Stochastic Analysis
An Introduction

Prof. Dr. Andreas Eberle

February 10, 2011

Contents

Chapter 1

Brownian Motion

This introduction to stochastic analysis starts with an introduction to Brownian motion. Brownian Motion is a diffusion process, i.e. a continuous-time Markov process $(B_t)_{t\geq 0}$ with continuous sample paths $t \mapsto B_t(\omega)$. In fact, it is the only nontrivial continuous-time process that is a Lévy process as well as a martingale and a Gaussian process. A rigorous construction of this process has been carried out first by N. Wiener in 1923. Already about 20 years earlier, related models had been introduced independently for financial markets by L. Bachelier [*Théorie de la spéculation*, Ann. Sci. École Norm. Sup. 17, 1900], and for the velocity of molecular motion by A. Einstein [*Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, Annalen der Physik 17, 1905*].
It has been a groundbreaking approach of K. Itô to construct general diffusion processes from Brownian motion, cf. [...]. In classical analysis, the solution of an ordinary differential equation $x'(t) = f(t, x(t))$ is a function, that can be approximated locally for t close to t_0 by the linear function $x(t_0) + f(t_0, x(t_0)) \cdot (t - t_0)$. Similarly, Itô showed, that a diffusion process behaves locally like a linear function of Brownian motion – the connection being described rigorously by a stochastic differential equation (SDE).

The fundamental rôle played by Brownian motion in stochastic analysis is due to the central limit Theorem. Similarly as the normal distribution arises as a universal scaling limit of standardized sums of independent, identically distributed, square integrable random variables, Brownian motion shows up as a universal scaling limit of Random Walks with square integrable increments.

1.1 From Random Walks to Brownian Motion

To motivate the definition of Brownian motion below, we first briefly discuss discrete-time stochastic processes and possible continuous-time scaling limits on an informal level.

A standard approach to model stochastic dynamics in discrete time is to start from a sequence of random variables $\eta_1, \eta_2, \ldots$ defined on a common probability space $(\Omega, \mathcal{A}, P)$. The random variables η_n describe the stochastic influences (*noise*) on the system. Often they are assumed to be *independent and identically distributed (i.i.d.)*. In this case the collection (η_n) is also called a **white noise**, where as a **colored noise** is given by dependent random variables. A stochastic process $X_n, n = 0, 1, 2, \ldots,$ taking values in $\mathbb{R}^d$ is then defined recursively on $(\Omega, \mathcal{A}, P)$ by

$$X_{n+1} \quad = \quad X_n + \Phi_{n+1}(X_n, \eta_{n+1}), \qquad n = 0, 1, 2, \ldots. \tag{1.1.1}$$

Here the Φ_n are measurable maps describing the *random law of motion*. If X_0 and $\eta_1, \eta_2, \ldots$ are independent random variables, then the process (X_n) is a Markov chain with respect to P.

Now let us assume that the random variables η_n are independent and identically distributed taking values in $\mathbb{R}$, or, more generally, $\mathbb{R}^d$. The easiest type of a nontrivial stochastic dynamics as described above is the Random Walk $S_n = \sum_{i=1}^{n} \eta_i$ which satisfies

$$S_{n+1} \quad = \quad S_n + \eta_{n+1} \qquad \text{for } n = 0, 1, 2, \ldots.$$

Since the noise random variables η_n are the increments of the Random Walk (S_n), the law of motion (1.1.1) in the general case can be rewritten as

$$X_{n+1} - X_n \quad = \quad \Phi_{n+1}(X_n, S_{n+1} - S_n), \qquad n = 0, 1, 2, \ldots. \tag{1.1.2}$$

This equation is a difference equation for (X_n) driven by the stochastic process (S_n).

Our aim is to carry out a similar construction as above for stochastic dynamics in continuous time. The stochastic difference equation (1.1.2) will then eventually be replaced by a *stochastic differential equation (SDE)*. However, before even being able to think about how to write down and make sense of such an equation, we have to identify a continuous-time stochastic process that takes over the rôle of the Random Walk. For this purpose, we first determine possible scaling limits of Random Walks when the time steps tend to 0. It will turn out that if the increments are square integrable and the size of the increments goes to 0 as the length of the time steps tends to 0, then by the Central Limit Theorem there is essentially only one possible limit process in continuous time: Brownian motion.

Central Limit Theorem

Suppose that $Y_{n,i} : \Omega \to \mathbb{R}^d, 1 \leq i \leq n < \infty$, are identically distributed, square-integrable random variables on a probability space $(\Omega, \mathcal{A}, P)$ such that $Y_{n,1}, \ldots, Y_{n,n}$ are independent for each $n \in \mathbb{N}$. Then the rescaled sums

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_{n,i} - E[Y_{n,i}])$$

converge in distribution to a multivariate normal distribution $N(0, C)$ with covariance matrix

$$C_{kl} \quad = \quad \mathrm{Cov}[Y_{n,i}^{(k)}, Y_{n,i}^{(l)}].$$

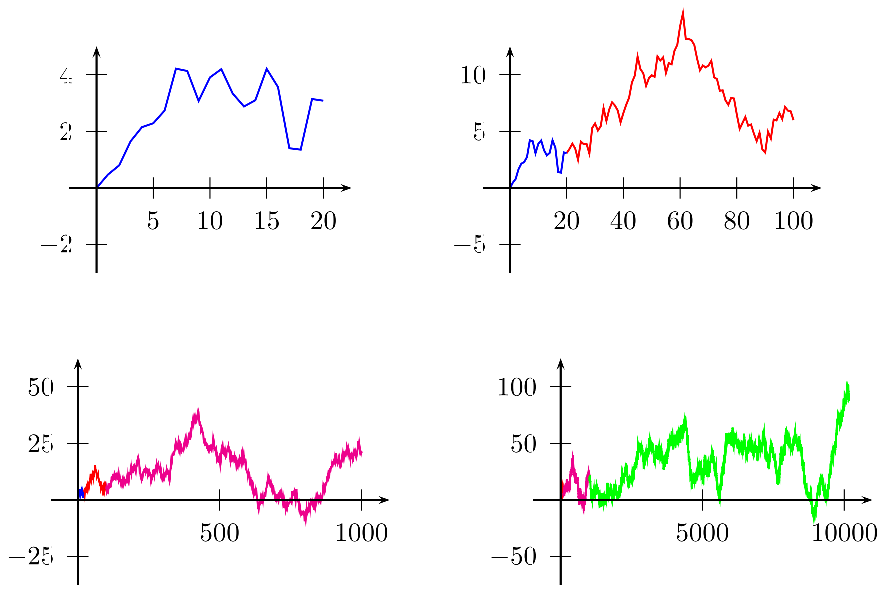
To see, how the CLT determines the possible scaling limits of Random Walks, let us consider a one-dimensional Random Walk

$$S_n = \sum_{i=1}^{n} \eta_i, \qquad n = 0, 1, 2, \ldots,$$

on a probability space $(\Omega, \mathcal{A}, P)$ with independent increments $\eta_i \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ normalized such that

$$E[\eta_i] \quad = \quad 0 \qquad \text{and} \qquad \mathrm{Var}[\eta_i] \quad = \quad 1. \tag{1.1.3}$$

Plotting many steps of the Random Walk seems to indicate that there is a limit process with continuous sample paths after appropriate rescaling:



To see what appropriate means, we fix a positive integer m, and try to define a rescaled Random Walk $S_t^{(m)}$ $(t = 0, 1/m, 2/m, \ldots)$ with time steps of size $1/m$ by

$$S_{k/m}^{(m)} \quad = \quad c_m \cdot S_k \qquad (k = 0, 1, 2, \ldots)$$

for some constants $c_m > 0$. If t is a multiple of $1/m$, then

$$\mathrm{Var}[S_t^{(m)}] \quad = \quad c_m^2 \cdot \mathrm{Var}[S_{mt}] \quad = \quad c_m^2 \cdot m \cdot t.$$

Hence in order to achieve convergence of $S_t^{(m)}$ as $m \to \infty$, we should choose c_m proportional to $m^{-1/2}$. This leads us to define a continuous time process $(S_t^{(m)})_{t \geq 0}$ by

$$S_t^{(m)}(\omega) \quad := \quad \frac{1}{\sqrt{m}} S_{mt}(\omega) \qquad \text{whenever } t = k/m \text{ for some integer } k,$$

and by linear interpolation for $t \in \left(\frac{k-1}{m}, \frac{k}{m}\right]$.

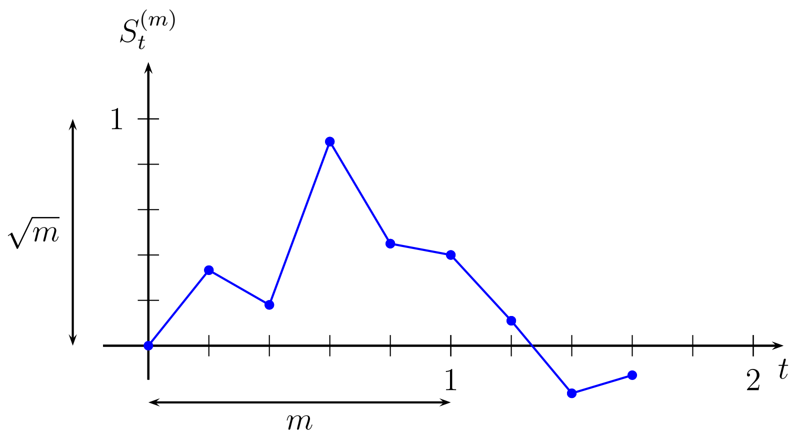


Figure 1.1: Rescaling of a Random Walk.

Clearly,

$$E[S_t^{(m)}] = 0 \qquad \text{for all } t \geq 0,$$

and

$$\mathrm{Var}[S_t^{(m)}] = \frac{1}{m}\mathrm{Var}[S_{mt}] = t$$

whenever t is a multiple of $1/m$. In particular, the expectation values and variances for a fixed time t do not depend on m. Moreover, if we fix a partition $0 \leq t_0 < t_1 < \ldots < t_n$ such that each t_i is a multiple of $1/m$, then the increments

$$S_{t_{i+1}}^{(m)} - S_{t_i}^{(m)} = \frac{1}{\sqrt{m}}\left(S_{mt_{i+1}} - S_{mt_i}\right), \qquad i = 0,1,2,\ldots,n-1, \tag{1.1.4}$$

of the rescaled process $(S_t^{(m)})_{t\geq 0}$ are independent centered random variables with variances $t_{i+1} - t_i$. If t_i is not a multiple of $1/m$, then a corresponding statement holds approximately with an error that should be negligible in the limit $m \to \infty$. Hence, if the rescaled Random Walks $(S_t^{(m)})_{t\geq 0}$ converge in distribution to a limit process $(B_t)_{t\geq 0}$, then $(B_t)_{t\geq 0}$ should have *independent increments* $B_{t_{i+1}} - B_{t_i}$ *over disjoint time intervals with mean* 0 *and variances* $t_{i+1} - t_i$.

It remains to determine the precise distributions of the increments. Here the Central Limit Theorem applies. In fact, we can observe that by (1.1.4) each increment

$$S_{t_{i+1}}^{(m)} - S_{t_i}^{(m)} = \frac{1}{\sqrt{m}}\sum_{k=mt_i+1}^{mt_{i+1}} \eta_k$$

of the rescaled process is a rescaled sum of $m \cdot (t_{i+1} - t_i)$ i.i.d. random variables with mean 0 and variance 1. Therefore, the CLT implies that the distributions of the increments converge weakly to a normal distribution:

$$S^{(m)}_{t_{i+1}} - S^{(m)}_{t_i} \quad \xrightarrow{\mathcal{D}} \quad N(0, t_{i+1} - t_i).$$

Hence if a limit process (B_t) exists, then it should have *independent, normally distributed increments*.

Our considerations motivate the following definition:

Definition (Brownian Motion).

(1). Let $a \in \mathbb{R}$. A continuous-time stochastic process $B_t : \Omega \to \mathbb{R}$, $t \geq 0$, defined on a probability space $(\Omega, \mathcal{A}, P)$, is called a ***Brownian motion (starting in** a**)*** *if and only if*

(a) $B_0(\omega) \;=\; a$ *for each $\omega \in \Omega$.*

(b) For any partition $0 \leq t_0 < t_1 < \ldots < t_n$, the increments $B_{t_{i+1}} - B_{t_i}$ are independent random variables with distribution

$$B_{t_{i+1}} - B_{t_i} \quad \sim \quad N(0, t_{i+1} - t_i).$$

(c) P-almost every sample path $t \mapsto B_t(\omega)$ is continuous.

(2). An $\mathbb{R}^d$-valued stochastic process $B_t(\omega) = (B_t^{(1)}(\omega), \ldots, B_t^{(d)}(\omega))$ is called a multi-dimensional Brownian motion if and only if the component processes $(B_t^{(1)}), \ldots, (B_t^{(d)})$ are independent one-dimensional Brownian motions.

Thus the increments of a d-dimensional Brownian motion are independent over disjoint time intervals and have a multivariate normal distribution:

$$B_t - B_s \quad \sim \quad N(0, (t-s) \cdot I_d) \qquad \text{for any } 0 \leq s \leq t.$$

Remark. (1). *Continuity:* Continuity of the sample paths has to be assumed separately: If $(B_t)_{t\geq 0}$ is a one-dimensional Brownian motion, then the modified process $(\widetilde{B}_t)_{t\geq 0}$ defined by $\widetilde{B}_0 = B_0$ and

$$\widetilde{B}_t \quad = \quad B_t \cdot I_{\{B_t \in \mathbb{R}\setminus\mathbb{Q}\}} \qquad \text{for } t > 0$$

has almost surely discontinuous paths. On the other hand, it satisfies (a) and (b) since the distributions of $(\widetilde{B}_{t_1}, \ldots, \widetilde{B}_{t_n})$ and $(B_{t_1}, \ldots, B_{t_n})$ coincide for all $n \in \mathbb{N}$ and $t_1, \ldots, t_n \geq 0$.

(2). *Spatial Homogeneity:* If $(B_t)_{t\geq 0}$ is a Brownian motion starting at 0, then the translated process $(a + B_t)_{t\geq 0}$ is a Brownian motion starting at a.

(3). *Existence:* There are several constructions and existence proofs for Brownian motion. In Section 1.3 below we will discuss in detail the Wiener-Lévy construction of Brownian motion as a random superposition of infinitely many deterministic paths. This explicit construction is also very useful for numerical approximations. A more general (but less constructive) existence proof is based on Kolmogorov's extension Theorem, cf. e.g. [Klenke].

(4). *Functional Central Limit Theorem:* The construction of Brownian motion as a scaling limit of Random Walks sketched above can also be made rigorous. *Donsker's invariance principle* is a functional version of the central limit Theorem which states that the rescaled Random Walks $(S_t^{(m)})$ converge in distribution to a Brownian motion. As in the classical CLT the limit is universal, i.e., it does not depend on the distribution of the increments η_i provided (1.1.3) holds, cf. Section **??**.

Brownian motion as a Lévy process.

The definition of Brownian motion shows in particular that Brownian motion is a *Lévy process*, i.e., it has stationary independent increments (over disjoint time intervals). In fact, the analogues of Lévy processes in discrete time are Random Walks, and it is rather obvious, that all scaling limits of Random Walks should be Lévy processes. Brownian motion is the only Lévy process L_t in continuous time with paths such that $E[L_1] = 0$ and $\mathrm{Var}[L_1] = 1$. The normal distribution of the increments follows under these assumptions by an extension of the CLT, cf. e.g. [Breiman: Probability]. A simple example of a Lévy process with non-continuous paths is the Poisson process. Other examples are α-stable processes which arise as scaling limits of Random Walks when the increments are not square-integrable. Stochastic analysis based on general Lévy processes has attracted a lot of interest recently.

Let us now consider consider a Brownian motion $(B_t)_{t\geq 0}$ starting at a fixed point $a \in \mathbb{R}^d$, defined on a probability space $(\Omega, \mathcal{A}, P)$. The information on the process up to time t is encoded in the σ-algebra

$$\mathcal{F}_t^B \quad = \quad \sigma(B_s \mid 0 \leq s \leq t)$$

generated by the process. The independence of the increments over disjoint intervals immediately implies:

Lemma 1.1. *For any $0 \leq s \leq t$, the increment $B_t - B_s$ is independent of $\mathcal{F}_s^B$.*

Proof. For any partition $0 = t_0 \leq t_1 \leq \ldots \leq t_n = s$ of the interval $[0, s]$, the increment $B_t - B_s$ is independent of the σ-algebra

$$\sigma(B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}})$$

generated by the increments up to time s. Since

$$B_{t_k} \quad = \quad B_{t_0} + \sum_{i=1}^{k} (B_{t_i} - B_{t_{i-1}})$$

and B_{t_0} is constant, this σ-algebra coincides with $\sigma(B_{t_0}, B_{t_1}, \ldots, B_{t_n})$. Hence $B_t - B_s$ is independent of all finite subcollections of $(B_u \,|\, 0 \leq u \leq s)$ and therefore independent of $\mathcal{F}_s^B$. □

Brownian motion as a Markov process.

As a process with stationary increments, Brownian motion is in particular a time-homogeneous Markov process. In fact, we have:

Theorem 1.2 (Markov property). *A Brownian motion $(B_t)_{t\geq 0}$ in $\mathbb{R}^d$ is a time-homogeneous Markov process with transition densities*

$$p_t(x, y) \quad = \quad (2\pi t)^{-d/2} \cdot \exp\left(-\frac{|x-y|^2}{2t}\right), \qquad t > 0, \quad x, y \in \mathbb{R}^d,$$

i.e., for any Borel set $A \subseteq \mathbb{R}^d$ and $0 \leq s < t$,

$$P[B_t \in a \mid \mathcal{F}_s^B] \quad = \quad \int_A p_{t-s}(B_s, y)\, dy \qquad P\text{-almost surely.}$$

Proof. For $0 \leq s < t$ we have $B_t = B_s + (B_t - B_s)$ where B_s is $\mathcal{F}_s^B$-measurable, and $B_t - B_s$ is independent of $\mathcal{F}_s^B$ by Lemma 1.1. Hence

$$\begin{aligned} P[B_t \in A \mid \mathcal{F}_s^B](\omega) \;&=\; P[B_s(\omega) + B_t - B_s \in A] \quad = \quad N(B_s(\omega), (t-s) \cdot I_d)[A] \\ =\; &\int_A (2\pi(t-s))^{-d/2} \cdot \exp\left(-\frac{|y - B_s(\omega)|^2}{2(t-s)}\right) dy \qquad P\text{-almost surely.} \end{aligned}$$

□

Remark (Heat equation as backward equation and forward equation). The transition function of Brownian motion is the *heat kernel* in $\mathbb{R}^d$, i.e., it is the fundamental solution of the heat equation

$$\frac{\partial u}{\partial t} \quad = \quad \frac{1}{2}\Delta u.$$

More precisely, $p_t(x, y)$ solves the initial value problem

$$\begin{aligned} \frac{\partial}{\partial t} p_t(x, y) \;&=\; \frac{1}{2}\Delta_x p_t(x, y) \qquad &&\text{for any } t > 0, x, y \in \mathbb{R}^d, \\ \lim_{t \searrow 0} \int p_t(x, y) f(y)\, dy \;&=\; f(x) \qquad &&\text{for any } f \in C_b(\mathbb{R}^d), x \in \mathbb{R}^d, \end{aligned} \tag{1.1.5}$$

where $\Delta_x = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}$ denotes the action of the Laplace operator on the x-variable. The equation (1.1.5) can be viewed as a version of *Kolmogorov's backward equation* for Brownian motion as a time-homogeneous Markov process, which states that for each $t > 0, y \in \mathbb{R}^d$ and $f \in C_b(\mathbb{R}^d)$, the function

$$v(s, x) \quad = \quad \int p_{t-s}(x, y) f(y)\, dy$$

solves the terminal value problem

$$\frac{\partial v}{\partial s}(s, x) \quad = \quad -\frac{1}{2}\Delta_x v(s, x) \qquad \text{for } s \in [0, t), \qquad \lim_{s \nearrow t} v(s, x) \;=\; f(x). \tag{1.1.6}$$

Note that by the Markov property, $v(s,x) = (p_{t-s}f)(x)$ is a version of the conditional expectation $E[f(B_t) \mid B_s = x]$. Therefore, the backward equation describes the dependence of the expectation value on starting point and time.

By symmetry, $p_t(x,y)$ also solves the initial value problem

$$\begin{aligned} \frac{\partial}{\partial t} p_t(x,y) &= \frac{1}{2}\Delta_y p_t(x,y) && \text{for any } t>0, \quad \text{and} \quad x,y\in\mathbb{R}^d, \\ \lim_{t\searrow 0}\int g(x)p_t(x,y)\,dx &= g(y) && \text{for any } g\in C_b(\mathbb{R}^d), y\in\mathbb{R}^d. \end{aligned} \tag{1.1.7}$$

The equation (1.1.7) is a version of *Kolmogorov's forward equation*, stating that for $g \in C_b(\mathbb{R}^d)$, the function $u(t,y) = \int g(x)p_t(x,y)\,dx$ solves

$$\frac{\partial u}{\partial t}(t,y) = \frac{1}{2}\Delta_y u(t,y) \qquad \text{for } t>0, \qquad \lim_{t\searrow 0} u(t,y) = g(y). \tag{1.1.8}$$

The forward equation describes the forward time evolution of the transition densities $p_t(x,y)$ for a given starting point x.

The Markov property enables us to compute the marginal distributions of Brownian motion:

Corollary 1.3 (Finite dimensional marginals). *Suppose that $(B_t)_{t\ge 0}$ is a Brownian motion starting at $x_0 \in \mathbb{R}^d$ defined on a probability space $(\Omega, \mathcal{A}, P)$. Then for any $n \in \mathbb{N}$ and $0 = t_0 < t_1 < t_2 < \ldots < t_n$, the joint distribution of $B_{t_1}, B_{t_2}, \ldots, B_{t_n}$ is absolutely continuous with density*

$$\begin{aligned} f_{B_{t_1},\ldots,B_{t_n}}(x_1,\ldots,x_n) &= p_{t_1}(x_0,x_1)p_{t_2-t_1}(x_1,x_2)p_{t_3-t_2}(x_2,x_3)\cdots p_{t_n-t_{n-1}}(x_{n-1},x_n) \\ &= \prod_{i=1}^n (2\pi(t_i-t_{i-1}))^{-d/2}\cdot\exp\left(-\frac{1}{2}\sum_{i=1}^n \frac{|x_i-x_{i-1}|^2}{t_i-t_{i-1}}\right) \end{aligned} \tag{1.1.9}$$

Proof. By the Markov property and induction on n, we obtain

$$
\begin{aligned}
& P[B_{t_1} \in A_1, \ldots, B_{t_n} \in A_n] \\
&= E[P[B_{t_n} \in A_n \mid \mathcal{F}^B_{t_{n-1}}] \,;\, B_{t_1} \in A_1, \ldots, B_{t_{n-1}} \in A_{n-1}] \\
&= E[p_{t_n - t_{n-1}}(B_{t_{n-1}}, A_n) \,;\, B_{t_1} \in A_1, \ldots, B_{t_{n-1}} \in A_{n-1}] \\
&= \int\limits_{A_1} \cdots \int\limits_{A_{n-1}} p_{t_1}(x_0, x_1) p_{t_2 - t_1}(x_1, x_2) \cdots \\
&\quad \cdot p_{t_{n-1} - t_{n-2}}(x_{n-2}, x_{n-1}) p_{t_n - t_{n-1}}(x_{n-1}, A_n)\, dx_{n-1} \cdots dx_1 \\
&= \int\limits_{A_1} \cdots \int\limits_{A_n} \left(\prod_{i=1}^n p_{t_i - t_{i-1}}(x_{n-1}, x_n) \right) dx_n \cdots dx_1
\end{aligned}
$$

for all $n \geq 0$ and $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R}^d)$. □

Remark (Brownian motion as a Gaussian process). The corollary shows in particular that Brownian motion is a Gaussian process, i.e., all the marginal distributions in (1.1.9) are multivariate normal distributions. We will come back to this important aspect in the next section.

Wiener Measure

The distribution of Brownian motion could be considered as a probability measure on the product space $(\mathbb{R}^d)^{[0,\infty)}$ consisting of all maps $x : [0,\infty) \to \mathbb{R}^d$. A disadvantage of this approach is that the product space is far too large for our purposes: It contains extremely irregular paths $x(t)$, although at least almost every path of Brownian motion is continuous by definition. Actually, since $[0,\infty)$ is uncountable, the subset of all continuous paths is not even measurable w.r.t. the product σ-algebra on $(\mathbb{R}^d)^{[0,\infty)}$.

Instead of the product space, we will directly consider the distribution of Brownian motion on the continuous path space $C([0,\infty), \mathbb{R}^d)$. For this purpose, we fix a Brownian motion $(B_t)_{t\geq 0}$ starting at $x_0 \in \mathbb{R}^d$ on a probability space $(\Omega, \mathcal{A}, P)$, and we **assume** that **every** sample path $t \mapsto B_t(\omega)$ is continuous. This assumption can always be fulfilled by modifying a given Brownian motion on a set of measure zero. The full process $(B_t)_{t\geq 0}$ can then be interpreted as a single path-space valued random variable (or a *"random path"*).

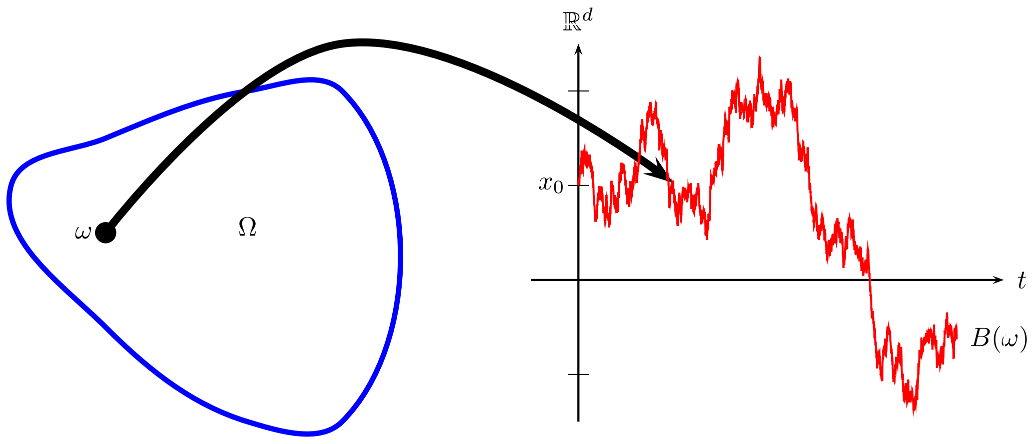


Figure 1.2: $B : \Omega \to C([0,\infty), \mathbb{R}^d), \quad B(\omega) = (B_t(\omega))_{t\geq 0}$.

We endow the space of continuous paths $x : [0,\infty) \to \mathbb{R}^d$ with the σ-algebra

$$\mathcal{B} = \sigma(X_t \mid t \geq 0)$$

generated by the coordinate maps

$$X_t : C([0,\infty), \mathbb{R}^d) \to \mathbb{R}^d, \quad X_t(x) = x_t, \qquad t \geq 0.$$

Note that we also have

$$\mathcal{B} = \sigma(X_t \mid t \in \mathcal{D})$$

for any dense subset $\mathcal{D}$ of $[0,\infty)$, because $X_t = \lim_{s\to t} X_s$ for each $t \in [0,\infty)$ by continuity. Furthermore, it can be shown that $\mathcal{B}$ is the Borel σ-algebra on $C([0,\infty), \mathbb{R}^d)$ endowed with the topology of uniform convergence on finite intervals.

Theorem 1.4 (Distribution of Brownian motion on path space). *The map $B : \Omega \to C([0,\infty), \mathbb{R}^d)$ is measurable w.r.t. the σ-algebras $\mathcal{A}/\mathcal{B}$. The distribution $P \circ B^{-1}$ of B is the unique probability measure μ_{x_0} on $(C([0,\infty), \mathbb{R}^d), \mathcal{B})$ with marginals*

$$\begin{aligned} &\mu_{x_0}\left[\{x \in C([0,\infty), \mathbb{R}^d) : x_{t_1} \in A_1, \ldots, x_{t_n} \in A_n\}\right] \qquad (1.1.10) \\ &= \prod_{i=1}^{n} (2\pi(t_i - t_{i-1}))^{-d/2} \int_{A_1} \cdots \int_{A_n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \frac{|x_i - x_{i-1}|^2}{t_i - t_{i-1}}\right) dx_n \cdots dx_1 \end{aligned}$$

for any $n \in \mathbb{N}$, $0 < t_1 < \ldots < t_n$, and $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R}^d)$.

Definition. *The probability measure μ_{x_0} on the path space $C([0,\infty),\mathbb{R}^d)$ determined by (1.1.10) is called **Wiener measure** (with start in x_0).*

Remark (Uniqueness in distribution). The Theorem asserts that the path space distribution of a Brownian motion starting at a given point x_0 is the corresponding Wiener measure. In particular, it is uniquely determined by the marginal distributions in (1.1.9).

Proof of Theorem 1.4. For $n \in \mathbb{N}, 0 < t_1 < \ldots < t_n$, and $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R}^d)$, we have

$$\begin{aligned} B^{-1}(\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\}) &= \{\omega : X_{t_1}(B(\omega)) \in A_1, \ldots, X_{t_n}(B(\omega)) \in A_n\} \\ &= \{B_{t_1} \in A_1, \ldots, B_{t_n} \in A_n\} \quad \in \quad \mathcal{A}. \end{aligned}$$

Since the cylinder sets of type $\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\}$ generate the σ-algebra $\mathcal{B}$, the map B is $\mathcal{A}/\mathcal{B}$-measurable. Moreover, by corollary 1.3, the probabilities

$$P[B \in \{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\}] \quad = \quad P[B_{t_1} \in A_1, \ldots, B_{t_n} \in A_n],$$

are given by the right hand side of (1.1.10). Finally, the measure μ_{x_0} is uniquely determined by (1.1.10), since the system of cylinder sets as above is stable under intersections and generates the σ-algebra $\mathcal{B}$. □

Definition (Canonical model for Brownian motion.). *By (1.1.10), the coordinate process*

$$X_t(x) \quad = \quad x_t, \qquad t \geq 0,$$

*on $C([0,\infty),\mathbb{R}^d)$ is a Brownian motion starting at x_0 w.r.t. Wiener measure μ_{x_0}. We refer to the stochastic process $(C([0,\infty),\mathbb{R}^d), \mathcal{B}, \mu_{x_0}, (X_t)_{t\geq 0})$ as the **canonical model for Brownian motion starting at** x_0.*

1.2 Brownian Motion as a Gaussian Process

We have already verified that Brownian motion is a Gaussian process, i.e., the finite dimensional marginals are multivariate normal distributions. We will now exploit this fact more thoroughly.

Multivariate normals

Let us first recall some basics on normal random vectors:

Definition. *Suppose that $m \in \mathbb{R}^n$ is a vector and $C \in \mathbb{R}^{n\times n}$ is a symmetric non-negative definite matrix. A random variable $Y : \Omega \to \mathbb{R}^n$ defined on a probability space $(\Omega, \mathcal{A}, P)$ has a* ***multivariate normal distribution*** $\mathbf{N}(\mathbf{m}, \mathbf{C})$ ***with mean*** $\mathbf{m}$ ***and covariance matrix*** $\mathbf{C}$ *if and only if its characteristic function is given by*

$$E[e^{ip\cdot Y}] \quad = \quad e^{ip\cdot m - \frac{1}{2}p\cdot Cp} \qquad \textit{for any } p \in \mathbb{R}^n. \tag{1.2.1}$$

If C is non-degenerate, then a multivariate normal random variable Y is absolutely continuous with density

$$f_Y(x) \quad = \quad (2\pi \det C)^{-1/2} \exp\left(-\frac{1}{2}(x-m)\cdot C^{-1}(x-m)\right).$$

A degenerate normal distribution with vanishing covariance matrix is a Dirac measure:

$$N(m, 0) \quad = \quad \delta_m.$$

Differentiating (1.2.1) w.r.t. p shows that for a random variable $Y \sim N(m, C)$, the mean vector is m and $C_{i,j}$ is the covariance of the components Y_i and Y_j. Moreover, the following important facts hold:

Theorem 1.5 (Properties of normal random vectors).

(1). A random variable $Y : \Omega \to \mathbb{R}^n$ has a multivariate normal distribution if and only if any linear combination

$$p \cdot Y \quad = \quad \sum_{i=1}^n p_i Y_i, \qquad p \in \mathbb{R}^n,$$

of the components Y_i has a one dimensional normal distribution.

(2). Any affine function of a normally distributed random vector Y is again normally distributed:

$$Y \sim N(m, C) \quad \Longrightarrow \quad AY + b \sim N(Am + b, ACA^\top)$$

for any $d \in \mathbb{N}, A \in \mathbb{R}^{d\times n}$ and $b \in \mathbb{R}^d$.

(3). *If* $Y = (Y_1, \dots, Y_n)$ *has a multivariate normal distribution, and the components* $Y_1, \dots, Y_n$ *are uncorrelated random variables, then* $Y_1, \dots, Y_n$ *are independent.*

Proof. (1). follows easily from the definition.

(2). For $Y \sim N(m, C)$, $A \in \mathbb{R}^{d \times n}$ and $b \in \mathbb{R}^d$ we have

$$\begin{aligned} E[e^{ip\cdot(AY+b)}] &= e^{ip\cdot b} E[e^{i(A^\top p)\cdot Y}] \\ &= e^{ip\cdot b} e^{i(A^\top p)\cdot m - \frac{1}{2}(A^\top p)\cdot CA^\top p} \\ &= e^{ip\cdot(Am+b) - \frac{1}{2}p\cdot ACA^\top} \qquad \text{for any } p \in \mathbb{R}^d, \end{aligned}$$

i.e., $AY + b \sim N(Am + b, ACA^\top)$.

(3). If $Y_1, \dots, Y_n$ are uncorrelated, then the covariance matrix $C_{i,j} = \mathrm{Cov}[Y_i, Y_j]$ is a diagonal matrix. Hence the characteristic function

$$E[e^{ip\cdot Y}] = e^{ip\cdot m - \frac{1}{2}p\cdot Cp} = \prod_{k=1}^{n} e^{im_k p_k - \frac{1}{2}C_{k,k}p_k^2}$$

is a product of characteristic functions of one-dimensional normal distributions. Since a probability measure on $\mathbb{R}^n$ is uniquely determined by its characteristic function, it follows that the adjoint distribution of $Y_1, \dots, Y_n$ is a product measure, i.e. $Y_1, \dots, Y_n$ are independent.

□

If Y has a multivariate normal distribution $N(m, C)$ then for any $p, q \in \mathbb{R}^n$, the random variables $p \cdot Y$ and $q \cdot Y$ are normally distributed with means $p \cdot m$ and $q \cdot m$, and covariance

$$\mathrm{Cov}[p \cdot Y, q \cdot Y] = \sum_{i,j=1}^{n} p_i C_{i,j} q_j = p \cdot Cq.$$

In particular, let $\{e_1, \dots, e_n\} \subseteq \mathbb{R}^n$ be an orthonormal basis consisting of eigenvectors of the covariance matrix C. Then the components $e_i \cdot Y$ of Y in this basis are uncorrelated and therefore independent, jointly normally distributed random variables with variances given by the corresponding eigenvectors λ_i:

$$\mathrm{Cov}[e_i \cdot Y, e_j \cdot Y] = \lambda_i \delta_{i,j}, \qquad 1 \le i, j \le n. \tag{1.2.2}$$

Correspondingly, the contour lines of the density of a non-degenerate multivariate normal distribution $N(m, C)$ are ellipsoids with center at m and principal axes of length $\sqrt{\lambda_i}$ given by the eigenvalues e_i of the covariance matrix C.

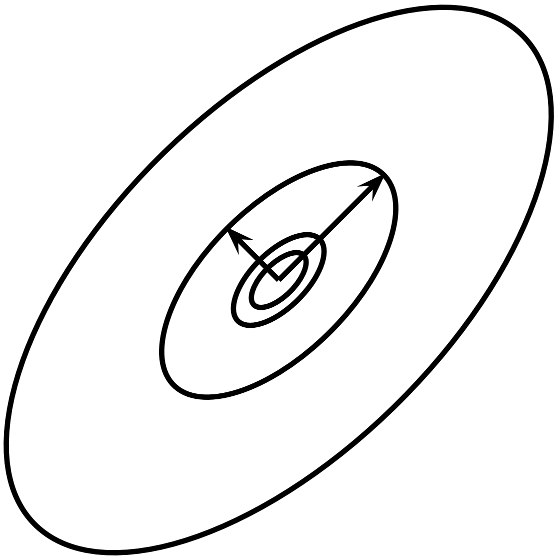


Figure 1.3: Level lines of the density of a normal random vector $Y \sim N\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\right)$.

Conversely, we can generate a random vector Y with distribution $N(m, C)$ from i.i.d. standard normal random variables $Z_1, \ldots, Z_n$ by setting

$$Y = m + \sum_{i=1}^{n} \sqrt{\lambda_i} Z_i e_i. \tag{1.2.3}$$

More generally, we have:

Corollary 1.6 (Generating normal random vectors). *Suppose that* $C = U \Lambda U^\top$ *with a matrix* $U \in \mathbb{R}^{n \times d}$, $d \in \mathbb{N}$, *and a diagonal matrix* $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^{d \times d}$ *with nonnegative entries* λ_i. *If* $Z = (Z_1, \ldots, Z_d)$ *is a random vector with i.i.d. standard normal random components* $Z_1, \ldots, Z_d$ *then*

$$Y = U \Lambda^{1/2} Z + m$$

has distribution $N(m, C)$.

Proof. Since $Z \sim N(0, I_d)$, the second assertion of Theorem 1.5 implies

$$Y \sim N(m, U\Lambda U^\top).$$

□

Choosing for U the matrix $(e_1, \ldots, e_n)$ consisting of the orthonormal eigenvectors $e_1, \ldots, e_n$ of C, we obtain (1.2.3) as a special case of the corollary. For computational purposes it is often more convenient to use the Cholesky decomposition

$$C = LL^\top$$

of the covariance matrix as a product of a lower triangular matrix L and the upper triangular transpose $L^\top$:

Algorithm 1.7 (**Simulation of multivariate normal random variables**).

Given: $m \in \mathbb{R}^n, C \in \mathbb{R}^{n\times n}$ symmetric and non-negative definite.

Output: Sample $y \sim N(m, C)$.

(1). Compute the Cholesky decomposition $C = LL^\top$.

(2). Generate independent samples $z_1, \ldots, z_n \sim N(0, 1)$ (e.g. by the Box-Muller method).

(3). Set $y := Lz + m$.

Gaussian processes

Let I be an arbitrary index set, e.g. $I = \mathbb{N}, I = [0, \infty)$ or $I = \mathbb{R}^n$.

