 1$; see Example 1.2.10. The straight solid line stands for the expectation function $\mu_X(t) = t$.*

Example 1.2.10 (Homogeneous Poisson process)
A stochastic process $(X_t, t \in [0,\infty))$ is called an *homogeneous Poisson process* or simply a *Poisson process with intensity* or *rate* $\lambda > 0$ if the following conditions are satisfied:

- It starts at zero: $X_0 = 0$.
- It has stationary, independent increments.
- For every $t > 0$, X_t has a Poisson $Poi(\lambda t)$ distribution; see Example 1.1.1 for the definition of the Poisson distribution.

Figure 1.2.9 shows several Poissonian sample paths.

Notice that, by stationarity of the increments, $X_t - X_s$ with $t > s$ has the same distribution as $X_{t-s} - X_0 = X_{t-s}$, i.e. a $Poi(\lambda(t-s))$ distribution.

An alternative definition of the Poisson process is given by

$$X_t = \#\{n : T_n \le t\}, \quad t > 0, \tag{1.10}$$

where $\#A$ denotes the number of elements of any particular set A, $T_n = Y_1 + \cdots + Y_n$ and (Y_i) is a sequence of iid exponential $Exp(\lambda)$ random variables with common distribution function

$$P(Y_1 \le x) = 1 - \mathrm{e}^{-\lambda x}, \quad x \ge 0.$$

This definition shows nicely what kind of sample path a Poisson process has. It is a pure jump function: it is constant on $[T_n, T_{n+1})$ and has upward jumps of size 1 at the random times T_n.

The rôle of the Poisson process and its modifications and ramifications is comparable with the rôle of Brownian motion. The Poisson process is a counting process; see (1.10). It has a large variety of applications in the most different fields. To name a few: for a given time interval $[0,t]$, X_t is a model for the number of

- telephone calls to be handled by an operator,
- customers waiting for service in a queue,
- claims arriving in an insurance portfolio. □

Notes and Comments

Introductions to the theory of stochastic processes are based on non-elementary facts from measure theory and functional analysis. Standard texts are Ash and Gardner (1975), Gikhman and Skorokhod (1975), Karlin and Taylor (1975,1981) and many others. An entertaining introduction to the theory of applied stochastic processes is Resnick (1992). Grimett and Stirzaker (1994) is an introduction "without burdening the reader with a great deal of measure theory".

1.3 Brownian Motion

1.3.1 Defining Properties

Brownian motion plays a central rôle in probability theory, the theory of stochastic processes, physics, finance,..., and also in this book. We start with the definition of this important process. Then we continue with some of its elementary properties.

A stochastic process $B = (B_t, t \in [0,\infty))$ is called (*standard*) *Brownian motion* or a *Wiener process* if the following conditions are satisfied:

- It starts at zero: $B_0 = 0$.
- It has stationary, independent increments; see p. 30 for the definition.
- For every $t > 0$, B_t has a normal $N(0,t)$ distribution.
- It has continuous sample paths: "no jumps".

See Figure 1.3.1 for a visualization of Brownian sample paths.

Brownian motion is named after the biologist Robert Brown whose research dates to the 1820s. Early in this century, Louis Bachelier (1900), Albert Einstein (1905) and Norbert Wiener (1923) began developing the mathematical theory of Brownian motion. The construction of Bachelier (1900) was erroneous but it captured many of the essential properties of the process. Wiener (1923) was the first to put Brownian motion on a firm mathematical basis.

Distribution, Expectation and Covariance Functions

The fidis of Brownian motion are multivariate Gaussian, hence B is a Gaussian process. Check this statement by observing that Brownian motion has

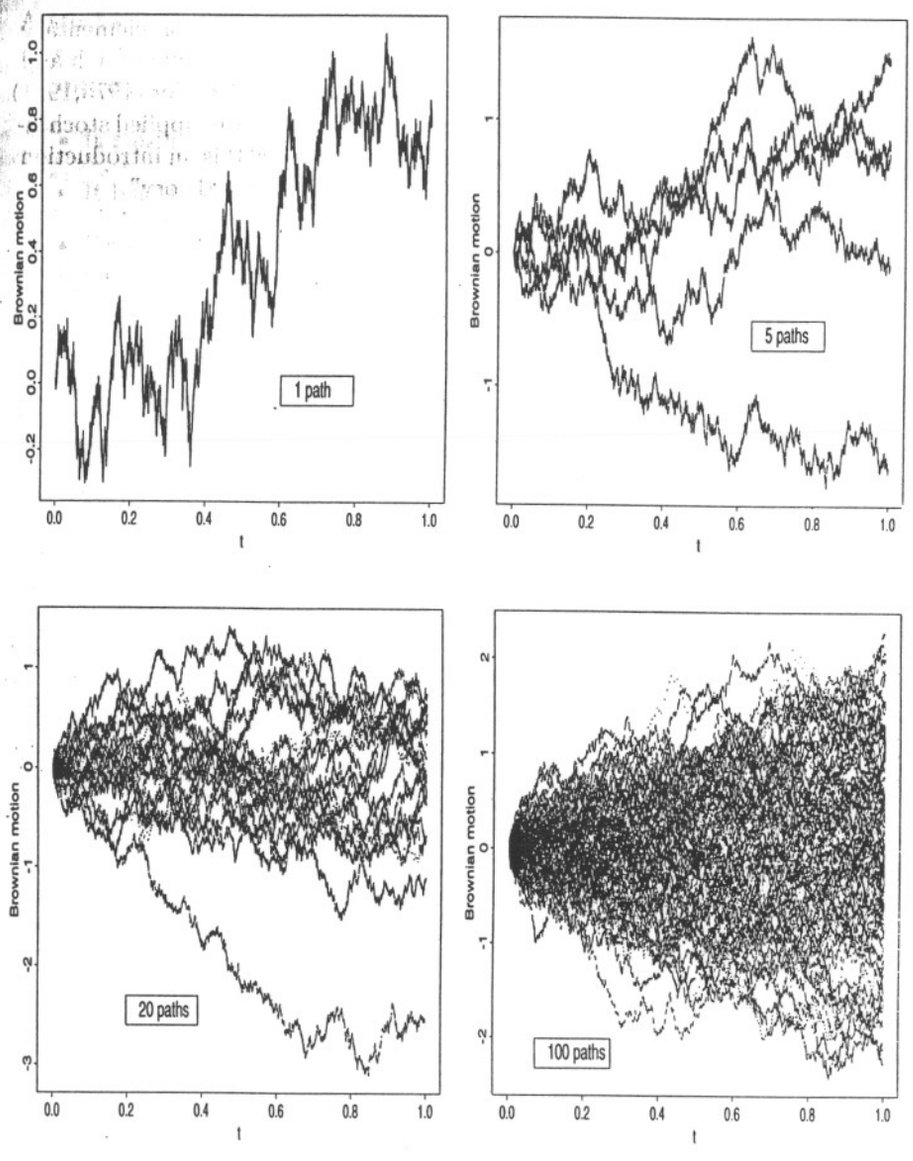

Figure 1.3.1 *Sample paths of Brownian motion on* $[0,1]$.

independent Gaussian increments and by using the formula for linear transformations of a Gaussian random vector; see p. 18.

> The random variables $B_t - B_s$ and B_{t-s} have an $N(0, t-s)$ distribution for $s < t$.

This follows from the stationarity of the increments. Indeed, $B_t - B_s$ has the same distribution as $B_{t-s} - B_0 = B_{t-s}$ which is normal with mean zero and variance $t - s$. Thus the variance is proportional to the length of the interval $[s,t]$. This means intuitively: the larger the interval, the larger the fluctuations of Brownian motion on this interval. This observation is also supported by simulated Brownian sample paths; see for example Figure 1.3.2.

Notice:

The distributional identity $B_t - B_s \stackrel{d}{=} B_{t-s}$ *does* not *imply pathwise identity: in general,*

$$B_t(\omega) - B_s(\omega) \neq B_{t-s}(\omega)\,.$$

It is worthwhile to compare Brownian motion with the Poisson process; see Example 1.2.10. Their definitions coincide insofar that they are processes with stationary, independent increments. The crucial difference is the kind of distribution of the increments. The requirement of the Poisson distribution makes the sample paths pure jump functions, whereas the Gaussian assumption makes the sample paths continuous.