Definition. *A collection* $(Y_t)_{t\in I}$ *of random variables* $Y_t : \Omega \to \mathbb{R}^d$ *defined on a probability space* $(\Omega, \mathcal{A}, P)$ *is called a* ***Gaussian process*** *if and only if the joint distribution of any finite subcollection* $Y_{t_1}, \ldots, Y_{t_n}$ *with* $n \in \mathbb{N}$ *and* $t_1, \ldots, t_n \in I$ *is a multivariate normal distribution.*

The distribution of a Gaussian process $(Y_t)_{t \in I}$ on the path space $\mathbb{R}^I$ or $C(I, \mathbb{R})$ endowed with the σ-algebra generated by the maps $x \mapsto x_t$, $t \in I$, is uniquely determined by the multinormal distributions of finite subcollections $Y_{t_1}, \ldots, Y_{t_n}$ as above, and hence by the expectation values

$$m(t) \quad = \quad E[Y_t], \qquad t \in I,$$

and the covariances

$$c(s,t) \quad = \quad \mathrm{Cov}[Y_s, Y_t], \qquad s,t \in I.$$

A Gaussian process is called **centered**, if $m(t) = 0$ for any $t \in I$.

Example ($\mathbf{AR(1)}$ process). The autoregressive process $(Y_n)_{n=0,1,2,\ldots}$ defined recursively by $Y_0 \sim N(0, v_0)$,

$$Y_n = \alpha Y_{n-1} + \varepsilon \eta_n \qquad \text{for } n \in \mathbb{N},$$

with parameters $v_0 > 0,\ \alpha, \varepsilon \in \mathbb{R},\ \eta_n$ i.i.d. $\sim N(0,1)$, is a centered Gaussian process. The covariance function is given by

$$c(n, n+k) \quad = \quad v_0 + \varepsilon^2 n \qquad \text{for any } n, k \geq 0 \quad \text{if } \alpha = 1,$$

and

$$c(n, n+k) \quad = \quad \alpha^k \cdot \left(\alpha^{2n} v_0 + (1 - \alpha^{2n}) \cdot \frac{\varepsilon^2}{1 - \alpha^2} \right) \qquad \text{for } n, k \geq 0 \quad \text{otherwise.}$$

This is easily verified by induction. We now consider some special cases:

$\alpha = \mathbf{0}$**:** In this case $Y_n = \varepsilon \eta_n$. Hence (Y_n) is a *white noise*, i.e., a sequence of independent normal random variables, and

$$\mathrm{Cov}[Y_n, Y_m] \quad = \quad \varepsilon^2 \cdot \delta_{n,m} \qquad \text{for any } n, m \geq 1.$$

$\alpha = \mathbf{1}$**:** Here $Y_n = Y_0 + \varepsilon \sum_{i=1}^{n} \eta_i$, i.e., the process (Y_n) is a *Gaussian Random Walk*, and

$$\mathrm{Cov}[Y_n, Y_m] \quad = \quad v_0 + \varepsilon^2 \cdot \min(n, m) \qquad \text{for any } n, m \geq 0.$$

We will see a corresponding expression for the covariances of Brownian motion.

$\alpha < 1$: For $\alpha < 1$, the covariances $\mathrm{Cov}[Y_n, Y_{n+k}]$ decay exponentially fast as $k \to \infty$. If $v_0 = \frac{\varepsilon^2}{1-\alpha^2}$, then the covariance function is translation invariant:

$$c(n, n+k) \quad = \quad \frac{\varepsilon^2 \alpha^k}{1-\alpha^2} \qquad \text{for any } n, k \geq 0.$$

Therefore, in this case the process (Y_n) is *stationary*, i.e., $(Y_{n+k})_{n\geq 0} \sim (Y_n)_{n\geq 0}$ for all $k \geq 0$.

Brownian motion is our first example of a nontrivial Gaussian process in continuous time. In fact, we have:

Theorem 1.8 (Gaussian characterization of Brownian motion). *A real-valued stochastic process $(B_t)_{t\in[0,\infty)}$ with continuous sample paths $t \mapsto B_t(\omega)$ and $B_0 = 0$ is a Brownian motion if and only if (B_t) is a centered Gaussian process with covariances*

$$\mathrm{Cov}[B_s, B_t] \quad = \quad \min(s,t) \qquad \textit{for any } s,t \geq 0. \tag{1.2.4}$$

Proof. For a Brownian motion (B_t) and $0 = t_0 < t_1 < \ldots < t_n$, the increments $B_{t_i} - B_{t_{i-1}}$, $1 \leq i \leq n$, are independent random variables with distribution $N(0, t_i - t_{i-1})$. Hence,

$$(B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}) \quad \sim \quad \bigotimes_{i=1}^{n} N(0, t_i - t_{i-1}),$$

which is a multinormal distribution. Since $B_{t_0} = B_0 = 0$, we see that

$$\begin{pmatrix} B_{t_1} \\ \vdots \\ B_{t_n} \end{pmatrix} \quad = \quad \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 1 & 0 & \ldots & 0 & 0 \\ & & \ddots & & & \\ & & & \ddots & & \\ 1 & 1 & 1 & \ldots & 1 & 0 \\ 1 & 1 & 1 & \ldots & 1 & 1 \end{pmatrix} \begin{pmatrix} B_{t_1} - B_{t_0} \\ \vdots \\ B_{t_n} - B_{t_{n-1}} \end{pmatrix}$$

also has a multivariate normal distribution, i.e., (B_t) is a Gaussian process. Moreover, since $B_t = B_t - B_0$, we have $E[B_t] = 0$ and

$$\mathrm{Cov}[B_s, B_t] \quad = \quad \mathrm{Cov}[B_s, B_s] + \mathrm{Cov}[B_s, B_t - B_s] \quad = \quad \mathrm{Var}[B_s] \quad = \quad s$$

for any $0 \leq s \leq t$, i.e., (1.2.4) holds.

Conversely, if (B_t) is a centered Gaussian process satisfying (1.2.4), then for any $0 = t_0 < t_1 < \ldots < t_n$, the vector $(B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}})$ has a multivariate normal distribution with

$$E[B_{t_i} - B_{t_{i-1}}] = E[B_{t_i}] - E[B_{t_{i-1}}] = 0, \qquad \text{and}$$

$$\begin{aligned} \operatorname{Cov}[B_{t_i} - B_{t_{i-1}}, B_{t_j} - B_{t_{j-1}}] &= \min(t_i, t_j) - \min(t_i, t_{j-1}) \\ &\quad - \min(t_{i-1}, t_j) + \min(t_{i-1}, t_{j-1}) \\ &= (t_i - t_{i-1}) \cdot \delta_{i,j} \qquad \text{for any } i, j = 1, \ldots, n. \end{aligned}$$

Hence by Theorem 1.5 (3), the increments $B_{t_i} - B_{t_{i-1}}, 1 \leq i \leq n$, are independent with distribution $N(0, t_i - t_{i-1})$, i.e., (B_t) is a Brownian motion. □

Symmetries of Brownian motion

A first important consequence of the Gaussian characterization of Brownian motion are several symmetry properties of Wiener measure:

Theorem 1.9 (Invariance properties of Wiener measure). *Let $(B_t)_{t\geq 0}$ be a Brownian motion starting at 0 defined on a probability space $(\Omega, \mathcal{A}, P)$. Then the following processes are again Brownian motions:*

(1). $(-B_t)_{t\geq 0}$ *(Reflection invariance)*

(2). $(B_{t+h} - B_h)_{t\geq 0}$ *for any* $h \geq 0$ *(Stationarity)*

(3). $(a^{-1/2} B_{at})_{t\geq 0}$ *for any* $a > 0$ *(Scale invariance)*

(4). The time inversion $(\widetilde{B}_t)_{t\geq 0}$ *defined by*

$$\widetilde{B}_0 = 0, \qquad \widetilde{B}_t = t \cdot B_{1/t} \quad \textit{for } t > 0.$$

Proof. The proofs of (1), (2) and (3) are left as an exercise to the reader. To show (4), we first note that for each $n \in \mathbb{N}$ and $0 \leq t_1 < \ldots < t_n$, the vector $(\widetilde{B}_{t_1}, \ldots, \widetilde{B}_{t_n})$ has a multivariate normal distribution since it is a linear transformation of $(B_{1/t_1}, \ldots, B_{1/t_n})$, $(B_0, B_{1/t_2}, \ldots, B_{1/t_n})$ respectively. Moreover,

$$\begin{aligned} E[\widetilde{B}_t] &= 0 \qquad \text{for any } t \geq 0, \\ \mathrm{Cov}[\widetilde{B}_s, \widetilde{B}_t] &= st \cdot \mathrm{Cov}[B_{1/s}, B_{1/t}] \\ &= st \cdot \min(\frac{1}{s}, \frac{1}{t}) \quad = \quad \min(t, s) \qquad \text{for any } s, t > 0, \quad \text{and} \\ \mathrm{Cov}[\widetilde{B}_0, \widetilde{B}_t] &= 0 \qquad \text{for any } t \geq 0. \end{aligned}$$

Hence $(\widetilde{B}_t)_{t \geq 0}$ is a centered Gaussian process with the covariance function of Brownian motion. By Theorem 1.8, it only remains to show that P-almost every sample path $t \mapsto \widetilde{B}_t(\omega)$ is continuous. This is obviously true for $t > 0$. Furthermore, since the finite dimensional marginals of the processes $(\widetilde{B}_t)_{t \geq 0}$ and $(B_t)_{t \geq 0}$ are multivariate normal distributions with the same means and covariances, the distributions of $(\widetilde{B}_t)_{t \geq 0}$ and $(B_t)_{t \geq 0}$ on the product space $\mathbb{R}^{(0,\infty)}$ endowed with the product σ-algebra generated by the cylinder sets agree. To prove continuity at 0 we note that the set

$$\left\{ x : (0, \infty) \to \mathbb{R} \;\middle|\; \lim_{\substack{t \searrow 0 \\ t \in \mathbb{Q}}} x_t = 0 \right\}$$

is measurable w.r.t. the product σ-algebra on $\mathbb{R}^{(0,\infty)}$. Therefore,

$$P\left[\lim_{\substack{t \searrow 0 \\ t \in \mathbb{Q}}} \widetilde{B}_t = 0\right] \quad = \quad P\left[\lim_{\substack{t \searrow 0 \\ t \in \mathbb{Q}}} B_t = 0\right] \quad = \quad 1.$$

Since $\widetilde{B}_t$ is almost surely continuous for $t > 0$, we can conclude that outside a set of measure zero,

$$\sup_{s \in (0,t)} |\widetilde{B}_s| \quad = \quad \sup_{s \in (0,t) \cap \mathbb{Q}} |\widetilde{B}_s| \quad \longrightarrow \quad 0 \qquad \text{as } t \searrow 0,$$

i.e., $t \mapsto \widetilde{B}_t$ is almost surely continuous at 0 as well. □

Remark (Long time asymptotics versus local regularity, LLN). The time inversion invariance of Wiener measure enables us to translate results on the long time asymptotics of Brownian motion ($t \nearrow \infty$) into local regularity results for Brownian paths ($t \searrow 0$) and vice versa. For example, the continuity of the process $(\widetilde{B}_t)$ at 0 is equivalent to the *law of large numbers*:

$$P\left[\lim_{t\to\infty} \frac{1}{t} B_t \;=\; 0\right] \quad = \quad P\left[\lim_{s\searrow 0} s B_{1/s} \;=\; 0\right] \quad = \quad 1.$$

At first glance, this looks like a simple proof of the LLN. However, the argument is based on the existence of a continuous Brownian motion, and the existence proof requires similar arguments as a direct proof of the law of large numbers.

Wiener measure as a Gaussian measure, path integral heuristics

Wiener measure (with start at 0) is the unique probability measure μ on the continuous path space $C([0,\infty),\mathbb{R}^d)$ such that the coordinate process

$$X_t : C([0,\infty),\mathbb{R}^d) \to \mathbb{R}^d, \qquad X_t(x) \;=\; x(t),$$

is a Brownian motion starting at 0. By Theorem 1.8, Wiener measure is a centered **Gaussian measure** on the infinite dimensional space $C([0,\infty),\mathbb{R}^d)$, i.e., for any $n \in \mathbb{N}$ and $t_1,\dots,t_n \in \mathbb{R}_+$, $(X_{t_1},\dots,X_{t_n})$ is normally distributed with mean 0. We now "derive" a heuristic representation of Wiener measure that is not mathematically rigorous but nevertheless useful:
Fix a constant $T > 0$. Then for $0 = t_0 < t_1 < \ldots < t_n \le T$, the distribution of $(X_{t_1},\dots,X_{t_n})$ w.r.t. Wiener measure is

$$\mu_{t_1,\dots,t_n}(dx_{t_1},\dots,dx_{t_n}) \quad = \quad \frac{1}{Z(t_1,\dots,t_n)} \exp\left(-\frac{1}{2}\sum_{i=1}^n \frac{|x_{t_i}-x_{t_{i-1}}|^2}{t_i - t_{i-1}}\right) \prod_{i=1}^n dx_{t_i}, \tag{1.2.5}$$

where $Z(t_1,\dots,t_n)$ is an appropriate finite normalization constant, and $x_0 := 0$. Now choose a sequence $(\tau_k)_{k\in\mathbb{N}}$ of partitions $0 = t_0^{(k)} < t_1^{(k)} < \ldots < t_{n(k)}^{(k)} = T$ of the interval

$[0,T]$ such that the mesh size $\max_i |t_{i+1}^{(k)} - t_i^{(k)}|$ tends to zero. Taking informally the limit in (1.2.5), we obtain the heuristic asymptotic representation

$$\mu(dx) \quad = \quad \frac{1}{Z_\infty} \exp\left(-\frac{1}{2}\int_0^T \left|\frac{dx}{dt}\right|^2 dt\right) \delta_0(dx_0) \prod_{t\in(0,T]} dx_t \tag{1.2.6}$$

for Wiener measure on continuous paths $x:[0,T]\to\mathbb{R}^d$ with a "normalizing constant" Z_∞. Trying to make the informal expression (1.2.6) rigorous fails for several reasons:

- The normalizing constant $Z_\infty = \lim_{k\to\infty} Z(t_1^{(k)},\ldots,t_{n(k)}^{(k)})$ is infinite.
- The integral $\int_0^T \left|\frac{dx}{dt}\right|^2 dt$ is also infinite for μ-almost every path x, since typical paths of Brownian motion are nowhere differentiable, cf. below.
- The product measure $\prod_{t\in(0,T]} dx_t$ can be defined on cylinder sets but an extension to the σ-algebra generated by the coordinate maps on $C([0,\infty),\mathbb{R}^d)$ does not exist.

Hence there are several infinities involved in the informal expression (1.2.6). These infinities magically balance each other such that the measure μ is well defined in contrast to all of the factors on the right hand side.

In physics, R. Feynman introduced correspondingly integrals w.r.t. "Lebesgue measure on path space", cf. e.g. the famous Feynman Lecture notes [...], or Glimm and Jaffe [...].

Although not mathematically rigorous, the heuristic expression (1.2.5) can be a very useful guide for intuition. Note for example that (1.2.5) takes the form

$$\mu(dx) \quad \propto \quad \exp(-\|x\|_H^2/2)\,\lambda(dx), \tag{1.2.7}$$

where $\|x\|_H = (x,x)_H^{1/2}$ is the norm induced by the inner product

$$(x,y)_H \quad = \quad \int_0^T \frac{dx}{dt}\frac{dy}{dt}dt \tag{1.2.8}$$

of functions $x, y : [0, T] \to \mathbb{R}^d$ vanishing at 0, and λ is a corresponding "infinite-dimensional Lebesgue measure" (which does not exist!). The vector space

$$H = \{x : [0,T] \to \mathbb{R}^d \;:\; x(0) = 0, x \text{ is absolutely continuous with } \frac{dx}{dt} \in L^2\}$$

is a Hilbert space w.r.t. the inner product (1.2.8). Therefore, (1.2.7) suggests to consider Wiener measure as a *standard normal distribution on* H. It turns out that this idea can be made rigorous although not as easily as one might think at first glance. The difficulty is that a standard normal distribution on an infinite-dimensional Hilbert space does not exist on the space itself but only on a larger space. In particular, we will see in the next sections that Wiener measure μ can indeed be realized on the continuous path space $C([0,T], \mathbb{R}^d)$, but μ-almost every path is not contained in H!

Remark (Infinite-dimensional standard normal distributions). The fact that a standard normal distribution on an infinite dimensional separable Hilbert space H can not be realized on the space H itself can be easily seen by contradiction: Suppose that μ is a standard normal distribution on H, and $e_n, n \in \mathbb{N}$, are infinitely many orthonormal vectors in H. Then by rotational symmetry, the balls

$$B_n = \left\{ x \in H \;:\; \|x - e_n\|_H < \frac{1}{2} \right\}, \qquad n \in \mathbb{N},$$

should all have the same measure. On the other hand, the balls are disjoint. Hence by σ-additivity,

$$\sum_{n=1}^{\infty} \mu[B_n] = \mu\left[\bigcup B_n\right] \leq \mu[H] = 1,$$

and therefore $\mu[B_n] = 0$ for all $n \in \mathbb{N}$. A scaling argument now implies

$$\mu[\{x \in H \;:\; \|x - h\| \leq \|h\|/2\}] = 0 \qquad \text{for all } h \in H,$$

and hence $\mu \equiv 0$.

1.3 The Wiener-Lévy Construction

In this section we discuss how to construct Brownian motion as a random superposition of deterministic paths. The idea already goes back to N. Wiener, who constructed

Brownian motion as a random Fourier series. The approach described here is slightly different and due to P. Lévy: The idea is to approximate the paths of Brownian motion on a finite time interval by their piecewise linear interpolations w.r.t. the sequence of dyadic partitions. This corresponds to a development of the Brownian paths w.r.t. Schauder functions ("wavelets") which turns out to be very useful for many applications including numerical simulations.

Our aim is to construct a one-dimensional Brownian motion B_t starting at 0 for $t \in [0,1]$. By stationarity and independence of the increments, a Brownian motion defined for all $t \in [0,\infty)$ can then easily be obtained from infinitely many independent copies of Brownian motion on $[0,1]$. We are hence looking for a random variable

$$B \;=\; (B_t)_{t\in[0,1]} \;:\Omega \;\longrightarrow\; C([0,1])$$

defined on a probability space $(\Omega,\mathcal{A},P)$ such that the distribution $P\circ B^{-1}$ is Wiener measure μ on the continuous path space $C([0,1])$.

A first attempt

Recall that μ_0 should be a kind of standard normal distribution w.r.t. the inner product

$$(x,y)_H \quad = \quad \int_0^1 \frac{dx}{dt}\frac{dy}{dt}\,dt \tag{1.3.1}$$

on functions $x,y:[0,1]\to\mathbb{R}$. Therefore, we could try to define

$$B_t(\omega) \quad := \quad \sum_{i=1}^{\infty} Z_i(\omega)e_i(t) \qquad \text{for } t\in[0,1] \text{ and } \omega\in\Omega, \tag{1.3.2}$$

where $(Z_i)_{i\in\mathbb{N}}$ is a sequence of independent standard normal random variables, and $(e_i)_{i\in\mathbb{N}}$ is an orthonormal basis in the Hilbert space

$$H \quad = \quad \{x:[0,1]\to\mathbb{R} \mid x(0)=0,\ x \text{ is absolutely continuous with } (x,x)_H<\infty\}. \tag{1.3.3}$$

However, the resulting series approximation does not converge in H:

Theorem 1.10. *Suppose* $(e_i)_{i\in\mathbb{N}}$ *is a sequence of orthonormal vectors in a Hilbert space* H *and* $(Z_i)_{i\in\mathbb{N}}$ *is a sequence of i.i.d. random variables with* $P[Z_i \neq 0] > 0$. *Then the series* $\sum\limits_{i=1}^{\infty} Z_i(\omega)e_i$ *diverges with probability* 1 *w.r.t. the norm on* H.

Proof. By orthonormality and by the law of large numbers,

$$\left\| \sum_{i=1}^{n} Z_i(\omega)e_i \right\|_H^2 \quad = \quad \sum_{i=1}^{n} Z_i(\omega)^2 \quad \longrightarrow \quad \infty$$

P-almost surely as $n \to \infty$. □

The Theorem again reflects the fact that a standard normal distribution on an infinite-dimensional Hilbert space can not be realized on the space itself.

To obtain a positive result, we will replace the norm

$$\|x\|_H \quad = \quad \left(\int_0^1 \left| \frac{dx}{dt} \right|^2 \, dt \right)^{\frac{1}{2}}$$

on H by the supremum norm

$$\|x\|_{\sup} \quad = \quad \sup_{t\in[0,1]} |x(t)|,$$

and correspondingly the Hilbert space H by the Banach space $C([0,1])$. Note that the supremum norm is weaker than the H-norm. In fact, for $x \in H$ and $t \in [0,1]$, the Cauchy-Schwarz inequality implies

$$|x(t)|^2 \quad = \quad \left| \int_0^t x'(s)\, ds \right|^2 \quad \leq \quad t \cdot \int_0^t |x'(s)|^2 \, ds \quad \leq \quad \|x\|_H^2,$$

and therefore

$$\|x\|_{\sup} \quad \leq \quad \|x\|_H \qquad \text{for any } x \in H.$$

There are two choices for an orthonormal basis of the Hilbert space H that are of particular interest: The first is the Fourier basis given by

$$e_0(t) \quad = \quad t, \qquad e_n(t) \quad = \quad \frac{\sqrt{2}}{\pi n} \sin(\pi n t) \qquad \text{for } n \geq 1.$$

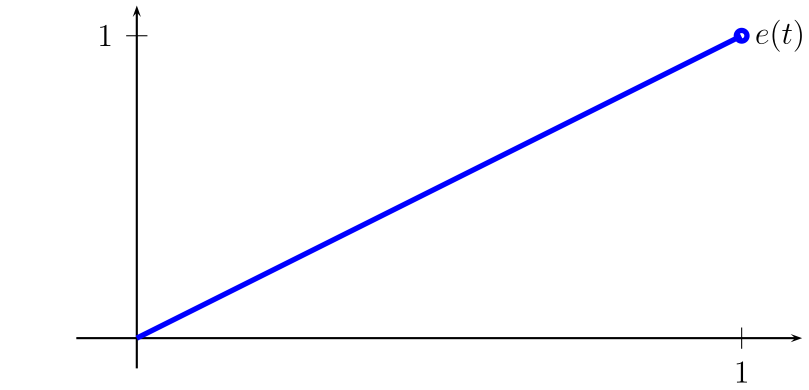
With respect to this basis, the series in (1.3.2) is a Fourier series with random coefficients. Wiener's original construction of Brownian motion is based on a *random Fourier series*. A second convenient choice is the basis of *Schauder functions* ("wavelets") that has been used by P. Lévy to construct Brownian motion. Below, we will discuss Lévy's construction in detail. In particular, we will prove that for the Schauder functions, the series in (1.3.2) converges almost surely w.r.t. the supremum norm towards a continuous (but not absolutely continuous) random path $(B_t)_{t\in[0,1]}$. It is then not difficult to conclude that $(B_t)_{t\in[0,1]}$ is indeed a Brownian motion.

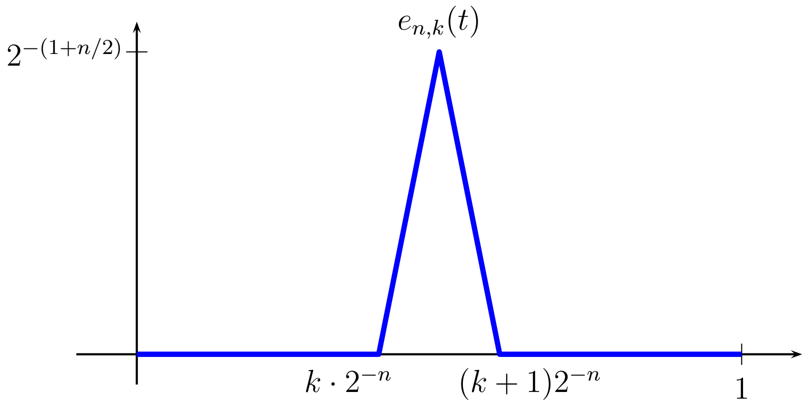
The Wiener-Lévy representation of Brownian motion

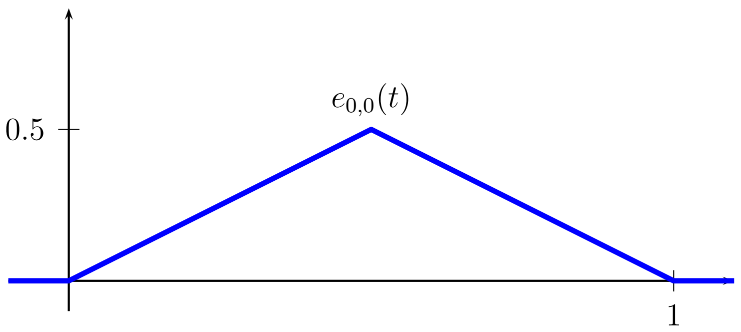
Before carrying out Lévy's construction of Brownian motion, we introduce the Schauder functions, and we show how to expand a given Brownian motion w.r.t. this basis of function space. Suppose we would like to approximate the paths $t \mapsto B_t(\omega)$ of a Brownian motion by their piecewise linear approximations adapted to the sequence of dyadic partitions of the interval $[0,1]$. An obvious advantage of this approximation over a Fourier expansion is that the values of the approximating functions at the dyadic points remain fixed once the approximating partition is fine enough. The piecewise linear approximations of a continuous function on $[0,1]$ correspond to a series expansion w.r.t. the base functions

$$\begin{aligned} e(t) &= t \quad \text{, and} \\ e_{n,k}(t) &= 2^{-n/2} e_{0,0}(2^n t - k), \quad n = 0,1,2,\ldots, k = 0,1,2,\ldots,2^{n-1}, \quad \text{, where} \end{aligned}$$

$$e_{0,0}(t) = \min(t, 1-t)^+ = \begin{cases} t & \text{for } t \in [0,1/2] \\ 1-t & \text{for } t \in (1/2,1] \\ 0 & \text{for } t \in \mathbb{R} \setminus [0,1] \end{cases} .$$







The functions $e_{n,k}$ $(n \geq 0, 0 \leq k < 2^n)$ are called **Schauder functions**. It is rather obvious that piecewise linear approximation w.r.t. the dyadic partitions corresponds to the expansion of a function $x \in C([0,1])$ with $x(0) = 0$ in the basis given by $e(t)$ and the Schauder functions. The normalization constants in defining the functions $e_{n,k}$ have been chosen in such a way that the $e_{n,k}$ are orthonormal w.r.t. the H-inner product introduced above.

Definition. *A sequence $(e_i)_{i\in\mathbb{N}}$ of vectors in an infinite-dimensional Hilbert space H is called an* ***orthonormal basis*** *(or* ***complete orthonormal system****) of H if and only if*

(1). Orthonormality: $(e_i, e_j) = \delta_{ij}$ *for any $i, j \in \mathbb{N}$, and*

(2). Completeness: Any $h \in H$ can be expressed as

$$h = \sum_{i=1}^{\infty} (h, e_i)_H e_i.$$

Remark (Equivalent characterizations of orthonormal bases). Let $e_i, i \in \mathbb{N}$, be orthonormal vectors in a Hilbert space H. Then the following conditions are equivalent:

(1). $(e_i)_{i\in\mathbb{N}}$ is an orthonormal basis of H.

(2). The linear span

$$\text{span}\{e_i \mid i \in \mathbb{N}\} = \left\{ \sum_{i=1}^{k} c_i e_i \,\middle|\, k \in \mathbb{N}, c_1, \ldots, c_k \in \mathbb{R} \right\}$$

is a dense subset of H.

(3). There is no element $x \in H, x \neq 0$, such that $(x, e_i)_H = 0$ for every $i \in \mathbb{N}$.

(4). For any element $x \in H$, *Parseval's relation*

$$\|x\|_H^2 = \sum_{i=1}^{\infty} (x, e_i)_H^2 \tag{1.3.4}$$

holds.

(5). For any $x, y \in H$,

$$(x, y)_H = \sum_{i=1}^{\infty} (x, e_i)_H (y, e_i)_H. \tag{1.3.5}$$

For the proofs we refer to any book on functional analysis, cf. e.g. [Reed and Simon: Methods of modern mathematical physics, Vol. I].

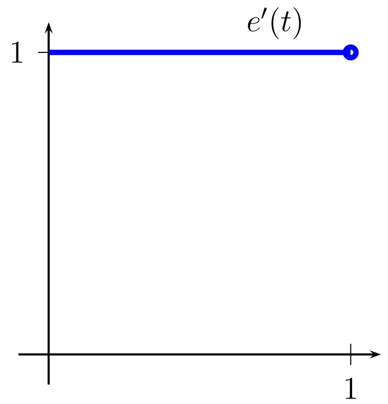
Lemma 1.11. *The Schauder functions e and $e_{n,k}$ $(n \geq 0, 0 \leq k < 2^n)$ form an orthonormal basis in the Hilbert space H defined by (1.3.3).*

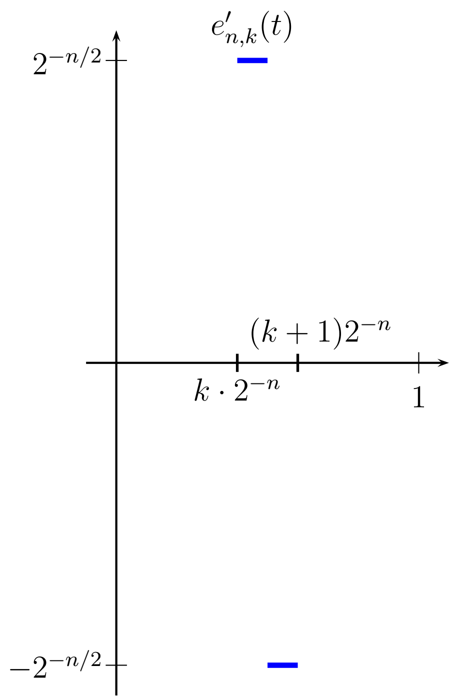
Proof. By definition of the inner product on H, the linear map d/dt which maps an absolutely continuous function $x \in H$ to its derivative $x' \in L^2(0,1)$ is an isometry from H onto $L^2(0,1)$, i.e.,

$$(x,y)_H \quad = \quad (x',y')_{L^2(0,1)} \qquad \text{for any } x,y \in H.$$

The derivatives of the Schauder functions are the Haar functions

$$\begin{aligned} e'(t) &\equiv 1, \\ e'_{n,k}(t) &= 2^{n/2}(I_{[k\cdot 2^{-n},(k+1/2)\cdot 2^{-n})}(t) - I_{[(k+1/2)\cdot 2^{-n},(k+1)\cdot 2^{-n})}(t)) \qquad \text{for a.e. } t. \end{aligned}$$





It is easy to see that these functions form an orthonormal basis in $L^2(0,1)$. In fact, orthonormality w.r.t. the L^2 inner product can be verified directly. Moreover, the linear span of the functions e' and $e'_{n,k}$ for $n = 0,1,\dots,m$ and $k = 0,1,\dots,2^n-1$ consists of all step functions that are constant on each dyadic interval $[j\cdot 2^{-(m+1)},(j+1)\cdot 2^{-(m+1)})$. An arbitrary function in $L^2(0,1)$ can be approximated by dyadic step functions w.r.t.

the L^2 norm. This follows for example directly from the L^2 martingale convergence Theorem, cf. ... below. Hence the linear span of e' and the Haar functions $e'_{n,k}$ is dense in $L^2(0,1)$, and therefore these functions form an orthonormal basis of the Hilbert space $L^2(0,1)$. Since $x \mapsto x'$ is an isometry from H onto $L^2(0,1)$, we can conclude that e and the Schauder functions $e_{n,k}$ form an orthonormal basis of H. □

The expansion of a function $x : [0,1] \to \mathbb{R}$ in the basis of Schauder functions can now be made explicit. The coefficients of a function $x \in H$ in the expansion are

$$(x,e)_H = \int_0^1 x'e'\,dt = \int_0^1 x'\,dt = x(1)-x(0) = x(1)$$

$$\begin{aligned}(x,e_{n,k})_H = \int_0^1 x'e'_{n,k}\,dt &= 2^{n/2}\int_0^1 x'(t)e'_{0,0}(2^n t-k)\,dt \\ &= 2^{n/2}\left[(x((k+\tfrac{1}{2})\cdot 2^{-n})-x(k\cdot 2^{-n}))-(x((k+1)\cdot 2^{-n})-x((k+\tfrac{1}{2})\cdot 2^{-n}))\right].\end{aligned}$$

Theorem 1.12. *Let $x \in C([0,1])$. Then the expansion*

$$x(t) = x(1)e(t) - \sum_{n=0}^{\infty}\sum_{k=0}^{2^n-1} 2^{n/2}\Delta_{n,k}x\cdot e_{n,k}(t),$$

$$\Delta_{n,k}x = \left[(x((k+1)\cdot 2^{-n})-x((k+\tfrac{1}{2})\cdot 2^{-n}))-(x((k+\tfrac{1}{2})\cdot 2^{-n})-x(k\cdot 2^{-n}))\right]$$

holds w.r.t. uniform convergence on $[0,1]$. For $x \in H$ the series also converges w.r.t. the stronger H-norm.

Proof. It can be easily verified that by definition of the Schauder functions, for each $m \in \mathbb{N}$ the partial sum

$$x^{(m)}(t) := x(1)e(t) - \sum_{n=0}^{m}\sum_{k=0}^{2^n-1} 2^{n/2}\Delta_{n,k}x\cdot e_{n,k}(t) \tag{1.3.6}$$

is the polygonal interpolation of $x(t)$ w.r.t. the $(m+1)$-th dyadic partition of the interval $[0,1]$. Since the function x is uniformly continuous on $[0,1]$, the polygonal interpolations converge uniformly to x. This proves the first statement. Moreover, for $x \in H$, the series is the expansion of x in the orthonormal basis of H given by the Schauder functions, and therefore it also converges w.r.t. the H-norm. □

Applying the expansion to the paths of a Brownian motions, we obtain:

Corollary 1.13 (Wiener-Lévy representation). *For a Brownian motion* $(B_t)_{t\in[0,1]}$ *the series representation*

$$B_t(\omega) \quad = \quad Z(\omega)e(t) + \sum_{n=0}^{\infty}\sum_{k=0}^{2^n-1} Z_{n,k}(\omega)e_{n,k}(t), \qquad t\in[0,1], \tag{1.3.7}$$

holds w.r.t. uniform convergence on $[0,1]$ *for* P*-almost every* $\omega \in \Omega$*, where*

$$Z \;:=\; B_1, \quad \textit{and} \quad Z_{n,k} \;:=\; -2^{n/2}\Delta_{n,k}B \qquad (n\geq 0, 0\leq k\leq 2^n-1)$$

are independent random variables with standard normal distribution.

Proof. It only remains to verify that the coefficients Z and $Z_{n,k}$ are independent with standard normal distribution. A vector given by finitely many of these random variables has a multivariate normal distribution, since it is a linear transformation of increments of the Brownian motion B_t. Hence it suffices to show that the random variables are uncorrelated with variance 1. This is left as an exercise to the reader. □

Lévy's construction of Brownian motion

The series representation (1.3.7) can be used to construct Brownian motion starting from independent standard normal random variables. The resulting construction does not only prove existence of Brownian motion but it is also very useful for numerical implementations:

Theorem 1.14 (P. Lévy 1948). *Let* Z *and* $Z_{n,k}$ $(n\geq 0, 0\leq k\leq 2^n-1)$ *be independent standard normally distributed random variables on a probability space* $(\Omega,\mathcal{A},P)$*. Then the series in (1.3.7) converges uniformly on* $[0,1]$ *with probability* 1*. The limit process* $(B_t)_{t\in[0,1]}$ *is a Brownian motion.*

The convergence proof relies on a combination of the Borel-Cantelli Lemma and the Weierstrass criterion for uniform convergence of series of functions. Moreover, we will need the following result to identify the limit process as a Brownian motion:

Lemma 1.15 (Parseval relation for Schauder functions). *For any* $s,t \in [0,1]$,

$$e(t)e(s) + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} e_{n,k}(t)e_{n,k}(s) \quad = \quad \min(t,s).$$

Proof. Note that for $g \in H$ and $s \in [0,1]$, we have

$$g(s) \quad = \quad g(s) - g(0) \quad = \quad \int_0^1 g' \cdot I_{(0,s)} \quad = \quad (g, h^{(s)})_H,$$

where $h^{(s)}(t) \;:=\; \int_0^t I_{(0,s)} \;=\; \min(s,t)$. Hence the Parseval relation (1.3.4) applied to the functions $h^{(s)}$ and $h^{(t)}$ yields

$$\begin{aligned} & e(t)e(s) + \sum_{n,k} e_{n,k}(t)e_{n,k}(s) \\ &= \; (e,h^{(t)})(e,h^{(s)}) + \sum_{n,k} (e_{n,k}, h^{(t)})(e_{n,k}, h^{(s)}) \\ &= \; (h^{(t)}, h^{(s)}) \quad = \quad \int_0^1 I_{(0,t)} I_{(0,s)} \quad = \quad \min(t,s). \end{aligned}$$

□

Proof of Theorem 1.14. We proceed in 4 steps:

(1). *Uniform convergence for P-a.e. ω:* By the Weierstrass criterion, a series of functions converges uniformly if the sum of the supremum norms of the summands is finite. To apply the criterion, we note that for any fixed $t \in [0,1]$ and $n \in \mathbb{N}$, only one of the functions $e_{n,k}$, $k = 0, 1, \ldots, 2^n - 1$, does not vanish at t. Moreover, $|e_{n,k}(t)| \leq 2^{-n/2}$. Hence

$$\sup_{t\in[0,1]} \left| \sum_{k=0}^{2^n-1} Z_{n,k}(\omega) e_{n,k}(t) \right| \quad \leq \quad 2^{-n/2} \cdot M_n(\omega), \tag{1.3.8}$$

where

$$M_n \quad := \quad \max_{0\leq k<2^n} |Z_{n,k}|.$$

We now apply the Borel-Cantelli Lemma to show that with probability 1, M_n grows at most linearly. Let Z denote a standard normal random variable. Then we have

$$P[M_n > n] \leq 2^n \cdot P[|Z| > n] \leq \frac{2^n}{n} \cdot E[|Z|\,;\,|Z| > n]$$
$$= \frac{2 \cdot 2^n}{n \cdot \sqrt{2\pi}} \int_n^\infty x e^{-x^2/2}\,dx = \sqrt{\frac{2}{\pi}} \frac{2^n}{n} \cdot e^{-n^2/2}$$

for any $n \in \mathbb{N}$. Since the sequence on the right hand side is summable, $M_n \leq n$ holds eventually with probability one. Therefore, the sequence on the right hand side of (1.3.8) is also summable for P-almost every ω. Hence, by (1.3.8) and the Weierstrass criterion, the partial sums

$$B_t^{(m)}(\omega) = Z(\omega)e(t) + \sum_{n=0}^{m} \sum_{k=0}^{2^n-1} Z_{n,k}(\omega)e_{n,k}(t), \qquad m \in \mathbb{N},$$

converge almost surely uniformly on $[0,1]$. Let

$$B_t = \lim_{m\to\infty} B_t^{(m)}$$

denote the almost surely defined limit.