It is immediate from the definition that Brownian motion has expectation function

$$\mu_B(t) = EB_t = 0\,, \quad t \geq 0\,,$$

and, since the increments $B_s - B_0 = B_s$ and $B_t - B_s$ are independent for $t > s$, it has covariance function (recall its definition from p. 28)

$$\begin{aligned} c_B(t,s) &= E\big[[(B_t - B_s) + B_s]\,B_s\big] = E[(B_t - B_s)\,B_s] + EB_s^2 \\ &= E(B_t - B_s)\,EB_s + s = 0 + s = s\,, \qquad 0 \leq s < t\,. \end{aligned}$$

Since a Gaussian process is characterized by its expectation and covariance functions (see Example 1.2.5), we can give an alternative definition:

> Brownian motion is a Gaussian process with
>
> $$\mu_B(t) = 0 \quad \text{and} \quad c_B(t,s) = \min(s,t)\,. \tag{1.11}$$

Path Properties: Non-Differentiability and Unbounded Variation

In what follows, we fix one sample path $B_t(\omega)$, $t \geq 0$, and consider its properties. We already know from the definition of Brownian motion that its sample paths are continuous. However, a glance at simulated Brownian paths immediately convinces us that these functions of t are extremely irregular: they oscillate wildly. The main reason is that the increments of B are independent. In particular, increments of Brownian motion on adjacent intervals are independent whatever the length of the intervals. Since we can imagine the sample path as constructed from its independent increments on adjacent intervals, it is rather surprising that continuity of the path results.

Thus:

How irregular is a Brownian sample path?

Before we answer this question we make a short excursion to a class of stochastic processes which contains Brownian motion as a special case. All members of this class have irregular sample paths.

A stochastic process $(X_t, t \in [0,\infty))$ is *H-self-similar* for some $H > 0$ if its fidis satisfy the condition

$$(T^H B_{t_1}, \ldots, T^H B_{t_n}) \stackrel{d}{=} (B_{Tt_1}, \ldots, B_{Tt_n}) \tag{1.12}$$

for every $T > 0$, any choice of $t_i \geq 0$, $i = 1, \ldots, n$, and $n \geq 1$.

Notice:

Self-similarity is a distributional, not *a pathwise property. In* (1.12), *one must not replace* $\stackrel{d}{=}$ *with* $=$.

Roughly speaking, self-similarity means that the properly scaled patterns of a sample path in any small or large time interval have a similar shape, but they are *not* identical. See Figure 1.3.2 for an illustration.

The sample paths of a self-similar process are nowhere differentiable; see Proposition A3.1 on p. 188. And here it comes:

Brownian motion is 0.5-self-similar, i.e.

$$\left(T^{1/2} B_{t_1}, \ldots, T^{1/2} B_{t_n}\right) \stackrel{d}{=} (B_{Tt_1}, \ldots, B_{Tt_n}) \tag{1.13}$$

for every $T > 0$, any choice of $t_i \geq 0$, $i = 1, \ldots, n$, and $n \geq 1$.

Hence its sample paths are nowhere differentiable.

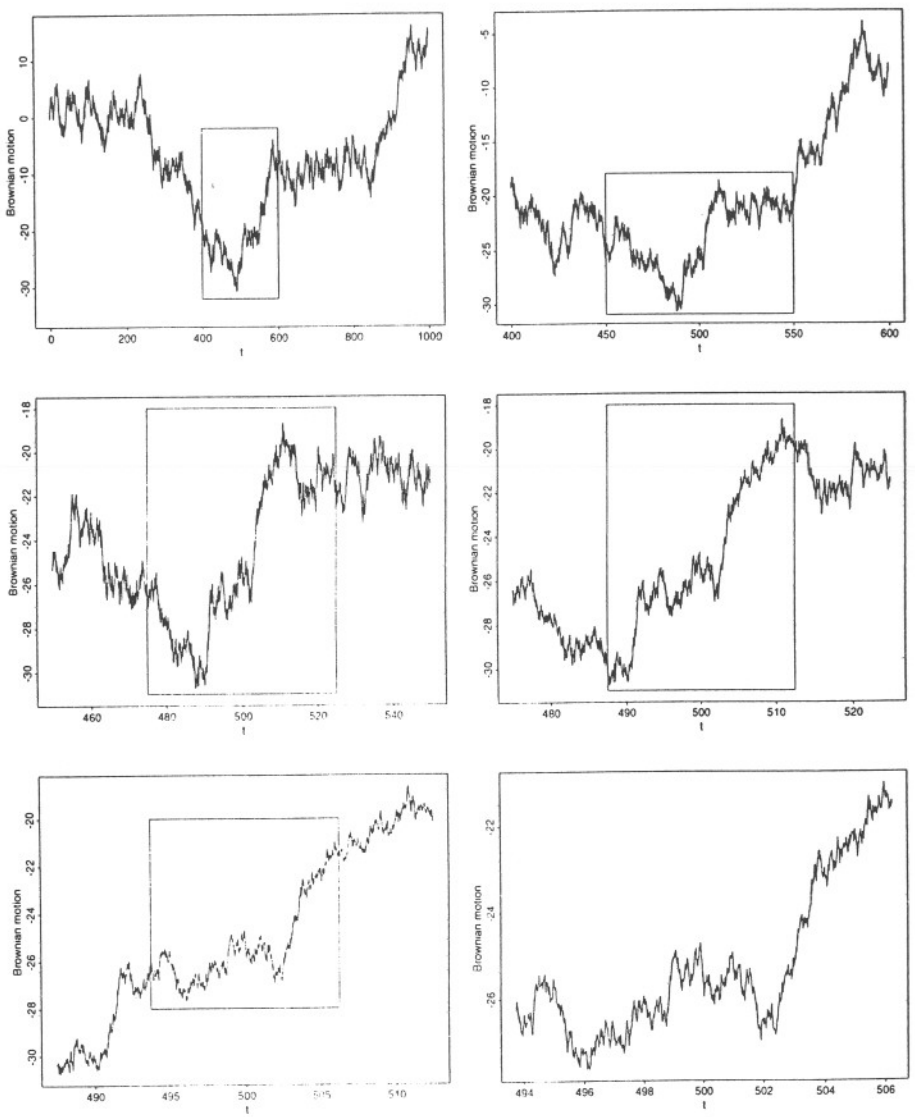

Figure 1.3.2 Self-similarity: *the same Brownian sample path on different scales. The shapes of the curves on different intervals look similar, but they are not simply scaled copies of each other.*

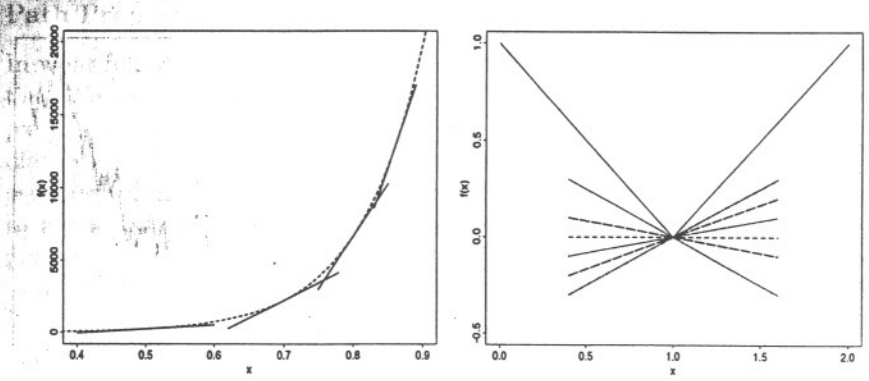

Figure 1.3.3 Left: *a differentiable function. At every point its graph can be approximated by a linear function which is the unique tangent at this point.* Right: *this function is not differentiable at $x = 1$. There are infinitely many tangents to the curve of the function at this point.*

One can easily check the distributional identity (1.13). Indeed, the left- and right-hand sides of (1.13) are Gaussian random vectors, and therefore it suffices to verify that they have the same expectation and covariance matrix. Check these properties by using (1.11).

Differentiability of a function f means that its graph is smooth. Indeed, if the limit

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

exists and is finite for some $x_0 \in (0, t)$, say, then we may write for small Δx

$$f(x_0 + \Delta x) = f(x_0) + f'(x_0)\Delta x + h(x_0, \Delta x)\Delta x\,,$$

where $h(x_0, \Delta x) \to 0$ as $\Delta x \to 0$. Hence, in a small neighborhood of x_0, the function f is roughly linear (as a function of Δx). This explains its smoothness. Alternatively, differentiability of f at x_0 implies that we have a unique tangent to the curve of the function f at this point; see Figure 1.3.3 for an illustration. In this figure you can also see a function which is not differentiable at one point.

Now try to imagine a nowhere differentiable function: the graph of this function changes its shape in the neighborhood of any point in a completely non-predictable way. You will admit that you cannot really imagine such a

function: it is physically impossible. Nevertheless, Brownian motion is considered as a very good approximation to many real-life phenomena. We will see in Section 1.3.3 that Brownian motion is a limit process of certain sum processes.

The self-similarity property of Brownian motion has a nice consequence for the simulation of its sample paths. In order to simulate a path on $[0, T]$ it suffices to simulate one path on $[0, 1]$, then scale the time interval by the factor T and the sample path by the factor $T^{1/2}$. Then we are done.

In some books one can find the claim that the limit

$$\lim_{\Delta t \to 0} \operatorname{var}\left(\frac{B_{t_0+\Delta t} - B_{t_0}}{\Delta t}\right) \tag{1.14}$$

does not exist and *therefore* the sample paths of Brownian motion are non-differentiable. It is easy to check (do it!) that the limit (1.14) does not exist, but without further theory it would be wrong to conclude from this distributional result that the paths of the process are non-differentiable.

The existence of a nowhere differentiable continuous function was discovered in the 19th century. Such a function was constructed by Weierstrass. It was considered as a curiosity, far away from any practical application. Brownian motion is a process with nowhere differentiable sample paths. Currently it is considered as one of those processes which have a multitude of applications in very different fields. One of them is stochastic calculus; see Chapters 2 and 3.

A further indication of the irregularity of Brownian sample paths is given by the following fact:

> Brownian sample paths do not have bounded variation on any finite interval $[0, T]$. This means that
>
> $$\sup_{\tau} \sum_{i=1}^{n} \left|B_{t_i}(\omega) - B_{t_{i-1}}(\omega)\right| = \infty\,,$$
>
> where the supremum (see p. 211 for its definition) is taken over all possible partitions $\tau : 0 = t_0 < \cdots < t_n = T$ of $[0, T]$.

A proof of this fact is provided by Proposition A3.2 on p. 189. We mention at this point that the unbounded variation and non-differentiability of Brownian sample paths are major reasons for the failure of classical integration methods, when applied to these paths, and for the introduction of stochastic calculus.

1.3.2 Processes Derived from Brownian Motion

The purpose of this section is to get some feeling for the distributional and pathwise properties of Brownian motion. If you want to start with Chapter 2 on stochastic calculus as soon as possible, you can easily skip this section and return to it whenever you need a reference to a property or definition.

Various Gaussian and non-Gaussian stochastic processes of practical relevance can be derived from Brownian motion. Below we introduce some of those processes which will find further applications in the course of this book. As before, $B = (B_t, t \in [0, \infty))$ denotes Brownian motion.

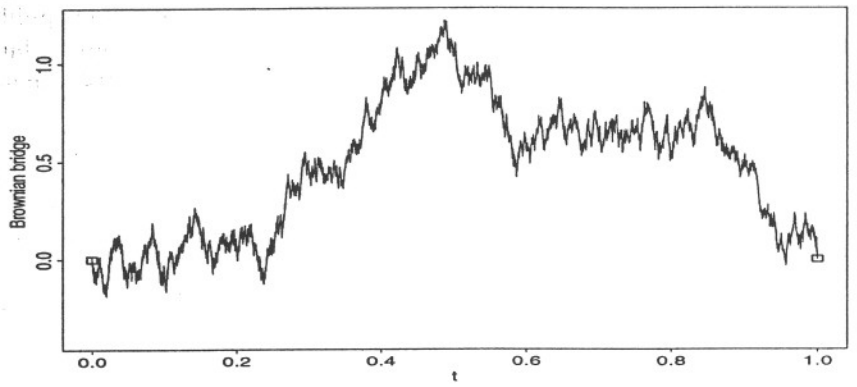

Figure 1.3.4 *A sample path of the Brownian bridge.*

Example 1.3.5 (Brownian bridge)
Consider the process

$$X_t = B_t - t\,B_1\,, \quad 0 \le t \le 1\,.$$

Obviously,

$$X_0 = B_0 - 0\,B_1 = 0 \quad \text{and} \quad X_1 = B_1 - 1\,B_1 = 0\,.$$

For this simple reason, the process X bears the name (standard) *Brownian bridge* or *tied down Brownian motion*. A glance at the sample paths of this "bridge" (see Figure 1.3.4) may or may not convince you that this name is justified.

Using the formula for linear transformations of Gaussian random vectors (see p. 18), one can show that the fidis of X are Gaussian. Verify this! Hence X is a Gaussian process. You can easily calculate the expectation and covariance functions of the Brownian bridge:

$$\mu_X(t) = 0 \quad \text{and} \quad c_X(t,s) = \min(t,s) - t\,s\,, \quad s,t \in [0,1]\,.$$

Since X is Gaussian, the Brownian bridge is characterized by these two functions.

The Brownian bridge appears as the limit process of the normalized *empirical distribution function* of a sample of iid uniform $U(0,1)$ random variables. This is a fundamental result from non-parametric statistics; it is the basis for numerous goodness-of-fit tests in statistics. See for example Shorack and Wellner (1986). □

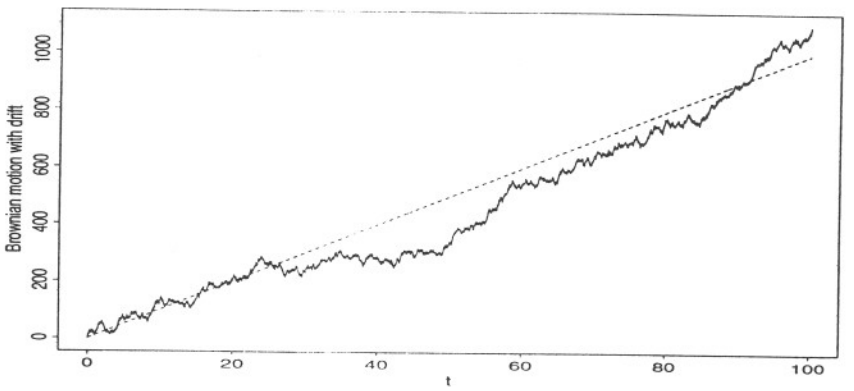

Figure 1.3.6 *A sample path of Brownian motion with drift* $X_t = 20\,B_t + 10\,t$ *on* $[0, 100]$. *The dashed line stands for the drift function* $\mu_X(t) = 10\,t$.

Example 1.3.7 (Brownian motion with drift)
Consider the process

$$X_t = \mu\,t + \sigma\,B_t\,, \quad t \ge 0\,,$$

for constants $\sigma > 0$ and $\mu \in \mathbb{R}$. Clearly, it is a Gaussian process (why?) with expectation and covariance functions

$$\mu_X(t) = \mu\,t \quad \text{and} \quad c_X(t,s) = \sigma^2 \min(t,s)\,, \quad s,t \ge 0\,.$$

The expectation function $\mu_X(t) = \mu t$ (the deterministic "drift" of the process) essentially determines the characteristic shape of the sample paths; see Figure 1.3.6 for an illustration. Therefore X is called *Brownian motion with (linear) drift.* □

With the fundamental discovery of Bachelier in 1900 that prices of risky assets (stock indices, exchange rates, share prices, etc.) can be well described by Brownian motion, a new area of applications of stochastic processes was born. However, Brownian motion, as a Gaussian process, may assume negative values, which is not a very desirable property of a price. In their celebrated papers from 1973, Black, Scholes and Merton suggested another stochastic process as a model for speculative prices. In Section 4.1 we consider their approach to the pricing of European call options in more detail. It is one of the promising and motivating examples for the use of stochastic calculus.

Example 1.3.8 (Geometric Brownian motion)
The process suggested by Black, Scholes and Merton is given by

$$X_t = \mathrm{e}^{\mu t + \sigma B_t}, \quad t \geq 0,$$

i.e. it is the exponential of Brownian motion with drift; see Example 1.3.7. Clearly, X is not a Gaussian process (why?).