(2). *L^2 convergence for fixed t:* We now want to prove that the limit process (B_t) is a Brownian motion, i.e., a continuous Gaussian process with $E[B_t] = 0$ and $\mathrm{Cov}[B_t, B_s] = \min(t,s)$ for any $t, s \in [0,1]$. To compute the covariances we first show that for a given $t \in [0,1]$ the series approximation $B_t^{(m)}$ of B_t converges also in L^2. Let $l, m \in \mathbb{N}$ with $l < m$. Since the $Z_{n,k}$ are independent (and hence uncorrelated) with variance 1, we have

$$E[(B_t^{(m)} - B_t^{(l)})^2] = E\left[\left(\sum_{n=l+1}^{m} \sum_{k=0}^{2^n-1} Z_{n,k}e_{n,k}(t)\right)^2\right] = \sum_{n=l+1}^{m} \sum_k e_{n,k}(t)^2.$$

The right hand side converges to 0 as $l, m \to \infty$ since $\sum_{n,k} e_{n,k}(t)^2 < \infty$ by Lemma 1.15. Hence $B_t^{(m)}, m \in \mathbb{N}$, is a Cauchy sequence in $L^2(\Omega, \mathcal{A}, P)$. Since $B_t = \lim_{m\to\infty} B_t^{(m)}$ almost surely, we obtain

$$B_t^{(m)} \overset{m\to\infty}{\longrightarrow} B_t \qquad \text{in } L^2(\Omega, \mathcal{A}, P).$$

(3). *Expectations and Covariances:* By the L^2 convergence we obtain for any $s,t \in [0,1]$:

$$E[B_t] \quad = \quad \lim_{m\to\infty} E[B_t^{(m)}] \quad = \quad 0, \qquad \text{and}$$

$$\begin{aligned} \mathrm{Cov}[B_t,B_s] \quad &= \quad E[B_tB_s] \quad = \quad \lim_{m\to\infty} E[B_t^{(m)}B_s^{(m)}] \\ &= \quad e(t)e(s) + \lim_{m\to\infty} \sum_{n=0}^{m}\sum_{k=0}^{2^n-1} e_{n,k}(t)e_{n,k}(s). \end{aligned}$$

Here we have used again that the random variables Z and $Z_{n,k}$ are independent with variance 1. By Parseval's relation (Lemma 1.15), we conclude

$$\mathrm{Cov}[B_t,B_s] \quad = \quad \min(t,s).$$

Since the process $(B_t)_{t\in[0,1]}$ has the right expectations and covariances, and, by construction, almost surely continuous paths, it only remains to show that (B_t) is a Gaussian process in oder to complete the proof:

(4). $(B_t)_{t\in[0,1]}$ *is a Gaussian process:* We have to show that $(B_{t_1},\ldots,B_{t_l})$ has a multivariate normal distribution for any $0 \le t_1 < \ldots < t_l \le 1$. By Theorem 1.5, it suffices to verify that any linear combination of the components is normally distributed. This holds by the next Lemma since

$$\sum_{j=1}^{l} p_j B_{t_j} \quad = \quad \lim_{m\to\infty} \sum_{j=1}^{l} p_j B_{t_j}^{(m)} \qquad P\text{-a.s.}$$

is an almost sure limit of normally distributed random variables for any $p_1,\ldots,p_l \in \mathbb{R}$.

Combining Steps $3,4$ and the continuity of sample paths, we conclude that $(B_t)_{t\in[0,1]}$ is indeed a Brownian motion. □

Lemma 1.16. *Suppose that $(X_n)_{n\in\mathbb{N}}$ is a sequence of normally distributed random variables defined on a joint probability space $(\Omega,\mathcal{A},P)$, and X_n converges almost surely to a random variable X. Then X is also normally distributed.*

Proof. Suppose $X_n \sim N(m_n, \sigma_n^2)$ with $m_n \in \mathbb{R}$ and $\sigma_n \in (0, \infty)$. By the Dominated Convergence Theorem,

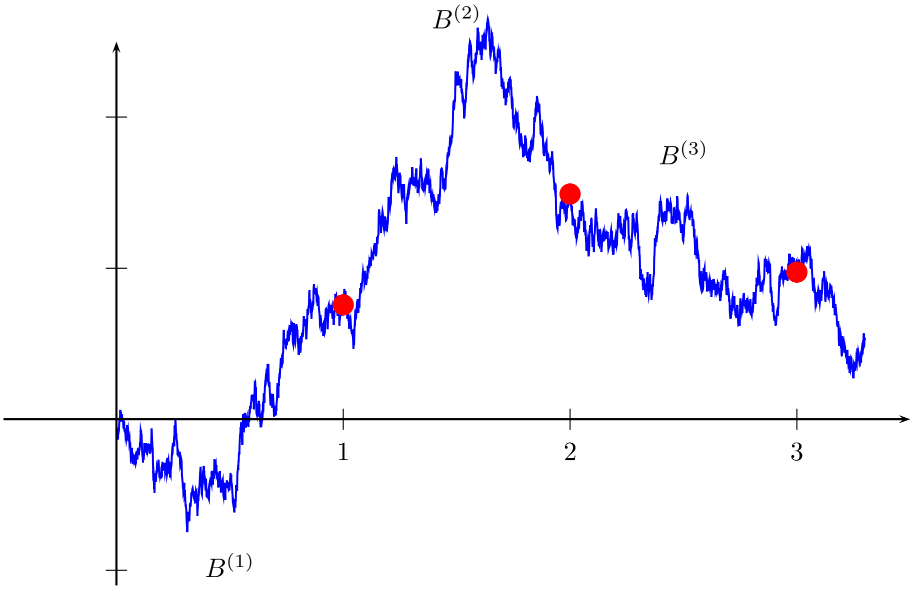
$$E[e^{ipX}] = \lim_{n\to\infty} E[e^{ipX_n}] = \lim_{n\to\infty} e^{ipm_n} e^{-\frac{1}{2}\sigma_n^2 p^2}.$$

The limit on the right hand side only exists for all p, if either $\sigma_n \to \infty$, or the sequences σ_n and m_n both converge to finite limits $\sigma \in [0, \infty)$ and $m \in \mathbb{R}$. In the first case, the limit would equal 0 for $p \neq 0$ and 1 for $p = 0$. This is a contradiction, since characteristic functions are always continuous. Hence the second case occurs, and, therefore

$$E[e^{ipX}] = e^{ipm - \frac{1}{2}\sigma^2 p^2} \qquad \text{for any } p \in \mathbb{R},$$

i.e., $X \sim N(m, \sigma^2)$. □

So far, we have constructed Brownian motion only for $t \in [0, 1]$. Brownian motion on any finite time interval can easily be obtained from this process by rescaling. Brownian motion defined for all $t \in \mathbb{R}_+$ can be obtained by joining infinitely many Brownian motions on time intervals of length 1:



Theorem 1.17. *Suppose that $B_t^{(1)}, B_t^{(2)}, \ldots$ are independent Brownian motions starting at* 0 *defined for* $t \in [0,1]$. *Then the process*

$$B_t \quad := \quad B_{t-\lfloor t \rfloor}^{(\lfloor t \rfloor + 1)} + \sum_{i=1}^{\lfloor t \rfloor} B_1^{(i)}, \qquad t \geq 0,$$

is a Brownian motion defined for $t \in [0, \infty)$.

The proof is left as an exercise.

1.4 The Brownian Sample Paths

In this section we study some properties of Brownian sample paths in dimension one. We show that a typical Brownian path is nowhere differentiable, and Hölder-continuous with parameter α if and only if $\alpha < 1/2$. Furthermore, the set $\Lambda_a = \{t \geq 0 : B_t = a\}$ of all passage times of a given point $a \in \mathbb{R}$ is a fractal. We will show that almost surely, Λ_a has Lebesgue measure zero but any point in Λ_a is an accumulation point of Λ_a.
We consider a one-dimensional Brownian motion $(B_t)_{t \geq 0}$ with $B_0 = 0$ defined on a probability space $(\Omega, \mathcal{A}, P)$. Then:

Typical Brownian sample paths are nowhere differentiable

For any $t \geq 0$ and $h > 0$, the difference quotient $\frac{B_{t+h} - B_t}{h}$ is normally distributed with mean 0 and standard deviation

$$\sigma[(B_{t+h} - B_t)/h] \quad = \quad \sigma[B_{t+h} - B_t]/h \quad = \quad 1/\sqrt{h}.$$

This suggests that the derivative

$$\frac{d}{dt} B_t \quad = \quad \lim_{h \searrow 0} \frac{B_{t+h} - B_t}{h}$$

does not exist. Indeed, we have the following stronger statement.

Theorem 1.18 (Paley, Wiener, Zygmund 1933). *Almost surely, the Brownian sample path $t \mapsto B_t$ is nowhere differentiable, and*

$$\limsup_{s \searrow t} \left| \frac{B_s - B_t}{s - t} \right| \quad = \quad \infty \qquad \textit{for any } t \geq 0.$$

Note that, since there are uncountably many $t \geq 0$, the statement is stronger than claiming only the almost sure non-differentiability for any given $t \geq 0$.

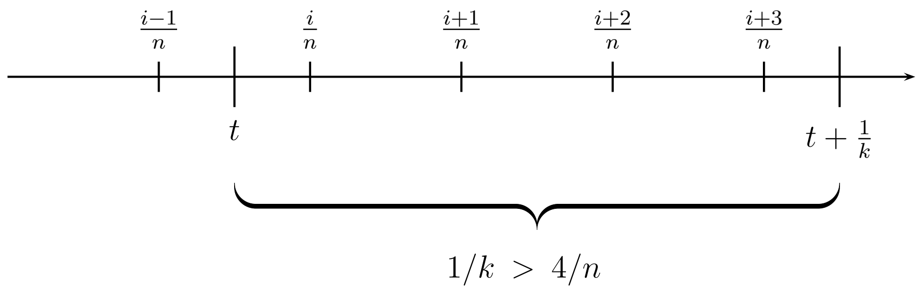
Proof. It suffices to show that the set

$$N = \left\{ \omega \in \Omega \;\middle|\; \exists\, t \in [0,T], k, L \in \mathbb{N}\ \forall\, s \in (t, t+\frac{1}{k}) \;:\; |B_s(\omega) - B_t(\omega)| \leq L|s-t| \right\}$$

is a null set for any $T \in \mathbb{N}$. Hence fix $T \in \mathbb{N}$, and consider $\omega \in N$. Then there exist $k, L \in \mathbb{N}$ and $t \in [0,T]$ such that

$$|B_s(\omega) - B_t(\omega)| \quad \leq \quad L \cdot |s-t| \qquad \text{holds for } s \in (t, t+\frac{1}{k}). \tag{1.4.1}$$

To make use of the independence of the increments over disjoint intervals, we note that for any $n > 4k$, we can find an $i \in \{1, 2, \ldots, nT\}$ such that the intervals $(\frac{i}{n}, \frac{i+1}{n})$, $(\frac{i+1}{n}, \frac{i+2}{n})$, and $(\frac{i+2}{n}, \frac{i+3}{n})$ are all contained in $(t, t+\frac{1}{k})$:



Hence by (1.4.1), the bound

$$\begin{aligned} \left| B_{\frac{j+1}{n}}(\omega) - B_{\frac{j}{n}}(\omega) \right| &\leq \left| B_{\frac{j+1}{n}}(\omega) - B_t(\omega) \right| + \left| B_t(\omega) - B_{\frac{j}{n}}(\omega) \right| \\ &\leq L \cdot (\frac{j+1}{n} - t) + L \cdot (\frac{j}{n} - t) \quad \leq \quad \frac{8L}{n} \end{aligned}$$

holds for $j = i, i+1, i+2$. Thus we have shown that N is contained in the set

$$\widetilde{N} \quad := \quad \bigcup_{k,L \in \mathbb{N}} \bigcap_{n > 4k} \bigcup_{i=1}^{nT} \left\{ \left| B_{\frac{j+1}{n}} - B_{\frac{j}{n}} \right| \leq \frac{8L}{n} \quad \text{for } j = i, i+1, i+2 \right\}.$$

We now prove $P[\widetilde{N}] = 0$. By independence and stationarity of the increments we have

$$\begin{aligned} & P\left[\left\{\left|B_{\frac{j+1}{n}} - B_{\frac{j}{n}}\right| \le \frac{8L}{n} \quad \text{for } j = i, i+1, i+2\right\}\right] \\ = \;& P\left[\left|B_{\frac{1}{n}}\right| \le \frac{8L}{n}\right]^3 \quad = \quad P\left[|B_1| \le \frac{8L}{\sqrt{n}}\right]^3 \qquad (1.4.2) \\ \le \;& \left(\frac{1}{\sqrt{2\pi}}\frac{16L}{\sqrt{n}}\right)^3 \quad = \quad \frac{16^3}{\sqrt{2\pi}^3} \cdot \frac{L^3}{n^{3/2}} \end{aligned}$$

for any i and n. Here we have used that the standard normal density is bounded from above by $1/\sqrt{2\pi}$. By (1.4.2) we obtain

$$\begin{aligned} & P\left[\bigcap_{n>4k} \bigcup_{i=1}^{nT} \left\{\left|B_{\frac{j+1}{n}} - B_{\frac{j}{n}}\right| \le \frac{8L}{n} \quad \text{for } j = i, i+1, i+2\right\}\right] \\ \le \;& \frac{16^3}{\sqrt{2\pi}^3} \cdot \inf_{n>4k} nTL^3/n^{3/2} \quad = \quad 0. \end{aligned}$$

Hence, $P[\widetilde{N}] = 0$, and therefore N is a null set. □

Hölder continuity

The statement of Theorem 1.18 says that a typical Brownian path is not Lipschitz continuous on any non-empty open interval. On the other hand, the Wiener-Lévy construction shows that the sample paths are continuous. We can almost close the gap between these two statements by arguing in both cases slightly more carefully:

Theorem 1.19. *The following statements hold almost surely:*

(1). For any $\alpha > 1/2$,

$$\limsup_{s \searrow t} \frac{|B_s - B_t|}{|s-t|^\alpha} \quad = \quad \infty \qquad \textit{for all } t \ge 0.$$

(2). For any $\alpha < 1/2$,

$$\sup_{\substack{s,t \in [0,T] \\ s \ne t}} \frac{|B_s - B_t|}{|s-t|^\alpha} \quad < \quad \infty \qquad \textit{for all } T > 0.$$

Hence a typical Brownian path is nowhere Hölder continuous with parameter $\alpha > 1/2$, but it is Hölder continuous with parameter $\alpha < 1/2$ on any finite interval. The critical case $\alpha = 1/2$ is more delicate, and will be briefly discussed below.

Proof of Theorem 1.19. The first statement can be shown by a similar argument as in the proof of Theorem 1.18. The details are left to the reader.

To prove the second statement for $T = 1$, we use the Wiener-Lévy representation

$$B_t \quad = \quad Z \cdot t + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} Z_{n,k} e_{n,k}(t) \qquad \text{for any } t \in [0,1]$$

with independent standard normal random variables $Z, Z_{n,k}$. For $t, s \in [0,1]$ we obtain

$$|B_t - B_s| \quad \leq \quad |Z| \cdot |t-s| + \sum_n M_n \sum_k |e_{n,k}(t) - e_{n,k}(s)|,$$

where $M_n := \max_k |Z_{n,k}|$ as in the proof of Theorem 1.14. We have shown above that by the Borel-Cantelli Lemma, $M_n \leq n$ eventually with probability one, and hence

$$M_n(\omega) \quad \leq \quad C(\omega) \cdot n$$

for some almost surely finite constant $C(\omega)$. Moreover, note that for each s, t and n, at most two summands in $\sum_k |e_{n,k}(t) - e_{n,k}(s)|$ do not vanish. Since $|e_{n,k}(t)| \leq \frac{1}{2} \cdot 2^{-n/2}$ and $|e'_{n,k}(t)| \leq 2^{n/2}$, we obtain the estimates

$$|e_{n,k}(t) - e_{n,k}(s)| \;\leq\; 2^{-n/2}, \qquad \text{and} \tag{1.4.3}$$

$$|e_{n,k}(t) - e_{n,k}(s)| \;\leq\; 2^{n/2} \cdot |t-s|. \tag{1.4.4}$$

For given $s, t \in [0,1]$, we now choose $N \in \mathbb{N}$ such that

$$2^{-N} \quad \leq \quad |t-s| \quad < \quad 2^{1-N}. \tag{1.4.5}$$

By applying (1.4.3) for $n > N$ and (1.4.4) for $n \leq N$, we obtain

$$|B_t - B_s| \quad \leq \quad |Z| \cdot |t-s| + 2C \cdot \left(\sum_{n=1}^{N} n 2^{n/2} \cdot |t-s| + \sum_{n=N+1}^{\infty} n 2^{-n/2} \right).$$

By (1.4.5) the sums on the right hand side can both be bounded by a constant multiple of $|t-s|^\alpha$ for any $\alpha < 1/2$. This proves that $(B_t)_{t \in [0,1]}$ is almost surely Hölder-continuous of order α. □

Law of the iterated logarithm

Khintchine's version of the law of the iterated logarithm is a much more precise statement on the local regularity of a typical Brownian path at a fixed time $s \geq 0$. It implies in particular that almost every Brownian path is not Hölder continuous with parameter $\alpha = 1/2$. We state the result without proof:

Theorem 1.20 (Khintchine 1924). *For $s \geq 0$, the following statements hold almost surely:*

$$\limsup_{t \searrow 0} \frac{B_{s+t} - B_s}{\sqrt{2t \log\log(1/t)}} = +1, \quad \textit{and} \quad \liminf_{t \searrow 0} \frac{B_{s+t} - B_s}{\sqrt{2t \log\log(1/t)}} = -1.$$

For the proof cf. e.g. Breiman, Probability, Section 12.9.
By a time inversion, the Theorem translates into a statement on the global asymptotics of Brownian paths:

Corollary 1.21. *The following statements hold almost surely:*

$$\limsup_{t \to \infty} \frac{B_t}{\sqrt{2t \log\log t}} = +1, \quad \textit{and} \quad \liminf_{t \to \infty} \frac{B_t}{\sqrt{2t \log\log t}} = -1.$$

Proof. This follows by applying the Theorem above to the Brownian motion $\widehat{B}_t = t \cdot B_{1/t}$. For example, substituting $h = 1/t$, we have

$$\limsup_{t \to \infty} \frac{B_t}{\sqrt{2t \log\log(t)}} = \limsup_{h \searrow 0} \frac{h \cdot B_{1/h}}{\sqrt{2h \log\log 1/h}} = +1$$

almost surely. □

The corollary is a continuous time analogue of Kolmogorov's law of the iterated logarithm for Random Walks stating that for $S_n = \sum_{i=1}^{n} \eta_i$, η_i i.i.d. with $E[\eta_i] = 0$ and $\mathrm{Var}[\eta_i] = 1$, one has

$$\limsup_{n \to \infty} \frac{S_n}{\sqrt{2n \log\log n}} = +1 \quad \text{and} \quad \liminf_{n \to \infty} \frac{S_n}{\sqrt{2n \log\log n}} = -1$$

almost surely. In fact, one way to prove Kolmogorov's LIL is to embed the Random Walk into a Brownian motion, cf. e.g. Rogers and Williams, Vol. I, Ch. 7 or Section 3.3

Passage times

We now study the set of passage times to a given level a for a one-dimensional Brownian motion $(B_t)_{t\geq 0}$. This set has interesting properties – in particular it is a random fractal. Fix $a \in \mathbb{R}$, and let

$$\Lambda_a(\omega) \quad = \quad \{t \geq 0 \, : \, B_t(\omega) = a\} \quad \subseteq \quad [0, \infty).$$

Assuming that every path is continuous, the random set $\Lambda_a(\omega)$ is *closed* for every ω. Moreover, scale invariance of Brownian motion implies a *statistical self similarity* property for the sets of passage times: Since the rescaled process $(c^{-1/2} B_{ct})_{t\geq 0}$ has the same distribution as $(B_t)_{t\geq 0}$ for any $c > 0$, we can conclude that the set valued random variable $c \cdot \Lambda_{a/\sqrt{c}}$ has the same distribution as Λ_a. In particular, Λ_0 is a *fractal* in the sense that

$$\Lambda_0 \quad \sim \quad c \cdot \Lambda_0 \qquad \text{for any } c > 0.$$

Moreover, by Fubini's Theorem one easily verifies that Λ_a has almost surely Lebesgue measure zero. In fact, continuity of $t \mapsto B_t(\omega)$ for any ω implies that $(t, \omega) \mapsto B_t(\omega)$ is product measurable (Exercise). Hence $\{(t, \omega) \, : \, B_t(\omega) = a\}$ is contained in the product σ-algebra, and

$$E[\lambda(\Lambda_a)] \quad = \quad E\left[\int_0^\infty I_{\{a\}}(B_t) \, dt\right] \quad = \quad \int_0^\infty P[B_t = a] \, dt \quad = \quad 0.$$

Theorem 1.22 (Unbounded oscillations, recurrence).

$$P\left[\sup_{t\geq 0} B_t \; = \; +\infty\right] \quad = \quad P\left[\inf_{t\geq 0} B_t \; = \; -\infty\right] \quad = \quad 1.$$

In particular, for any $a \in \mathbb{R}$, the random set Λ_a is almost surely unbounded, i.e. Brownian motion is recurrent.

Proof. By scale invariance,

$$\sup_{t\geq 0} B_t \quad \sim \quad c^{-1/2} \sup_{t\geq 0} B_{ct} \quad = \quad c^{-1/2} \sup_{t\geq 0} B_t \qquad \text{for any } c > 0.$$

Hence,

$$P\left[\sup_{t\geq 0} B_t \geq a\right] \; = \; P\left[\sup_{t\geq 0} B_t \geq a \cdot \sqrt{c}\right]$$

for any $c > 0$, and therefore $\sup B_t \in \{0, \infty\}$ almost surely. The first part of the assertion now follows since $\sup B_t$ is almost surely strictly positive. By reflection symmetry, we also obtain $\inf B_t = -\infty$ with probability one. □

The last Theorem makes a statement on the global structure of the set Λ_a. By invariance w.r.t. time inversion this again translates into a local regularity result:

Theorem 1.23 (Fine structure of Λ_a). *The set Λ_a is almost surely a **perfect set**, i.e., any $t \in \Lambda_a$ is an accumulation point of Λ_a.*

Proof. We prove the statement for $a = 0$, the general case being left as an exercise. We proceed in three steps:

STEP 1: *0 is almost surely an accumulation point of Λ_0:* This holds by time-reversal. Setting $\widehat{B}_t = t \cdot B_{1/t}$, we see that 0 is an accumulation point of Λ_0 if and only of for any $n \in \mathbb{N}$ there exists $t > n$ such that $\widehat{B}_t = 0$, i.e., if and only if the zero set of $\widehat{B}_t$ is unbounded. By Theorem 1.22, this holds almost surely.

STEP 2: *For any $s \geq 0$, $T_s := \min(\Lambda_a \cap [s, \infty)) = \min\{t \geq s : B_t = a\}$ is almost surely an accumulation point of Λ_a:* For the proof we need the strong Markov property of Brownian motion which will be proved in the next section. By Theorem 1.22, the random variable T_s is almost surely finite. Hence, by continuity, $B_{T_s} = a$ almost surely. The strong Markov property says that the process

$$\widetilde{B}_t \ := \ B_{T_s+t} - B_{T_s}, \qquad t \geq 0,$$

is again a Brownian motion starting at 0. Therefore, almost surely, 0 is an accumulation point of the zero set of $\widetilde{B}_t$ by Step 1. The claim follows since almost surely

$$\{t \geq 0 : \widetilde{B}_t = 0\} \ = \ \{t \geq 0 : B_{T_s+t} = B_{T_s}\} \ = \ \{t \geq T_s : B_t = a\} \ \subseteq \ \Lambda_a.$$

STEP 3: To complete the proof note that we have shown that the following properties hold with probability one:

(1). Λ_a is closed.

(2). $\min(\Lambda_a \cap [s,\infty))$ is an accumulation point of Λ_a for any $s \in \mathbb{Q}_+$.

Since $\mathbb{Q}_+$ is a dense subset of $\mathbb{R}_+$, (1) and (2) imply that any $t \in \Lambda_a$ is an accumulation point of Λ_a. In fact, for any $s \in [0,t] \cap \mathbb{Q}$, there exists an accumulation point of Λ_a in $(s,t]$ by (2), and hence t is itself an accumulation point.

□

Remark. It can be shown that the set Λ_a has Hausdorff dimension $1/2$.

1.5 Strong Markov property and reflection principle

In this section we prove a strong Markov property for Brownian motion. Before, we give another motivation for our interest in an extension of the Markov property to random times.

Maximum of Brownian motion

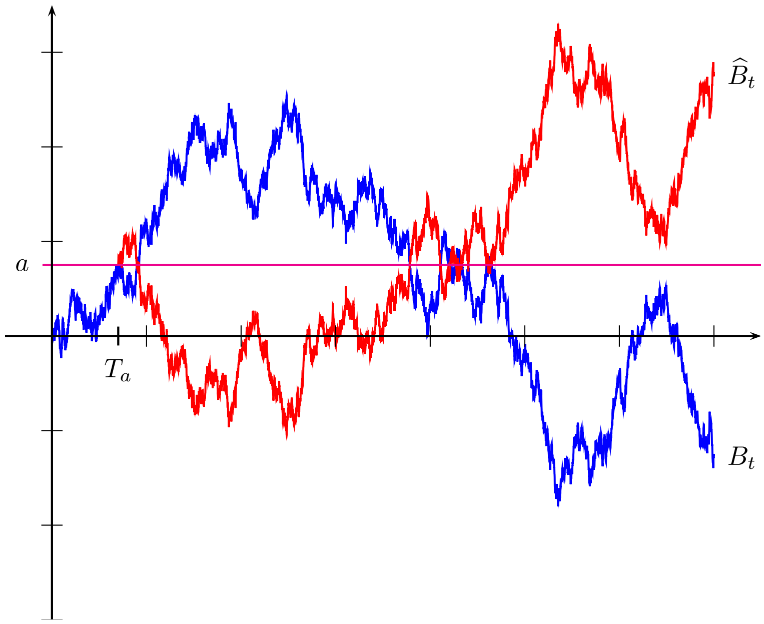
Suppose that $(B_t)_{t\geq 0}$ is a one-dimensional continuous Brownian motion starting at 0 defined on a probability space $(\Omega, \mathcal{A}, P)$. We would like to compute the distribution of the maximal value

$$M_s \quad = \quad \max_{t\in[0,s]} B_t$$

attained before a given time $s \in \mathbb{R}_+$. The idea is to proceed similarly as for Random Walks, and to reflect the Brownian path after the first passage time

$$T_a \quad = \quad \min\{t \geq 0 : B_t = a\}$$

to a given level $a > 0$:



It seems plausible (e.g. by the heuristic path integral representation of Wiener measure, or by a Random Walk approximation) that the reflected process $(\widehat{B}_t)_{t\geq 0}$ defined by

$$\widehat{B}_t \quad := \quad \begin{cases} B_t & \text{for } t \leq T_a \\ a - (B_t - a) & \text{for } t > T_a \end{cases}$$

is again a Brownian motion. At the end of this section, we will prove this reflection principle rigorously by the strong Markov property. Assuming the reflection principle is true, we can compute the distribution of M_s in the following way:

$$\begin{aligned} P[M_s \geq a] &= P[M_s \geq a, B_s \leq a] + P[M_s \geq a, B_s > a] \\ &= P[\widehat{B}_s \geq a] + P[B_s > a] \\ &= 2 \cdot P[B_s \geq a] \\ &= P[|B_s| \geq a]. \end{aligned}$$

Thus M_s has the same distribution as $|B_s|$.
Furthermore, since $M_s \geq a$ if and only if $\widehat{M_s} = \max\{\widehat{B}_t : t \in [0,s]\} \geq a$, we obtain the stronger statement

$$\begin{aligned} P[M_s \geq a, B_s \leq c] &= P[\widehat{M_s} \geq a, \widehat{B}_s \geq 2a - c] = P[\widehat{B}_s \geq 2a - c] \\ &= \frac{1}{\sqrt{2\pi s}} \int_{2a-c}^{\infty} \exp(-x^2/2s)\, dx \end{aligned}$$

for any $a \geq 0$ and $c \leq a$. As a consequence, we have:

Theorem 1.24 (Maxima of Brownian paths)**.**

(1). For any $s \geq 0$, the distribution of M_s is absolutely continuous with density

$$f_{M_s}(x) = \frac{2}{\sqrt{2\pi s}} \exp(-x^2/2s) \cdot I_{(0,\infty)}(x).$$

(2). The joint distribution of M_s and B_s is absolutely continuous with density

$$f_{M_s,B_s}(x,y) = 2\, \frac{2x - y}{\sqrt{2\pi s^3}} \exp\left(-\frac{(2x-y)^2}{2s}\right) I_{(0,\infty)}(x) I_{(-\infty,x)}(y).$$

Proof. (1) holds since $M_s \sim |B_s|$. For the proof of (2) we assume w.l.o.g. $s = 1$. The general case can be reduced to this case by the scale invariance of Brownian motion (Exercise). For $a \geq 0$ and $c \leq a$ let

$$G(a,c) := P[M_1 \geq a, B_1 \leq c].$$

By the reflection principle,

$$G(a,c) = P[B_1 \geq 2a - c] = 1 - \Phi(2a - c),$$

where Φ denotes the standard normal distribution function. Since $\lim_{a\to\infty} G(a,c) = 0$ and $\lim_{c\to-\infty} G(a,c) = 0$, we obtain

$$\begin{aligned} P[M_1 \geq a, B_1 \leq c] = G(a,c) &= -\int_{x=a}^{\infty} \int_{y=-\infty}^{c} \frac{\partial^2 G}{\partial x \partial y}(x,y)\, dydx \\ &= \int_{x=a}^{\infty} \int_{y=-\infty}^{c} 2 \cdot \frac{2x - y}{\sqrt{2\pi}} \cdot \exp\left(-\frac{(2x-y)^2}{2}\right) dydx. \end{aligned}$$

This implies the claim for $s = 1$, since $M_1 \geq 0$ and $B_1 \leq M_1$ by definition of M_1. □

The Theorem enables us to compute the distributions of the first passage times T_a. In fact, for $a > 0$ and $s \in [0, \infty)$ we obtain

$$\begin{aligned} P[T_a \leq s] \;=\; P[M_s \geq a] \;&=\; 2 \cdot P[B_s \geq a] \;=\; 2 \cdot P[B_1 \geq a/\sqrt{s}] \\ &=\; \sqrt{\frac{2}{\pi}} \int\limits_{a/\sqrt{s}}^{\infty} e^{-x^2/2}\, dx. \end{aligned} \tag{1.5.1}$$

Corollary 1.25 (Distribution of T_a). *For any $a \in \mathbb{R} \setminus \{0\}$, the distribution of T_a is absolutely continuous with density*

$$f_{T_a}(s) \;=\; \frac{|a|}{\sqrt{2\pi s^3}} \cdot e^{-a^2/2s}.$$

Proof. For $a > 0$, we obtain

$$f_{T_a}(s) \;=\; F'_{T_a}(s) \;=\; \frac{a}{\sqrt{2\pi s^3}} e^{-a^2/2s}$$

by (1.5.1). For $a < 0$ the assertion holds since $T_a \sim T_{-a}$ by reflection symmetry of Brownian motion. □

Next, we prove a strong Markov property for Brownian motion. Below we will then complete the proof of the reflection principle and the statements above by applying the strong Markov property to the passage time T_a.

Strong Markov property for Brownian motion

Suppose again that $(B_t)_{t\geq 0}$ is a d-dimensional continuous Brownian motion starting at 0 on a probability space $(\Omega, \mathcal{A}, P)$, and let

$$\mathcal{F}_t^B \;=\; \sigma(B_s \,:\, 0 \leq s \leq t), \qquad t \geq 0,$$

denote the σ-algebras generated by the process up to time t.

Definition. *A random variable $T : \Omega \to [0, \infty]$ is called an $(\mathcal{F}_t^B)$-**stopping time** if and only if*

$$\{T \leq t\} \;\in\; \mathcal{F}_t^B \qquad \textit{for any } t \geq 0.$$

Example. Clearly, for any $a \in \mathbb{R}$, the first passage time

$$T_a \;=\; \min\{t \geq 0 \,:\, B_t = a\}$$

to a level a is an $(\mathcal{F}_t^B)$-stopping time.
The σ-algebra $\mathcal{F}_T^B$ describing the information about the process up to a stopping time T is defined by

$$\mathcal{F}_T^B \;=\; \{A \in \mathcal{A} \,:\, A \cap \{T \leq t\} \in \mathcal{F}_t^B \quad \text{for any } t \geq 0\}.$$

Note that for $(\mathcal{F}_t^B)$ stopping times S and T with $S \leq T$ we have $\mathcal{F}_S^B \subseteq \mathcal{F}_T^B$, since for $t \geq 0$

$$A \cap \{S \leq t\} \in \mathcal{F}_t^B \qquad \Longrightarrow \qquad A \cap \{T \leq t\} \;=\; A \cap \{S \leq t\} \cap \{T \leq t\} \;\in\; \mathcal{F}_t^B.$$

For any constant $s \in \mathbb{R}_+$, the process $(B_{s+t} - B_s)_{t \geq 0}$ is a Brownian motion independent of $\mathcal{F}_s^B$.

A corresponding statement holds for stopping times:

Theorem 1.26 (Strong Markov property). *Suppose that T is an almost surely finite $(\mathcal{F}_t^B)$ stopping time. Then the process $(\widetilde{B}_t)_{t \geq 0}$ defined by*

$$\widetilde{B}_t \;=\; B_{T+t} - B_T \quad \text{if } T < \infty, \qquad 0 \quad \text{otherwise,}$$

is a Brownian motion independent of $\mathcal{F}_T^B$.

Proof. We first assume that T takes values only in $C \cup \{\infty\}$ where C is a countable subset of $[0, \infty)$. Then for $A \in \mathcal{F}_T^B$ and $s \in C$, we have $A \cap \{T = s\} \in \mathcal{F}_s^B$ and $\widetilde{B}_t = B_{t+s} - B_s$ on $A \cap \{T = s\}$. By the Markov property, $(B_{t+s} - B_s)_{t \geq 0}$ is a Brownian motion independent of $\mathcal{F}_s^B$. Hence for any measurable subset Γ of $C([0,\infty], \mathbb{R}^d)$, we have

$$\begin{aligned} P[\{(\widetilde{B}_t)_{t \geq 0} \in \Gamma\} \cap A] \;&=\; \sum_{s \in C} P[\{(B_{t+s} - B_s)_{t \geq 0} \in \Gamma\} \cap A \cap \{T = s\}] \\ &=\; \sum_{s \in C} \mu_0[\Gamma] \cdot P[A \cap \{T = s\}] \quad = \quad \mu_0[\Gamma] \cdot P[A] \end{aligned}$$

where μ_0 denotes the distribution of Brownian motion starting at 0. This proves the assertion for discrete stopping times.

For an arbitrary $(\mathcal{F}_t^B)$ stopping time T that is almost surely finite and $n \in \mathbb{N}$, we set $T_n = \frac{1}{n}\lceil nT \rceil$, i.e.,

$$T_n \;=\; \frac{k}{n} \quad \text{on} \quad \left\{ \frac{k-1}{n} < T \leq \frac{k}{n} \right\} \quad \text{for any } k \in \mathbb{N}.$$

Since the event $\{T_n = k/n\}$ is $\mathcal{F}_{k/n}^B$-measurable for any $k \in \mathbb{N}$, T_n is a discrete $(\mathcal{F}_t^B)$ stopping time. Therefore, $(B_{T_n+t} - B_{T_n})_{t\geq 0}$ is a Brownian motion that is independent of $\mathcal{F}_{T_n}^B$, and hence of the smaller σ-algebra $\mathcal{F}_T^B$. As $n \to \infty$, $T_n \to T$, and thus, by continuity,

$$\widetilde{B}_t \;=\; B_{T+t} - B_T \;=\; \lim_{n\to\infty} (B_{T_n+t} - B_{T_n}).$$

Now it is easy to verify that $(\widetilde{B}_t)_{t\geq 0}$ is again a Brownian motion that is independent of $\mathcal{F}_T^B$. □

A rigorous reflection principle

We now apply the strong Markov property to prove a reflection principle for Brownian motion. Consider a one-dimensional continuous Brownian motion $(B_t)_{t\geq 0}$ starting at 0. For $a \in \mathbb{R}$ let

$$\begin{aligned} T_a &= \min\{t \geq 0 \,:\, B_t = a\} && \text{(first passage time)}, \\ B_t^{T_a} &= B_{\min\{t, T_a\}} && \text{(process stopped at } T_a\text{)}, \quad \text{and} \\ \widetilde{B}_t &= B_{T_a+t} - B_{T_a} && \text{(process after } T_a\text{)}. \end{aligned}$$

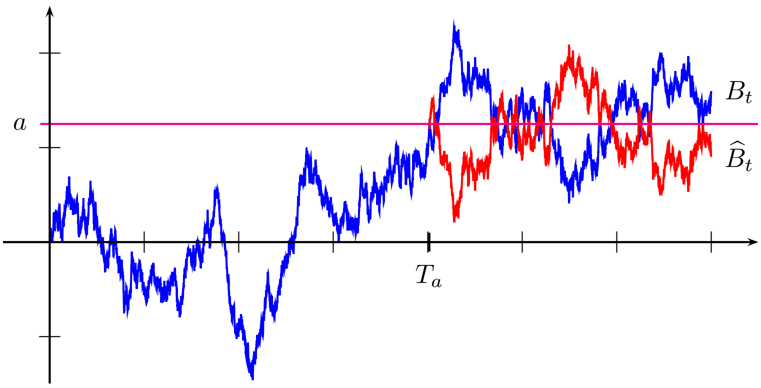
Theorem 1.27 (Reflection principle). *The joint distributions of the following random variables with values in* $\mathbb{R}_+ \times C([0,\infty)) \times C([0,\infty))$ *agree:*

$$(T_a, (B_t^{T_a})_{t\geq 0}, (\widetilde{B}_t)_{t\geq 0}) \quad \sim \quad (T_a, (B_t^{T_a})_{t\geq 0}, (-\widetilde{B}_t)_{t\geq 0})$$

Proof. By the strong Markov property, the process $\widetilde{B}$ is a Brownian motion starting at 0 independent of $\mathcal{F}_{T_a}$, and hence of T_a and $B^{T_a} = (B_t^{T_a})_{t\geq 0}$. Therefore,

$$P \circ (T_a, B^{T_a}, \widetilde{B})^{-1} \;=\; P \circ (T_a, B^{T_a})^{-1} \otimes \mu_0 \;=\; P \circ (T_a, B^{T_a}, -\widetilde{B})^{-1}.$$

□



As a consequence of the theorem, we can complete the argument given at the beginning of this section: The "shadow path" $\widehat{B}_t$ of a Brownian path B_t with reflection when reaching the level a is given by

$$\widehat{B}_t = \begin{cases} B_t^{T_a} & \text{for } t \leq T_a \\ a - \widetilde{B}_{t-T_a} & \text{for } t > T_a \end{cases},$$

whereas

$$B_t = \begin{cases} B_t^{T_a} & \text{for } t \leq T_a \\ a + \widetilde{B}_{t-T_a} & \text{for } t > T_a \end{cases}.$$

By the Theorem 1.27, $(\widehat{B}_t)_{t\geq 0}$ has the same distribution as $(B_t)_{t\geq 0}$. Therefore, and since $\max\limits_{t\in[0,s]} B_t \geq a$ if and only if $\max\limits_{t\in[0,s]} \widehat{B}_t \geq a$, we obtain for $a \geq c$:

$$\begin{aligned} P\left[\max_{t\in[0,s]} B_t \geq a, B_s \leq c\right] &= P\left[\max_{t\in[0,s]} \widehat{B}_t \geq a, \widehat{B}_s \geq 2a - c\right] \\ &= P\left[\widehat{B}_s \geq 2a - c\right] \\ &= \frac{1}{\sqrt{2\pi s}} \int_{2a-c}^{\infty} e^{-x^2/2s}\, dx. \end{aligned}$$

Chapter 2

Martingales

Classical analysis starts with studying convergence of sequences of real numbers. Similarly, stochastic analysis relies on basic statements about sequences of real-valued random variables. Any such sequence can be decomposed uniquely into a martingale, i.e., a real.valued stochastic process that is "constant on average", and a predictable part. Therefore, estimates and convergence theorems for martingales are crucial in stochastic analysis.