For the purpose of later use, we calculate the expectation and covariance functions of geometric Brownian motion. For readers, familiar with probability theory, you may recall that for an $N(0,1)$ random variable Z,

$$E\mathrm{e}^{\lambda Z} = \mathrm{e}^{\lambda^2/2}, \quad \lambda \in \mathbb{R}. \tag{1.15}$$

It is easily derived as shown below:

$$\begin{aligned} E\mathrm{e}^{\lambda Z} &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \mathrm{e}^{\lambda z} \mathrm{e}^{-z^2/2}\, dz \\ &= \mathrm{e}^{\lambda^2/2} \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \mathrm{e}^{-(z-\lambda)^2/2}\, dz \\ &= \mathrm{e}^{\lambda^2/2}. \end{aligned}$$

Here we used the fact that $(2\pi)^{-1/2} \exp\{-(z-\lambda)^2/2\}$ is the density of an $N(\lambda, 1)$ random variable.

From (1.15) and the self-similarity of Brownian motion it follows immediately that

$$\mu_X(t) = \mathrm{e}^{\mu t} E\mathrm{e}^{\sigma B_t} = \mathrm{e}^{\mu t} E\mathrm{e}^{\sigma t^{1/2} B_1} = \mathrm{e}^{(\mu + 0.5\,\sigma^2)t}. \tag{1.16}$$

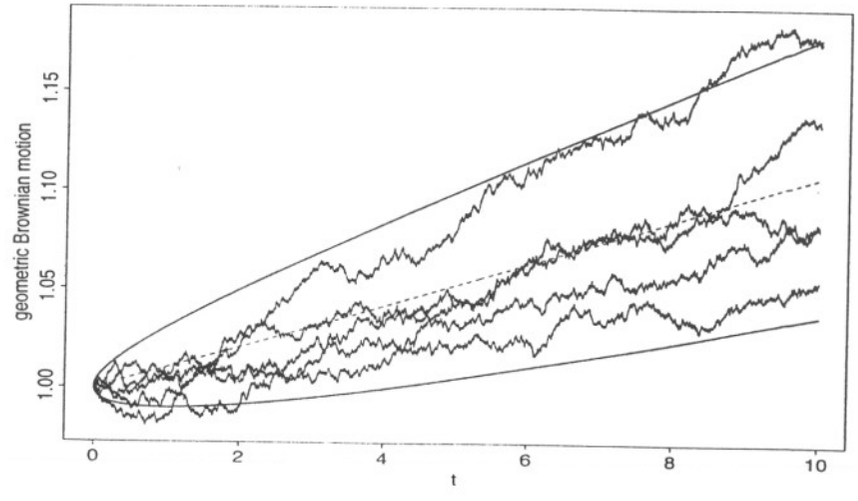

Figure 1.3.9 *Sample paths of geometric Brownian motion* $X_t = \exp\{0.01t + 0.01B_t\}$ *on* $[0, 10]$, *the expectation function* $\mu_X(t)$ *(dashed line) and the graphs of the functions* $\mu_X(t) \pm 2\sigma_X(t)$ *(solid lines). The latter curves have to be interpreted with care since the distributions of the* X_t*s are not normal.*

For $s \leq t$, $B_t - B_s$ and B_s are independent, and $B_t - B_s \stackrel{d}{=} B_{t-s}$. Hence

$$\begin{aligned} c_X(t,s) &= EX_tX_s - EX_tEX_s \qquad (1.17) \\ &= \mathrm{e}^{\mu(t+s)} E\mathrm{e}^{\sigma(B_t + B_s)} - \mathrm{e}^{(\mu + 0.5\,\sigma^2)(t+s)} \\ &= \mathrm{e}^{\mu(t+s)} E\mathrm{e}^{\sigma[(B_t - B_s) + 2B_s]} - \mathrm{e}^{(\mu + 0.5\,\sigma^2)(t+s)} \\ &= \mathrm{e}^{\mu(t+s)} E\mathrm{e}^{\sigma(B_t - B_s)} E\mathrm{e}^{2\sigma B_s} - \mathrm{e}^{(\mu + 0.5\,\sigma^2)(t+s)} \\ &= \mathrm{e}^{(\mu + 0.5\,\sigma^2)(t+s)} \left(\mathrm{e}^{\sigma^2 s} - 1\right). \end{aligned}$$

In particular, geometric Brownian motion has variance function

$$\sigma_X^2(t) = \mathrm{e}^{(2\mu + \sigma^2)t} \left(\mathrm{e}^{\sigma^2 t} - 1\right). \tag{1.18}$$

See Figure 1.3.9 for an illustration of various sample paths of geometric Brownian motion. □

Example 1.3.10 (Gaussian white and colored noise)
In statistics and time series analysis one often uses the name "white noise" for a sequence of iid or uncorrelated random variables. This is in contrast to physics, where white noise is understood as a certain derivative of Brownian sample paths. This does not contradict our previous remarks since this derivative is not obtained by ordinary differentiation. Since white noise is "physically impossible", one considers an approximation to it, called *colored noise*. It is a Gaussian process defined as

$$X_t = \frac{B_{t+h} - B_t}{h}, \quad t \geq 0, \tag{1.19}$$

where $h > 0$ is some fixed constant. Its expectation and covariance functions are given by

$$\mu_X(t) = 0 \quad \text{and} \quad c_X(t,s) = h^{-2}[(s+h) - \min(s+h, t)], \quad s \leq t.$$

Notice that $c_X(t,s) = 0$ if $t - s \geq h$, hence X_t and X_s are independent, but if $t - s < h$, $c_X(t,s) = h^{-2}[h - (t-s)]$. Since X is Gaussian and $c_X(t,s)$ is a function only of $t - s$, it is stationary (see Example 1.2.8).

Clearly, if B was differentiable, we could let h in (1.19) go to zero, and in the limit we would obtain the ordinary derivative of B at t. But, as we know, this argument is not applicable. The variance function $\sigma_X^2(t) = h^{-1}$ gives an indication that the fluctuations of colored noise become larger as h decreases. Simulated paths of colored noise look very much like the sample paths in Figure 1.2.6. □

1.3.3 Simulation of Brownian Sample Paths

This section is not necessary for the understanding of stochastic calculus. However, it will characterize Brownian motion as a distributional limit of partial sum processes (so-called functional central limit theorem). This observation will help you to understand the Brownian path properties (non-differentiability, unbounded variation) much better. A second objective of this section is to show that Brownian sample paths can easily be simulated by using standard software.

Using the almost unlimited power of modern computers, you can visualize the paths of almost every stochastic process. This is desirable because we like to see sample paths in order to understand the stochastic process better. On the other hand, simulations of the paths of stochastic processes are sometimes unavoidable if you want to say something about the distributional properties of such a process. In most cases, we cannot determine the exact distribution

of a stochastic process and its functionals (such as its maximum or minimum on a given interval). Then simulations and numerical techniques offer some alternative to calculate these distributions.

Simulation via the Functional Central Limit Theorem

From an elementary course in probability theory we know about the *central limit theorem* (*CLT*). It is a fundamental result: it explains why the normal distribution plays such an important rôle in probability theory and statistics. The CLT says that the properly normalized and centered partial sums of an iid finite variance sequence converge in distribution to a normal distribution. To be precise: let $Y_1, Y_2, \ldots$ be iid non-degenerate (i.e. non-constant) random variables with mean $\mu_Y = EY_1$ and variance $\sigma_Y^2 = \text{var}(Y_1)$. Define the partial sums

$$R_0 = 0, \quad R_n = Y_1 + \cdots + Y_n, \quad n \geq 1.$$

Recall that Φ denotes the distribution function of a standard normal random variable.

> If Y_1 has finite variance, then the sequence (Y_i) obeys the CLT, i.e.
>
> $$\sup_x \left| P\left(\frac{R_n - ER_n}{[\text{var}(R_n)]^{1/2}} \leq x \right) - \Phi(x) \right| \to 0 \quad \text{as} \quad n \to \infty.$$

Thus, for large n, the distribution of $(R_n - \mu_Y n)/(\sigma_Y^2 n)^{1/2}$ is approximately standard normal. This is an amazing fact, since the CLT holds independently of the distribution of the Y_is; all one needs is a finite variance σ_Y^2.

The CLT has an analogue for stochastic processes. Consider the process with continuous sample paths on [0, 1]:

$$S_n(t) = \begin{cases} (\sigma_Y^2 n)^{-1/2}(R_i - \mu_Y i), & \text{if } t = i/n, \, i = 0, \ldots, n, \\ \text{linearly interpolated}, & \text{elsewhere.} \end{cases} \tag{1.20}$$

In Figures 1.3.11 and 1.3.12 you will find realizations of S_n for various n.

Assume for the moment that the Y_is are iid $N(0,1)$ and consider the restriction of the process S_n to the points i/n. We immediately see that the following properties hold:

- S_n starts at zero: $S_n(0) = 0$.

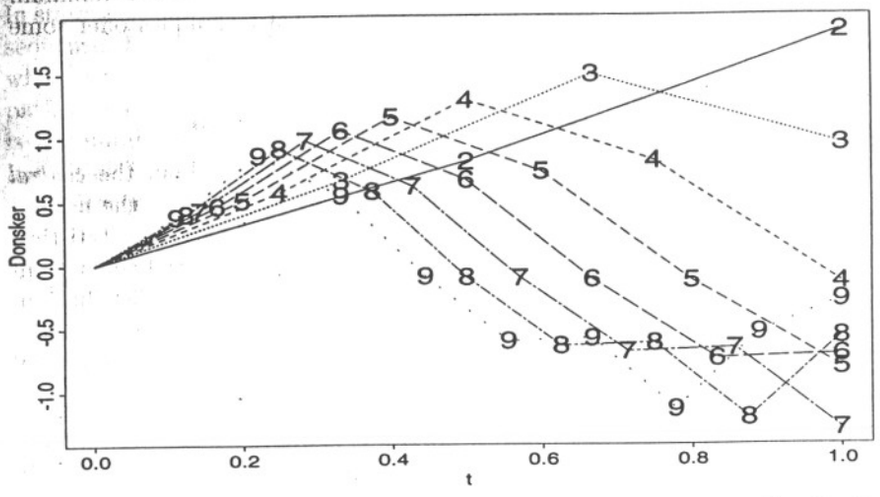

Figure 1.3.11 *Sample paths of the process S_n for one sequence of realizations $Y_1(\omega), \ldots, Y_9(\omega)$ and $n = 2, \ldots, 9$.*

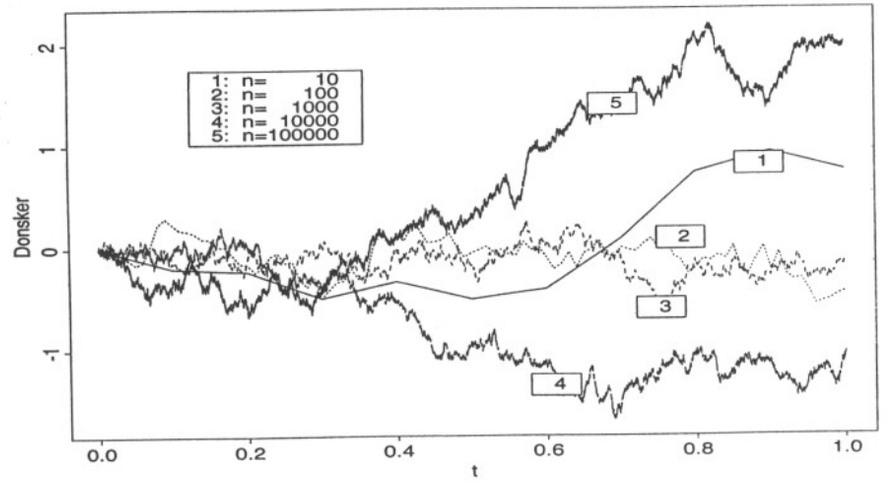

Figure 1.3.12 *Sample paths of the process S_n for different n and the same sequence of realizations $Y_1(\omega), \ldots, Y_{100,000}(\omega)$.*

- S_n has independent increments, i.e. for all integers $0 \leq i_1 \leq \cdots \leq i_m \leq n$, the random variables

$$S_n(i_2/n) - S_n(i_1/n), \ldots, S_n(i_m/n) - S_n(i_{m-1}/n)$$

  are independent.

- For every $0 \leq i \leq n$, $S_n(i/n)$ has a normal $N(0, i/n)$ distribution.

Thus, S_n and Brownian motion B on $[0,1]$, when restricted to the points i/n, have very much the same properties; cf. the definition of Brownian motion on p. 33. Naturally, the third property above is not valid if we drop the assumption that the Y_is are iid Gaussian. However, in an asymptotic sense the stochastic process S_n is close to Brownian motion:

> If Y_1 has finite variance, then the sequence (Y_i) obeys the *functional CLT*, also called *Donsker's invariance principle*, i.e. the processes S_n converge in distribution to Brownian motion B on $[0,1]$.

Convergence in distribution of S_n has a two-fold meaning. The first one is quite intuitive:

- The fidis of S_n converge to the corresponding fidis of B, i.e.

$$P(S_n(t_1) \leq x_1, \ldots, S_n(t_m) \leq x_m) \to P(B_{t_1} \leq x_1, \ldots, B_{t_m} \leq x_m) \tag{1.21}$$

  for all possible choices of $t_i \in [0,1]$, $x_i \in \mathbb{R}$, $i = 1, \ldots, m$, and all integers $m \geq 1$.

But convergence of the fidis is not sufficient for the convergence in distribution of stochastic processes. Fidi convergence determines the Gaussian limit distribution for every choice of finitely many fixed instants of time t_i, but stochastic processes are infinite-dimensional objects, and therefore unexpected events may happen. For example, the sample paths of the converging processes may fluctuate very wildly with increasing n, in contrast to the limiting process of Brownian motion which has continuous sample paths. In order to avoid such irregular behavior,

- a so-called *tightness* or *stochastic compactness* condition must be satisfied.

✱ Fortunately, the partial sum processes S_n are tight, but it is beyond the scope of this book to show it.

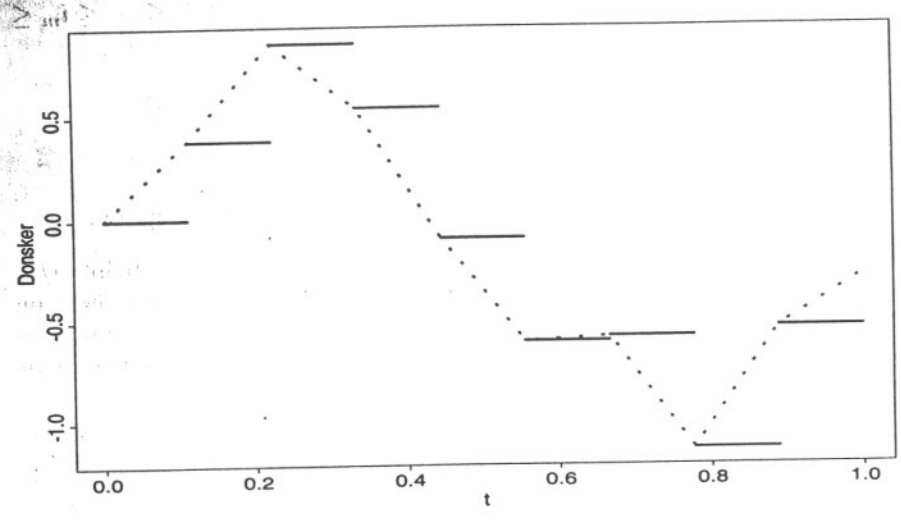

Figure 1.3.13 *One sample path of the processes S_9 (dotted line) and $\widetilde{S}_9$ (solid line) for the same sequence of realizations $Y_1(\omega), \ldots, Y_9(\omega)$. See* (1.20) *and* (1.22) *for the definitions of S_n and $\widetilde{S}_n$.*

What was said about the convergence of the processes S_n remains valid for the processes

$$\widetilde{S}_n(t) = (\sigma_Y^2 n)^{-1/2}(R_{[nt]} - \mu_Y\,[nt])\,, \quad 0 \le t \le 1\,, \tag{1.22}$$

where $[nt]$ denotes the integer part of the real number nt; see Figure 1.3.13 for an illustration. In contrast to S_n, the process $\widetilde{S}_n$ is constant on the intervals $[(i-1)/n, i/n)$ and has jumps at the points i/n. But S_n and $\widetilde{S}_n$ coincide at the points i/n, and the differences between these two processes are asymptotically negligible: the normalization $n^{1/2}$ makes the jumps of $\widetilde{S}_n$ arbitrarily small for large n. As for S_n, we can formulate the following functional CLT:

> If Y_1 has finite variance, then the sequence (Y_i) obeys the *functional CLT*, i.e. the processes $\widetilde{S}_n$ converge in distribution to Brownian motion B on $[0,1]$.