2.1 Definitions and Examples

We fix a probability space $(\Omega, \mathcal{A}, P)$. Moreover, we assume that we are given an increasing sequence $\mathcal{F}_n$ $(n = 0, 1, 2, \ldots)$ of sub-σ-algebras of $\mathcal{A}$. Intuitively, we often think of $\mathcal{F}_n$ as describing the information available to us at time n. Formally, we define:

Definition. *(1). A* ***filtration*** *on* $(\Omega, \mathcal{A})$ *is an increasing sequence*

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$$

of σ*-algebras* $\mathcal{F}_n \subseteq \mathcal{A}$.

(2). A stochastic process $(X_n)_{n \geq 0}$ *is* ***adapted*** *to a filtration* $(\mathcal{F}_n)_{n \geq 0}$ *iff each* X_n *is* $\mathcal{F}_n$*-measurable.*

Example. (1). The **canonical filtration** $(\mathcal{F}_n^X)$ generated by a stochastic process (X_n) is given by

$$\mathcal{F}_n^X \;=\; \sigma(X_0, X_1, \ldots, X_n).$$

If the filtration is not specified explicitly, we will usually consider the canonical filtration.

(2). Alternatively, filtrations containing additional information are of interest, for example the filtration

$$\mathcal{F}_n \;=\; \sigma(Z, X_0, X_1, \ldots, X_n)$$

generated by the process (X_n) and an additional random variable Z, or the filtration

$$\mathcal{F}_n \;=\; \sigma(X_0, Y_0, X_1, Y_1, \ldots, X_n, Y_n)$$

generated by the process (X_n) and a further process (Y_n). Clearly, the process (X_n) is adapted to any of these filtrations. In general, (X_n) is adapted to a filtration $(\mathcal{F}_n)$ if and only if $\mathcal{F}_n^X \subseteq \mathcal{F}_n$ for any $n \geq 0$.

Martingales and Supermartingales

We can now formalize the notion of a real-valued stochastic process that is constant (respectively decreasing, increasing) on average:

Definition. *(1). A sequence of real-valued random variables $M_n : \Omega \to \mathbb{R}$ $(n = 0, 1, \ldots)$ on the probability space $(\Omega, \mathcal{A}, P)$ is called a **martingale w.r.t. the filtration** $(\mathcal{F}_n)$ if and only if*

(a) (M_n) is adapted w.r.t. $(\mathcal{F}_n)$.

(b) M_n is integrable for any $n \geq 0$.

(c) $E[M_n \mid \mathcal{F}_{n-1}] \;=\; M_{n-1}$ for any $n \in \mathbb{N}$.

*(2). Similarly, (M_n) is called a **supermartingale** (resp. a **submartingale**) w.r.t. $(\mathcal{F}_n)$, if and only if (a) holds, the positive part M_n^+ (resp. the negative part M_n^-) is integrable for any $n \geq 0$, and (c) holds with "$=$" replaced by "$\leq$", "$\geq$" respectively.*

Condition (c) in the martingale definition can equivalently be written as

(c') $E[M_{n+1} - M_n \mid \mathcal{F}_n] \;=\; 0 \qquad$ for any $n \in \mathbb{N}$,

and correspondingly with "$=$" replaced by "$\leq$" or "$\geq$" for super- or submartingales.

Intuitively, a martingale is a fair game, i.e., M_{n-1} is the best prediction (w.r.t. the mean square error) for the next value M_n given the information up to time $n-1$. A supermartingale is "decreasing on average", a submartingale is "increasing on average", and a martingale is both "decreasing" and "increasing", i.e., "constant on average." In particular, by induction on n, a martingale satisfies

$$E[M_n] = E[M_0] \qquad \text{for any } n \geq 0.$$

Similarly, for a supermartingale, the expectation values $E[M_n]$ are decreasing. More generally, we have:

Lemma 2.1. *If* (M_n) *is a martingale (respectively a supermartingale) w.r.t. a filtration* $(\mathcal{F}_n)$ *then*

$$E[M_{n+k} \mid \mathcal{F}_n] \overset{(\leq)}{=} M_n \qquad P\text{-a.s. for any } n, k \geq 0.$$

Proof. By induction on k: The assertion holds for $k = 0$, since M_n is $\mathcal{F}_n$-measurable. Moreover, the assertion for $k - 1$ implies

$$\begin{aligned} E[M_{n+k} \mid \mathcal{F}_n] &= E\big[E[M_{n+k} \mid \mathcal{F}_{n+k-1}] \,\big|\, \mathcal{F}_n\big] \\ &= E[M_{n+k-1} \mid \mathcal{F}_n] \;=\; M_n \qquad P\text{-a.s.} \end{aligned}$$

by the tower property for conditional expectations. □

Remark (**Supermartingale Convergence Theorem**). A key fact in analysis is that any lower bounded decreasing sequence of real numbers converges to its infimum. The counterpart of this result in stochastic analysis is the Supermartingale Convergence Theorem: Any lower bound supermartingale converges almost surely, c.f. Theorem **??** below.

Some fundamental examples

a) Sums of independent random variables

A Random Walk

$$S_n \;=\; \sum_{i=1}^{n} \eta_i, \qquad n = 0, 1, 2, \ldots,$$

with independent increments $\eta_i \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ is a martingale w.r.t. to the filtration

$$\mathcal{F}_n \;=\; \sigma(\eta_1, \ldots, \eta_n) \;=\; \sigma(S_0, S_1, \ldots, S_n)$$

if and only if the increments η_i are centered random variables. In fact, for any $n \in \mathbb{N}$,

$$E[S_n - S_{n-1} \mid \mathcal{F}_{n-1}] \;=\; E[\eta_n \mid \mathcal{F}_{n-1}] \;=\; E[\eta_n]$$

by independence of the increments.
Correspondingly, (S_n) is an $(\mathcal{F}_n)$ supermartingale if and only if $E[\eta_i] \leq 0$ for any $i \in \mathbb{N}$.

b) Products of independent non-negative random variables

A stochastic process

$$M_n \;=\; \prod_{i=1}^{n} Y_i, \qquad n = 0, 1, 2, \ldots,$$

with independent non-negative factors $Y_i \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ is a martingale respectively a supermartingale w.r.t. the filtration

$$\mathcal{F}_n \;=\; \sigma(Y_1, \ldots, Y_n)$$

if and only if $E[Y_i] = 1$ for any $i \in \mathbb{N}$, or $E[Y_i] \leq 1$ for any $i \in \mathbb{N}$ respectively. In fact, as Z_n is $\mathcal{F}_n$-measurable and Y_{n+1} is independent of $\mathcal{F}_n$, we have

$$E[M_{n+1} \mid \mathcal{F}_n] \;=\; E[M_n \cdot Y_{n+1} \mid \mathcal{F}_n] \;=\; M_n \cdot E[Y_{n+1}] \qquad \text{for any } n \geq 0.$$

Martingales and supermartingales of this type occur naturally in stochastic growth models. For the supermartingale property, integrability of the factors is not required.

Example (Exponential martingales). Consider a Random Walk $S_n = \sum_{i=1}^n \eta_i$ with i.i.d. increments η_i, and let

$$Z(\lambda) \;=\; E[\exp(\lambda\eta_i)], \qquad \lambda \in \mathbb{R},$$

denote the moment generating function of the increments. Then for any $\lambda \in \mathbb{R}$ with $Z(\lambda) < \infty$, the process

$$M_n^\lambda \;:=\; e^{\lambda S_n}/Z(\lambda)^n \;=\; \prod_{i=1}^n \big(e^{\lambda\eta_i}/Z(\lambda)\big)$$

is a martingale. This martingale can be used to prove exponential bounds for Random Walks, cf. e.g. Chernov's theorem [“Einführung in die Wahrscheinlichkeitstheorie”, Theorem 8.3] or the applications of the maximal inequality in Section 2.4 below.

Example (CRR model of stock market). In the Cox-Ross-Rubinstein binomial model of mathematical finance, the price of an asset is changing during each period either by a factor $1 + a$ or by a factor $1 + b$ with $a, b \in (-1, \infty)$ such that $a < b$. We can model the price evolution in N periods by a stochastic process

$$S_n = S_0 \cdot \prod_{i=1}^n X_i, \qquad n = 0, 1, 2, \dots, N,$$

defined on $\Omega = \{1 + a, 1 + b\}^N$, where the initial price S_0 is a given constant, and $X_i(\omega) = \omega_i$. Taking into account a constant interest rate $r > 0$, the discounted stock price after n periods is

$$\widetilde{S}_n \;=\; S_n/(1+r)^n \;=\; S_0 \cdot \prod_{i=1}^n \frac{X_i}{1+r}.$$

A probability measure P on Ω is called a **martingale measure** if the discounted stock price is a martingale w.r.t. P and the filtration $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Martingale measures are important for option pricing under no arbitrage assumptions, cf. Section 2.3 below. For $1 \le n \le N$,

$$E[\widetilde{S}_n \mid \mathcal{F}_{n-1}] \;=\; E\left[\widetilde{S}_{n-1} \cdot \frac{X_n}{1+r} \;\middle|\; \mathcal{F}_{n-1}\right] \;=\; \widetilde{S}_{n-1} \cdot \frac{E[X_n \mid \mathcal{F}_{n-1}]}{1+r}.$$

Hence $(\widetilde{S}_n)$ is an $(\mathcal{F}_n)$ martingale w.r.t. P if and only if

$$E[X_n \mid \mathcal{F}_{n-1}] \quad = \quad 1+r \qquad \text{for any } 1 \le n \le N. \tag{2.1.1}$$

On the other hand, since in the CRR model X_n only takes the values $1+a$ and $1+b$, we have

$$\begin{aligned} E[X_n \mid \mathcal{F}_{n-1}] &= (1+a) \cdot P[X_n = 1+a \mid \mathcal{F}_{n-1}] + (1+b) \cdot P[X_n = 1+b \mid \mathcal{F}_{n-1}] \\ &= 1 + a + (b-a) \cdot P[X_n = 1+b \mid \mathcal{F}_{n-1}]. \end{aligned}$$

Therefore, by (2.1.1), $(\widetilde{S}_n)$ is a martingale if and only if

$$P[X_n = 1+b \mid \mathcal{F}_{n-1}] \;=\; \frac{r-a}{b-a} \qquad \text{for any } n = 1, \ldots, N,$$

i.e., if and only if the growth factors $X_1, \ldots, X_N$ are independent with

$$P[X_n = 1+b] \;=\; \frac{r-a}{b-a} \qquad \text{and} \qquad P[X_n = 1+a] \;=\; \frac{b-r}{b-a}. \tag{2.1.2}$$

Hence for $r \notin [a,b]$, a martingale measure does not exist, and for $r \in [a,b]$, the product measure P on Ω satisfying $(2.1.2)$ is the unique martingale measure. Intuitively this is plausible: If $r < a$ or $r > b$ respectively, then the stock price is always growing more or less than the discount factor $(1+r)^n$, so the discounted stock price can not be a martingale. If, on the other hand, $a < r < b$, then $(\widetilde{S}_n)$ is a martingale provided the growth factors are independent with

$$\frac{P[X_n = 1+b]}{P[X_n = 1+a]} \quad = \quad \frac{(1+r)-(1+a)}{(1+b)-(1+r)}.$$

We remark, however, that uniqueness of the martingale measure only follows from (2.1.1) since we have assumed that each X_n takes only two possible values (binomial model). In a corresponding trinomial model there are infinitely many martingale measures!

c) Successive prediction values

Let F be an integrable random variable and $(\mathcal{F}_n)$ a filtration on a probability space $(\Omega, \mathcal{A}, P)$. Then the process

$$M_n \;:=\; E[F \mid \mathcal{F}_n], \qquad n = 0, 1, 2, \ldots,$$

of successive prediction values for F based on the information up to time n is a martingale. Indeed, by the tower property for conditional expectations, we have

$$E[M_n \mid \mathcal{F}_{n-1}] \;=\; E\big[E[F \mid \mathcal{F}_n] \,\big|\, \mathcal{F}_{n-1}\big] \;=\; E\big[F \,\big|\, \mathcal{F}_{n-1}\big] \;=\; M_{n-1}$$

almost surely for any $n \in \mathbb{N}$.

Remark (**Representing martingales as successive prediction values**). The class of martingales that have a representation as successive prediction values almost contains general martingales. In fact, for an arbitrary $(\mathcal{F}_n)$ martingale (M_n) and any finite integer $m \geq 0$, the representation

$$M_n \;=\; E[M_m \mid \mathcal{F}_n]$$

holds for any $n = 0, 1, \ldots, m$. Moreover, the L^1 Martingale Convergence Theorem implies that under appropriate uniform integrability assumptions, the limit $M_\infty = \lim_{n\to\infty} M_n$ exists in $\mathcal{L}^1$, and the representation

$$M_n \;=\; E[M_\infty \mid \mathcal{F}_n]$$

holds for any $n \geq 0$, cf. Section **??** below.

d) Functions of martingales

By Jensen's inequality for conditional expectations, convex functions of martingales are submartingales, and concave functions of martingales are supermartingales:

Theorem 2.2. *Suppose that $(M_n)_{n\geq 0}$ is an $(\mathcal{F}_n)$ martingale, and $u : \mathbb{R} \to \mathbb{R}$ is a convex function that is bounded from below. Then $(u(M_n))$ is an $(\mathcal{F}_n)$ submartingale.*

Proof. Since u is lower bounded, $u(M_n)^-$ is integrable for any n. Jensen's inequality for conditional expectations now implies

$$E[u(M_{n+1}) \mid \mathcal{F}_n] \;\geq\; u\big(E[M_{n+1} \mid \mathcal{F}_n]\big) \;=\; u(M_n)$$

almost surely for any $n \geq 0$. □

Example. If (M_n) is a martingale then $(|M_n|^p)$ is a submartingale for any $p \geq 1$.

e) Functions of Markov chains

Let $p(x, dy)$ be a stochastic kernel on a measurable space $(S, \mathcal{S})$.

Definition. *(1). A stochastic process $(X_n)_{n\geq 0}$ with state space $(S, \mathcal{S})$ defined on the probability space $(\Omega, \mathcal{A}, P)$ is called a* ***(time-homogeneous) Markov chain with transition kernel*** p ***w.r.t. the filtration*** $(\mathcal{F}_\mathbf{n})$*, if and only if*

(a) (X_n) *is* $(\mathcal{F}_n)$ *adapted, and*

(b) $P[X_{n+1} \in B \mid \mathcal{F}_n] \;=\; p(X_n, B)$ *holds P-almost surely for any $B \in \mathcal{S}$ and $n \geq 0$.*

(2). A measurable function $h : S \to \mathbb{R}$ is called ***superharmonic*** *(resp.* ***subharmonic****) w.r.t. p if and only if the integrals*

$$(ph)(x) \;:=\; \int p(x, dy)h(y), \qquad x \in S,$$

exist, and

$$(ph)(x) \leq h(x) \qquad \textit{(respectively } (ph)(x) \geq h(x)\textit{)}$$

holds for any $x \in S$.

The function h is called ***harmonic*** *iff it is both super- and subharmonic, i.e., iff*

$$(ph)(x) \;=\; h(x) \qquad \textit{for any } x \in S.$$

By the tower property for conditional expectations, any $(\mathcal{F}_n)$ Markov chain is also a Markov chain w.r.t. the canonical filtration generated by the process.

Example (Classical Random Walk on $\mathbb{Z}^d$). The standard Random Walk $(X_n)_{n\geq 0}$ on $\mathbb{Z}^d$ is a Markov chain w.r.t. the filtration $\mathcal{F}_n^X = \sigma(X_0, \ldots, X_n)$ with transition probabilities $p(x, x+e) = 1/2d$ for any unit vector $e \in \mathbb{Z}^d$.

The coordinate processes $(X_n^i)_{n\geq 0},\ i = 1, \ldots, d$, are Markov chains w.r.t. the same filtration with transition probabilities

$$\overline{p}(x, x+1) \;=\; \overline{p}(x, x-1) \;=\; \frac{1}{2d}, \quad \overline{p}(x, x) \;=\; \frac{2d-2}{2d}.$$

A function $h : \mathbb{Z}^d \to \mathbb{R}$ is harmonic w.r.t. p if and only if

$$(ph)(x) - h(x) \;=\; \frac{1}{2d}\sum_{i=1}^{d}\big(h(x+e_i) - 2h(x) + h(x-e_i)\big) \;=\; 0$$

for any $x \in \mathbb{Z}^d$, i.e., if h solves the discrete Laplace equation

$$\Delta_{\mathbb{Z}^d} h(x) \;=\; \sum_{i=1}^{d}((h(x+e_i) - h(x)) - (h(x) - h(x-e_i))) \;=\; 0.$$

Similarly, h is superharmonic if and only if $\Delta_{\mathbb{Z}^d} \le 0$.
A function $h : \mathbb{Z} \to \mathbb{R}$ is harmonic w.r.t. $\overline{p}$ if and only if $h(x) = ax + b$ with $a, b \in \mathbb{R}$, and h is superharmonic if and only if it is concave.

It is easy to verify that (super-)harmonic functions of Markov chains are (super-)martingales:

Theorem 2.3. *Suppose that (X_n) is a $(\mathcal{F}_n)$ Markov chain. Then the real-valued process*

$$M_n \;:=\; h(X_n), \quad n = 0, 1, 2, \ldots,$$

is a martingale (resp. a supermartingale) w.r.t. $(\mathcal{F}_n)$ for every harmonic (resp. superharmonic) function $h : S \to \mathbb{R}$ such that $h(X_n)$ (resp. $h(X_n^-)$) is integrable for each n.

Proof. Clearly, (M_n) is again $(\mathcal{F}_n)$ adapted. Moreover,

$$E[M_{n+1} \mid \mathcal{F}_n] \;=\; E[h(X_{n+1}) \mid \mathcal{F}_n] \;=\; (ph)(X_n) \qquad P\text{-a.s.}$$

The assertion now follows immediately from the definitions. □

Below, we will show how to construct more general martingales from Markov chains, cf. Theorem 2.5. At first, however, we consider a simple example that demonstrates the usefulness of martingale methods in analyzing Markov chains:

Example (Multinomial resampling). Consider a population of N individuals (replicas) with a finite number of possible types, where the types $y_n^{(1)}, \ldots, y_n^{(N)}$ of the individuals in the n-th generation are determined recursively by the following algorithm:

for i:=1,N **do**

1. generate $u \sim \mathrm{Unif}\{1, \dots, N\}$
2. $y_n^{(i)} := y_{n-1}^{(u)}$

end for

Hence each individual selects its type randomly and independently according to the relative frequencies of the types in the previous generation. The model is relevant both for stochastic algorithms (cf. e.g. [Cappé, Moulines, Ryden]) and as a basic model in evolutionary biology, cf. [Ethier, Kurtz].

The number X_n of individuals of a given type in generation n is a Markov chain with state space $S = \{0, 1, \dots, N\}$ and transition kernel

$$p(k, \bullet) \;=\; \mathrm{Bin}(N, k/N).$$

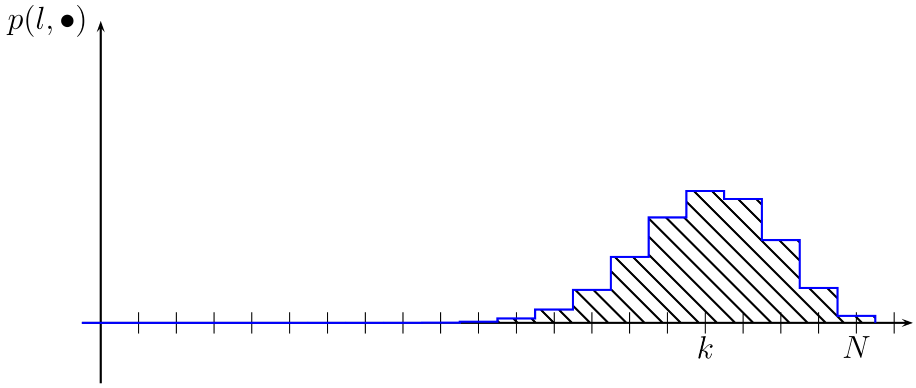


Figure 2.1: Transition function of (X_n).

Moreover, as the average of this binomial distribution is k, the function $h(x) = x$ is harmonic, and the expected number of individuals in generation $n+1$ given $X_0, \dots, X_n$ is

$$E[X_{n+1} \mid X_0, \dots, X_n] \;=\; X_n.$$

Hence, the process (X_n) is a bounded martingale. The Martingale Convergence Theorem now implies that the limit $X_\infty = \lim X_n$ exists almost surely, cf. Section **??** below. Since X_n takes discrete values, we can conclude that $X_n = X_\infty$ eventually with probability one. In particular, X_∞ is almost surely an absorbing state. Hence

$$P\big[X_n = 0 \text{ or } X_n = N \text{ eventually}\big] \;=\; 1. \tag{2.1.3}$$

In order to compute the probabilities of the events "$X_n = 0$ eventually" and "$X_n = N$ eventually" we can apply the Optional Stopping Theorem for martingales, cf. Section 2.3 below. Let

$$T := \min\{n \geq 0 : X_n = 0 \text{ or } X_n = N\}, \quad \min\emptyset := \infty,$$

denote the first hitting time of the absorbing states. If the initial number X_0 of individuals of the given type is k, then by the Optional Stopping Theorem,

$$E[X_T] = E[X_0] = k.$$

Hence by (2.1.3) we obtain

$$\begin{aligned} P\big[X_n = N \text{ eventually}\big] &= P[X_T = N] = \frac{1}{N}E[X_T] = \frac{k}{N}, \qquad \text{and} \\ P\big[X_n = N \text{ eventually}\big] &= 1 - \frac{k}{N} = \frac{N-k}{N}. \end{aligned}$$

Hence eventually all individuals have the same type, and a given type occurs eventually with probability determined by its initial relative frequency in the population.

2.2 Doob Decomposition and Martingale Problem

We will show now that any adapted sequence of real-valued random variables can be decomposed into a martingale and a predictable process. In particular, the variance process of a martingale (M_n) is the predictable part in the corresponding Doob decomposition of the process (M_n^2). The Doob decomposition for functions of Markov chains implies the Martingale Problem characterization of Markov chains.

Doob Decomposition

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $(\mathcal{F}_n)_{n\geq 0}$ a filtration on $(\Omega, \mathcal{A})$.

Definition. *A stochastic process $(A_n)_{n\geq 0}$ is called* ***predictable w.r.t.*** $(\mathcal{F}_\mathbf{n})$ *if and only if A_0 is constant and A_n is measurable w.r.t. $\mathcal{F}_{n-1}$ for any $n \in \mathbb{N}$.*

Intuitively, the value $A_n(\omega)$ of a predictable process can be predicted by the information available at time $n-1$.

Theorem 2.4. *Every $(\mathcal{F}_n)$ adapted sequence of integrable random variables Y_n $(n \geq 0)$ has a unique decomposition (up to modification on null sets)*

$$Y_n \quad = \quad M_n + A_n \tag{2.2.1}$$

into a $(\mathcal{F}_n)$ martingale (M_n) and a predictable process (A_n) such that $A_0 = 0$. Explicitly, the decomposition is given by

$$A_n \quad = \quad \sum_{k=1}^{n} E[Y_k - Y_{k-1} \,|\, \mathcal{F}_{k-1}], \qquad \textit{and} \quad M_n \; = \; Y_n - A_n. \tag{2.2.2}$$

Remark. (1). The increments $E[Y_k - Y_{k-1} | \mathcal{F}_{k-1}]$ of the process (A_n) are the predicted increments of (Y_n) given the previous information.

(2). The process (Y_n) is a supermartingale (resp. a submartingale) if and only if the predictable part (A_n) is decreasing (resp. increasing).

Proof of Theorem 2.4. Uniqueness: For any decomposition as in (2.2.1) we have

$$Y_k - Y_{k-1} \quad = \quad M_k - M_{k-1} + A_k - A_{k-1} \qquad \text{for any } k \in \mathbb{N}.$$

If (M_n) is a martingale and (A_n) is predictable then

$$E[Y_k - Y_{k-1} \,|\, \mathcal{F}_{k-1}] \quad = \quad E[A_k - A_{k-1} \,|\, \mathcal{F}_{k-1}] \quad = \quad A_k - A_{k-1} \qquad P\text{-a.s.}$$

This implies that (2.2.2) holds almost surely if $A_0 = 0$.

Existence: Conversely, if (A_n) and (M_n) are defined by (2.2.2) then (A_n) is predictable with $A_0 = 0$ and (M_n) is a martingale, since

$$E[M_k - M_{k-1} \,|\, \mathcal{F}_{k-1}] \; = \; 0 \qquad P\text{-a.s. for any } k \in \mathbb{N}.$$

□

Conditional Variance Process

Consider a martingale (M_n) such that M_n is square integrable for any $n \geq 0$. Then, by Jensen's inequality, (M_n^2) is a submartingale and can again be decomposed into a martingale $(\widetilde{M}_n)$ and a predictable process $\langle M \rangle_n$ such that $\langle M \rangle_0 = 0$:

$$M_n^2 \quad = \quad \widetilde{M}_n \, + \, \langle M \rangle_n \qquad \text{for any } n \geq 0.$$

The increments of the predictable process are given by

$$\begin{aligned} \langle M \rangle_k - \langle M \rangle_{k-1} &= E[M_k^2 - M_{k-1}^2 \mid \mathcal{F}_{k-1}] \\ &= E\big[(M_k - M_{k-1})^2 \,\big|\, \mathcal{F}_{k-1}\big] + 2 \cdot E\big[M_{k-1} \cdot (M_k - M_{k-1}) \,\big|\, \mathcal{F}_{k-1}\big] \\ &= \mathrm{Var}\big[M_k - M_{k-1} \,\big|\, \mathcal{F}_{k-1}\big] \qquad \text{for any } k \in \mathbb{N}. \end{aligned}$$

Here we have used in the last step that $E[M_k - M_{k-1} \mid \mathcal{F}_{k-1}]$ vanishes since (M_n) is a martingale.

Definition. *The predictable process*

$$\langle M \rangle_n \quad := \quad \sum_{k=1}^{n} \mathrm{Var}\left[M_k - M_{k-1} \mid \mathcal{F}_{k-1}\right], \qquad n \geq 0,$$

is called the ***conditional variance process*** *of the square integrable martingale* (M_n).

Example (Random Walks). If $M_n = \sum_{i=1}^{n} \eta_i$ is a sum of independent centered random variables η_i and $\mathcal{F}_n = \sigma(\eta_1, \ldots, \eta_n)$ then the conditional variance process is given by $\langle M \rangle_n = \sum_{i=1}^{n} \mathrm{Var}[\eta_i]$.

Remark (Quadratic variation). The quadratic variation of a square integrable martingale (M_n) is the process $[M]_n$ defined by

$$[M]_n \, = \, \sum_{k=1}^{n} (M_k - M_{k-1})^2, \qquad n \geq 0.$$

It is easy to verify that $M_n^2 - [M]_n$ is again a martingale (Exercise), however, $[M]_n$ is not adapted. For continuous martingales in continuous time, the quadratic variation and the conditional variance process coincide. In discrete time or for discontinuous martingales they are usually different.

The conditional variance process is crucial for generalizations of classical limit theorems such as the Law of Large Numbers of the Central Limit Theorem from sums of independent random variables to martingales. A direct consequence of the fact that $M_n^2 - \langle M \rangle_n$ is a martingale is that

$$E[M_n^2] \;=\; E[M_0^2] + E[\langle M \rangle_n] \qquad \text{for any } n \geq 0.$$

This can often be used to derive L^2-estimates for martingales.

Example (Discretizations of stochastic differential equations). Consider an ordinary differential equation

$$\frac{dX_t}{dt} \;=\; b(X_t), \qquad t \geq 0, \tag{2.2.3}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a given vector field. In order to take into account unpredictable effects on a system, one is frequently interested in studying random perturbations of the dynamics (2.2.3) of type

$$dX_t \quad = \quad b(X_t)\,dt + \text{ “noise”} \qquad t \geq 0, \tag{2.2.4}$$

with a random noise term. The solution $(X_t)_{t\geq 0}$ of such a stochastic differential equation (SDE) is a stochastic process in continuous time defined on a probability space $(\Omega, \mathcal{A}, P)$ where also the random variables describing the noise effects are defined. The vector field b is called the (deterministic) “drift”. We will need further preparations to make sense of general SDE, but we can also consider time discretizations.

For simplicity let us assume $d = 1$. Let $b, \sigma : \mathbb{R} \to \mathbb{R}$ be continuous functions, and let $(\eta_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables $\eta_i \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ describing the noise effects. We assume

$$E[\eta_i] \;=\; 0 \qquad \text{and} \qquad \mathrm{Var}[\eta_i] = 1 \qquad \text{for any } i \in \mathbb{N}.$$

Here, the values 0 and 1 are just a convenient normalization, but it is an important assumption that the random variables are independent with finite variances. Given an initial value $x_0 \in \mathbb{R}$ and a fine discretization step size $h > 0$, we now define a stochastic process $(X_n^{(h)})$ in discrete time by $X_0^{(h)} = x_0$, and

$$X_{k+1}^{(h)} - X_k^{(h)} \quad = \quad b(X_k^{(h)}) \cdot h + \sigma(X_k^{(h)})\sqrt{h}\eta_{k+1}, \qquad \text{for } k = 0, 1, 2, \ldots \tag{2.2.5}$$

One should think of $X_k^{(h)}$ as an approximation for the value of the process (X_t) at time $t = k \cdot h$. The equation (2.2.5) can be rewritten as

$$X_n^{(h)} = x_0 + \sum_{k=0}^{n-1} b(X_k^{(h)}) \cdot h + \sum_{k=0}^{n-1} \sigma(X_k^{(h)}) \cdot \sqrt{h} \cdot \eta_{k+1}. \tag{2.2.6}$$

To understand the scaling factors h and $\sqrt{h}$ we note first that if $\sigma \equiv 0$ then (2.2.5) respectively (2.2.6) is the Euler discretization of the ordinary differential equation (2.2.3). Furthermore, if $b \equiv 0$ and $\sigma \equiv 1$, then the *diffusive scaling* by a factor $\sqrt{h}$ in the second term ensures that the process $X_{\lfloor t/h \rfloor}^{(h)}, t \geq 0$, converges in distribution as $h \searrow 0$. Indeed, the functional central limit theorem (Donsker's invariance principle) implies that the limit process in this case is a Brownian motion $(B_t)_{t \geq 0}$. In general, (2.2.6) is an Euler discretization of a stochastic differential equation of type

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dB_t$$

where $(B_t)_{t \geq 0}$ is a Brownian motion, cf. Section **??** below. Let $\mathcal{F}_n = \sigma(\eta_1, \ldots, \eta_n)$ denote the filtration generated by the random variables η_i. The following exercise summarizes basic properties of the process $X^{(h)}$:

Exercise. (1). Prove that the process $X^{(h)}$ is a time-homogeneous $\mathcal{F}_n$ Markov chain with transition kernel

$$p(x, \bullet) = N(x + b(x)h, \sigma(x)^2 h)[\bullet].$$

(2). Show that the Doob decomposition $X^{(h)} = M^{(h)} + A^{(h)}$ is given by

$$A_n^{(h)} = \sum_{k=0}^{n-1} b(X_k^{(h)}) \cdot h \tag{2.2.7}$$

$$M_n^{(h)} = x_0 + \sum_{k=0}^{n-1} \sigma(X_k^{(h)}) \sqrt{h} \eta_{k+1}$$

and the conditional variance process of the martingale part is

$$\langle M^{(h)} \rangle_n = \sum_{k=0}^{n-1} \sigma(X_k^{(h)})^2 \cdot h. \tag{2.2.8}$$

(3). Conclude that

$$E[(M_n^{(h)} - x_0)^2] \quad = \quad \sum_{k=0}^{n-1} E[\sigma(X_k^{(h)})^2] \cdot h. \tag{2.2.9}$$

The last equation can be used to derive bounds for the process $(X^{(h)})$ in an efficient way. In particular, the L^2 maximal inequality for martingales implies that

$$P\left[\max_{0\le l\le n} |M_l^{(h)} - x_o| \ge c\right] \quad \le \quad \frac{1}{c^2} \cdot E[(M_n^{(h)} - x_0)^2] \qquad \text{for any } c > 0, \tag{2.2.10}$$

cf. Section **??** below. By combining this estimate with (2.2.7) and (2.2.8), we obtain the upper bounds

$$P\left[\max_{0\le l\le n} \left| X_l^{(h)} - x_0 - \sum_{k=0}^{l-1} b(X_k^{(h)}) \cdot h \right| \ge c\right] \quad \le \quad \frac{h}{c^2} \sum_{k=0}^{n-1} E[\sigma(X_k^{(h)})^2]$$

for $n \in \mathbb{N}$ and $c > 0$. Under appropriate assumptions on b and σ, these bounds can be applied inductively to control for example the deviation of $X^{(h)}$ from the Euler discretization of the deterministic o.d.e. $dX_t/dt = b(X_t)$.

Martingale problem

For a Markov chain (X_n) we obtain a Doob decomposition

$$f(X_n) \;=\; M_n^{[f]} + A_n^{[f]} \tag{2.2.11}$$

for any function f on the state space such that $f(X_n)$ is integrable for each n. computation of the predictable part leads to the following general result:

Theorem 2.5. *Let p be a stochastic kernel on a measurable space $(S, \mathcal{S})$. then for an $(\mathcal{F}_n)$ adapted stochastic process $(X_n)_{n\ge 0}$ with state space $(S, \mathcal{S})$ the following statements are equivalent:*

(1). (X_n) is a time homogeneous $(\mathcal{F}_n)$ Markov chain with transition kernel p.

(2). (X_n) is a ***solution of the martingale problem for the operator*** *$\mathscr{L} = p - I$, i.e., there is a decomposition*

$$f(X_n) \;=\; M_n^{[f]} \;+\; \sum_{k=0}^{n-1} (\mathscr{L}f)(X_k), \qquad n \ge 0,$$

where $(M_n^{[f]})$ *is an* $(\mathcal{F}_n)$ *martingale for any function* $f : S \to \mathbb{R}$ *such that* $f(X_n)$ *is integrable for each* n*, or, equivalently, for any bounded function* $f : S \to \mathbb{R}$*.*

In particular, we see once more that if f is bounded and is harmonic, i.e., $\mathscr{L} f = 0$, then $f(X_n)$ is a martingale, and if f is lower bounded and superharmonic ($\mathscr{L} f \le 0$), then $f(X_n)$ is a supermartingale. The theorem hence extends Theorem 2.3 above.

Proof. The implication "(i)⇒(ii)" is just the Doob decomposition for $f(X_n)$. In fact, by Theorem 2.3, the predictable part is given by

$$\begin{aligned} A_n^{[f]} &= \sum_{k=0}^{n-1} E[f(X_{k+1}) - f(X_k) \mid \mathcal{F}_k] \\ &= \sum_{k=0}^{n-1} E[pf(X_k) - f(X_k) \mid \mathcal{F}_k] \quad = \quad \sum_{k=0}^{n-1} (\mathscr{L} f)(X_k), \end{aligned}$$

and $M_n^{[f]} = f(X_n) - A_n^{[f]}$ is a martingale.
To prove the converse implication "(ii)⇒(i)" suppose that $M_n^{[f]}$ is a martingale for any bounded $f : S \to \mathbb{R}$. then

$$\begin{aligned} 0 &= E[M_{n+1}^{[f]} - M_n^{[f]} \mid \mathcal{F}_n] \\ &= E[f(X_{n+1}) - f(X_n) \mid \mathcal{F}_n] - ((pf)(X_n) - f(X_n)) \\ &= E[f(X_{n+1}) \mid \mathcal{F}_n] - (pf)(X_n) \end{aligned}$$

almost surely for any bounded function f. Hence (X_n) is an $(\mathcal{F}_n)$ Markov chain with transition kernel p. □

Example (One dimensional Markov chains). Suppose that under P_x, the process (X_n) is a time homogeneous Markov chain with state space $S = \mathbb{R}$ or $S = \mathbb{Z}$, initial state $X_0 = x$, and transition kernel p. assuming $X_n \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ for each n, we define the "*drift*" and the "*fluctuations*" of the process by

$$\begin{aligned} b(x) &:= E_x[X_1 - X_0] \\ a(x) &= \mathrm{Var}_x[X_1 - X_0]. \end{aligned}$$

We now compute the Doob decomposition of X_n. Choosing $f(x) = x$ we have

$$(p - I)f(x) \quad = \quad \int y \, p(x, dy) - x \quad = \quad E_x[X_1 - X_0] = b(x).$$

Hence by Theorem 2.4,

$$X_n \quad = \quad M_n + \sum_{k=0}^{n-1} b(X_k) \tag{2.2.12}$$

with a $(\mathcal{F}_n)$ martingale (M_n).

To obtain detailed information on M_n, we compute the variance process: By (2.2.12) and the Markov property, we obtain

$$\langle M \rangle_n \;=\; \sum_{k=0}^{n-1} \mathrm{Var}[M_{k+1} - M_k \mid \mathcal{F}_k] \;=\; \sum_{k=0}^{n-1} \mathrm{Var}[X_{k+1} - X_k \mid \mathcal{F}_k] \quad = \quad \sum_{k=0}^{n-1} a(X_k).$$

Therefore

$$M_n^2 \quad = \quad \widetilde{M}_n + \sum_{k=0}^{n-1} a(X_k) \tag{2.2.13}$$

with another $(\mathcal{F}_n)$ martingale $(\widetilde{M}_n)$. The functions $a(x)$ and $b(x)$ can now be used in connection with fundamental results for martingales as e.g. the maximal inequality (cf. **??** below) to derive bounds for Markov chains in an efficient way.

2.3 Gambling strategies and stopping times

Throughout this section, we fix a filtration $(\mathcal{F}_n)_{n\geq 0}$ on a probability space $(\Omega, \mathcal{A}, P)$.

Martingale transforms

Suppose that $(M_n)_{n\geq 0}$ is a martingale w.r.t. $(\mathcal{F}_n)$, and $(C_n)_{n\in\mathbb{N}}$ is a predictable sequence of real-valued random variables. For example, we may think of C_n as the state in the n-th round of a fair game, and of the martingale increment $M_n - M_{n-1}$ as the net gain (resp. loss) per unit stake. In this case, the capital I_n of a player with gambling strategy (C_n) after n rounds is given recursively by

$$I_n \quad = \quad I_{n-1} + C_n \cdot (M_n - M_{n-1}) \qquad \text{for any } n \in \mathbb{N},$$

i.e.,

$$I_n \quad = \quad I_0 + \sum_{k=1}^{n} C_k \cdot (M_k - M_{k-1}).$$

Definition. *The stochastic process $C_\bullet M$ defined by*

$$(C_\bullet M)_n \;:=\; \sum_{k=1}^{n} C_k \cdot (M_k - M_{k-1}) \qquad \textit{for any } n \geq 0,$$

*is called the **martingale transform** of the martingale $(M_n)_{n\geq 0}$ w.r.t. the predictable sequence $(C_k)_{k\geq 1}$, or the discrete stochastic integral of (C_n) w.r.t. (M_n).*

The process $C_\bullet M$ is a time-discrete version of the stochastic integral $\int\limits_0^t C_s \, dM_s$ for continuous-time processes C and M, cf. **??** below.