Since $\widetilde{S}_n$ is a jump process, the notion of convergence in distribution becomes even more complicated than for S_n. We refrain from discussing details.

Thus we have a first simple tool for simulating Brownian sample paths in our hands:

Plot the paths of the processes S_n or $\widetilde{S}_n$ for sufficiently large n, and you get a reasonable approximation to Brownian sample paths.

Also notice:

Since Brownian motion appears as a distributional limit you will see completely different graphs for different values of n, for the same sequence of realizations $Y_i(\omega)$.

From the self-similarity property we also know how to obtain an approximation to the sample paths of Brownian motion on any interval $[0,T]$:

Simulate one path of S_n or $\widetilde{S}_n$ on $[0,1]$, then scale the time interval by the factor T and the sample path by the factor $T^{1/2}$.

Standard software (such as Splus, Mathematica, Matlab, etc.) provides you with quick and reliable algorithms for generating random numbers of standard distributions. Random number generators are frequently based on natural processes, for example radiation, or on algebraic methods. The generated "random" numbers can be considered as "pseudo" realizations $Y_i(\omega)$ of iid random variables Y_i.

For practical purposes, you may want to choose the realizations $Y_i(\omega)$ (or, as you like, the random numbers $Y_i(\omega)$) from an appropriate distribution. If you are interested in "good" approximations to the Gaussian distribution of Brownian motion, you would generate the $Y_i(\omega)$s from a Gaussian distribution. If you are forced to simulate many sample paths of S_n or $\widetilde{S}_n$ in a short period of time, you would perhaps choose the $Y_i(\omega)$s as realizations of iid Bernoulli random variables, i.e. $P(Y_i = \pm 1) = 0.5$, or of iid uniform random variables.

Simulation via Series Representations

Recall from a course on calculus that every continuous 2π-periodic function f on $\mathbb{R}$ (i.e. $f(x + 2\pi) = f(x)$ for all $x \in \mathbb{R}$) has a Fourier series representation, i.e. it can be written as an infinite series of trigonometric functions.

Since Brownian sample paths are continuous functions, we can try to expand them in a Fourier series. However, the paths are random functions: for different ω we obtain different functions. This means that the coefficients of this Fourier series are random variables, and since the process is Gaussian, they must be Gaussian as well.

The following representation of Brownian motion on the interval $[0, 2\pi]$ is called *Paley–Wiener representation*: let $(Z_n, n \ge 0)$ be a sequence of iid

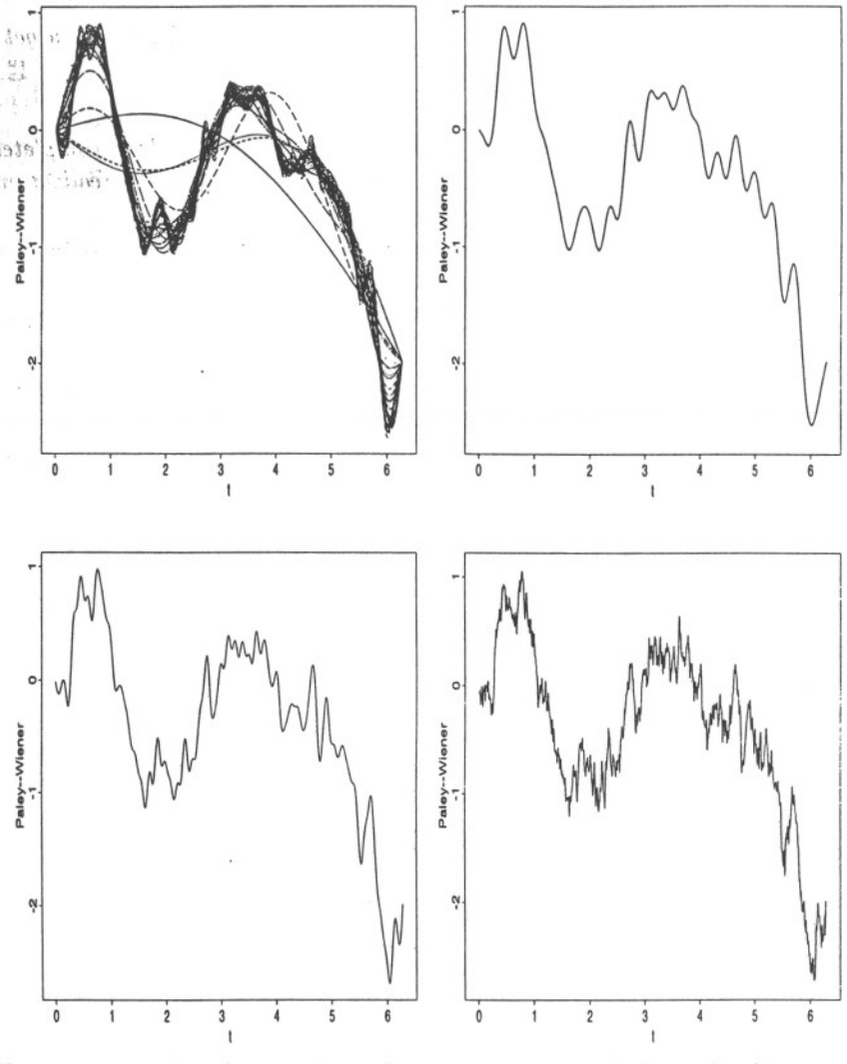

Figure 1.3.14 *Simulation of one Brownian sample path from the discretization* (1.24) *of the Paley–Wiener representation with* $N = 1{,}000$. Top left: *all paths for* $M = 2, \ldots, 40$. Top right: *the path only for* $M = 40$. Bottom left: $M = 100$. Bottom right: $M = 800$.

$N(0,1)$ random variables, then

$$B_t(\omega) = Z_0(\omega)\frac{t}{(2\pi)^{1/2}} + \frac{2}{\pi^{1/2}}\sum_{n=1}^{\infty} Z_n(\omega)\frac{\sin(nt/2)}{n}, \quad t \in [0, 2\pi]. \tag{1.23}$$

This series converges for every fixed t, but also uniformly for $t \in [0, 2\pi]$, i.e. the rate of convergence is comparable for all t. For an application of this formula, one has to decide about the number M of sine functions and the number N of discretization points at which the sine functions will be evaluated. This amounts to calculating the values

$$Z_0(\omega)\frac{t_j}{(2\pi)^{1/2}} + \frac{2}{\pi^{1/2}}\sum_{n=1}^{M} Z_n(\omega)\frac{\sin(nt_j/2)}{n}, \tag{1.24}$$

$$t_j = \frac{2\pi j}{N}, \quad j = 0, 1, \ldots, N.$$

The problem of choosing the "right" values for M and N is similar to the choice of the sample size n in the functional CLT; it is difficult to give a simple rule of thumb for the choices of M and N.

In Figure 1.3.14 you can see that the shape of the sample paths does not change very much if one switches from $M = 100$ to $M = 800$ sine functions. A visual inspection in the case $M = 100$ gives the impression that the sample path is still too smooth. This is not completely surprising since a sum of M sine functions is differentiable; only in the limit (as $M \to \infty$) do we gain a non-differentiable sample path.

The Paley–Wiener representation is just one of infinitely many possible series representations of Brownian motion. Another well-known such representation is due to Lévy. In the *Lévy representation*, the sine functions are replaced by certain polygonal functions (the Schauder functions).