Example (**Martingale strategy**)**.** One origin of the word "martingale" is the name of a well-known gambling strategy: In a standard coin-tossing game, the stake is doubled each time a loss occurs, and the player stops the game after the first time he wins. If the net gain in n rounds with unit stake is given by a standard Random Walk

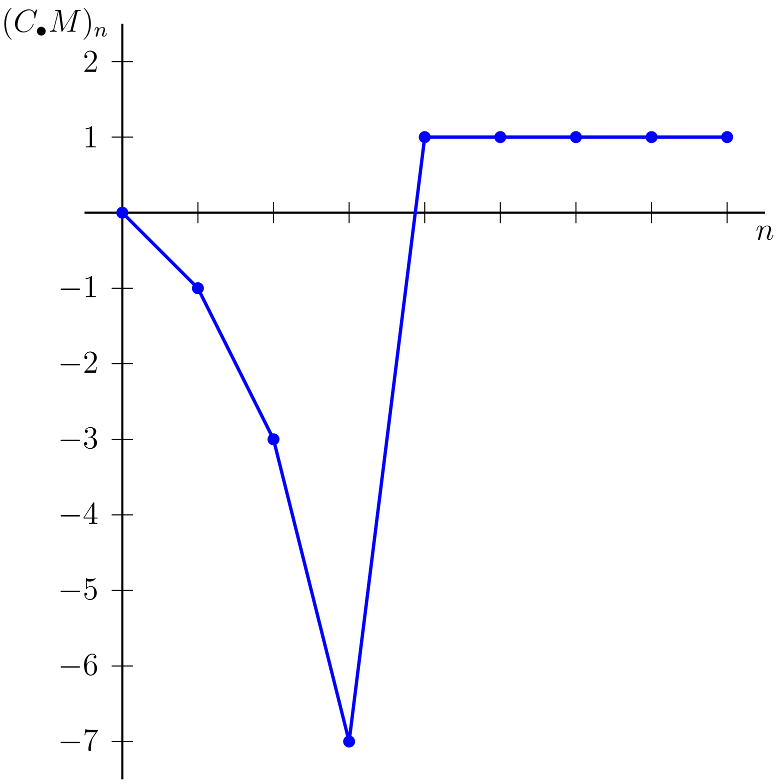
$$M_n \;=\; \eta_1 + \ldots + \eta_n, \qquad \eta_i \text{ i.i.d. with } P[\eta_i = 1] = P[\eta_i = -1] = 1/2,$$

then the stake in the n-th round is

$$C_n \;=\; 2^{n-1} \qquad \text{if } \eta_1 = \ldots = \eta_{n-1} = -1 \text{, and} \qquad C_n \;=\; 0 \quad \text{otherwise.}$$

Clearly, with probability one, the game terminates in finite time, and at that time the player has always won one unit, i.e.,

$$P[(C_\bullet M)_n = 1 \quad \text{eventually}] \quad = \quad 1.$$



At first glance this looks like a safe winning strategy, but of course this would only be the case, if the player had unlimited capital and time available.

Theorem 2.6 (You can't beat the system!). *In $(M_n)_{n\geq 0}$ is an $(\mathcal{F}_n)$ martingale, and $(C_n)_{n\geq 1}$ is predictable with $C_n \cdot (M_n - M_{n-1}) \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ for any $n \geq 1$, then $C_\bullet M$ is again an $(\mathcal{F}_n)$ martingale.*

Proof. For $n \geq 1$ we have

$$\begin{aligned} E[(C_\bullet M)_n - (C_\bullet M)_{n-1} \mid \mathcal{F}_{n-1}] &= E[C_n \cdot (M_n - M_{n-1}) \mid \mathcal{F}_{n-1}] \\ &= C_n \cdot E[M_n - M_{n-1} \mid \mathcal{F}_{n-1}] \quad = \quad 0 \qquad P\text{-a.s.} \end{aligned}$$

□

The theorem shows that a fair game (a martingale) can not be transformed by choice of a clever gambling strategy into an unfair (or "superfair") game. In economic models this fact is crucial to exclude the existence of arbitrage possibilities (riskless profit).

Example (Martingale strategy, cont.). For the classical martingale strategy, we obtain

$$E[(C_\bullet M)_n] \;=\; E[(C_\bullet M)_0] \;=\; 0 \qquad \text{for any } n \geq 0$$

by the martingale property, although

$$\lim_{n\to\infty} (C_\bullet M)_n \;=\; 1 \qquad P\text{-a.s.}$$

This is a classical example showing that the assertion of the dominated convergence theorem may not hold if the assumptions are violated.

Remark. (1). The integrability assumptions in Theorem 2.6 is always satisfied if the random variables C_n are bounded, or if both C_n and M_n are square-integrable for any n.

(2). A corresponding statement holds for supermartingales if in addition $C_n \geq 0$ is assured:

If (M_n) is an $(\mathcal{F}_n)$ supermartingale and $(C_n)_{n\geq 1}$ is non-negative and predictable with $C_n \cdot (M_n - M_{n-1}) \in \mathcal{L}^1$ for any n, then $C_\bullet M$ is again a supermartingale.

The proof is left as an exercise.

Example (Financial market model with one risky asset). Suppose an investor is holding in the time interval $(n-1, n)$ Φ_n units of an asset with price S_n per unit at time n. We assume that (S_n) is an adapted and (Φ_n) a predictable stochastic process w.r.t. a filtration $(\mathcal{F}_n)$. If the investor always puts his remaining capital onto a bank account with guaranteed interest rate r ("riskless asset") then the change of his capital V_n during the time interval $(n-1, n)$ is given by

$$V_n \;=\; V_{n-1} + \Phi_n \cdot (S_n - S_{n-1}) + (V_{n-1} - \Phi_n \cdot S_{n-1}) \cdot r. \tag{2.3.1}$$

Considering the discounted quantity $\widetilde{V}_n \;=\; V_n/(1+r)^n$, we obtain the equivalent recursion

$$\widetilde{V}_n \;=\; \widetilde{V}_{n-1} + \Phi_n \cdot (\widetilde{S}_n - \widetilde{S}_{n-1}) \qquad \text{for any } n \geq 1. \tag{2.3.2}$$

In fact, (2.3.1) holds if and only if

$$V_n - (1+r)V_{n-1} \;=\; \Phi_n \cdot (S_n - (1+r)S_{n-1}),$$

which is equivalent to (2.3.2). Therefore, the discounted capital at time n is given by

$$\widetilde{V}_n \;=\; V_0 + (\Phi_\bullet \widetilde{S})_n.$$

By Theorem 2.6, we can conclude that, if the discounted price progress $(\widetilde{S}_n)$ is an $(\mathcal{F}_n)$ martingale w.r.t. a given probability measure, then $(\widetilde{V}_n)$ is a martingale as well. In this case, we obtain in particular

$$E[\widetilde{V}_n] \;=\; V_0,$$

or, equivalently,

$$E[V_n] \;=\; (1+r)^n V_0 \qquad \text{for any } n \geq 0. \tag{2.3.3}$$

This fact, together with the existence of a martingale measure, can now be used for option pricing under a

No-Arbitrage assumption:
Suppose that the pay off of an option at time N is given by an $(\mathcal{F}_N)$-measurable random variable F. For example, the payoff of a European call option with strike price K based on the asset with price process (S_n) is $S_N - K$ if the price S_n at maturity exceeds K, and 0 otherwise, i.e.,

$$F \;=\; (S_N - K)^+.$$

Let us assume further that the option can be *replicated by a hedging strategy* (Φ_n), i.e., there exists a $\mathcal{F}_0$-measurable random variable V_0 and a predictable sequence of random variables $(\Phi_n)_{1\leq n\leq N}$ such that

$$F \;=\; V_N$$

is the value at time N of a portfolio with initial value V_0 w.r.t. the trading strategy (Φ_n). Then, assuming the non-existence of arbitrage possibilities, the option price at time 0 has to be V_0, since otherwise one could construct an arbitrage strategy by selling the option and investing money in the stock market with strategy (Φ_n), or conversely.
Therefore, if a martingale measure exists (i.e., an underlying probability measure such that the discounted stock price $(\widetilde{S}_n)$ is a martingale), then the no-arbitrage price of the option at time 0 can be computed by (2.3.3) where the expectation is taken w.r.t. the martingale measure.

The following exercise shows how this works out in the Cox-Ross-Rubinstein binomial model, cf. **??** below:

Exercise (No-Arbitrage Pricing in the CRR model). Consider the CRR binomial model, i.e., $\Omega = \{1+a, 1+b\}^N$ with $-1 < a < r < b < \infty$, $X_i(\omega_1, \ldots, \omega_N) = \omega_i$, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, and

$$S_n \;=\; S_0 \cdot \prod_{i=1}^{n} X_i, \quad n = 0, 1, \ldots, N,$$

where S_0 is a constant.

(1). *Completeness of the CRR model:* Prove that for any function $F : \Omega \to \mathbb{R}$ there exists a constant V_0 and a predictable sequence $(\Phi_n)_{1 \le n \le N}$ such that $F = V_N$ where $(V_n)_{1 \le n \le N}$ is defined by (2.3.1), or, equivalently,

$$\frac{F}{(1+r)^N} \;=\; \widetilde{V}_N \;=\; V_0 + (\Phi_\bullet \widetilde{S})_N.$$

Hence in the CRR model, any $\mathcal{F}_N$-measurable function F can be replicated by a predictable trading strategy. Market models with this property are called *complete*.

Hint: Prove inductively that for $n = N, N-1, \ldots, 0$, $\widetilde{F} = F/(1+r)^N$ can be represented as

$$\widetilde{F} = \widetilde{V}_n + \sum_{i=n+1}^{N} \Phi_i \cdot (\widetilde{S}_i - \widetilde{S}_{i-1})$$

with an $\mathcal{F}_n$-measurable function $\widetilde{V}_n$ and a predictable sequence $(\Phi_i)_{n+1 \le i \le N}$.

(2). *Option pricing:* Derive a general formula for the no-arbitrage price of an option with pay off function $F : \Omega \to \mathbb{R}$ in the European call option with maturity N and strike K explicitly.

Stopped Martingales

One possible strategy for controlling a fair game is to terminate the game at a time depending on the previous development.

Definition. *A random variable $T : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ is called a* ***stopping time*** *w.r.t. the filtration $(\mathcal{F}_n)$ if and only if*

$$\{T = n\} \in \mathcal{F}_n \qquad \text{for any } n \geq 0.$$

Clearly, T is a stopping time if and only if the event $\{T \leq n\}$ is contained in $\mathcal{F}_n$ for any $n \geq 0$. This condition is more adequate for late generalizations to continuous time.

Example. (1). The **first hitting time**

$$T_A(\omega) = \min\{n \geq 0 : X_n(\omega) \in A\} \quad (\min \emptyset := \infty)$$

of a measurable set A of the state space by an $(\mathcal{F}_n)$ adapted stochastic process is always an $(\mathcal{F}_n)$ stopping time. For example, if on decides to sell an asset as soon as the price S_n exceeds a given level $\lambda > 0$ then the selling time

$$T_{(\lambda_\infty)} = \min\{n \geq 0 : S_n > \lambda\}$$

is a stopping time w.r.t. $\mathcal{F}_n = \sigma(S_0, S_1, \ldots, S_n)$.

(2). The **last visit time**

$$L_A(\omega) = \sup\{n \geq 0 : X_n(\omega) \in A\}, \quad (\sup \emptyset := 0)$$

is in general not a stopping time (Exercise).

Now let us consider an $(\mathcal{F}_n)$-adapted stochastic process $(M_n)_{n\geq 0}$, and a $(\mathcal{F}_n)$-stopping time T on the probability space $(\Omega, \mathcal{A}, P)$. The process stopped at time T is defined as $(M_{T\wedge n})_{n\geq 0}$ where

$$M_{T\wedge n}(\omega) = M_{T(\omega)\wedge n}(\omega) = \begin{cases} M_n(\omega) & \text{for } n \leq T(\omega), \\ M_{T(\omega)}(\omega) & \text{for } n \geq T(\omega). \end{cases}$$

For example, the process stopped at a hitting time T_A gets stuck at the first time it enters the set A.

Theorem 2.7 (Optional Stopping Theorem,Version 1). *If $(M_n)_{n\geq 0}$ is a martingale (resp. a supermartingale) w.r.t. $(\mathcal{F}_n)$, and T is an $(\mathcal{F}_n)$-stopping time, then the stopped process $(M_{T\wedge n})_{n\geq 0}$ is again an $(\mathcal{F}_n)$-martingale (resp. supermartingale). In particular, we have*

$$E[M_{T\wedge n}] \overset{(\leq)}{=} E[M_0] \qquad \textit{for any } n \geq 0.$$

Proof. Consider the following strategy:

$$C_n = I_{T\geq n} = 1 - I_{T\leq n-1},$$

i.e., we put a unit stake and quit playing at time T. Since T is a stopping time, the sequence (C_n) is predictable. Moreover,

$$M_{T\wedge n} - M_0 = (C_\bullet M)_n \qquad \text{for any } n \geq 0. \tag{2.3.4}$$

In fact, for the increments of the stopped process we have

$$M_{T\wedge n} - M_{T\wedge(n-1)} = \begin{cases} M_n - M_{n-1} & \text{if } T \geq n \\ 0 & \text{if } T \leq n-1 \end{cases} = C_n \cdot (M_n - M_{n-1}),$$

and (2.3.4) follows by summing over n. Since the sequence (C_n) is predictable, bounded and non-negative, the process $C_\bullet M$ is a martingale, supermartingale respectively, provided the same holds for M. □

Remark (IMPORTANT). (1). In general, it is NOT TRUE that

$$E[M_T] = E[M_0], \quad E[M_t] \leq E[M_0] \quad \text{respectively}. \tag{2.3.5}$$

Suppose for example that (M_n) is the classical Random Walk starting at 0 and $T = T_{\{1\}}$ is the first hitting time of the point 1. Then, by recurrence of the Random Walk, $T < \infty$ and $M_t = 1$ hold almost surely although $M_0 = 0$.

(2). If, on the other hand, T is a *bounded stopping time*, then there exists $n \in \mathbb{N}$ such that $T(\omega) \leq n$ for any ω. In this case the optional stopping theorem implies

$$E[M_T] = E[M_{T\wedge n}] \overset{(\leq)}{=} E[M_0].$$

Optional Stopping Theorems

Stopping times occurring in applications are typically not bounded. Therefore, we need more general conditions guaranteeing that (2.3.5) holds nevertheless. A first general criterion is obtained by applying the Dominated Convergence Theorem:

Theorem 2.8 (Optional Stopping Theorem, Version 2). *Suppose that (M_n) is a (super-) martingale w.r.t. $(\mathcal{F}_n)$, T is a $(\mathcal{F}_n)$-stopping time with $P[T < \infty] = 1$, and there exists a random variable $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ such that*

$$|M_{T\wedge n}|_{-} \leq Y \qquad P\text{-almost surely for any } n \in \mathbb{N}.$$

Then

$$E[M_T] \overset{(\leq)}{=} E[M_0].$$

Proof. Since $P[T < \infty] = 1$, we have

$$M_T = \lim_{n\to\infty} M_{T\wedge n} \qquad P\text{-almost surely.}$$

Hence by the Theorem 2.7 and the dominated convergence theorem,

$$E[M_0] \overset{(\leq)}{=} E[M_{T\wedge n}] \overset{n\to\infty}{\longrightarrow} E[M_T].$$

□

Remark (Weakening the assumptions). Instead of the existence of an integrable random variable Y dominating the random variables $M_{T\wedge n}$, $n \in \mathbb{N}$, it is enough to assume that these random variables are **uniformly integrable**, i.e.,

$$\sup_{n\in\mathbb{N}} E\big[|M_{T\wedge n}|\,;\,|M_{T\wedge n}| \geq c\big] \quad \to \quad 0 \qquad \text{as } c \to \infty.$$

A corresponding generalization of the Dominated Convergence Theorem is proven in Section **??** below.

Example (Classical Ruin Problem). Let $a, b, x \in \mathbb{Z}$ with $a < x < b$. We consider the classical Random Walk

$$X_n = x + \sum_{i=1}^{n} \eta_i, \qquad \eta_i \text{ i.i.d. with } P[\eta_i = \pm 1] = \frac{1}{2},$$

with initial value $X_0 = x$. We now show how to apply the optional stopping theorem to compute the distributions of the exit time

$$T(\omega) \;=\; \min\{n \geq 0 \;:\; X_n(\omega) \notin (a,b)\},$$

and the exit point X_T. These distributions can also be computed by more traditional methods (first step analysis, reflection principle), but martingales yield an elegant and general approach.

(1). *Ruin probability* $r(x) = P[X_T = a]$.
The process (X_n) is a martingale w.r.t. the filtration $\mathcal{F}_n = \sigma(\eta_1, \ldots, \eta_n)$, and $T < \infty$ almost surely holds by elementary arguments. As the stopped process $X_{T\wedge n}$ is bounded $(a \leq X_{T\wedge n} << b)$, we obtain

$$x \;=\; E[X_0] \;=\; E[X_{T\wedge n}] \overset{n\to\infty}{\to} E[X_T] \;=\; a \cdot r(x) + b \cdot (1 - r(x))$$

by the Optional Stopping Theorem and the Dominated Convergence Theorem. Hence

$$r(x) \quad = \quad \frac{b-x}{a-x}. \tag{2.3.6}$$

(2). *Mean exit time from* (a,b).
To compute the expectation value $E[T]$, we apply the Optional Stopping Theorem to the $(\mathcal{F}_n)$ martingale

$$M_n \;:=\; X_n^2 - n.$$

By monotone and dominated convergence, we obtain

$$\begin{aligned} x^2 \quad &= \quad E[M_0] \;=\; E[M_{T\wedge n}] \;=\; E[X_{T\wedge n}^2] - E[T \wedge n] \\ &\overset{n\to\infty}{\longrightarrow} \quad E[X_T^2] - E[T]. \end{aligned}$$

Therefore, by (2.3.6),

$$\begin{aligned} E[T] \quad &= \quad E[X_T^2] - x^2 \;=\; a^2 \cdot r(x) + b^2 \cdot (1 - r(x)) - x^2 \\ &= \quad (b-x)\cdot(x-a). \end{aligned} \tag{2.3.7}$$

(3). *Mean passage time of b is infinite.*

The first passage time $T_b = \min\{n \geq 0 \ : \ X_n = b\}$ is greater or equal than the exit time from the interval (a, b) for any $a < x$. Thus by (2.3.7), we have

$$E[T_b] \ \geq \ \lim_{a \to -\infty} (b - x) \cdot (x - a) \ = \ \infty,$$

i.e., T_b is **not integrable**! These and some other related passage times are important examples of random variables with a heavy-tailed distribution and infinite first moment.

(4). *Distribution of passage times.*

We now compute the distribution of the first passage time T_b explicitly in the case $x = 0$ and $b = 1$. Hence let $T = T_1$. As shown above, the process

$$M_n^\lambda \ := \ e^{\lambda X_n} / (\cosh \lambda)^n, \qquad n \geq 0,$$

is a martingale for each $\lambda \in \mathbb{R}$. Now suppose $\lambda > 0$. By the optional Sampling Theorem,

$$1 \ = \ E[M_0^\lambda] \ = \ E[M_{T \wedge n}^\lambda] \ = \ E[e^{\lambda X_{T \wedge n}} / (\cosh \lambda)^{T \wedge n}] \tag{2.3.8}$$

for any $n \in \mathbb{N}$. As $n \to \infty$, the integrands on the right hand side converge $e^\lambda (\cosh \lambda)^{-T} \cdot I_{\{T < \infty\}}$.

Moreover, they are uniformly bounded by e^λ, since $X_{T \wedge n} \leq 1$ for any n. Hence by the Dominated Convergence Theorem, the expectation on the right hand side of (2.3.8) converges to $E[e^\lambda / (\cosh \lambda)^T \ ; \ T < \infty]$, and we obtain the identity

$$E[(\cosh \lambda)^{-T} \ ; \ T < \infty] \ = \ e^{-\lambda} \qquad \text{for any } \lambda > 0. \tag{2.3.9}$$

Taking the limit as $\lambda \searrow 0$, we see that $P[T < \infty] = 1$. Taking this into account, and substituting $s = 1/\cosh \lambda$ in (2.3.9), we can now compute the generating function of T explicitly:

$$E[s^T] \ = \ e^{-\lambda} \ = \ (1 - \sqrt{1 - s^2})/s \qquad \text{for any } s \in (0, 1). \tag{2.3.10}$$

Developing both sides into a power series finally yields

$$\sum_{n=0}^{\infty} s^n \cdot P[T = n] \ = \ \sum_{m=0}^{\infty} (-1)^{m+1} \binom{1/2}{m} s^{2m-1}.$$

Therefore, the distribution of the first passage time of 1 is given by $P[T = 2m] = 0$ and

$$P[T = 2m-1] = (-1)^{m+1}\binom{1/2}{m} = (-1)^{m+1}\cdot\frac{1}{2}\cdot\left(-\frac{1}{2}\right)\cdots\left(\frac{1}{2}-m+1\right)/m!$$

for any $m \geq 0$.

For non-negative supermartingales, we can apply Fatou's Lemma instead of the Dominated Convergence Theorem to pass to the limit as $n \to \infty$ in the Stopping Theorem. The advantage is that no integrability assumption is required. Of course, the price to pay is that we only obtain an inequality:

Theorem 2.9 (Optional Stopping Theorem, Version 3). *If (M_n) is a non-negative supermartingale w.r.t. $(\mathcal{F}_n)$, then*

$$E[M_0] \;\geq\; E[M_T\,;\,T < \infty]$$

holds for any $(\mathcal{F}_n)$ stopping time T.

Proof. Since $M_T = \lim_{n\to\infty} M_{T\wedge n}$ on $\{T < \infty\}$, and $M_T \geq 0$, Theorem 2.7 combined with Fatou's Lemma implies

$$E[M_0] \;\geq\; \liminf_{n\to\infty} E[M_{T\wedge n}] \;\geq\; E\left[\liminf_{n\to\infty} M_{T\wedge n}\right] \;\geq\; E[M_T\,;\,T < \infty].$$

□

Example (Markov chains and Dirichlet problem). Suppose that w.r.t. the probability measure P_x, the process (X_n) is a time-homogeneous Markov chain with measurable state space $(S, \mathcal{S})$, transition kernel p, and start in x. Let $D \in \mathcal{S}$ be a measurable subset of the state space, and $f : D^C \to \mathbb{R}$ a measurable function (the given "boundary values"), and let

$$T \;=\; \min\{n \geq 0 \,:\, X_n \in D^C\}$$

denote the first exit time of the Markov chain from D. By conditioning on the first step of the Markov chain, one can show that if f is non-negative or bounded, then the function

$$h(x) \;=\; E_x[f(X_T)\,;\,T < \infty], \qquad (x \in S),$$

is a solution of the *Dirichlet problem*

$$\begin{aligned} (ph)(x) &= h(x) && \text{for } x \in D, \\ h(x) &= f(x) && \text{for } x \in D^C, \end{aligned}$$

cf. e.g. [Wahrscheinlichkeitstheorie + Stochastische Prozesse]. By considering the martingale $h(X_{T\wedge n})$ for a function h that is harmonic on D, we obtain a converse statement:

Exercise (Uniqueness of the Dirichlet problem). Suppose that $P_x[T < \infty] = 1$ for any $x \in S$.

(1). Prove that $h(X_{T\wedge n})$ is a martingale w.r.t. P_x for any bounded solution h of the Dirichlet problem and any $x \in S$.

(2). Conclude that if f is bounded, then

$$h(x) = E_x[f(X_T)] \tag{2.3.11}$$

is the unique bounded solution of the Dirichlet problem.

(3). Similarly, show that for any non-negative f, the function h defined by (2.3.11) is the minimal non-negative solution of the Dirichlet problem.

We finally state a version of the Optional Stopping Theorem that applies in particular to martingales with bounded increments:

Corollary 2.10 (Optional Stopping for martingales with bounded increments). *Suppose that (M_n) is a $(\mathcal{F}_n)$ martingale, and there exists a finite constant $K \in (0,\infty)$ such that*

$$E[|M_{n+1} - M_n| \mid \mathcal{F}_n] \le K \qquad P\text{-almost surely for any } n \ge 0. \tag{2.3.12}$$

Then for any $(\mathcal{F}_n)$ stopping time T with $E[T] < \infty$, we have

$$E[M_T] = E[M_0].$$

Proof. For any $n \geq 0$,

$$|M_{T\wedge n}| \;\leq\; |M_0| + \sum_{i=0}^{\infty} |M_{i+1} - M_i| \cdot I_{\{T>i\}}.$$

Let Y denote the expression on the right hand side. We will show that Y is an integrable random variable – this implies the assertion by Theorem 2.8. To verify integrability of Y note that the event $\{T > i\}$ is contained in $\mathcal{F}_i$ for any $i \geq 0$ since T is a stopping time. therefore and by (2.3.12),

$$E[|M_{i+1} - M_i|\,;\,T > i] \;=\; E[E[|M_{i+1} - M_i|\,|\,\mathcal{F}_i]\,;\,T > i] \;\leq\; k \cdot P[T > i].$$

Summing over i, we obtain

$$E[Y] \;\leq\; E[|M_0|] + k \cdot \sum_{i=0}^{\infty} P[T > i] \;=\; E[|M_0|] + k \cdot E[T] \;<\; \infty$$

by the assumptions. □

Exercise. Prove that the expectation value $E[T]$ of a stopping time T is finite if there exist constants $\varepsilon > 0$ and $k \in \mathbb{N}$ such that

$$P[T \leq n + k \mid \mathcal{F}_n] \;>\; \varepsilon \qquad P\text{-a.s. for any } n \in \mathbb{N}.$$

Wald's identity for random sums

We finally apply the Optional Stopping Theorem to sums of independent random variables with a random number T of summands. The point is that we do not assume that T is independent of the summands but only that is a stopping time w.r.t. the filtration generated by the summands.

Let $S_n = \eta_1 + \ldots + \eta_n$ with i.i.d. random variables $\eta_i \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Denoting by m the expectation values of the increments η_i, the process

$$M_n \;=\; S_n - n \cdot m$$

is a martingale w.r.t. $\mathcal{F}_n = \sigma(\eta_1, \ldots, \eta_n)$. By applying Corollary 2.10 to this martingale, we obtain:

Theorem 2.11 (Wald's identity). *Suppose that T is an $(\mathcal{F}_n)$ stopping time with $E[T] < \infty$. Then*

$$E[S_T] \;=\; m \cdot E[T].$$

Proof. For any $n \geq 0$, we have

$$E[|M_{n+1} - M_n| \mid \mathcal{F}_n] \;=\; E[|\eta_{n+1} - m| \,|\mathcal{F}_n] \;=\; E[|\eta_{n+1} - m|]$$

by the independence of the η_i. As the η_i are identically distributed and integrable, the right hand side is a finite constant. Hence Corollary 2.10 applies, and we obtain

$$0 \;=\; E[M_0] \;=\; E[M_t] \;=\; E[S_t] - m \cdot E[T].$$

□

2.4 Maximal inequalities

For a standard Random Walk $S_n = \eta_1 + \ldots + \eta_n$, η_i i.i.d. with $P[\eta_i = \pm 1] = 1/2$, the reflection principle implies the identity

$$\begin{aligned} P[\max(S_0, S_1, \ldots) \geq c] &= P[S_n \geq c] + P[S_n \geq c+1] \\ &= P[|S_n| > c] + \frac{1}{2} \cdot P[|S_n| = c] \end{aligned}$$

for any $c \in \mathbb{N}$. In combination with the Markov-Čebyšev inequality this can be used to control the running maximum of the Random Walk in terms of the moments of the last value S_n.

Maximal inequalities are corresponding estimates for $\max(M_0, M_1, \ldots, M_n)$ or $\sup_{k \geq 0} M_k$ when (M_n) is a sub- or supermartingale respectively. These estimates are an important tool in stochastic analysis. They are a consequence of the Optional Stopping Theorem.

Doob's inequality

We first prove the basic version of maximal inequalities for sub- and supermartingales:

Theorem 2.12 (Doob).

(1). Suppose that $(M_n)_{n\ge 0}$ is a non-negative supermartingale. Then

$$P\left[\sup_{k\ge 0} M_k \ge c\right] \;\le\; \frac{1}{c}\cdot E[M_0] \qquad \textit{for any } c>0.$$

(2). Suppose that $(M_n)_{n\ge 0}$ is a non-negative submartingale. Then

$$P\left[\max_{0\le k\le n} M_k \ge c\right] \;\le\; \frac{1}{c}\cdot E\left[M_n\,;\,\max_{0\le k\le n} M_k \ge c\right] \;\le\; \frac{1}{c}\cdot E[M_n] \quad \textit{for any } c>0.$$

Proof. (1). For $c>0$ we consider the stopping time

$$T_c \;=\; \min\{k\ge 0\;:\; M_k\ge c\}, \qquad \min\emptyset = \infty.$$

Note that $T_c<\infty$ whenever $\sup M_k > c$. Hence by the version of the Optional Stopping Theorem for non-negative supermartingales, we obtain

$$P[\sup M_k > c] \;\le\; P[T_c<\infty] \;\le\; \frac{1}{c}E[M_{T_c}\,;\,T_c<\infty] \;\le\; \frac{1}{c}E[M_0].$$

Here we have used in the second and third step that (M_n) is non-negative. Replacing c by $c-\varepsilon$ and letting ε tend to zero we can conclude

$$P[\sup M_k\ge c] \;=\; \lim_{\varepsilon\searrow 0} P[\sup M_k > c-\varepsilon] \;\le\; \liminf_{\varepsilon\searrow 0}\frac{1}{c-\varepsilon}E[M_0] \;=\; \frac{1}{c}\cdot E[M_0].$$

(2). For a non-negative supermartingale, we obtain

$$\begin{aligned} P\left[\max_{0\le k\le n} M_k\ge c\right] &= P[T_c\le n] \;\le\; \frac{1}{c}E[M_{T_c}\,;\,T_c\le n] \\ &= \frac{1}{c}\sum_{k=0}^{n} E[M_k\,;\,T_c=k] \;\le\; \frac{1}{c}\sum_{k=0}^{n} E[M_n\,;\,T_c=k] \\ &= \frac{1}{c}\cdot E[M_n\,;\,T_c\le n]. \end{aligned}$$

Here we have used in the second last step that $E[M_k\,;\,T_c=k]\le E[M_n\,;\,T_c=k]$ since (M_n) is a supermartingale and $\{T_c=k\}$ is in $\mathcal{F}_k$.

□

First consequences of Doob's maximal inequality for submartingales are extensions of the classical Markov- Čebyšev inequalities:

Corollary 2.13. *(1). Suppose that $(M_n)_{n\geq 0}$ is an arbitrary submartingale (not necessarily non-negative!). Then*

$$P\left[\max_{k\leq n} M_k \geq c\right] \;\leq\; \frac{1}{c}E\left[M_n^+ \,;\, \max_{k\leq n M_k\geq c}\right] \qquad \textit{for any } c>0, \textit{ and}$$
$$P\left[\max_{k\leq n} M_k \geq c\right] \;\leq\; e^{-\lambda c}E\left[e^{\lambda M_n} \,;\, \max_{k\leq n} M_k\geq c\right] \qquad \textit{for any } \lambda, c>0.$$

(2). If (M_n) is a martingale then, moreover, the estimates

$$P\left[\max_{k\leq n}|M_k|\geq c\right] \;\leq\; \frac{1}{c^p}E\left[|M_n|^p \,;\, \max_{k\leq n}|M_k|\geq c\right]$$

hold for any $c>0$ and $p\in[1,\infty)$.

Proof. The corollary follows by applying the maximal inequality to the non-negative submartingales $M_n^+, \exp(\lambda M_n), |M_n|^p$ respectively. These processes are indeed submartingales, as the functions $x\mapsto x^+$ and $x\mapsto\exp(\lambda x)$ are convex and non-decreasing for any $\lambda>0$, and the functions $x\mapsto|x|^p$ are convex for any $p\geq 1$. □

L^p inequalities

The last estimate in Corollary 2.13 can be used to bound the L^p norm of the running maximum of a martingale in terms of the L^p-norm of the last value. The resulting bound, known as Doob's L^p-inequality, is crucial for stochastic analysis. We first remark:

Lemma 2.14. *If $Y:\Omega\to\mathbb{R}_+$ is a non-negative random variable, and $G(y)=\int_0^y g(x)dx$ is the integral of a non-negative function $g:\mathbb{R}_+\to\mathbb{R}_+$, then*

$$E[G(Y)] \;=\; \int_0^\infty g(c)\cdot P[Y\geq c]\,dc.$$

Proof. By Fubini's theorem we have

$$\begin{aligned} E[G(Y)] &= E\left[\int_0^Y g(c)\,dc\right] \\ &= E\left[\int_0^\infty I_{[0,Y]}(c)g(c)\,dc\right] \\ &= \int_0^\infty g(c)\cdot P[Y\geq c]\,dc. \end{aligned}$$

□

Theorem 2.15 (Doob's L^p inequality). *Suppose that $(M_n)_{n\geq 0}$ is a martingale, and let*

$$M_n^* := \max_{k\leq n}|M_k|, \qquad \textit{and} \qquad M^* := \sup_k |M_k|.$$

Then, for any $p,q\in(1,\infty)$ such that $\frac{1}{p}+\frac{1}{q}=1$, we have

$$\|M_n^*\|_{L^p} \leq q\cdot\|M_n\|_{L^p}, \qquad \textit{and} \qquad \|M^*\|_{L^p} \leq q\cdot\sup_n\|M_n\|_{L^p}.$$

In particular, if (M_n) is bounded in L^p then M^* is contained in L^p.

Proof. By Lemma 2.14, Corollary 2.13 applied to the martingales M_n and $(-M_n)$, and Fubini's theorem, we have

$$\begin{aligned} E[(M_n^*)^p] &\overset{2.14}{=} \int_0^\infty pc^{p-1}\cdot P[M_n^*\geq c]\,dc \\ &\overset{2.13}{\leq} \int_0^\infty pc^{p-2}E[|M_n|\,;\,M_n^*\geq c]\,dc \\ &\overset{\text{Fub.}}{=} E\left[|M_n|\cdot\int_0^{M_n^*} pc^{p-2}\,dp\right] \\ &= \frac{p}{p-1}E[|M_n|\cdot(M_n^*)^{p-1}] \end{aligned}$$

for any $n\geq 0$ and $p\in(1,\infty)$. Setting $q=\frac{p}{p-1}$ and applying Hölder's inequality to the right hand side, we obtain

$$E[(M_n^*)^p] \leq q\cdot\|M_n\|_{L^p}\cdot\|(M_n^*)^{p-1}\|_{L^q} = q\cdot\|M_n\|_{L^p}\cdot E[(M_n^*)^p]^{1/q},$$

i.e.,

$$\|M_n^*\|_{L^p} \;=\; E[(M_n^*)^p]^{1-1/q} \;\leq\; q \cdot \|M_n\|_{L^p}. \tag{2.4.1}$$

This proves the first inequality. The second inequality follows as $n \to \infty$, since

$$\begin{aligned}\|M^*\|_{L^p} &= \left\|\lim_{n\to\infty} M_n^*\right\|_{L^p} \\ &= \liminf_{n\to\infty} \|M_n^*\|_{L^p} \\ &\leq q \cdot \sup_{n\in\mathbb{N}} \|M_n\|_{L^p}\end{aligned}$$

by Fatou's Lemma. □

Hoeffding's inequality

For a standard Random Walk (S_n) starting at 0, the reflection principle combined with Bernstein's inequality implies the upper bound

$$P[\max(S_0,\ldots,S_n) \geq c] \;\leq\; 2 \cdot P[S_n \geq c] \;\leq\; 2 \cdot \exp(-2c^2/n)$$

for any $n \in \mathbb{N}$ and $c \in (0,\infty)$. A similar inequality holds for arbitrary martingales with bounded increments:

Theorem 2.16 (Azuma, Hoeffding). *Suppose that (M_n) is a martingale such that*

$$|M_n - M_{n-1}| \;\leq\; a_n \qquad P\text{-almost surely}$$

for a sequence (a_n) of non-negative constants. Then

$$P\left[\max_{k\leq n}(M_k - M_0) \geq c\right] \;\leq\; \exp\left(-\frac{1}{2}c^2 \Big/ \sum_{i=1}^n a_i^2\right) \tag{2.4.2}$$

for any $n \in \mathbb{N}$ and $c \in (0,\infty)$.

Proof. W.l.o.g. we may assume $M_0 = 0$. Let $Y_n = M_n - M_{n-1}$ denote the martingale increments. We will apply the exponential form of the maximal inequality. For $\lambda > 0$ and $n \in \mathbb{N}$, we have,

$$E[e^{\lambda M_n}] \;=\; E\left[\prod_{i=1}^n e^{\lambda Y_i}\right] \;=\; E\left[e^{\lambda M_{n-1}} \cdot E\left[e^{\lambda Y_n} \mid \mathcal{F}_{n-1}\right]\right]. \tag{2.4.3}$$

To bound the conditional expectation, note that

$$e^{\lambda Y_n} \;\le\; \frac{1}{2}\frac{a_n - Y_n}{a_n}e^{-\lambda a_n} + \frac{1}{2}\frac{a_n + Y_n}{a_n}e^{\lambda a_n}$$

holds almost surely, since $x \mapsto \exp(\lambda x)$ is a convex function, and $-a_n \le Y_n \le a_n$. Indeed, the right hand side is the value at Y_n of the line connecting the points $(-a_n, \exp(-\lambda a_n))$ and $(a_n, \exp(\lambda a_n))$. Since (M_n) is a martingale, we have

$$E[Y_n|\mathcal{F}_{n-1}] \;=\; 0,$$

and therefore

$$E[e^{\lambda Y_n} \mid \mathcal{F}_{n-1}] \;\le\; \left(e^{-\lambda a_n} + e^{\lambda a_n}\right) \;=\; \cosh(\lambda a_n) \;\le\; e^{(\lambda a_n)^2/2}$$

almost surely. Now, by (2.4.3), we obtain

$$E[e^{\lambda Y_n}] \;\le\; E[e^{\lambda M_{n-1}}] \cdot e^{(\lambda a_n)^2/2}.$$

Hence, by induction on n,

$$E[e^{\lambda M_n}] \;\le\; \exp\left(\frac{1}{2}\lambda^2\sum_{i=1}^n a_i^2\right) \qquad \text{for any } n \in \mathbb{N}, \tag{2.4.4}$$

and, by the exponential maximal inequality (cf. Corollary 2.13),

$$P[\max_{k\le n} M_k \ge c] \;\le\; \exp\left(-\lambda c + \frac{1}{2}\lambda^2\sum_{i=1}^n a_i^2\right) \tag{2.4.5}$$

holds for any $n \in \mathbb{N}$ and $c, \lambda > 0$.

For a given c and n, the expression on the right hand side of (2.4.5) is minimal for $\lambda = c/\sum_{i=1}^n a_i^2$. Choosing λ correspondingly, we finally obtain the upper bound (2.4.2). □

Hoeffding's concentration inequality has numerous applications, for example in the analysis of algorithms, cf. [Mitzenmacher, Upful: Probability and Computing]. Here, we just consider one simple example to illustrate the way it typically is applied:

Example (Pattern Matching). Suppose that $X_1, X_2, \ldots, X_n$ is a sequence of i.i.d. random variables ("letters") taking value sin a finite set S (the underlying "alphabet"), and let

$$N \;=\; \sum_{i=0}^{n-l} I_{\{X_{i+1}=a_1, X_{i+2}=a_x, \ldots, X_{i+l}=a_l\}} \tag{2.4.6}$$

denote the number of occurrences of a given "word" $a_1a_2 \cdots a_l$ with l letters in the random text. In applications, the "word" could for example be a DNA sequence. We easily obtain

$$E[N] \;=\; \sum_{i=0}^{n-l} P[X_{i+k} = a_k \text{ for } k = 1, \ldots, l] \;=\; (n-l+1)/|S|^l. \tag{2.4.7}$$

To estimate the fluctuations of the random variable N around its mean value, we consider the martingale

$$M_i \;=\; E[N \mid \sigma(X_1, \ldots, X_i)], \qquad (i = 0, 1, \ldots, n)$$

with initial value $M_0 = E[N]$ and terminal value $M_n = N$. Since at most l of the summands in (2.4.6) are not independent of i, and each summand takes values 0 and 1 only, we have

$$|M_i - M_{i-1}| \;\leq\; l \qquad \text{for each } i = 0, 1, \ldots, n.$$

Therefore, by Hoeffding's inequality, applied in both directions, we obtain

$$P[|N - E[N]| \geq c] \;=\; P[|M_n - M_0| \geq c] \;\leq\; 2\exp(-c^2/(2nl^2))$$

for any $c > 0$, or equivalently,

$$P[|N - E[N]| \geq \varepsilon \cdot l\sqrt{n}] \;\leq\; 2 \cdot \exp(-\varepsilon^2/2) \qquad \text{for any } \varepsilon > 0. \tag{2.4.8}$$

The equation (2.4.7) and the estimate (2.4.8) show that N is highly concentrated around its mean if l is small compared to $\sqrt{n}$.

Chapter 3

Martingales of Brownian Motion

The notion of a martingale, sub- and super martingale in continuous time can be defined similarly as in the discrete parameter case. Fundamental results such as optional stopping theorem or the maximal inequality carry over from discrete parameter to continuous time martingales under additional regularity conditions as, for example, continuity of the sample paths. Similarly as for Markov chains in discrete time, martingale methods can be applied to to derive expressions and estimates for probabilities and expectation values of Brownian motion in a clear and efficient way.

We start with the definition of martingales in continuous time. Let $(\Omega, \mathcal{A}, P)$ denote a probability space.

Definition. *(1). A continuous-time* ***filtration*** *on* $(\Omega, \mathcal{A})$ *is a family* $(\mathcal{F}_t)_{t\in[0,\infty)}$ *of* σ*-algebras* $\mathcal{F}_t \subseteq \mathcal{A}$ *such that* $\mathcal{F}_s \subseteq \mathcal{F}_t$ *for any* $0 \le s \le t$.

(2). A real-valued stochastic process $(M_t)_{t\in[0,\infty)}$ *on* $(\Omega, \mathcal{A}, P)$ *is called a* ***martingale*** *(or* ***super****-,* ***submartingale****) w.r.t. a filtration* $(\mathcal{F}_t)$ *if and only if*

(a) (M_t) *is adapted w.r.t.* $(\mathcal{F}_t)$*, i.e.,* M_t *is* $\mathcal{F}_t$ *measurable for any* $t \ge 0$.

(b) For any $t \ge 0$*, the random variable* M_t *(resp.* M_t^+, M_t^-*) is integrable.*

(c) $E[M_t \mid \mathcal{F}_s] \overset{(\le,\ge)}{=} M_s$ P*-a.s. for any* $0 \le s \le t$.

3.1 Some fundamental martingales of Brownian Motion

In this section, we identify some important martingales that are functions of Brownian motion. Let $(B_t)_{t\geq 0}$ denote a d-dimensional Brownian motion defined on $(\Omega, \mathcal{A}, P)$.

Filtrations

Any stochastic process $(X_t)_{t\geq 0}$ in continuous time generates a filtration

$$\mathcal{F}_t^X = \sigma(X_s \mid 0 \leq s \leq t), \qquad t \geq 0.$$

However, not every hitting time that we are interested in is a stopping time w.r.t. this filtration. For example, for one-dimensional Brownian motion (B_t), the first hitting time $T = \inf\{t \geq 0 : B_t > c\}$ of the *open* interval (c, ∞) is not an $(\mathcal{F}_t^B)$ stopping time. An intuitive explanation for this fact is that for $t \geq 0$, the event $\{T \leq t\}$ is not contained in $\mathcal{F}_t^B$, since for a path with $B_s \leq c$ on $[0,t]$ and $B_t = c$, we can not decide at time t, if the path will enter the interval (c, ∞) in the next instant. For this and other reasons, we consider the right-continuous filtration

$$\mathcal{F}_t := \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}^B, \qquad t \geq 0,$$

that takes into account "infinitesimal information on the future development."

Exercise. Prove that the first hitting time $T_A = \inf\{t \geq 0 \ : \ B_t \in A\}$ of a set $A \subseteq \mathbb{R}^d$ is an $(\mathcal{F}_t^B)$ stopping time if A is closed, whereas T_A is a $(\mathcal{F}_t)$ stopping time but not necessarily a $(\mathcal{F}_t^B)$ stopping time if A is open.

It is easy to verify that the d-dimensional Brownian motion (B_t) is also a Brownian motion w.r.t. the right-continuous filtration $(\mathcal{F}_t)$:

Lemma 3.1. *For any $0 \leq s < t$, the increment $B_t - B_s$ is independent of $\mathcal{F}_s$ with distribution $N(0, (t-s) \cdot I_d)$.*

Proof. Since $t \mapsto B_t$ is almost surely continuous, we have

$$B_t - B_s = \lim_{\substack{\varepsilon \searrow 0 \\ \varepsilon \in \mathbb{Q}}} (B_t - B_{s+\varepsilon}) \qquad P\text{-a.s.} \tag{3.1.1}$$

For small $\varepsilon > 0$ the increment $B_t - B_{s+\varepsilon}$ is independent of $\mathcal{F}^B_{s+\varepsilon}$, and hence independent of $\mathcal{F}_s$. Therefore, by (3.1.1), $B_t - B_s$ is independent of $\mathcal{F}_s$ as well. □

Another filtration of interest is the completed filtration $(\mathcal{F}^P_t)$. A σ-algebra $\mathcal{F}$ is called **complete** w.r.t. a probability measure P iff it contains all subsets of P-measure zero sets. The **completion** of a σ-algebra $\mathcal{A}$ w.r.t. a probability measure P on $(\Omega, \mathcal{A})$ is the complete σ-algebra

$$\mathcal{A}^P \;=\; \{A \subseteq \Omega \mid \exists A_1, A_2 \in \mathcal{A} \;:\; A_1 \subseteq A \subseteq A_2, P[A_2 \setminus A_1] = 0\}$$

generated by all sets in $\mathcal{A}$ and all subsets of P-measure zero sets in $\mathcal{A}$.

It can be shown that the completion $(\mathcal{F}^P_t)$ of the right-continuous filtration $(\mathcal{F}_t)$ is again right-continuous. The assertion of Lemma 3.1 obviously carries over to the completed filtration.

Remark (The "usual conditions")**.** Some textbooks on stochastic analysis consider only complete right-continuous filtrations. A filtration with these properties is said to **satisfy the usual conditions**. A disadvantage of completing the filtration, however, is that $(\mathcal{F}^P_t)$ depends on the underlying probability measure P (or, more precisely, on its null sets). This can cause problems when considering several non-equivalent probability measures at the same time.

Brownian Martingales

We now identify some basic martingales of Brownian motion:

Theorem 3.2. *For a d-dimensional Brownian motion (B_t) the following processes are martingales w.r.t. each of the filtrations $(\mathcal{F}^B_t)$, $(\mathcal{F}_t)$ and $(\mathcal{F}^P_t)$:*

(1). The coordinate processes $B^{(i)}_t$, $1 \le i \le d$.

(2). $B^{(i)}_t - B^{(j)}_t - t \cdot \delta_{ij}$ for any $1 \le i, j \le d$.

(3). $\exp(\alpha \cdot B_t - \frac{1}{2}|\alpha|^2 t)$ for any $\alpha \in \mathbb{R}^d$.

*The processes $M^\alpha_t = \exp(\alpha \cdot B_t - \frac{1}{2}|\alpha|^2 t)$ are called **exponential martingales**.*

Proof. We only prove the second assertion for $d = 1$ and the right-continuous filtration $(\mathcal{F}_t)$. The verification of the remaining statements is left as an exercise.
For $d = 1$, since B_t is normally distributed, the $\mathcal{F}_t$-measurable random variable $B_t^2 - t$ is integrable for any t. Moreover, by Lemma 3.1,

$$\begin{aligned} E[B_t^2 - B_s^2 \mid \mathcal{F}_s] &= E[(B_t - B_s)^2 \mid \mathcal{F}_s] + 2B_s \cdot E[B_t - B_s \mid \mathcal{F}_s] \\ &= E[(B_t - B_s)^2] + 2B_s \cdot E[B_t - B_s] \;=\; t - s \end{aligned}$$

almost surely. Hence

$$E[B_t^2 - t \mid \mathcal{F}_s] \;=\; B_s^2 - s \qquad P\text{-a.s. for any } 0 \le s \le t,$$

i.e., $B_t^2 - t$ is a $(\mathcal{F}_t)$ martingale. □

Remark (Doob decomposition, variance process of Brownian motion). For a one-dimensional Brownian motion (B_t), the theorem yields the Doob decomposition

$$B_t^2 \;=\; M_t + t$$

of the submartingale (B_t^2) into a martingale (M_t) and the increasing adapted process $\langle B \rangle_t = t$.
A Doob decomposition of the process $f(B_t)$ for general functions $f \in C^2(\mathbb{R})$ will be obtained below as a consequence of Itô's celebrated formula. It states that

$$f(B_t) - f(B_0) \;=\; \int_0^t f'(B_s)\,dB_s + \frac{1}{2}\int_0^t f''(B_s)\,ds \tag{3.1.2}$$

where the first integral is an Itô stochastic integral, cf. Section **??** below. If, for example, f' is bounded, then the Itô integral is a martingale as a function of t. If f is convex then $f(B_t)$ is a submartingale and the second integral is an increasing adapted process in t. It is a consequence of (3.1.2) that Brownian motion solves the martingale problem for the operator $\mathscr{L}f = f''$ with domain $\mathrm{Dom}(\mathscr{L}) = \{f \in C^2(\mathbb{R}) \;:\; f' \text{ bounded}\}$.
Itô's formula (3.1.2) can be extended to the multi-dimensional case. The second derivative is the replaced by the Laplacian $\Delta f = \sum_{i=1}^d \frac{\partial^2 f}{\partial x_i^2}$. The multi-dimensional Itô formula implies that a sub- or superharmonic function of d-dimensional Brownian motion is a sub- or supermartingale respectively, if appropriate integrability conditions hold. We now give a direct proof of this fact by the mean value property:

Lemma 3.3 (Mean value property for harmonic function in $\mathbb{R}^d$). *Suppose that $h \in C^2(\mathbb{R}^d)$ is a **(super-)harmonic function**, i.e.,*

$$\Delta h(x) \overset{(\leq)}{=} 0 \qquad \text{for any } x \in \mathbb{R}^d.$$

Then for any $x \in \mathbb{R}^d$ and any rotationally invariant probability measure μ on $\mathbb{R}^d$,

$$\int h(x+y)\,\mu(dy) \overset{(\leq)}{=} h(x). \tag{3.1.3}$$

Proof. By the classical mean value property, $h(x)$ is equal to (resp. greater or equal than) the average value $\fint_{\partial B_r(x)} h$ of h on any sphere $\partial B_r(x)$ with center at x and radius $r > 0$, cf. e.g. [Königsberger: Analysis II]. Moreover, if μ is a rotationally invariant probability measure then the integral in (3.1.3) is an average of average values over spheres:

$$\int h(x+y)\,\mu(dy) = \int \fint_{\partial B_r(x)} h\,\mu_R(dr) \overset{(\leq)}{=} h(x),$$

where μ_R is the distribution of $R(x) = |x|$ under μ. □

Theorem 3.4. *If $h \in C^2(\mathbb{R}^d)$ is a (super-) harmonic function then $(h(B_t))$ is a (super-) martingale w.r.t. $(\mathcal{F}_t)$ provided $h(B_t)$ (resp. $h(B_t)^+$) is integrable for any $t \geq 0$.*

Proof. By Lemma 3.1 and the mean value property, we obtain

$$\begin{aligned} E[h(B_t) \mid \mathcal{F}_s](\omega) &= E[h(B_s + B_t - B_s) \mid \mathcal{F}_s](\omega) \\ &= E[h(B_s(\omega) + B_t - B_s)] \\ &= \int h(B_s(\omega) + y)\, N(0, (t-s)\cdot I)(dy) \\ &\overset{(\leq)}{=} h(B_s(\omega)) \end{aligned}$$

for any $0 \leq s \leq t$ and P-almost every ω. □

3.2 Optional Sampling Theorem and Dirichlet problem

Optional Sampling and Optional Stopping

The optional stopping theorem can be easily extended to continuous time martingales with continuous sample paths. We directly prove a generalization:

Theorem 3.5 (**Optional Sampling Theorem**). *Suppose that $(M_t)_{t\in[0,\infty]}$ is a martingale w.r.t. an arbitrary filtration $(\mathcal{F}_t)$ such that $t \mapsto M_t(\omega)$ is continuous for P-almost every ω. Then*

$$E[M_T \mid \mathcal{F}_S] \;=\; M_S \qquad P\text{-almost surely} \tag{3.2.1}$$

for any bounded $(\mathcal{F}_t)$ stopping times S and T with $S \leq T$.

We point out that an additional assumption on the filtration (e.g. right-continuity) is not required in the theorem. Stopping times and the σ-algebra $\mathcal{F}_S$ are defined for arbitrary filtrations on complete analogy to the definitions for the filtration $(\mathcal{F}_t^B)$ in Section 1.5.

Remark (**Optional Stopping**). By taking expectation values in the Optional Sampling Theorem, we obtain

$$E[M_T] \;=\; E[E[M_T \mid \mathcal{F}_0]] \;=\; E[M_0]$$

for any bounded stopping time T. For unbounded stopping times,

$$E[M_T] \;=\; E[M_0]$$

holds by dominated convergence provided $T < \infty$ almost surely, and the random variables $M_{T\wedge n}, n \in \mathbb{N}$, are uniformly integrable.

Proof of Theorem 3.5. We verify the defining properties of the conditional expectation in (3.4) by approximating the stopping times by discrete random variable:

(1). *M_S has a $\mathcal{F}_S$-measurable modification:* For $n \in \mathbb{N}$ let $\widetilde{S}_n = 2^{-n}\lfloor 2^n S \rfloor$, i.e.,

$$\widetilde{S}_n \;=\; k \cdot 2^{-n} \qquad \text{on} \qquad \{k \cdot 2^{-n} \leq S < (k+1)2^{-n} \text{ for any } k = 0, 1, 2, \ldots\}.$$

We point out that in general, $\widetilde{S}_n$ is **not** a stopping time w.r.t. $(\mathcal{F}_t)$. Clearly, the sequence $(\widetilde{S}_n)_{n\in\mathbb{N}}$ is *increasing* with $S = \lim S_n$. By almost sure continuity

$$M_S \;=\; \lim_{n\to\infty} M_{\widetilde{S}_n} \qquad P\text{-almost surely}. \tag{3.2.2}$$

On the other hand, each of the random variables $M_{\widetilde{S}_n}$ is $\mathcal{F}_S$-measurable. In fact,

$$M_{\widetilde{S}_n} \cdot I_{S\leq t} \;=\; \sum_{k:k\cdot 2^{-n}\leq t} M_{k\cdot 2^{-n}} \cdot I_{k2^{-n}\leq S<(k+1)2^{-n} \text{ and } S\leq t}$$

is $\mathcal{F}_t$-measurable for any $t \geq 0$ since S is an $(\mathcal{F}_t)$ stopping time. Therefore, by (3.2.2), the random variable $\widetilde{M}_S := \limsup_{n\to\infty} M_{\widetilde{S}_n}$ is an $\mathcal{F}_S$-measurable modification of M_S.

(2). $E[M_T; A] = E[M_S; A]$ *for any* $A \in \mathcal{F}_S$*:* For $n \in \mathbb{N}$, the discrete random variables $T_n = 2^{-n} \cdot \lceil 2^n T \rceil$, cf. the proof of Theorem 1.26. In particular, $\mathcal{F}_S \subseteq \mathcal{F}_{S_n} \subseteq \mathcal{F}_{T_n}$. Furthermore, (T_n) and (S_n) are *decreasing* sequences with $T = \lim T_n$ and $S = \lim S_n$. As T and S are bounded random variables by assumption, the sequences (T_n) and (S_n) are *uniformly bounded* by a finite constant $c \in (0, \infty)$. Therefore, we obtain

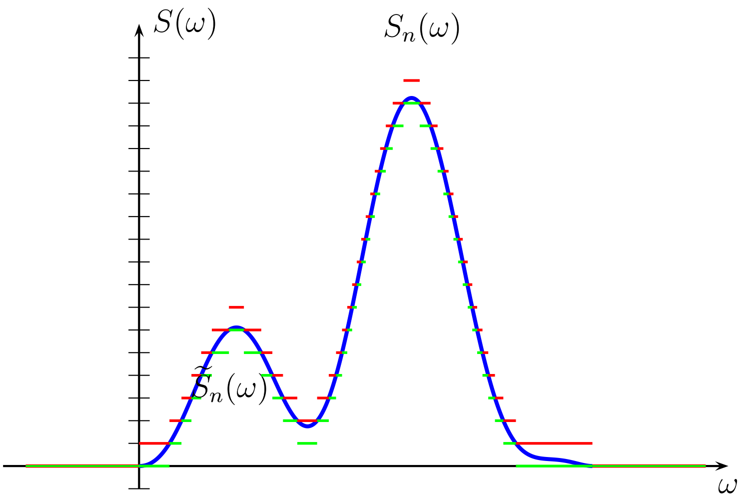


Figure 3.1: Two ways to approximate a continuous stopping time.

$$\begin{aligned} E[M_{T_n}\,;\,A] &= \sum_{k:k\cdot 2^{-n}\leq c} E[M_{k\cdot 2^{-n}}\,;\,A\cap\{T_n = k\cdot 2^{-n}\}] \\ &= \sum_{k:k\cdot 2^{-n}\leq c} E[M_c\,;\,A\cap\{T_n = k\cdot 2^{-n}\}] \\ &= E[M_c\,;\,A] \qquad \text{for any } A\in\mathcal{F}_{T_n}, \end{aligned} \tag{3.2.3}$$

and similarly

$$E[M_{S_n}\,;\,A] \quad = \quad E[M_c\,;\,A] \qquad \text{for any } A \in \mathcal{F}_{S_n}. \tag{3.2.4}$$

In (3.2.3) we have used that (M_t) is an $(\mathcal{F}_t)$ martingale, and $A \cap \{T_n = k \cdot 2^{-n}\} \in \mathcal{F}_{k \cdot 2^{-n}}$. A set $A \in \mathcal{F}_S$ is contained both in $\mathcal{F}_{T_n}$ and $\mathcal{F}_{S_n}$. This by (3.2.3) and (3.2.4),

$$E[M_{T_n}\,;\,A] \;=\; E[M_{S_n}\,;\,A] \qquad \text{for any } n \in \mathbb{N} \text{ and } A \in \mathcal{F}_S. \tag{3.2.5}$$

As $n \to \infty$, $M_{T_n} \to M_T$ and $M_{S_n} \to M_S$ almost surely by continuity. It remains to show that the expectations in (3.2.5) also converge. To this end note that by (3.2.3) and (3.2.4),

$$M_{T_n} \;=\; E[M_c \,|\, \mathcal{F}_{T_n}] \quad \text{and} \quad M_{S_n} \;=\; E[M_c \,|\, \mathcal{F}_{S_n}] \qquad P\text{-almost surely.}$$

We will prove in Section **??** that any family of conditional expectations of a given random variable w.r.t. different σ-algebras is uniformly integrable, and that for uniformly integrable random variables a generalized Dominated Convergence Theorem holds, cf. **??**. Therefore, we finally obtain

$$\begin{aligned} E[M_T\,;\,A] &= E[\lim M_{T_n}\,;\,A] \;=\; \lim E[M_{T_n}\,;\,A] \\ &= \lim E[M_{S_n}\,;\,A] \;=\; E[\lim M_{S_n}\,;\,A] \;=\; E[M_S\,;\,A], \end{aligned}$$

completing the proof of the theorem.

□

Remark (Measurability and completion). In general, the random variable M_S is not necessarily $\mathcal{F}_S$-measurable. However, we have shown in the proof that M_S always has an $\mathcal{F}_S$-measurable modification $\widetilde{M}_S$. If the filtration contains all measure zero sets, then this implies that M_S itself is $\mathcal{F}_S$-measurable and hence a version of $E[M_T \mid \mathcal{F}_S]$.

Ruin probabilities and passage times revisited

Similarly as for Random Walks, the Optional Sampling Theorem can be applied to compute distributions of passage times and hitting probabilities for Brownian motion. For a one-dimensional Brownian motion (B_t) starting at 0 and $a, b > 0$ let $T = \inf\{t \geq 0 \,:\, B_t \notin (-b, a)\}$ and $T_a = \inf\{t \geq 0 \,:\, B_t = a\}$ denote the first exit time to the point

a respectively. In Section 1.5 we have computed the distribution of T_a by the reflection principle. This and other results can be recovered by applying optional stopping to the basic martingales of Brownian motion. The advantage of this approach is that it can be carried over to other diffusion processes.

Exercise. Prove by optional stopping:

(1). *Ruin probabilities:* $P[B_T = a] = b/(a+b), \quad P[B_T = -b] = a/(a+b)$,

(2). *Mean exit time:* $E[T] = a \cdot b$ and $E[T_a] = \infty$,

(3). *Laplace transform of passage times:* $E[\exp(-sT_a)] = \exp(-a\sqrt{2s})$ for any $s > 0$.

Conclude that the distribution of T_a on $(0, \infty)$ is absolutely continuous with density

$$f_{T_a}(t) \;=\; a \cdot (2\pi t^3)^{-1/2} \cdot \exp(-a^2/2t).$$

Exit distributions and Dirichlet problem

Applying optional stopping to harmonic functions of a multidimensional Brownian motion yields a generalization of the mean value property and a stochastic representation for solutions of the Dirichlet problem.

Suppose that $h \in C^2(\mathbb{R}^d)$ is a harmonic function and that $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion starting at x w.r.t. the probability measure P_x. Assuming that

$$E_x[h(B_t)] \;<\; \infty \qquad \text{for any } t \geq 0,$$

the mean value property for harmonic functions implies that $h(B_t)$ is a martingale under P_x w.r.t. the right continuous filtration $(\mathcal{F}_t)$, cf. Theorem 3.4. Since the filtration is right continuous, the first hitting time $T = \inf\{t \geq 0 \mid B_t \in \mathbb{R}^d \setminus D\}$ of the complement of an open Domain $D \subseteq \mathbb{R}^d$ is a stopping time w.r.t. $(\mathcal{F}_t)$. Therefore, by Theorem 3.5 and the remark below, we obtain

$$E_x[h(B_{T\wedge n})] \;=\; E_x[h(B_0)] \;=\; h(x) \qquad \text{for any } n \in \mathbb{N}. \tag{3.2.6}$$

Now let us assume in addition that the domain D is bounded or, more generally, that the function h is bounded on D. Then the sequence of random variables $h(B_{T\wedge n})\,(n \in \mathbb{N})$ is uniformly bounded because $B_{T\wedge n}$ takes values in the closure $\overline{D}$ for any $n \in \mathbb{N}$. Applying the Dominated Convergence Theorem to (3.2.6), we obtain the integral representation

$$h(x) \;=\; E_x[h(B_T)] \;=\; \int_{\partial D} h(y)\mu_x(dy) \tag{3.2.7}$$

where $\mu_x = P_x \circ B_T^{-1}$ denotes the exit distribution from D for Brownian motion starting at x.

More generally, one can prove:

Theorem 3.6 (Stochastic representation for solutions of the Dirichlet problem). *Suppose that D is a bounded open subset of $\mathbb{R}^d$, f is a continuous function on the boundary ∂D, and $h \in C^2(D) \cap C(\overline{D})$ is a solution of the Dirichlet problem*

$$\begin{aligned} \Delta h(x) &= 0 \qquad \text{for } x \in D, \\ h(x) &= f(x) \qquad \text{for } x \in \partial D. \end{aligned} \tag{3.2.8}$$

Then

$$h(x) \;=\; E_x[f(B_T)] \qquad \text{for any } x \in \overline{D}. \tag{3.2.9}$$

We have already proven (3.2.9) under the additional assumption that h can be extended to a harmonic function on $\mathbb{R}^d$ satisfying $E_x[h(B_T)] < \infty$ for all $t \geq 0$. The proof in the general case requires localization techniques, and will be postponed to Section **??** below.

The representations (3.2.7) and (3.2.9) have several important aspects and applications:

Generalized mean value property for harmonic functions For any bounded domain $D \subseteq \mathbb{R}^d$ and any $x \in D$, $h(x)$ is the average of the boundary values of h on ∂D w.r.t. the measure μ_x.

Monte Carlo Methods The stochastic representation (3.2.9) can be used as the basis of a Monte Carlo method for computing the harmonic function $h(x)$ numerically by simulating N sample paths of Brownian motion starting at x and estimating the expected value by the corresponding empirical average. Although in many cases classical numerical methods are more efficient, the Monte Carlo method can be useful in high dimensional cases. Furthermore, it carries over to far more general situations.

Computation of exit probabilities conversely, if the Dirichlet problem (3.2.8) has a unique solution h, then computation of h (for example by standard numerical methods) enables us to obtain the expected values in (3.2.8). In particular, the exit probability $h(x) = P_x[B_T \in A]$ on a subset $A \subseteq \partial D$ is informally given as the solution of the Dirichlet problem

$$\Delta h \;=\; 0 \quad \text{on } D, \qquad h = I_a \quad \text{on } \partial D.$$

This can be made rigorous under regularity assumptions. The full exit distribution is the *harmonic measure*, i.e., the probability measure μ_x such that the representation (3.2.7) holds for any function $h \in C^2(D) \cap C(\overline{D})$ with $\Delta h = 0$ on D. For simple domains, the harmonic measure can be computed explicitly.

Example (**Exit distribution for balls**)**.** The exit distribution from the unit ball $D = \{y \in \mathbb{R}^d \;:\; |y| < 1\}$ for Brownian motion stopping at a point $x \in \mathbb{R}^d$ with $|x| < 1$ is given by

$$\mu_x(dy) \;=\; \frac{1 - |x^2|}{|x-y|^2}\nu(dy)$$

where ν denotes the normalized surface measure on the unit sphere $S^{d-1} = \{y \in \mathbb{R}^d \;:\; |y| = 1\}$.

In fact, it is well known and can be verified by explicit computation, that for any $f \in C(S^{d-1})$, the function

$$h(x) \;=\; \int f(y)\, \mu_x(dy)$$

is harmonic on D with boundary values $\lim\limits_{x \to y} h(x) = f(y)$ for any $y \in S^{d-1}$. Hence by (3.2.9)

$$E_x[f(B_t)] \;=\; \int f(y)\, \mu_x(dy)$$

holds for any $f \in C(S^{d-1})$, and thus by a standard approximation argument, for any indicator function of a measurable subset of S^{d-1}.

3.3 Maximal inequalities and the Law of the Iterated Logarithm

The extension of Doob's maximal inequality to the continuous time case is straightforward: As a first application, we give a proof for the upper bound in the law of the iterated logarithm.

Maximal inequalities in continuous time

Theorem 3.7 (Doob's L^p inequality in continuous time). *Suppose that $(M_t)_{t\in[0,\infty)}$ is a martingale with almost surely right continuous sample paths $t \mapsto M_t(\omega)$. Then the following estimates hold for any $a \in [0,\infty)$, $p \in [1,\infty)$, $q \in (1,\infty]$ with $\frac{1}{p}+\frac{1}{q}=1$, and $c > 0$:*

(1). $P\left[\sup_{t\in[0,a]} |M_t| \geq c\right] \quad \leq \quad c^{-p} \cdot E[|M_a|^p],$

(2). $\left\|\sup_{t\in[0,a]} |M_t|\right\|_{L^p} \quad \leq \quad q \cdot \|M_a\|_{L^p}.$

Remark. The same estimates hold for non-negative submartingales.

Proof. Let (τ_n) denote an increasing sequence of partitions of the interval $[0,a]$ such that the mesh size $|\tau_n|$ goes to 0 as $n \to \infty$. By Corollary 2.13 applied to the discrete time martingale $(M_t)_{t\in\tau_n}$, we obtain

$$P\left[\max_{t\in\tau_n} |M_t| \geq c\right] \leq E[|M_a|^p]/c^p \qquad \text{for any } n \in \mathbb{N}.$$

Moreover, as $n \to \infty$,

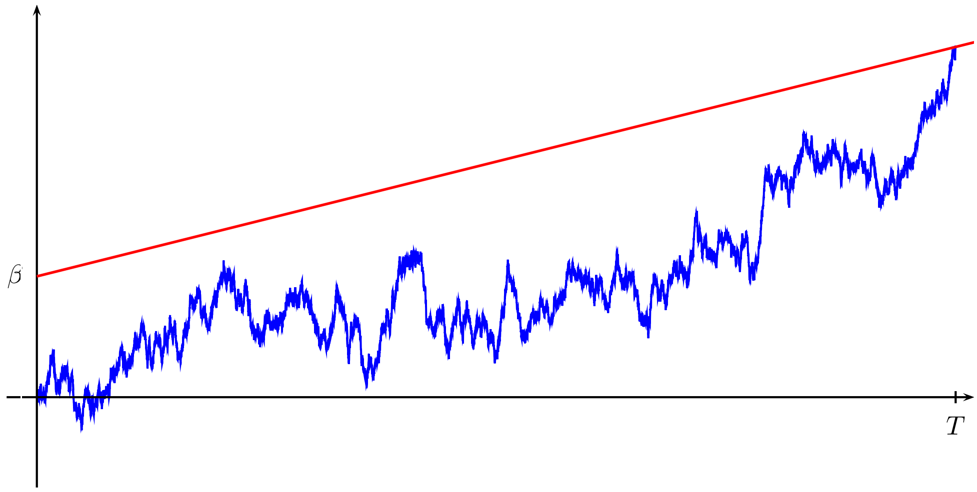
$$\max_{t\in\tau_n} |M_t| \quad \nearrow \quad \sup_{t\in[0,a]} |M_t| \qquad \text{almost surely}$$

by right continuity of the sample paths. Hence

$$P\left[\sup_{t\in[0,a]}|M_t| > c\right] = P\left[\bigcup_n \left\{\max_{t\in\tau_n}|M_t| > c\right\}\right]$$
$$= \lim_{n\to\infty} P\left[\max_{t\in\tau_n}|M_t| > c\right] \leq E[|M_a|^p]/c^p.$$

The first assertion now follows by replacing c by $c-\varepsilon$ and letting ε tend to 0. The second assertion follows similarly from Theorem 2.15. □

As a first application of the maximal inequality to Brownian motion, we derive an upper bound for the probability that the graph of one-dimensional Brownian motion passes a line in $\mathbb{R}^2$:



Lemma 3.8 (Passage probabilities for lines). *For a one-dimensional Brownian motion* (B_t) *starting at* 0 *we have*

$$P[B_t \geq \beta + \alpha t/2 \quad \textit{for some } t \geq 0] \leq \exp(-\alpha\beta) \qquad \textit{for any } \alpha,\beta > 0.$$

Proof. Applying the maximal inequality to the exponential martingale

$$M_t^\alpha = \exp(\alpha B_t - \alpha^2 t/2)$$

yields

$$\begin{aligned} P[B_t \geq \beta + \alpha t/2 \quad \text{for some } t \in [0,a]] &= P\left[\sup_{t\in[0,a]} (B_t - \alpha t/2) \geq \beta\right] \\ &= P\left[\sup_{t\in[0,a]} M_t^\alpha \geq \exp(\alpha\beta)\right] \\ &\leq \exp(-\alpha\beta)\cdot E[M_a^\alpha] \;=\; \exp(-\alpha\beta) \end{aligned}$$

for any $a > 0$. The assertion follows in the limit as $a \to \infty$. □

With slightly more effort, it is possible to compute the passage probability and the distribution of the first passage time of a line explicitly, cf. **??** below.

Application to LIL

A remarkable consequence of Lemma 3.8 is a simplified proof for the upper bound half of the Law of the Iterated Logarithm:

Theorem 3.9 (LIL, upper bound). *For a one-dimensional Brownian motion* (B_t) *starting at* 0*,*

$$\limsup_{t\searrow 0} \frac{B_t}{\sqrt{2t\log\log t^{-1}}} \quad\leq\quad +1 \qquad P\text{-almost surely.} \tag{3.3.1}$$

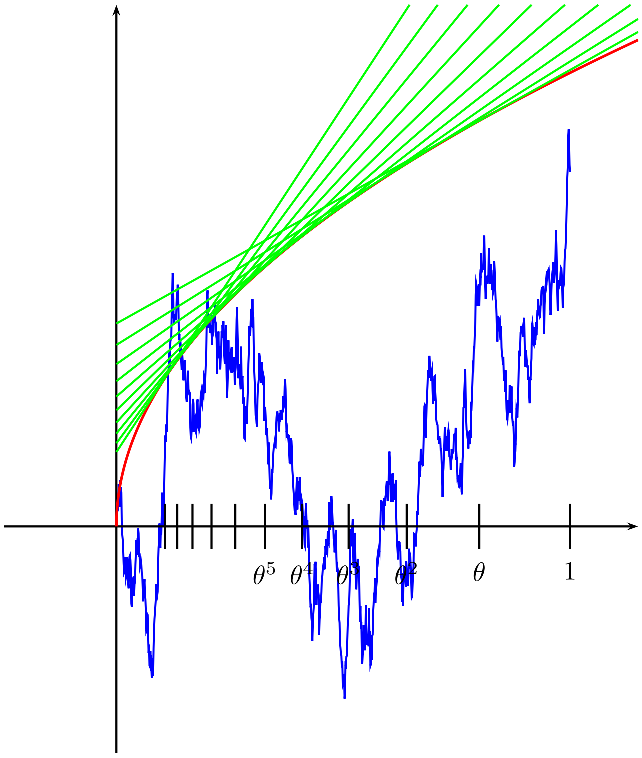
Proof. Let $\delta > 0$. We would like to show that almost surely,

$$B_t \;\leq\; (1+\delta)h(t) \qquad \text{for sufficiently small } t > 0,$$

where $h(t) := \sqrt{2t\log\log t^{-1}}$. Fix $\theta \in (0,1)$. The idea is to approximate the function $h(t)$ by affine functions

$$l_n(t) \;=\; \beta_n + \alpha_n t/2$$

on each of the intervals $[\theta^b, \theta^{n-1}]$, and to apply the upper bounds for the passage probabilities from the lemma.

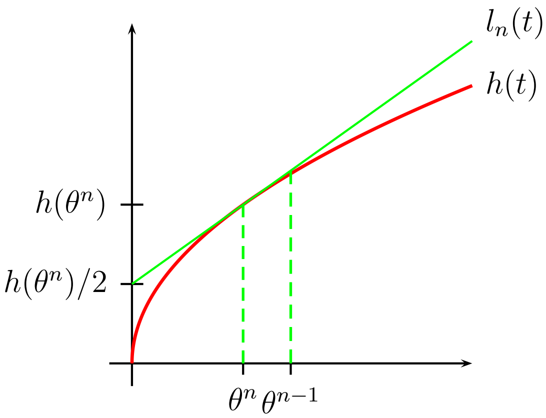


We choose α_n and β_n in a such way that $l_n(\theta^n) = h(\theta^n)$ and $l_n(0) = h(\theta^n)/2$, i.e.,

$$\beta_n = h(\theta^n)/2 \quad \text{and} \quad \alpha_n = h(\theta^n)/\theta^n.$$

For this choice we have $l_n(\theta^n) \geq \theta \cdot l_n(\theta^{n-1})$, and hence

$$\begin{aligned} l_n(t) \leq l_n(\theta^{n-1}) &\leq \frac{l_n(\theta^n)}{\theta} \\ = \frac{h(\theta^n)}{\theta} &\leq \frac{h(t)}{\theta} \qquad \text{for any } t \in [\theta^n, \theta^{n-1}]. \end{aligned} \tag{3.3.2}$$



We now want to apply the Borel-Cantelli lemma to show that with probability one, $B_t \leq (1+\delta)l_n(t)$ for large n. By Lemma 3.8,

$$\begin{aligned} P[B_t \geq (1+\delta)l_n(t) \quad \text{for some } t \geq 0] &\leq \exp(-\alpha_n\beta_n \cdot (1+\delta)^2) \\ &= \exp\left(-\frac{h(\theta^n)^2}{2\theta^n} \cdot (1+\delta)^2\right). \end{aligned}$$

Choosing $h(t) = \sqrt{2t \log\log t^{-1}}$, the right hand side is equal to a constant multiple of $n^{-(1+\delta)^2}$, which is a summable sequence. Note that we do not have to know the precise form of $h(t)$ in advance to carry out the proof – we just choose $h(t)$ in such a way that the probabilities become summable!
Now, by Borel-Cantelli, for P-almost every ω there exists $N(\omega) \in \mathbb{N}$ such that

$$B_t(\omega) \quad \leq \quad (1+\delta)l_n(t) \qquad \text{for any } t \in [0,1] \text{ and } n \geq N(\omega). \tag{3.3.3}$$

By (3.3.2), the right hand side of (3.3.3) is dominated by $(1+\delta)h(t)/\theta$ for $t \in [\theta^n, \theta^{n-1}]$. Hence

$$B_t \leq \frac{1+\delta}{\theta}h(t) \qquad \text{for any } t \in \bigcup_{n\geq N} [\theta^n, \theta^{n-1}],$$

i.e., for any $t \in (0, \theta^{N-1})$, and therefore,

$$\limsup_{t\searrow 0} \frac{B_t}{h(t)} \leq \frac{1+\delta}{\theta} \qquad P\text{-almost surely.}$$

The assertion then follows in the limit as $\theta \nearrow 1$ and $\delta \searrow 0$. □

Since $(-B_t)$ is again a Brownian motion starting at 0, the upper bound (3.3.1) also implies

$$\liminf_{t\searrow 0} \frac{B_t}{\sqrt{2t\log\log t^{-1}}} \geq -1 \qquad P\text{-almost surely.} \tag{3.3.4}$$

The converse bounds are actually easier to prove since we can use the independence of the increments and apply the second Borel-Cantelli Lemma. We only mention the key steps and leave the details as an exercise:

Example (Complete proof of LIL). Prove the Law of the Iterated Logarithm:

$$\limsup_{t\searrow 0} \frac{B_t}{h(t)} = +1 \qquad \text{and} \qquad \liminf_{t\searrow 0} \frac{B_t}{h(t)} = -1$$

where $h(t) = \sqrt{2t\log\log t^{-1}}$. Proceed in the following way:

(1). Let $\theta \in (0,1)$ and consider the increments $Z_n = B_{\theta^n} - B_{\theta^{n+1}}, n \in \mathbb{N}$. Show that if $\varepsilon > 0$, then

$$P[Z_n > (1-\varepsilon)h(\theta^n) \quad \text{infinitely often}] = 1.$$

(Hint: $\int_x^\infty \exp(-z^2/2)dz \leq x^{-1}\exp(-x^2/2)$)

(2). Conclude that by (3.3.4),

$$\limsup_{t\searrow 0} \frac{B_t}{h(t)} \geq 1-\varepsilon \qquad P\text{-almost surely for any } \varepsilon > 0,$$

and complete the proof of the LIL by deriving the lower bounds

$$\limsup_{t\searrow 0} \frac{B_t}{h(t)} \geq 1 \quad \text{and} \quad \liminf_{t\searrow 0} \frac{B_t}{h(t)} \leq -1 \qquad P\text{-almost surely.} \tag{3.3.5}$$

Chapter 4

Martingale Convergence Theorems

The strength of martingale theory is partially due to powerful general convergence theorems that hold for martingales, sub- and supermartingales. In this chapter, we study convergence theorems with different types of convergence including almost sure, L^2 and L^1 convergence, and consider first applications.