To be precise, first define the *Haar functions* H_n on $[0,1]$ as follows:

$$H_1(t) = 1,$$

$$H_{2^{m+1}}(t) = \begin{cases} 2^{m/2}, & \text{if } t \in \left[1 - \frac{2}{2^{m+1}}, 1 - \frac{1}{2^{m+1}}\right), \\ -2^{m/2}, & \text{if } t \in \left[1 - \frac{1}{2^{m+1}}, 1\right], \\ 0, & \text{elsewhere}, \end{cases}$$

$$H_{2^m+k}(t) = \begin{cases} 2^{m/2}, & \text{if } t \in \left[\frac{k-1}{2^m}, \frac{k-1}{2^m} + \frac{1}{2^{m+1}}\right), \\ -2^{m/2}, & \text{if } t \in \left[\frac{2k-1}{2^{m+1}}, \frac{k}{2^m}\right), \\ 0, & \text{elsewhere}, \end{cases}$$

$$k = 1, \ldots, 2^m - 1; \quad m = 0, 1, \ldots .$$

From these functions define the system of the *Schauder functions* on $[0,1]$ by integrating the Haar functions:

$$\widetilde{H}_n(t) = \int_0^t H_n(s)\, ds, \quad n = 1, 2, \ldots .$$

Figures 1.3.15 and 1.3.16 show the graphs of H_n and $\widetilde{H}_n$ for the first n. A series representation for a Brownian sample path on $[0,1]$ is then given by

$$B_t(\omega) = \sum_{n=1}^{\infty} Z_n(\omega) \widetilde{H}_n(t), \quad t \in [0,1], \tag{1.25}$$

where the convergence of this series is uniform for $t \in [0,1]$ and the $Z_n(\omega)$s are realizations of an iid $N(0,1)$ sequence (Z_n). As for simulations of Brownian motion via sine functions, one has to choose a truncation point M of the infinite series (1.25). In Figure 1.3.17 we show how a Brownian sample path is approximated by the superposition of the first M terms in the series representation (1.25). In contrast to Figure 1.3.14, the polygonal shape of the Schauder functions already anticipates the irregular behavior of a Brownian path (its non-differentiability) for relatively small M.

The Paley–Wiener and Lévy representations are just two of infinitely many possible series representations of Brownian motion. They are special cases of the so-called *Lévy–Ciesielski representation*. Ciesielski showed that Brownian motion on $[0,1]$ can be represented in the form

$$B_t(\omega) = \sum_{n=1}^{\infty} Z_n(\omega) \int_0^t \phi_n(x) dx, \quad t \in [0,1],$$

where Z_n are iid $N(0,1)$ random variables and (ϕ_n) is a complete orthonormal function system on $[0,1]$.

Notes and Comments

Brownian motion is the best studied stochastic process. Various books are devoted to it, for example Borodin and Salminen (1996), Hida (1980), Karatzas

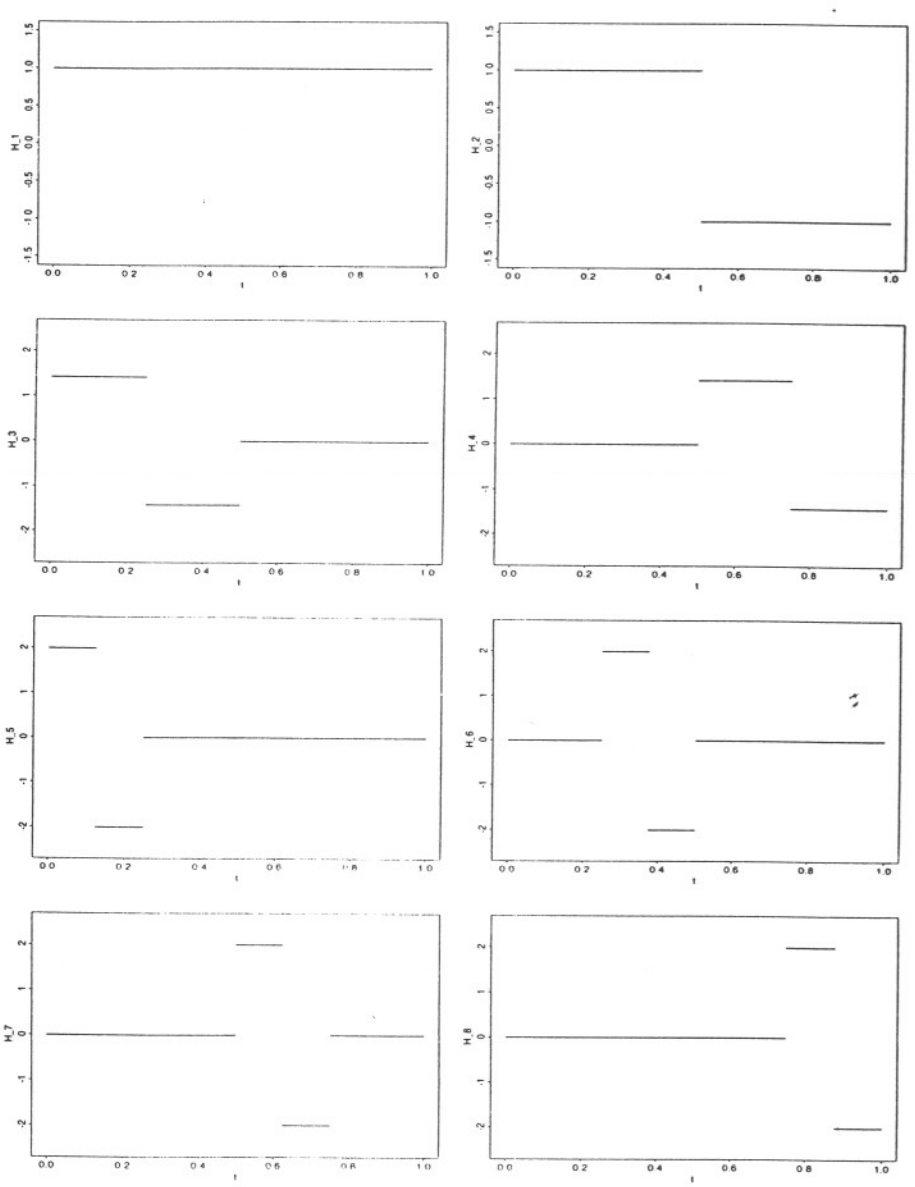

Figure 1.3.15 *The Haar functions* $H_1, \ldots, H_8$.