At first, we will again focus on discrete-parameter martingales – the results can later be easily extended to continuous martingales.

4.1 Convergence in L^2

Already when proving the Law of Large Numbers, L^2 convergence is much easier to show than, for example, almost sure convergence. The situation is similar for martingales: A necessary and sufficient condition for convergence in the Hilbert space $L^2(\Omega, \mathcal{A}, P)$ can be obtained by elementary methods.

Martingales in L^2

Consider a discrete-parameter martingale $(M_n)_{n\geq 0}$ w.r.t. a filtration $(\mathcal{F}_n)$ on a probability space $(\Omega, \mathcal{A}, P)$. Throughout this section we assume:

Assumption (Square integrability). $E[M_n^2] < \infty$ *for any* $n \geq 0$.

We start with an important remark

Lemma 4.1. *The increments* $Y_n = M_n - M_{n-1}$ *of a square-integrable martingale are centered and orthogonal in* $L^2(\Omega, \mathcal{A}, P)$ *(i.e. uncorrelated).*

Proof. By definition of a martingale, $E[Y_n \,|\, \mathcal{F}_{n-1}] = 0$ for any $n \geq 0$. Hence $E[Y_n] = 0$ and $E[Y_m Y_n] = E[Y_m \cdot E[Y_n \mid \mathcal{F}_{n-1}]] = 0$ for $0 \leq m < n$. □

Since the increments are also orthogonal to M_0 by an analogue argument, a square integrable martingale sequence consists of partial sums of a sequence of uncorrelated random variables:

$$M_n \;=\; M_0 + \sum_{k=1}^{n} Y_k \qquad \text{for any } n \geq 0.$$

The convergence theorem

The central result of this section shows that an L^2-bounded martingale (M_n) can **always** be extended to $n \in \{0, 1, 2, \ldots\} \cup \{\infty\}$:

Theorem 4.2 (**L^2** Martingale Convergence Theorem)**.** *The martingale sequence* (M_n) *converges in* $L^2(\Omega, \mathcal{A}, P)$ *as* $n \to \infty$ *if and only if it is bounded in* L^2 *in the sense that*

$$\sup_{n \geq 0} E[M_n^2] < \infty. \tag{4.1.1}$$

In this case, the representation

$$M_n \;=\; E[M_\infty \mid \mathcal{F}_n]$$

holds almost surely for any $n \geq 0$*, where* M_∞ *denotes the limit of* M_n *in* $L^2(\Omega, \mathcal{A}, P)$*.*

We will prove in the next section that (M_n) does also converge almost surely to M_∞. An analogue result to Theorem 4.2 holds with L^2 replaced by L^p for any $p \in (1, \infty)$ but bot for $p = 1$, cf. Section **??** below.

Proof. (1). Let us first note that

$$E[(M_n - M_m)^2] \;=\; E[M_n^2] - E[M_m^2] \qquad \text{for } 0 \leq m \leq n. \tag{4.1.2}$$

In fact,

$$\begin{aligned} E[M_n^2] - E[M_m^2] &= E[(M_n - M_m)(M_n + M_m)] \\ &= E[(M_n - M_m)^2] + 2E[M_m \cdot (M_n - M_m)], \end{aligned}$$

and the last term vanishes since the increment $M_n - M_m$ is orthogonal to M_m in L^2.

(2). To prove that (4.1.1) is sufficient for L^2 convergence, note that the sequence $(E[M_n^2])_{n\geq 0}$ is increasing by (4.1.2). If (4.1.1) holds then this sequence is bounded, and hence a Cauchy sequence. Therefore, by (4.1.2), (M_n) is a Cauchy sequence in L^2. Convergence now follows by completeness of $L^2(\Omega, \mathcal{A}, P)$.

(3). Conversely, if (M_n) converges in L^2 to a limit M_∞, then the L^2 norms are bounded. Moreover, by Jensen's inequality,

$$E[M_n \mid \mathcal{F}_k] \;\longrightarrow\; E[M_\infty \mid \mathcal{F}_k] \qquad \text{in } L^2(\Omega, \mathcal{A}, P) \text{ as } n \to \infty$$

for each fixed $k \geq 0$. As (M_n) is a martingale, we have $E[M_n \mid \mathcal{F}_k] = M_k$ for $n \geq k$, and hence

$$M_k \;=\; E[M_\infty \mathcal{F}_k] \qquad P\text{-almost surely.}$$

□

Remark (Functional analytic interpretation of L^2 convergence theorem). The assertion of the L^2 martingale convergence theorem can be rephrased as a purely functional analytic statement:

An infinite sum $\sum_{k=1}^{\infty} Y_k$ of orthogonal vectors Y_k in the Hilbert space $L^2(\Omega, \mathcal{A}, P)$ is convergent if and only if the sequence of partial sums $\sum_{k=1}^{n} Y_k$ is bounded.

How can boundedness in L^2 be verified for martingales? Writing the martingale (M_n) as the sequence of partial sums of its increments $Y_n = M_n - M_{n-1}$, we have

$$E[M_n^2] \;=\; \left(M_0 + \sum_{k=1}^{n} Y_k, M_0 + \sum_{k=1}^{n} Y_k\right)_{L^2} \;=\; E[M_0^2] + \sum_{k=1}^{n} E[Y_k^2]$$

by orthogonality of the increments and M_0. Hence

$$\sup_{n\geq 0} E[M_n^2] \;=\; E[M_0^2] + \sum_{k=1}^{\infty} E[Y_k^2].$$

Alternatively, we have $E[M_n^2] = E[M_0^2] + E[\langle M\rangle_n]$. Hence by monotone convergence

$$\sup_{n\geq 0} E[M_n^2] \;=\; E[M_0^2] + E[\langle M\rangle_\infty]$$

where $\langle M\rangle_\infty \;=\; \sup\langle M\rangle_n$.

Summability of sequences with random signs

As a first application we study the convergence of series with coefficients with random signs. In an introductory analysis course it is shown as an application of the integral and Leibniz criterion for convergence of series that

$$\sum_{n=1}^{\infty} n^{-\alpha} \text{ converges} \quad \Longleftrightarrow \quad \alpha > 1 \quad \text{, whereas}$$
$$\sum_{n=1}^{\infty} (-1)^n n^{-\alpha} \text{ converges} \quad \Longleftrightarrow \quad \alpha > 0.$$

Therefore, it seems interesting to see what happens if the signs are chosen randomly. The L^2 martingale convergence theorem yields:

Corollary 4.3. *Let (a_n) be a real sequence. If (ε_n) is a sequence of independent random variables on $(\Omega, \mathcal{A}, P)$ with $P[\varepsilon_n = +1] = P[\varepsilon_n = -1] = 1/2$, then*

$$\sum_{n=1}^{\infty} \varepsilon_n a_n \quad \textit{converges in } L^2(\Omega, \mathcal{A}, P) \quad \Longleftrightarrow \quad \sum_{n=1}^{\infty} < \infty.$$

Proof. The sequence $M_n = \sum_{k=1}^{n} \varepsilon_k a_k$ of partial sums is a martingale with

$$\sup_{n\geq 0} E[M_n^2] \;=\; \sum_{k=1}^{\infty} E[\varepsilon_k^2 a_k^2] \;=\; \sum_{k=1}^{\infty} a_k^2.$$

□

Example. The series $\sum_{n=1}^{\infty} \varepsilon_n \cdot n^{-\alpha}$ converges in L^2 if and only if $\alpha > \frac{1}{2}$.

Remark (Almost sure asymptotics). By the Supermartingale Convergence Theorem (cf. Theorem 4.5 below), the series $\sum \varepsilon_n a_n$ also converges almost surely if $\sum a_n^2 < \infty$. On the other hand, if $\sum a_n^2 \;=\; \infty$ then the series of partial sums has almost surely unbounded oscillations:

Exercise. Suppose that $\sum a_n = \infty$, and let $M_n = \sum_{k=1}^{n} \varepsilon_k a_k$, ε_k i.i.d. with $P[\varepsilon_k = \pm 1] = \frac{1}{2}$.

(1). Compute the conditional variance process $\langle M \rangle_n$.

(2). For $c > 0$ let $T_c = \inf\{n \geq 0 \ : \ |M_n| \geq c\}$. Apply the Optional Stopping Theorem to the martingale in the Doob decomposition of (M_n^2), and conclude that $P[T_c = \infty] = 0$.

(3). Prove that (M_n) has almost surely unbounded oscillations.

$\mathbf{L^2}$ convergence in continuous time

The L^2 convergence theorem directly extends to the continuous-parameter case.

Theorem 4.4. *Let $a \in (0, \infty]$. If $(M_t)_{t \in [0,a)}$ is a martingale w.r.t. a filtration $(\mathcal{F}_t)_{t \in [0,a)}$ such that*

$$\sup_{t \in [0,a)} E[M_t^2] \ < \ \infty$$

then $M_a = \lim_{t \nearrow a} M_t$ exists in $L^2(\Omega, \mathcal{A}, P)$ and $(M_t)_{t \in [0,a]}$ is again a square-integrable martingale.

Proof. Choose any increasing subsequence $t_n \in [0, a)$ such that $t_n \to a$. Then (M_{t_n}) is a L^2-bounded discrete-parameter martingale, hence the limit $M_a = \lim M_{t_n}$ exists in L^2, and

$$M_{t_n} \ = \ E[M_a \mid \mathcal{F}_{t_n}] \qquad \text{for any } n \in \mathbb{N}. \tag{4.1.3}$$

For any $t \in [0, a)$, there exists $n \in \mathbb{N}$ with $t_n \in (t, a)$. Hence

$$M_t \ = \ E[M_{t_n} \mid \mathcal{F}_t] \ = \ E[M_a \mid \mathcal{F}_t]$$

by (4.1.3) and the tower property. In particular, $(M_t)_{t \in [0,a]}$ is a square-integrable martingale. By orthogonality of the increments,

$$E[(M_a - M_{t_n})^2] \ = \ E[(M_a - M_t)^2] + E[(M_t - M_{t_n})^2] \ \geq \ E[(M_a - M_t)^2]$$

whenever $t_n \leq t \leq a$. Since $M_{t_n} \to M_a$ in L^2, we obtain

$$\lim_{t \nearrow a} E[(M_a - M_t)^2] \;=\; 0.$$

□

Remark. (1). Note that in the proof it is enough to consider one fixed sequence $t_n \nearrow a$.

(2). To obtain almost sure convergence, an additional regularity condition on the sample paths is required, e.g. right-continuity, cf. below. This assumption is not needed for L^2 convergence.

4.2 Almost sure convergence of supermartingales

Let $(Z_n)_{n \geq 0}$ be a discrete-parameter supermartingale w.r.t. a filtration $(\mathcal{F}_n)_{n \geq 0}$ on a probability space $(\Omega, \mathcal{A}, P)$. The following theorem yields a stochastic counterpart to the fact that any lower bounded decreasing sequence of reals converges to a finite limit:

Theorem 4.5 (**Supermartingale Convergence Theorem, Doob**). *If* $\sup\limits_{n \geq 0} E[Z_n^-] < \infty$ *then* (Z_n) *converges almost surely to an integrable random variable* $Z_\infty \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. *In particular, supermartingales that are uniformly bounded from above converge almost surely to an integrable random variable.*

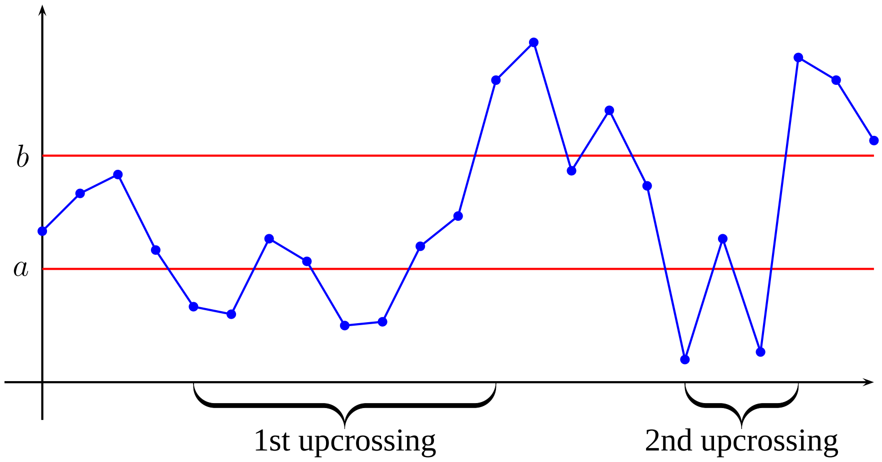
Remark.

(1). Although the limit is integrable, L^1 convergence does **not** hold in general, cf. Section **??** below.

(2). The condition $\sup E[Z_n^-] < \infty$ holds if and only if (Z_n) is bounded in L^1. Indeed, as $E[Z_n^+] < \infty$ by our definition of a supermartingale, we have

$$E[\,|Z_n|\,] \;=\; E[Z_n] + 2E[Z_n^-] \;\leq\; E[Z_0] + 2E[Z_n^-] \qquad \text{for any } n \geq 0.$$

For proving the Supermartingale Convergence Theorem, we introduce the number $U^{(a,b)}(\omega)$ of upcrossings of an interval (a, b) by the sequence $Z_n(\omega)$, cf. below for the exact definition.



Note that if $U^{(a,b)}(\omega)$ is finite for any non-empty bounded interval $[a, b]$ then $\limsup Z_n(\omega)$ and $\liminf Z_n(\omega)$ coincide, i.e., the sequence $(Z_n(\omega))$ converges. Therefore, to show almost sure convergence of (Z_n), we derive an upper bound for $U^{(a,b)}$. We first prove this key estimate and then complete the proof of the theorem.

Doob's upcrossing inequality

For $n \in \mathbb{N}$ and $a, b \in \mathbb{R}$ with $a < b$ we define the number $U_n^{(a,b)}$ of upcrossings of the interval $[a, b]$ before time n by

$$U_n^{(a,b)} = \max\big\{k \geq 0 \,|\, \exists\, 0 \leq s_1 < t_1 < s_2 < t_2 \ldots < s_k < t_k \leq n \; : \; Z_{s_i}(\omega) \leq a, Z_{t_i}(\omega) \geq b\big\}.$$

Lemma 4.6 (Doob). *If (Z_n) is a supermartingale then*

$$(b-a) \cdot E[U_n^{(a,b)}] \;\leq\; E[(Z_n - a)^-] \qquad \textit{for any } a < b \textit{ and } n \geq 0.$$

Proof. We may assume $E[Z_n^-] < \infty$ since otherwise there is nothing to prove. The key idea is to set up a previsible gambling strategy that increases our capital by $(b - a)$ for each completed upcrossing. Since the net gain with this strategy should again be a supermartingale this yields an upper bound for the average number of upcrossings. Here is the strategy:

repeat

- Wait until $Z_k \leq a$.
- Then play unit stakes until $Z_k \geq b$.
-

The stake C_k in round k is

$$C_1 \;=\; \begin{cases} 1 & \text{if } Z_0 \leq a \\ 0 & \text{otherwise} \end{cases}$$

and

$$C_k \;=\; \begin{cases} 1 & \text{if } (C_{k-1} = 1 \text{ and } Z_{k-1} \leq b) \text{ or } (C_{k-1} = 0 \text{ and } Z_{k-1} \leq a) \\ 0 & \text{otherwise} \end{cases}.$$

Clearly, (C_k) is a previsible, bounded and non-negative sequence of random variables. Moreover, $C_k \cdot (Z_k - Z_{k-1})$ is integrable for any $k \leq n$, because C_k is bounded and

$$E\big[\,|Z_k|\,\big] \;=\; 2E[Z_k^+] - E[Z_k] \;\leq\; 2E[Z_k^+] - E[Z_n] \leq 2E[Z_k^+] - E[Z_n^-]$$

for $k \leq n$. Therefore, by Theorem 2.6 and the remark below, the process

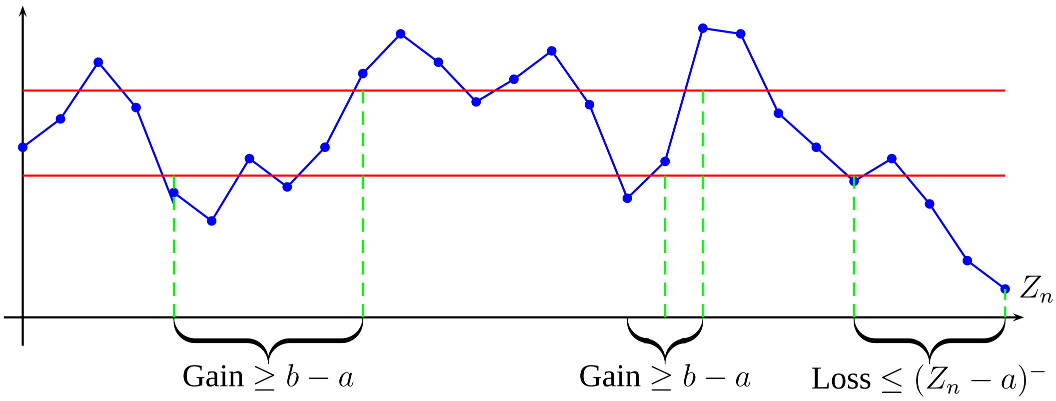
$$(C_\bullet Z)_k \;=\; \sum_{i=1}^{k} C_i \cdot (Z_i - Z_{i-1}), \qquad 0 \leq k \leq n,$$

is again a supermartingale.

Clearly, the value of the process $C_\bullet Z$ increases by at least $(b-a)$ units during each completed upcrossing. Between upcrossing periods, the value of $(C_\bullet Z)_k$ is constant. Finally, if the final time n is contained in an upcrossing period, then the process can decrease by at most $(Z_n - a)^-$ units during that last period (since Z_k might decrease before the next upcrossing is completed). Therefore, we have

$$(C_\bullet Z)_n \quad \geq \quad (b-a) \cdot U_n^{(a,b)} - (Z_n - a)^-, \qquad \text{i.e.,}$$

$$(b-a) \cdot U_n^{(a,b)} \quad \geq \quad (C_\bullet Z)_n + (Z_n - a)^-.$$



Since $C_\bullet Z$ is a supermartingale with initial value 0, we obtain the upper bound

$$(b-a)E[U_n^{(a,b)}] \;\leq\; E[(C_\bullet Z)_n] + E[(Z_n - a)^-] \;\leq\; E[(Z_n - a)^-].$$

□

Proof of Doob's Convergence Theorem

We can now complete the proof of Theorem 4.5.

Proof. Let

$$U^{(a,b)} \;=\; \sup_{n\in\mathbb{N}} U_n^{(a,b)}$$

denote the total number of upcrossings of the supermartingale (Z_n) over an interval (a, b) with $-\infty < a < b < \infty$. By the upcrossing inequality and monotone convergence,

$$E[U^{(a,b)}] \;\lim_{n\to\infty} E[U_n^{(a,b)}] \;\leq\; \frac{1}{b-a} \cdot \sup_{n\in\mathbb{N}} E[(Z_n - a)^-]. \tag{4.2.1}$$

Assuming $\sup E[Z_n^-] < \infty$, the right hand side of (4.2.1) is finite since $(Z_n - a)^- \leq |a| + Z_n^-$. Therefore,

$$U^{(a,b)} \;<\; \infty \qquad P\text{-almost surely,}$$

and hence the event

$$\{\liminf Z_n \;\neq\; \limsup Z_n\} \;=\; \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} \{U^{(a,b)} = \infty\}$$

has probability zero. This proves almost sure convergence.

It remains to show that the almost sure limit $Z_\infty = \lim Z_n$ is an integrable random variable (in particular, it is finite almost surely). This holds true as, by the remark below Theorem 4.5, $\sup E[Z_n^-] < \infty$ implies that (Z_n) is bounded in L^1, and therefore

$$E[\,|Z_\infty|\,] \;=\; E[\lim\,|Z_n|\,] \;\le\; \liminf E[\,|Z_n|\,] \;<\; \infty$$

by Fatou's lemma. □

Examples and first applications

We now consider a few prototypic applications of the almost sure convergence theorem:

Example (1. Sums of i.i.d. random variables). Consider a Random Walk

$$S_n = \sum_{i=1}^n \eta_i$$

on $\mathbb{R}$ with centered and bounded increments

$$\eta_i \;\text{ i.i.d. with } |\eta_i| \le c \text{ and } E[\eta_i] = 0, \quad c \in \mathbb{R}.$$

Suppose that $P[\eta_i \neq 0] > 0$. Then there exists $\varepsilon > 0$ such that $P[|\eta_i| \ge \varepsilon] > 0$. As the increments are i.i.d., the event $\{|\eta_i| \ge \varepsilon\}$ occurs infinitely often with probability one. Therefore, almost surely the martingale (S_n) does not converge as $n \to \infty$.

Now let $a \in \mathbb{R}$. We consider the first hitting time

$$T_a \;=\; \inf\{t \ge 0 \,:\, S_n \ge a\}$$

of the interval $[a, \infty)$. By the Optional Stopping Theorem, the stopped Random Walk $(S_{T_a \wedge n})_{n \ge 0}$ is again a martingale. Moreover, as $S_k < a$ for any $k < T_a$ and the increments are bounded by c, we obtain the upper bound

$$S_{T_a \wedge n} \;<\; a + c \qquad \text{for any } n \in \mathbb{N}.$$

Therefore, the stopped Random Walk converges almost surely by the Supermartingale Convergence Theorem. As (S_n) does not converge, we can conclude that

$$P[T_a < \infty] \;=\; 1 \qquad \text{for any } a > 0, \text{ i.e., } \limsup S_n \;=\; \infty \quad \text{almost surely.}$$

Since (S_n) is also a submartingale, we obtain

$$\liminf S_n \; = \; -\infty \qquad \text{almost surely}$$

by an analogue argument. A generalization of this result is given in Theorem 4.7 below.

Remark (Almost sure vs. $\mathbf{L^p}$ convergence). In the last example, the stopped process does not converge in L^p for any $p \in [1, \infty)$. In fact,

$$\lim_{n\to\infty} E[S_{T_a \wedge n}] \; = \; E[S_{T_a}] \; \geq \; a \qquad \text{whereas} \qquad E[S_0] = 0.$$

Example (2. Products of non-negative i.i.d. random variables). Consider a growth process

$$Z_n \quad = \quad \prod_{i=1}^{n} Y_i$$

with i.i.d. factors $Y_i \geq 0$ with finite expectation $\alpha \in (0, \infty)$. Then

$$M_n \quad = \quad Z_n / \alpha^n$$

is a martingale. By the almost sure convergence theorem, a finite limit M_∞ exists almost surely, because $M_n \geq 0$ for all n. For the almost sure asymptotics of (Z_n), we distinguish three different cases:

(1). $\alpha < 1$*:* In this case,

$$Z_n \quad = \quad M_n \cdot \alpha^n$$

converges to 0 exponentially fast with probability one.

(2). $\alpha = 1$*:* Here (Z_n) is a martingale and converges almost surely to a finite limit. If $P[Y_i \neq 1] > 0$ then there exists $\varepsilon > 0$ such that $Y_i \geq 1 + \varepsilon$ infinitely often with probability one. This is consistent with convergence of (Z_n) only if the limit is zero. Hence, if (Z_n) is not almost surely constant, then also in the critical case $Z_n \to 0$ almost surely.

(3). $\alpha > 1$ *(supercritical):* In this case, on the set $\{M_\infty > 0\}$,

$$Z_n \; = \; M_n \cdot \alpha^n \quad \sim \quad M_\infty \cdot \alpha^n,$$

i.e., (Z_n) grows exponentially fast. The asymptotics on the set $\{M_\infty = 0\}$ is not evident and requires separate considerations depending on the model.

Although most of the conclusions in the last example could have been obtained without martingale methods (e.g. by taking logarithms), the martingale approach has the advantage of carrying over to far more general model classes. These include for example branching processes or exponentials of continuous time processes.

Example (3. Boundary behaviors of harmonic functions). Let $D \subseteq \mathbb{R}^d$ be a bounded open domain, and let $h : D \to \mathbb{R}$ be a harmonic function on D that is bounded from below:

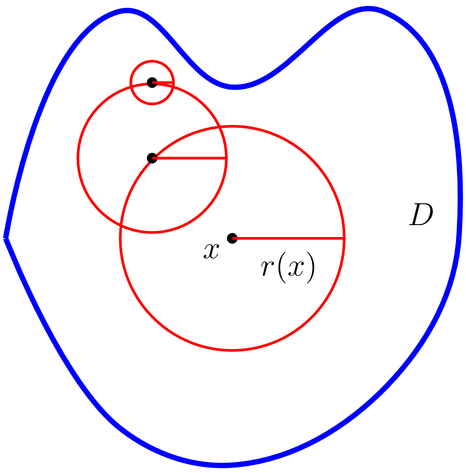
$$\Delta h(x) = 0 \quad \text{for any } x \in D, \quad \inf_{x\in D} h(x) > -\infty. \tag{4.2.2}$$

To study the asymptotic behavior of $h(x)$ as x approaches the boundary ∂D, we construct a Markov chain (X_n) such that $h(X_n)$ is a martingale: Let $r : D \to (0, \infty)$ be a continuous function such that

$$0 < r(x) < \operatorname{dist}(x, \partial D) \qquad \text{for any } x \in D, \tag{4.2.3}$$

and let (X_n) w.r.t P_x denote the canonical time-homogeneous Markov chain with state space D, initial value x, and transition probabilities

$$p(x, dy) = \text{Uniform distribution on } \{y \in \mathbb{R}^d : |y - x| = r(x)\}.$$



By (4.2.3), the function h is integrable w.r.t. $p(x, dy)$, and, by the mean value property,

$$(ph)(x) = h(x) \qquad \text{for any } x \in D.$$

Therefore, the process $h(X_n)$ is a martingale w.r.t. P_x for each $x \in D$. As $h(X_n)$ is lower bounded by (4.2.2), the limit as $n \to \infty$ exists P_x-almost surely by the Supermartingale Convergence Theorem. In particular, since the coordinate functions $x \mapsto x_i$ are also harmonic and lower bound on $\overline{D}$, the limit $X_\infty = \lim_{n\to\infty} X_n$ exists P_x almost surely. Moreover, X_∞ is in ∂D, because r is bounded from below by a strictly positive constant on any compact subset of D.

Summarizing we have shown:

(1). *Boundary regularity:* If h is harmonic and bounded from below on D then the limit $\lim_{n\to\infty} h(X_n)$ exists along almost every trajectory X_n to the boundary ∂D.

(2). *Representation of h in terms of boundary values:* If h is continuous on $\overline{D}$, then $h(X_n) \to h(X_\infty)$ P_x-almost surely and hence

$$h(x) = \lim_{n\to\infty} E_x[h(X_n)] = E[h(X_\infty)],$$

i.e., the distribution of X_∞ w.r.t. P_x is the harmonic measure on ∂D.

Note that, in contrast to classical results from analysis, the first statement holds without any smoothness condition on the boundary ∂D. Thus, although boundary values of h may not exist in the classical sense, they still do exist almost every trajectory of the Markov chain!

Martingales with bounded increments and a Generalized Borel-Cantelli Lemma

Another application of the almost sure convergence theorem is a generalization of the Borel-Cantelli lemmas. We first prove a dichotomy for the asymptotic behavior of martingales with L^1-bounded increments:

Theorem 4.7 (Asymptotics of martingales with L^1 bounded increments). *Suppose that (M_n) is a martingale, and there exists an integrable random variable Y such that*

$$|M_n - M_{n-1}| \leq Y \qquad \textit{for any } n \in \mathbb{N}.$$

Then for P-almost every ω, the following dichotomy holds:

Either: *The limit* $\lim_{n\to\infty} M_n(\omega)$ *exists in* $\mathbb{R}$,

or: $\limsup_{n\to\infty} M_n(\omega) = +\infty$ *and* $\liminf_{n\to\infty} M_n(\omega) = -\infty$.

The theorem and its proof are a generalization of Example 1 above.

Proof. For $a \in (-\infty, 0)$ let $T_a = \min\{n \geq 0 \;:\; M_n \geq a\}$. By the Optional Stopping Theorem, $(M_{T_a \wedge n})$ is a martingale. moreover,

$$M_{T_a \wedge n} \;\geq\; \min(M_0, a - Y) \qquad \text{for any } n \geq 0,$$

and the right hand side is an integrable random variable. Therefore, (M_n) converges almost surely on $\{T_a = \infty\}$. Since this holds for any $a < 0$, we obtain almost sure convergence on the set

$$\{\liminf\; M_n \;>\; -\infty\} \;=\; \bigcup_{\substack{a<0 \\ a\in\mathbb{Q}}} \{T_a = \infty\}.$$

Similarly, almost sure convergence follows on the set $\{\limsup M_n < \infty\}$. ☐

Now let $(\mathcal{F}_n)_{n\geq 0}$ be an arbitrary filtration. As a consequence of Theorem 4.7 we obtain:

Corollary 4.8 (Generalized Borel-Cantelli Lemma). *If* (A_n) *is a sequence of events with* $A_n \in \mathcal{F}_n$ *for any* n*, then the equivalence*

$$\omega \in A_n \quad \textit{infinitely often} \qquad \Longleftrightarrow \qquad \sum_{n=1}^{\infty} P[A_n \mid \mathcal{F}_{n-1}](\omega) \;=\; \infty$$

holds for almost every $\omega \in \Omega$.

Proof. Let $S_n = \sum_{k=1}^{n} I_{A_k}$ and $T_n = \sum_{k=1}^{n} E[I_{A_n} \mid \mathcal{F}_{k-1}]$. Then S_n and T_n are almost surely increasing sequences. Let $S_\infty = \sup S_n$ and $T_\infty = \sup T_n$ denote the limits on $[0, \infty]$. the claim is that almost surely,

$$S_\infty \;=\; \infty \qquad \Longleftrightarrow \qquad T_\infty \;=\; \infty. \tag{4.2.4}$$

to prove (4.2.4) we note that $S_n - T_n$ is a martingale with bounded increments. Therefore, almost surely, $S_n - T_n$ converges to a finite limit, or ($\limsup(S_n - T_n) = \infty$ and $\liminf(S_n - T_n) = -\infty$). In the first case, (4.2.4) holds. In the second case, $S_\infty = \infty$ and $T_\infty = \infty$, so (4.2.4) holds, too. ☐

The assertion of Corollary 4.8 generalizes both classical Borel-Cantelli Lemmas: If (A_n) is an arbitrary sequence of events in a probability space $(\Omega, \mathcal{A}, P)$ then we can consider the filtration $\mathcal{F}_n = \sigma(A_1, \ldots, A_n)$. By Corollary 4.8 we obtain:

1st ***Borel-Cantelli Lemma:*** If $\sum P[A_n] < \infty$ then $\sum P[A_n \,|\, \mathcal{F}_{n-1}] < \infty$ almost surely, and therefore

$$P[A_n \quad \text{infinitely often}] \;=\; 0.$$

2nd ***Borel-Cantelli Lemma:*** If $\sum P[A_n] = \infty$ and the A_n are independent then $\sum P[A_n \mid \mathcal{F}_{n-1}] \;=\; \sum P[A_n] \;=\; \infty$ almost surely, and therefore

$$P[A_n \quad \text{infinitely often}] \;=\; 1.$$

Upcrossing inequality and convergence theorem in continuous time

The upcrossing inequality and the supermartingale convergence theorem carry over immediately to the continuous time case if we assume right continuity (or left continuity) of the sample paths. Let $t_0 \in (0, \infty]$, and let $(Z_s)_{s\in[0,t_0)}$ be a supermartingale in continuous time w.r.t. a filtration $(\mathcal{F}_s)$. We define the number of upcrossings of (Z_s) over an interval (a, b) before time t as the supremum of the number of upcrossings of all time discretizations $(Z_s)_{s\in\pi}$ is a partition of the interval $[0, t]$:

$$U_t^{(a,b)}[Z] \;:=\; \sup_{\substack{\pi\subset[0,t]\\ \text{finite}}} U^{(a,b)}[(Z_s)_{s\in\pi}].$$

Note that if (Z_s) has right-continuous sample paths and (π_n) is a sequence of partitions of $[0, t]$ such that $0, t \in \pi_0$, $\pi_n \subset \pi_{n+1}$ and mesh$(\pi_n) \to 0$ then

$$U_t^{(a,b)}[Z] \;=\; \lim_{n\to\infty} U^{(a,b)}[(Z_s)_{s\in\pi_n}].$$

Theorem 4.9. *Suppose that $(Z_s)_{s\in[0,t_0)}$ is a right continuous supermartingale.*

(1). Upcrossing inequality: For any $t \in [0, t_0)$ and $a < b$,

$$E[U_t^{(a,b)}] \;\leq \frac{1}{b-a} E[(Z_t - a)^-].$$

(2). *Convergence Theorem: if* $\sup_{s\in[0,t_0)} E[Z_s^-] < \infty$*, then the limit* $Z_{t_0} = \lim_{s\nearrow t_0} Z_s$ *exists almost surely, and* Z_{t_0} *is an integrable random variable.*

Proof. (1). By the upcrossing inequality in discrete time,

$$E[U^{(a,b)}[(Z_s)_{s\in\pi_n}]] \leq E[(Z_t - u)^-] \qquad \text{for any } n \in \mathbb{N}$$

where (π_n) is a sequence of partitions as above. The assertion now follows by the Monotone Convergence Theorem.

(2). The almost sure convergence can now be proven in the same way as in the discrete time case.

□

More generally than stated above, the upcrossing inequality also implies that for a right-continuous supermartingale $(Z_s)_{s\in[0,t_0)}$ all the left limits $\lim_{s\nearrow t} Z_s, t \in [0, t_0)$, exist *simultaneously* with probability one. Thus almost every sample path is *càdlàg* (continue à droite limites a gauche, i.e., right continuous with left limits). By similar arguments, the existence of a modification with right continuous (and hence càdlàg) sample paths can be proven for *any* supermartingale (Z_s) provided the filtration is right continuous and complete, and $s \mapsto E[Z_s]$ is right continuous, cf. e.g. [Revuz/Yor, Ch.II,§2].

4.3 Uniform integrability and L^1 convergence

The Supermartingale Convergence Theorem shows that every supermartingale (Z_n) that is bounded in L^1, converges almost surely to an integrable limit Z_∞. However, L^1 convergence does not necessarily hold:

Example. (1). Suppose that $Z_n = \prod_{i=1}^n Y_i$ where the Y_i are i.i.d. with $E[Y_i] = 1$, $P[Y_i \neq 1] > 0$. Then, $Z_n \to 0$ almost surely, cf. Example 2 in Section 4.2. On the other hand, L^1 convergence does not hold as $E[Z_n] = 1$ for any n.

(2). Similarly, the exponential martingale $M_t = \exp(B_t - t/2)$ of a Brownian motion converges to 0 almost surely, but $E[M_t] = 1$ for any t.

L^1 convergence of martingales is of interest because it implies that a martingale sequence (M_n) can be extended to $n = \infty$, and the random variables M_n are given as conditional expectations of the limit M_∞. Therefore, we now prove a generalization of the Dominated Convergence Theorem that leads to a necessary and sufficient condition for L^1 convergence.

Uniform integrability

Let $(\Omega, \mathcal{A}, P)$ be a probability space. The key condition required to deduce L^1 convergence from convergence in probability is uniform integrability. To motivate the definition we first recall two characterizations of integrable random variables:

Lemma 4.10. *If $X : \Omega \to \mathbb{R}$ is an integrable random variable on $(\Omega, \mathcal{A}, P)$, then*

(1). $\lim\limits_{c\to\infty} E[|X|\,;\,|X| \geq c] = 0,$ *and*

(2). for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$E[|X|\,;\,A] \;<\; \varepsilon \quad \text{for any } A \in \mathcal{A} \text{ with } P[A] < \delta.$$

The second statement says that the positive measure

$$Q(A) \;=\; E[|X|\,;\,A], \qquad A \in \mathcal{A},$$

with relative density $|X|$ w.r.t. P is **absolutely continuous** w.r.t. P in the following sense: *For any $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$P[A] \;<\; \delta \quad \Rightarrow \quad Q(A) \;<\; \varepsilon.$$

Proof. (1). For an integrable random variable X the first assertion holds by the Monotone Convergence Theorem, since $|X| \cdot I_{\{|X|\geq c\}} \searrow 0$ as $c \nearrow \infty$.

(2). Let $\varepsilon > 0$. By (1),

$$\begin{aligned} E[|X|\,;\,A] &= E[|X|\,;\,A \cap \{|X| \geq c\}] + E[|X|\,;\,A \cap \{|X| \leq c\}] \\ &\leq E[|X|\,;\,\{|X| \geq c\}] + c \cdot P[A] \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \quad = \quad \varepsilon \end{aligned}$$

provided $c \in (0, \infty)$ is chosen appropriately and $P[A] < \varepsilon/2c$.