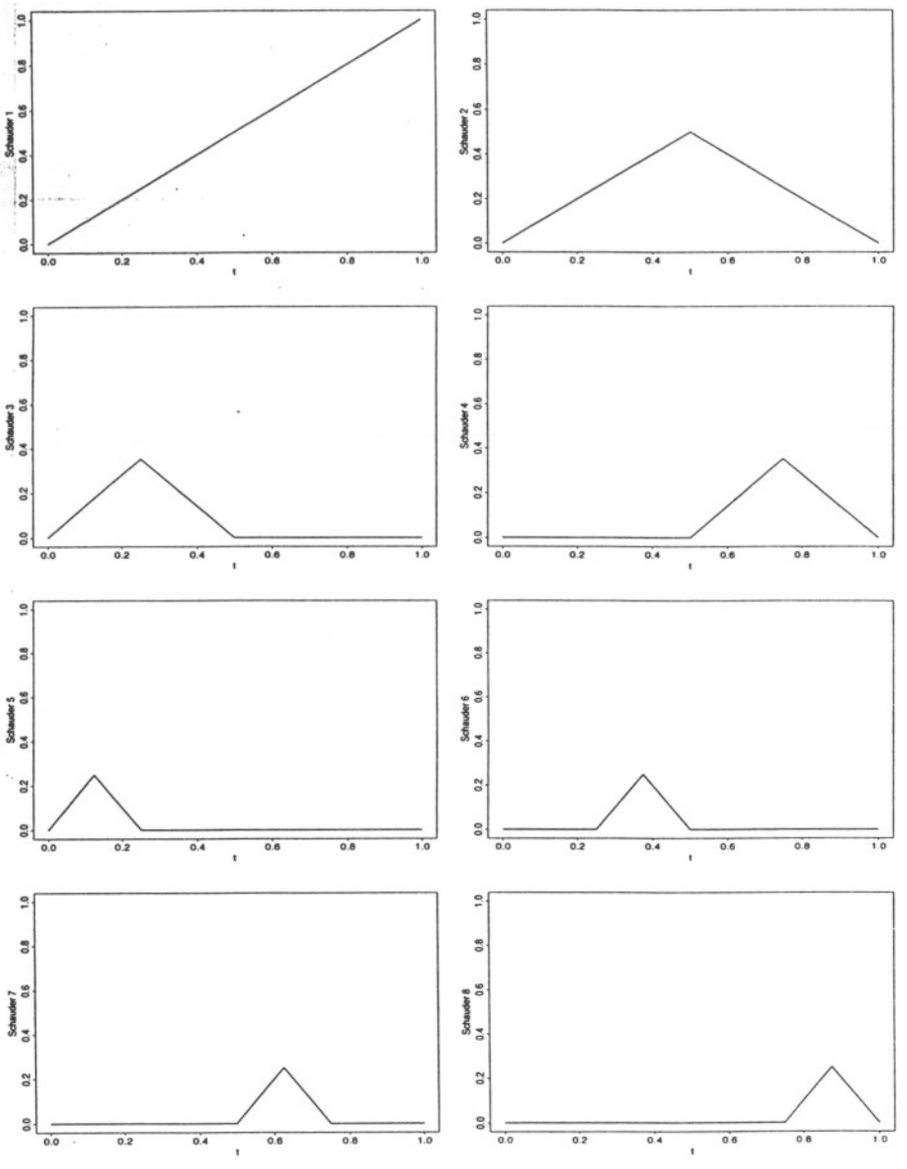

Figure 1.3.16 *The Schauder functions* $\widetilde{H}_1, \ldots, \widetilde{H}_8$.

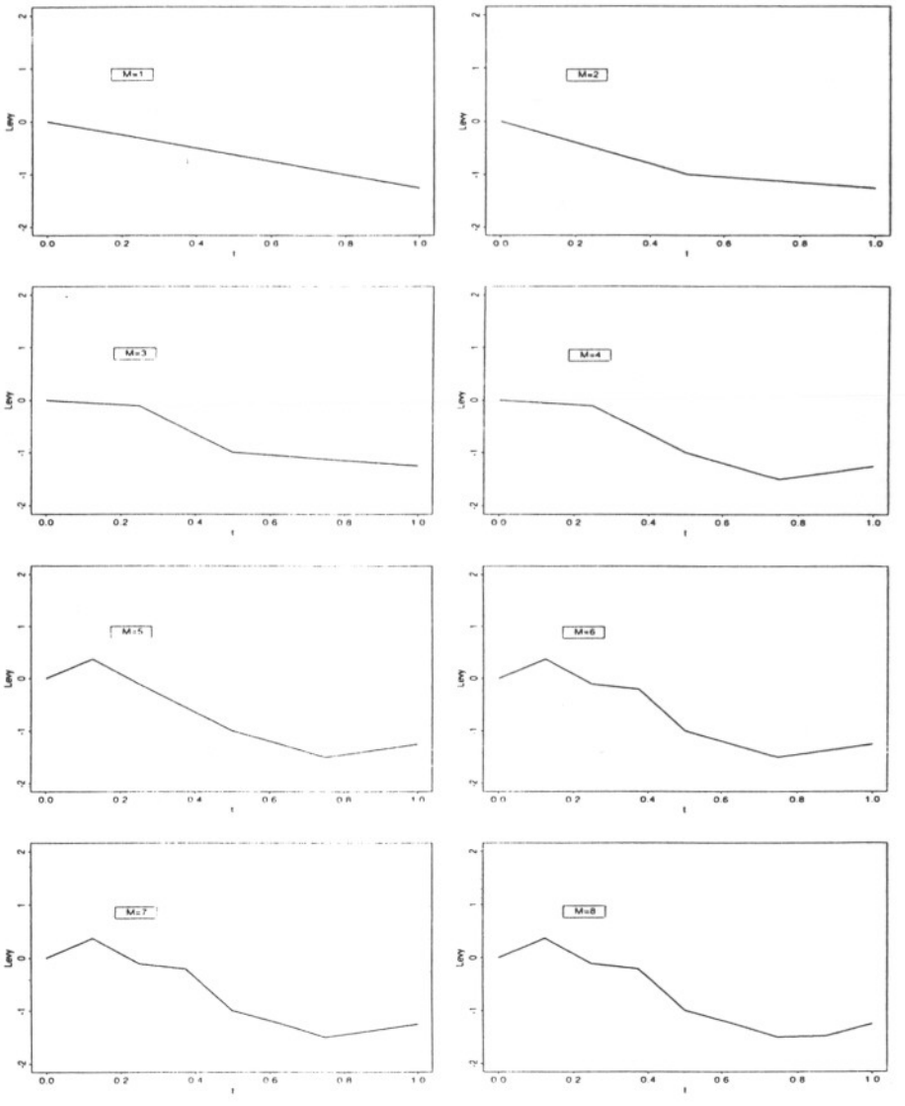

Figure 1.3.17 *The first steps in the construction of one Brownian sample path from the Lévy representation* (1.25) *via M Schauder functions,* $M = 1, \ldots, 8$.

and Shreve (1988) and Revuz and Yor (1991). The reader of these books must be familiar with the theory of stochastic processes, functional analysis, special functions and measure theory. Every textbook on stochastic processes also contains at least one chapter about Brownian motion; see the references on p. 33.

In addition to the non-differentiability and unbounded variation, Brownian motion has many more exciting path and distributional properties. Hida (1980) is a good reference to read about them.

The functional CLT is to be found in advanced textbooks on stochastic processes and the convergence of probability measures; see for example Billingsley (1968) or Pollard (1984). The series representations of Brownian motion can be found in Hida (1980); see also Ciesielski (1965).

1.4 Conditional Expectation

You cannot avoid this section; it contains material which is essential for the understanding of martingales, and more generally, Itô stochastic integrals.

If you are not interested in details you may try to read from one box to another. At the end of Section 1.4 you should know:

- the σ-field generated by a random variable, a random vector or a stochastic process; see Section 1.4.2,
- the conditional expectation of a random variable given a σ-field; see Section 1.4.3,
- the most common rules for calculating conditional expectations; see Section 1.4.4.

You should start with Section 1.4.1, where an example of a conditional expectation is given. It will give you some motivation for the abstract notion of conditional expectation given a σ-field, and every time when you get lost in this section, you should return to Section 1.4.1 and try to figure out what the general theory says in this concrete case.

1.4.1 Conditional Expectation under Discrete Condition

From an elementary course on probability theory we know the *conditional probability of A given B*, i.e.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

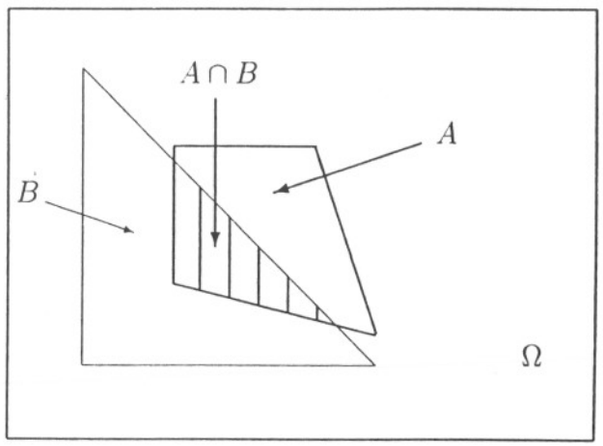

Figure 1.4.1 *The classical conditional probability: if we know that B occurred, we assign the new probability 1 to it. Events outside B cannot occur, hence they have the new probability 0.*

Clearly,

$$P(A \mid B) = P(A) \quad \text{if and only if } A \text{ and } B \text{ are independent.}$$

For the definition of $P(A \mid B)$ it is crucial that $P(B)$ is positive. It is the objective of Section 1.4.3 to relax this condition.

The probability $P(A \mid B)$ can be interpreted as follows. Assume the event B occurred. This is additional information which substantially changes the underlying probability measure. In particular, we assign the new probabilities 0 to B^c (we know that B^c will not happen) and 1 to B. The event B becomes our new probability space Ω', say. All events of interest are now subsets of Ω': $A \cap B \subset \Omega'$. In order to get a new probability measure on Ω' we have to normalize the old probabilities $P(A \cap B)$ by $P(B)$. In sum, the occurrence of B makes our original space Ω shrink to Ω', and the original probabilities $P(A)$ have to be replaced with $P(A \mid B)$.

Given that $P(B) > 0$, we can define the *conditional distribution function of a random variable X given B*

$$F_X(x \mid B) = \frac{P(X \le x\,,\, B)}{P(B)}, \quad x \in \mathbb{R},$$