□

Uniform integrability means that properties (1) and (2) hold uniformly for a family of random variables:

Definition. *A family $\{X_i : i \in I\}$ of random variables on $(\Omega, \mathcal{A}, P)$ is called* ***uniformly integrable*** *if and only if*

$$\sup_{i \in I} E[|X_i| \; ; \; |X_i| \geq c] \quad \longrightarrow \quad 0 \qquad \text{as } c \to \infty.$$

Exercise. Prove that $\{X_i : i \in I\}$ is uniformly integrable if and only if $\sup E[|X_i|; A] < \infty$ and the measures $Q_i(A) = E[|X_i| \; ; \; A]$ are **uniformly absolutely continuous**, i.e., for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$P[A] \; < \; \delta \quad \Rightarrow \quad \sup_{i \in I} E[|X_i| \; ; \; A] \; < \; \varepsilon.$$

We will prove below that convergence in probability plus uniform integrability is equivalent to L^1 convergence. Before, we state two lemmas giving sufficient conditions for uniform integrability (and hence for L^1 convergence) that can often be verified in applications:

Lemma 4.11. *A family $\{X_i : i \in I\}$ of random variables is uniformly integrable if one of the following conditions holds:*

(1). There exists an integrable random variable Y such that

$$|X_i| \; \leq \; Y \qquad \text{for any } i \in I.$$

(2). There exists a measurable function $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\lim_{x \to \infty} \frac{g(x)}{x} \; = \; \infty \qquad \text{and} \qquad \sup_{i \in I} E[g(|X_i|)] \; < \; \infty.$$

Proof. (1). If $|X_i| \leq Y$ then

$$\sup_{i \in I} E[|X_i| \; ; \; |X_i| \geq c] \; \leq \; E[Y \; ; \; Y \geq c].$$

The right hand side converges to 0 as $c \to \infty$ if Y is integrable.

(2). The second condition implies uniform integrability, because

$$\sup_{i\in I} E[|X_i| \,;\, |X_i| \geq c] \;\leq\; \sup_{y\geq c} \frac{y}{g(y)} \cdot \sup_{i\in I} E[g(|X_i|)].$$

□

The first condition in Lemma 4.11 is the classical assumption in the Dominated Convergence Theorem. The second condition holds in particular if

$$\sup_{i\in I} E[|X_i|^p] \;<\; \infty \qquad \text{for some } p > 1 \text{ } (\mathbf{L^p}\textbf{ boundedness})$$

or, if

$$\sup_{i\in I} E[|X_i|(\log|X_i|)^+] \;<\; \infty \qquad \textbf{entropy condition}$$

is satisfied. Boundedness in L^1, however, does not imply uniform integrability, cf. any counterexample to the Dominated Convergence Theorem.

The next observation is crucial for the application of uniform integrability to martingales:

Lemma 4.12 (**Conditional expectations are uniformly integrable**). *If X is an integrable random variable on $(\Omega, \mathcal{A}, P)$ then the family*

$$\{E[X \mid \mathcal{F}] \;:\; \mathcal{F} \subseteq \mathcal{A} \quad \sigma\text{-algebras}\}$$

of all conditional expectations of X given sub-σ-algebras of $\mathcal{A}$ is uniformly integrable.

Proof. By Lemma 4.10, for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$E[|E[X|\mathcal{F}]|; E[X|\mathcal{F}] \geq c] \leq E[E[|X||\mathcal{F}]; E[X|\mathcal{F}] \geq c] = E[|X|; E[X|\mathcal{F}] \geq c] < \varepsilon \tag{4.3.1}$$

holds for $c > 0$ with $P[|E[X \mid \mathcal{F}]| \geq c] < \delta$. Since

$$P[|E[X \mid \mathcal{F}]| \geq c] \;\leq\; \frac{1}{c} E[|E[X \mid \mathcal{F}]|] \;\leq\; \frac{1}{c} E[\,|X|\,],$$

(4.3.1) holds simultaneously for all σ-algebras $\mathcal{F} \subseteq \mathcal{A}$ if c is sufficiently large. □

Definitive version of Lebesgue's Dominated Convergence Theorem

Theorem 4.13. *Suppose that $(X_n)_{n\in\mathbb{N}}$ is a sequence of integrable random variables. Then (X_n) converges to a random variable X w.r.t. the L^1 norm if and only if X_n converges to X in probability and the family $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable.*

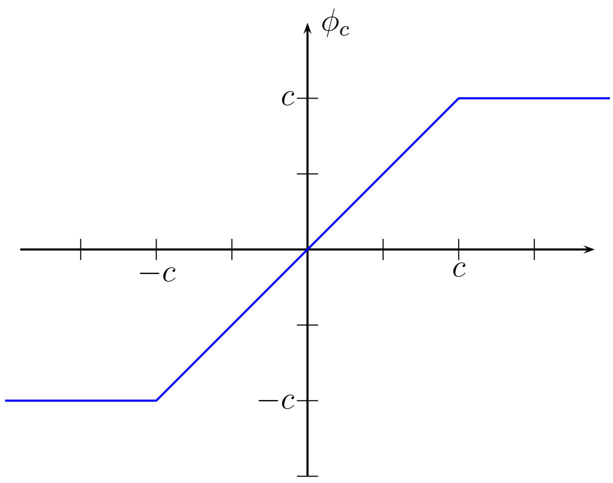
Proof. (1). We first prove the "if" part of the assertion under the additional assumption that the random variables $|X_n|$ are uniformly bounded by a finite constant c:

$$\begin{aligned} E[\,|X_n - X|\,] &= E[\,|X_n - X|\,;\,|X_n - X| > \varepsilon] + E[\,|X_n - X|\,;\,|X_n - X| \le \varepsilon] \\ &\le 2c \cdot P[\,|X_n - X| > \varepsilon]. \end{aligned} \tag{4.3.2}$$

Here we have used that $|X_n| \le c$ and hence $|X| \le c$ with probability one, because a subsequence of (X_n) converges almost surely to X. For sufficiently large n, the right hand side of (4.3.2) is smaller than 2ε. Therefore, $E[\,|X_n - X|\,] \to 0$ as $n \to \infty$.

(2). To prove the "if" part under the uniform integrability condition, we consider the cut-off-functions

$$\phi_c(x) = (x \wedge c) \vee (-c)$$



For $c \in (0, \infty)$, the function $\phi_c : \mathbb{R} \to \mathbb{R}$ is a contraction. Therefore,

$$|\phi_c(X_n) - \phi_c(X)| \le |X_n - X| \qquad \text{for any } n \in \mathbb{N}.$$

If $X_n \to X$ in probability then $\phi_c(X_n) \to \phi_c(X)$ in probability, by (1),

$$E[\,|\phi_c(X_n) - \phi_c(X)|\,] \longrightarrow 0 \qquad \text{for any } c > 0. \tag{4.3.3}$$

We would like to conclude that $E[\,|X_n - X|\,] \to 0$ as well. Since (X_n) is uniformly integrable, and a subsequence converges to X almost surely, we have $E[\,|X|\,] \liminf E[\,|X_n|\,] < \infty$ by Fatou's Lemma. We now estimate

$$\begin{aligned} E[\,|X_n - X|\,] &\leq E[\,|X_n - \phi_c(X_n)|\,] + E[\,|\phi_c(X_n) - \phi_c(X)|\,] + E[\,|\phi_c(X) - X|\,] \\ &\leq E[\,|X_n|\,;\,|X_n| \geq c] + E[\,|\phi_c(X_n) - \phi_c(X)|\,] + E[\,|X|\,;\,|X| \geq c]. \end{aligned}$$

Let $\varepsilon > 0$ be given. Choosing c large enough, the first and the last summand on the right hand side are smaller than $\varepsilon/3$ for all n by uniform integrability of $\{X_n : n \in \mathbb{N}\}$ and integrability of X. Moreover, by (4.3.3), there exists $n_0(c)$ such that the middle term is smaller than $\varepsilon/3$ for $n \geq n_0(c)$. Hence $E[\,|X_n - X|\,] < \varepsilon$ for $n \geq n_0$, i.e. $X_n \to X$ in L^1.

(3). Now suppose conversely that $X_n \to X$ in L^1. Then $X_n \to X$ in probability by Markov's inequality. To prove uniform integrability, we observe that

$$E[\,|X_n|\,;\,A] \leq E[\,|X|\,;\,A] + E[\,|X - X_n|\,] \qquad \text{for any } n \in \mathbb{N} \text{ and } A \in \mathcal{A}.$$

For $\varepsilon > 0$, there exist $n_0(\varepsilon) \in \mathbb{N}$ and $\delta(\varepsilon) > 0$ such that

$$\begin{aligned} E[\,|X - X_n|\,] &< \varepsilon/2 \qquad \text{for any } n > n_0, \text{ and} \\ E[\,|X|\,;\,A] &< \varepsilon/2 \qquad \text{whenever } P[A] < \delta, \end{aligned}$$

cf. Lemma 4.10. Hence, if $P[A] < \delta$ then $\sup_{n \geq n_0} E[\,|X_n|\,;\,A] < \varepsilon$.
Moreover, again by Lemma 4.10, there exist $\delta_1, \ldots, \delta_{n_0} > 0$ such that for $n \leq n_0$,

$$E[\,|X_n|\,;\,A] < \varepsilon \qquad \text{if} \quad P[A] < \delta_n.$$

Choosing $\widetilde{\delta} = \min(\delta, \delta_1, \delta_2, \ldots, \delta_{n_0})$, we obtain

$$\sup_{n \in \mathbb{N}} E[\,|X_n|\,;\,A] < \varepsilon \qquad \text{whenever} \quad P[A] < \widetilde{\delta}.$$

Therefore, $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable by the exercise below the definition of uniform integrability on page 130.

□

$\mathbf{L^1}$ convergence of martingales

If X is an integrable random variable and $(\mathcal{F}_n)$ is a filtration then $M_n = E[X \mid \mathcal{F}_n]$ is a martingale w.r.t. $(\mathcal{F}_n)$. The next result shows that an arbitrary martingale can be represented in this way if and only if it is uniformly integrable:

Theorem 4.14 ($\mathbf{L^1}$ Martingale Convergence Theorem). *Suppose that $(M_n)_{n\geq 0}$ is a martingale w.r.t. a filtration $(\mathcal{F}_n)$. Then the following statements are equivalent:*

(1). $\{M_n : n \geq 0\}$ is uniformly integrable.

(2). The sequence (M_n) converges w.r.t. the L^1 norm.

(3). There exists an integrable random variable X such that

$$M_n = E[X \mid \mathcal{F}_n] \qquad \textit{for any } n \geq 0.$$

Proof.

(3) ⇒ (1) holds by Lemma 4.12.

(1) ⇒ (2): If the sequence (M_n) is uniformly integrable then it is bounded in L^1 because

$$\sup_n E[\,|M_n|\,] \leq \sup_n E[\,|M_n|\,;\,|M_n| \geq c] + c \leq 1 + c$$

for $c \in (0, \infty)$ sufficiently large. Therefore, the limit $M_\infty = \lim M_n$ exists almost surely and in probability by the almost sure convergence theorem. Uniform integrability then implies

$$M_n \to M_\infty \quad \text{in } L^1$$

by Theorem 4.13.

(2) ⇒ (3): If M_n converges to a limit M_∞ in L^1 then

$$M_n = E[M_\infty \mid \mathcal{F}_n] \qquad \text{for any } n \geq 0.$$

In fact, M_n is a version of the conditional expectation since it is $\mathcal{F}_n$-measurable and

$$E[M_\infty\,;\,A] \;=\; \lim_{k\to\infty} E[M_k\,;\,A] \;=\; E[M_n\,;\,A] \qquad \text{for any } A\in\mathcal{F}_n \tag{4.3.4}$$

by the martingale property.

□

A first consequence of the L^1 convergence theorem is a limit theorem for conditional expectations:

Corollary 4.15. *If X is an integrable random variable and $(\mathcal{F}_n)$ is a filtration then*

$$E[X \mid \mathcal{F}_n] \;\to\; E[X \mid \mathcal{F}_\infty] \qquad \textit{almost surely and in } L^1,$$

where $\mathcal{F}_\infty := \sigma(\bigcup \mathcal{F}_n)$.

Proof. Let $M_n := E[X \mid \mathcal{F}_n]$. By the almost sure and the L^1 martingale convergence theorem, the limit $M_\infty = \lim M_n$ exists almost surely and in L^1. To obtain a measurable function that is defined everywhere, we set $M_\infty := \limsup M_n$. It remains to verify, that M_∞ is a version of the conditional expectation $E[X \mid \mathcal{F}_\infty]$. Clearly, M_∞ is measurable w.r.t. $\mathcal{F}_\infty$. Moreover, for $n \geq 0$ and $A \in \mathcal{F}_n$,

$$E[M_\infty\,;\,A] \;=\; E[M_n\,;\,A] \;=\; E[X\,;\,A]$$

by (4.3.4). Since $\bigcup \mathcal{F}_n$ is stable under finite intersections,

$$E[M_\infty\,;\,A] \;=\; E[X\,;\,A]$$

holds for all $A \in \sigma(\bigcup \mathcal{F}_n)$ as well. □

Example (Existence of conditional expectations). The common existence proof for conditional expectations relies either on the Radon-Nikodym Theorem or on the existence of orthogonal projections onto closed subspaces of the Hilbert space L^2. Martingale convergence can be used to give an alternative existence proof. Suppose that X is an integrable random variable on a probability space $(\Omega, \mathcal{A}, P)$ and $\mathcal{F}$ is a **separable**

sub-σ-algebra of $\mathcal{A}$, i.e., there exists a countable collection $(A_i)_{i\in\mathbb{N}}$ of events $A_i \in \mathcal{A}$ such that $\mathcal{F} = \sigma(A_i \mid i \in \mathbb{N})$. Let

$$\mathcal{F}_n = \sigma(A_1, \dots, A_n), \qquad n \geq 0.$$

Note that for each $n \geq 0$, there exist finitely many atoms $B_1, \dots, B_k \in \mathcal{A}$ (i.e. disjoint events with $\bigcup B_i = \Omega$) such that $\mathcal{F}_n = \sigma(B_, \dots, B_k)$. Therefore, the conditional expectation given $\mathcal{F}_n$ can be defined in an elementary way:

$$E[X \mid \mathcal{F}_n] := \sum_{i \,:\, P[B_i]\neq 0} E[X \mid B_i] \cdot I_{B_i}.$$

Moreover, by Corollary 4.15, the limit $M_\infty = \lim E[X \mid \mathcal{F}_n]$ exists almost surely and in L^1, and M_∞ is a version of the conditional expectation $E[X \mid \mathcal{F}]$.

You might (and should) object that the proofs of the martingale convergence theorems require the existence of conditional expectations. Although this is true, it is possible to state the necessary results by using only elementary conditional expectations, and thus to obtain a more constructive proof for existence of conditional expectations given separable σ-algebras.

Another immediate consequence of Corollary 4.15 is an extension of Kolmogorov's 0-1 law:

Corollary 4.16 (0-1 Law of P.Lévy). *If $(\mathcal{F}_n)$ is a filtration of $(\Omega, \mathcal{A}, P)$ then*

$$P[A \mid \mathcal{F}_n] \longrightarrow I_A \qquad P\text{-almost surely} \tag{4.3.5}$$

for any event $A \in \sigma(\bigcup \mathcal{F}_n)$.

Example (Kolmogorov's 0-1 Law). Suppose that $\mathcal{F}_n = \sigma(\mathcal{A}_1, \dots, \mathcal{A}_n)$ with independent σ-algebras $\mathcal{A}_i \subseteq \mathcal{A}$. If A is a **tail event**, i.e., A is in $\sigma(\mathcal{A}_{n+1}, \mathcal{A}_{n+2}, \dots)$ for any $n \in \mathbb{N}$, then A is independent of $\mathcal{F}_n$ for any n. Therefore, the corollary implies that $P[A] = I_A$ P-almost surely, i.e.,

$$P[A] = 0 \qquad \text{for any tail event } A.$$

The L^1 Martingale Convergence Theorem also implies that any martingale that is L^p bounded for some $p \in (1, \infty)$ converges in L^p:

Exercise ($\mathbf{L^p}$ Martingale Convergence Theorem). Let (M_n) be an $(\mathcal{F}_n)$ martingale with $\sup E[\,|M_n|^p\,] < \infty$ for some $p \in (1, \infty)$.

(1). Prove that (M_n) converges almost surely and in L^1, and $M_n = E[M_\infty \mid \mathcal{F}_n]$ for any $n \geq 0$.

(2). Conclude that $|M_n - M_\infty|^p$ is uniformly integrable, and $M_n \to M_\infty$ in L^p.

Note that uniform integrability of $|M_n|^p$ holds automatically and has not to be assumed.

Backward Martingale Convergence

We finally remark that Doob's upcrossing inequality can also be used to prove that the conditional expectations $E[X \mid \mathcal{F}_n]$ of an integrable random variable given a *decreasing* sequence $(\mathcal{F}_n)$ of σ-algebras converge almost surely to $E[X \mid \bigcap \mathcal{F}_n]$. For the proof one considers the martingale $M_{-n} = E[X \mid \mathcal{F}_n]$ indexed by the negative integers:

Exercise (Backward Martingale Convergence Theorem and LLN). Let $(\mathcal{F}_n)_{n\geq 0}$ be a *decreasing* sequence of sub-σ-algebras on a probability space $(\Omega, \mathcal{A}, P)$.

(1). Prove that for any random variable $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, the limit $M_{-\infty}$ of the sequence $M_{-n} := E[X \mid \mathcal{F}_n]$ as $n \to -\infty$ exists almost surely and in L^1, and

$$M_{-\infty} \;=\; E[X \mid \bigcap \mathcal{F}_n] \qquad \text{almost surely.}$$

(2). Now let (X_n) be a sequence of i.i.d. random variables in $\mathcal{L}^1(\Omega, \mathcal{A}, P)$, and let $\mathcal{F}_n = \sigma(S_n, S_{n+1}, \ldots)$ where $S_n = X_1 + \ldots + X_n$. Prove that

$$E[X_1 \mid \mathcal{F}_n] \;=\; \frac{S_n}{n},$$

and conclude that the strong Law of Large Numbers holds:

$$\frac{S_n}{n} \;\longrightarrow\; E[X_1] \qquad \text{almost surely.}$$

4.4 Local and global densities of probability measures

A thorough understanding of absolute continuity and relative densities of probability measures is crucial at many places in stochastic analysis. Martingale convergence yields an elegant approach to these issues including a proof of the Radon-Nikodym and the Lebesgue Decomposition Theorem. We first recall the definition of absolute continuity.

Absolute Continuity

Suppose that P and Q are probability measures on a measurable space $(\Omega, \mathcal{A})$, and $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{A}$.

Definition. *(1). The measure Q is called* ***absolutely continuous w.r.t.*** P ***on the σ-algebra*** $\mathcal{F}$ *if and only if $Q[A] = 0$ for any $A \in \mathcal{F}$ with $P[A] = 0$.*

(2). The measures Q and P are called ***singular on*** $\mathcal{F}$ *if and only if there exists $A \in \mathcal{F}$ such that $P[A] = 0$ and $Q[A^C] = 0$.*

We use the notations $Q \ll P$ for absolute continuity of Q w.r.t. P, $Q \approx P$ for mutual absolute continuity, and $Q \perp\!\!\!\perp P$ for singularity of Q and P. The definitions above extend to signed measures.

Example. The Dirac measure $S_{1/2}$ is obviously singular w.r.t. Lebesgue measure $\lambda_{(0,1]}$ on the Borel σ-algebra $\mathcal{B}((0,1])$. However, $\delta_{1/2}$ is absolutely continuous w.r.t. $\lambda_{(0,1]}$ on each of the σ-algebras $\mathcal{F}_n = \sigma(\mathcal{D}_n)$ generated by the dyadic partitions $\mathcal{D}_n = \{(k \cdot 2^{-n}, (k+1)2^{-n}] \, : \, 0 \leq k < 2^n\}$, and $\mathcal{B}([0,1)) = \sigma(\bigcup \mathcal{D}_n)$.

The next lemma clarifies the term "absolute continuity."

Lemma 4.17. *The probability measure Q is absolutely continuous w.r.t. P on the σ-algebra $\mathcal{F}$ if and only if for any $\varepsilon > 0$ there exists $\delta > 0$ such that for $A \in \mathcal{F}$,*

$$P[A] < \delta \quad \Rightarrow \quad Q[A] < \varepsilon. \tag{4.4.1}$$

Proof. The "if" part is obvious. If $P[A] = 0$ and (4.4.1) holds for each $\varepsilon > 0$ with δ depending on ε then $Q[A] < \varepsilon$ for any $\varepsilon > 0$, and hence $Q[A] = 0$.

To prove the "only if" part, we suppose that there exists $\varepsilon > 0$ such that (4.4.1) does not hold for any $\delta > 0$. Then there exists a sequence (A_n) of events in $\mathcal{F}$ such that

$$Q[A_n] \;\geq\; \varepsilon \qquad \text{and} \qquad P[A_n] \;\leq\; 2^{-n}.$$

Hence, by the Borel-Cantelli-Lemma,

$$P[A_n \quad \text{infinitely often}] \;=\; 0,$$

whereas

$$Q[A_n \quad \text{infinitely often}] \;=\; Q\left[\bigcap_n \bigcup_{m\geq n} A_m\right] \;=\; \lim_{n\to\infty} Q\left[\bigcup_{m\geq n} A_m\right] \;\geq\; \varepsilon.$$

Therefore Q is not absolutely continuous w.r.t. P. □

Example (**Absolute continuity on** $\mathbb{R}$)**.** A probability measure μ on a real interval is absolutely continuous w.r.t. Lebesgue measure if and only if the distribution function $F(t) = \mu[(-\infty, t]]$ satisfies:

$$\text{For any } \varepsilon > 0 \text{ there exists } \delta > 0 \text{ such that for } n \in \mathbb{N}$$
$$\sum_{i=1}^{n} |b_i - a_i| \;<\; \varepsilon \;\;\Rightarrow\;\; |F(b_i) - F(a_i)| \;<\; \delta, \tag{4.4.2}$$

cf. e.g. [Billingsley: Probability and Measures].

Definition. *A function $F : (a, b) \subset \mathbb{R} \to \mathbb{R}$ is called **absolutely continuous** iff (4.4.2) holds.*

The Radon-Nikodym Theorem states that absolute continuity is equivalent to the existence of a relative density.

Theorem 4.18 (Radon-Nikodym)**.** *The probability measure Q is absolutely continuous w.r.t. P on the σ-algebra $\mathcal{F}$ if and only if there exists a non-negative random variable $Z \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ such that*

$$Q[A] \;=\; \int_A Z\, dP \qquad \textit{for any } A \in \mathcal{F}. \tag{4.4.3}$$

The relative density Z of Q w.r.t. P on $\mathcal{F}$ is determined by (4.4.3) uniquely up to modification on P-measure zero sets. It is also called the **Radon-Nikodym derivative** or the **likelihood ratio** of Q w.r.t. P on $\mathcal{F}$. We use the notation

$$Z = \left.\frac{dP}{dQ}\right|_{\mathcal{F}},$$

and omit the $\mathcal{F}$ when the choice of the σ-algebra is clear.

Example (Finitely generated σ-algebra). Suppose that the σ-algebra $\mathcal{F}$ is generated by finitely many disjoint atoms $B_1, \ldots, B_k$ with $\Omega = \bigcup B_i$. Then Q is absolutely continuous w.r.t. P if and only if for any i,

$$P[B_i] = 0 \quad \Longrightarrow \quad Q[B_i] = 0.$$

In this case, the relative density is given by

$$\left.\frac{dP}{dQ}\right|_{\mathcal{F}} = \sum_{i\,:\,P[B_i]>0} \frac{Q[B_i]}{P[B_i]} \cdot I_{B_i}.$$

From local to global densities

Let $(\mathcal{F}_n)$ be a given filtration on $(\Omega, \mathcal{A})$.

Definition. *The measure Q is called **locally absolutely continuous** w.r.t. P and the filtration $(\mathcal{F}_n)$ if and only if Q is absolutely continuous w.r.t. P on the σ-algebra $\mathcal{F}_n$ for each n.*

Example (Dyadic partitions). Any probability measure on the unit interval $[0, 1]$ is locally absolutely continuous w.r.t. Lebesgue measure on the filtration $\mathcal{F}_n = \sigma(\mathcal{D}_n)$ generated by the dyadic partitions of the unit interval. The Radon-Nikodym derivative on $\mathcal{F}_n$ is the dyadic difference quotient defined by

$$\left.\frac{d\mu}{d\lambda}\right|_{\mathcal{F}_n}(x) = \frac{\mu[((k-1)\cdot 2^{-n}, k\cdot 2^{-n})]}{\lambda[((k-1)\cdot 2^{-n}, k\cdot 2^{-n})]} = \frac{F(k\cdot 2^{-n}) - F((k-1)\cdot 2^{-n})}{2^{-n}} \tag{4.4.4}$$

for $x \in ((k-1)2^{-n}, k2^{-n}]$.

Example (Product measures). If $Q = \bigotimes_{i=1}^{\infty} \nu$ and $P = \bigotimes_{i=1}^{\infty} \mu$ are infinite products of probability measures ν and μ, and ν is absolutely continuous w.r.t. μ with density ϱ, then Q is locally absolutely continuous w.r.t. P on the filtration

$$\mathcal{F}_n \; = \; \sigma(X_1, \ldots, X_n)$$

generated by the coordinate maps $X_i(\omega) = \omega_i$. The local relative density is

$$\left.\frac{dP}{dQ}\right|_{\mathcal{F}_n} \; = \; \prod_{i=1}^{n} \varrho(X_i)$$

However, if $\nu \neq \mu$, then Q is not absolutely continuous w.r.t. P on $\mathcal{F}_\infty = \sigma(X_1, X_2, \ldots)$, since by the LLN, $n^{-1} \sum_{i=1}^{n} I_A(X_i)$ converges Q almost surely to $\nu[A]$ and P-almost surely to $\mu[A]$.

Now suppose that Q is locally absolutely continuous w.r.t. P on a filtration $(\mathcal{F}_n)$ with relative densities

$$Z_n \; = \; \left.\frac{dQ}{dP}\right|_{\mathcal{F}_n}.$$

The L^1 martingale convergence theorem can be applied to study the existence of a global density on the σ-algebra

$$\mathcal{F}_\infty \; = \; \sigma(\bigcup \mathcal{F}_n).$$

Let $Z_\infty := \limsup Z_n$.

Theorem 4.19 (Convergence of local densities, Lebesgue decomposition).

(1). The sequence (Z_n) of successive relative densities is an $(\mathcal{F}_n)$-martingale w.r.t. P. In particular, (Z_n) converges P-almost surely to Z_∞, and Z_∞ is integrable w.r.t. P.

(2). The following statements are equivalent:

(a) (Z_n) is uniformly integrable w.r.t. P.

(b) Q is absolutely continuous w.r.t. P on $\mathcal{F}_\infty$.

(c) $Q[A] = \int_A Z_\infty \, dP$ for any P on $\mathcal{F}_\infty$.

(3). In general, the decomposition $Q = Q_a + Q_s$ holds with

$$Q_a[A] \;=\; \int_A Z_\infty \, dP, \qquad Q_s[A] \;=\; Q[A \cap \{Z_\infty = \infty\}]. \tag{4.4.5}$$

Q_a ans Q_s are positive measure with $Q_a \ll P$ and $Q_s \perp\!\!\!\perp P$.

The decomposition $Q = Q_a + Q_s$ into an absolutely continuous and a singular part is called the **Lebesgue decomposition** of the measure Q w.r.t. P on the σ-algebra $\mathcal{F}_\infty$.

Proof. (1). For $n \geq 0$, the density Z_n is in $\mathcal{L}^1(\Omega, \mathcal{F}_n, P)$, and

$$E_P[Z_n \,;\, A] \;=\; Q[A] \;=\; E_P[Z_{n+1} \,;\, A] \qquad \text{for any } A \in \mathcal{F}_n.$$

Hence $Z_n = E_P[Z_{n+1} \mid \mathcal{F}_n]$, i.e., (Z_n) is a martingale w.r.t. P. Since $Z_n \geq 0$, the martingale converges P-almost surely, and the limit is integrable.

(2). (a) $\Rightarrow$ (c): If (Z_n) is uniformly integrable w.r.t. P, then

$$Z_n \;=\; E_P[Z_\infty \mid \mathcal{F}_n] \qquad P\text{-almost surely for any } n,$$

by the L^1 convergence theorem. Hence for $A \in \mathcal{F}_n$,

$$Q[A] \;=\; E_P[Z_n \,;\, A] \;=\; E_P[Z_\infty \,;\, A].$$

This shows that $Q[A] = E_P[Z_\infty \,;\, A]$ holds for any $A \in \bigcup \mathcal{F}_n$, and thus for any $A \in \mathcal{F}_\infty = \sigma(\bigcup \mathcal{F}_n)$.

(c) $\Rightarrow$ (b) is evident.

(b) $\Rightarrow$ (a): If $Q \ll P$ on $\mathcal{F}_\infty$ then Z_n converges also Q-almost surely to a finite limit Z_∞. Hence (Z_n) is Q-almost surely bounded, and therefore

$$\begin{aligned} \sup_n E_P[\,|Z_n| \,;\, |Z_n| \geq c] \;&=\; \sup_n E_P[Z_n \,;\, Z_n \geq c] \;=\; \sup_n Q[Z_n \geq c] \\ &\leq\; Q[\sup Z_n \geq c] \quad \longrightarrow \quad 0 \end{aligned}$$

as $c \to \infty$, i.e., (Z_n) is uniformly integrable w.r.t. P.

(3). In general, $Q_a[A] = E_P[Z_\infty\,;\,A]$ is a positive measure on $\mathcal{F}_\infty$ with $Q_a \le Q$, since for $n \ge 0$ and $A \in \mathcal{F}_n$,

$$Q_a[A] \;=\; E_P[\liminf_{k\to\infty} Z_k\,;\,A] \;\le\; \liminf_{k\to\infty} E_P[Z_k\,;\,A] \;=\; E_P[Z_n\,;\,A] \;=\; Q[A]$$

by Fatou's Lemma and the martingale property.

It remains to show that

$$Q_a[A] \;=\; Q[A\cap\{Z_\infty<\infty\}] \qquad \text{for any } A\in\mathcal{F}_\infty. \tag{4.4.6}$$

If (4.4.6) holds, then $Q = Q_a + Q_s$ with Q_s defined by (4.4.5). In particular, Q_s is then singular w.r.t. P, since $P[Z_\infty = \infty] = 0$ and $Q_s[Z_\infty = \infty] = 0$, whereas Q_a is absolutely continuous w.r.t. P by definition.

Since $Q_a \le Q$, it suffices to verify (4.4.6) for $A = \Omega$. Then

$$(Q-Q_a)[A\cap\{Z_\infty<\infty\}] \;=\; (Q-Q_a)[Z_\infty<\infty] \;=\; 0,$$

and therefore

$$Q[A\cap\{Z_\infty<\infty\}] \;=\; Q_a[A\cap\{Z_\infty<\infty\}] \;=\; Q_a[A]$$

for any $A\in\mathcal{F}_\infty$.

To prove (4.4.6) for $A = \Omega$ we observe that for $c\in(0,\infty)$,

$$\begin{aligned} Q\left[\limsup_{n\to\infty} Z_n < c\right] \;&\le\; \limsup_{n\to\infty} Q[Z_n<c] \;=\; \limsup_{n\to\infty} E_P[Z_n\,;\,Z_n<c] \\ &\le\; E_P\left[\limsup_{n\to\infty} Z_n\cdot I_{\{Z_n<c\}}\right] \;\le\; E_P[Z_\infty] \;=\; Q_a[\Omega] \end{aligned}$$

by Fatou's Lemma. As $c\to\infty$, we obtain

$$Q[Z_\infty<\infty] \;\le\; Q_a[\Omega] \;=\; Q_a[Z_\infty<\infty] \;\le\; Q[Z_\infty<\infty]$$

and hence (4.4.6) with $A=\Omega$. This completes the proof

□

As a first consequence of Theorem 4.19, we prove the Radon-Nikodym Theorem on a separable σ-algebra $\mathcal{A}$. Let P and Q be probability measures on $(\Omega,\mathcal{A})$ with $Q \ll P$.

Proof of the Radon-Nikodym Theorem for separable σ-algebras. We fix a filtration $(\mathcal{F}_n)$ consisting of finitely generated σ-algebras $\mathcal{F}_n \subseteq \mathcal{A}$ with $\mathcal{A} = \sigma(\bigcup \mathcal{F}_n)$. Since Q is absolutely continuous w.r.t. P, the local densities $Z_n = dQ/dP|_{\mathcal{F}_n}$ on the finitely generated σ-algebras $\mathcal{F}_n$ exist, cf. the example above. Hence by Theorem 4.19,

$$Q[A] \; = \; \int_A Z_\infty \, dP \qquad \text{for any } A \in \mathcal{A}.$$

□

The approach above can be generalized to probability measures that are not absolutely continuous:

Exercise (Lebesgue decomposition, Lebesgue densities). Let P and Q be arbitrary (not necessarily absolutely continuous) probability measures on $(\Omega, \mathcal{A})$. A **Lebesgue density** of Q w.r.t. P is a random variable $Z : \Omega \to [0, \infty]$ such that $Q = Q_a + Q_s$ with

$$Q_a[A] \; = \; \int_A Z \, dP, \quad Q_s[A] \; = \; Q[A \cap \{Z = \infty\}] \qquad \text{for any } A \in \mathcal{A}.$$

The goal of the exercise is to prove that a Lebesgue density exists if the σ-algebra $\mathcal{A}$ is separable.

(1). Show that if Z is a Lebesgue density of Q w.r.t. P then $1/Z$ is a Lebesgue density of P w.r.t. Q. Here $1/\infty := 0$ and $1/0 := \infty$.

From now on suppose that the σ-algebra is separable with $\mathcal{A} = \sigma(\bigcup \mathcal{F}_n)$ where $(\mathcal{F}_n)$ is a filtration consisting of σ-algebras generated by finitely many atoms.

(1). Write down Lebesgue densities Z_n of Q w.r.t. P on each $\mathcal{F}_n$. Show that

$$Q[Z_n = \infty \quad \text{and} \quad Z_{n+1} < \infty] \; = \; 0 \qquad \text{for any } n,$$

and conclude that (Z_n) is a non-negative supermartingale under P, and $(1/Z_n)$ is a non-negative supermartingale and Q.

(2). Prove that the limit $Z_\infty = \lim Z_n$ exists both P-almost surely and Q-almost surely, and $P[Z_\infty < \infty] = 1$ and $Q[Z_\infty > 0] = 1$.

(3). conclude that Z_∞ is a Lebesgue density of P w.r.t. Q on $\mathcal{A}$, and $1/Z_\infty$ is a Lebesgue density of Q w.r.t. P on $\mathcal{A}$.

Derivations of monotone functions

Suppose that $F : [0,1] \to \mathbb{R}$ is a monotone and right-continuous function. After an appropriate linear transformation we may assume that F is non decreasing with $F(0) = 0$ and $F(1) = 1$. Let μ denote the probability measure with distribution function F. By the example above, the Radon-Nikodym derivative of μ w.r.t. Lebesgue measure on the σ-algebra $\mathcal{F}_n = \sigma(\mathcal{D}_n)$ generated by the n-th dyadic partition of the unit interval is given by the dyadic difference quotients (4.4.4) of F. By Theorem 4.19, we obtain a version of Lebesgue's Theorem on derivatives of monotone functions:

Corollary 4.20 (Lebesgue's Theorem). *Suppose that $F : [0,1] \to \mathbb{R}$ is monotone (and right continuous). Then the dyadic derivative*

$$F'(t) \;=\; \lim_{n\to\infty} \left.\frac{d\mu}{d\lambda}\right|_{\mathcal{F}_n}(t)$$

exists for almost every t and F' is an integrable function on $(0,1)$. Furthermore, if F is absolutely continuous then

$$F(s) \;=\; \int_0^s F'(t)\,dt \qquad \text{for all } s \in [0,1]. \tag{4.4.7}$$

Remark. Right continuity is only a normalization and can be dropped from the assumptions. Moreover, the assertion extends to function of bounded variation since these can be represented as the difference of two monotone functions, cf. **??** below. Similarly, (4.4.7) also holds for absolutely continuous functions that are not monotone. See e.g. [Elstrodt: Maß- und Integrationstheorie] for details.

Absolute continuity of infinite product measures

Suppose that $\Omega = \bigtimes_{i=1}^{\infty} S_i$, and

$$Q \;=\; \bigotimes_{i=1}^{\infty} \nu_i \qquad \text{and} \qquad P \;=\; \bigotimes_{i=1}^{\infty} \mu_i$$

are products of probability measures ν_i and μ_i defined on measurable spaces $(S_i, \mathcal{S}_i)$. We assume that ν_i and μ_i are mutually absolutely continuous for every $i \in \mathbb{N}$. Denoting by $X_k : \Omega \to S_k$ the evaluation of the k-th coordinate, the product measures are mutually absolutely continuous on each of the σ-algebras

$$\mathcal{F}_n \;=\; \sigma(X_1, \ldots, X_n), \qquad n \in \mathbb{N},$$

with relative densities

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_n} \;=\; Z_n \quad \text{and} \quad \left.\frac{dP}{dQ}\right|_{\mathcal{F}_n} \;=\; 1/Z_n,$$

where

$$Z_n \;=\; \prod_{i=1}^{n} \frac{d\nu_i}{d\mu_i}(X_i) \quad \in \;(0,\infty) \qquad P\text{-almost surely.}$$

In particular, (Z_n) is a martingale under P, and $(1/Z_n)$ is a martingale under Q. Let $\mathcal{F}_\infty = \sigma(X_1, X_2, \ldots)$ denote the product σ-algebra.

Theorem 4.21 (Kakutani's dichotomy). *The infinite product measures Q and P are either mutually absolutely continuous with relative density Z_∞. More precisely, the following statements are equivalent:*

(1). $Q \ll P$ *on* $\mathcal{F}_\infty$.

(2). $Q \approx P$ *on* $\mathcal{F}_\infty$.

(3). $\prod_{i=1}^{\infty} \int \sqrt{\frac{d\nu_i}{d\mu_i}} d\mu_i \;>\; 0.$

(4). $\sum_{i=1}^{\infty} d_H^2(\nu_i, \mu_i) \;<\; \infty.$

Here the squared Hellinger distance $d_H^2(\nu_i, \mu_i)$ of mutually absolutely continuous probability measures ν and μ is defined by

$$\begin{aligned} d_H^2 \;&=\; \frac{1}{2} \int \left(\sqrt{\frac{d\nu}{d\mu}} - 1 \right)^2 d\mu \;=\; \frac{1}{2} \int \left(\sqrt{\frac{d\mu}{d\nu}} - 1 \right)^2 d\nu \\ &=\; 1 - \int \sqrt{\frac{d\nu}{d\mu}}\, d\mu \;=\; 1 - \int \sqrt{\frac{d\mu}{d\nu}}\, d\nu. \end{aligned}$$

Remark. (1). If mutual absolutely continuity holds then the relative densities on $\mathcal{F}_\infty$ are

$$\frac{dQ}{dP} = \lim_{n\to\infty} Z_n \quad P\text{-almost surely, and} \quad \frac{dP}{dQ} = \lim_{n\to\infty} \frac{1}{Z_n} \quad Q\text{-almost surely.}$$

(2). If ν and μ are absolutely continuous w.r.t. a measure dx then

$$d_H^2(\nu,\mu) = \frac{1}{2}\int \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx = 1 - \int \sqrt{f(x)g(x)}\, dx.$$

Proof. (1) $\iff$ (3): For $i \in \mathbb{N}$ let $Y_i := \frac{d\nu_i}{d\mu_i}(X_i)$. Then the random variables Y_i are independent under both P and Q with $E_P[Y_i] = 1$, and

$$Z_n = Y_1 \cdot Y_2 \cdots Y_n.$$

By Theorem 4.19, the measure Q is absolutely continuous w.r.t. P if and only if the martingale (Z_n) is uniformly integrable. To obtain a sharp criterion for uniform integrability we switch from L^1 to L^2, and consider the non-negative martingale

$$M_n = \frac{\sqrt{Y_1}}{\beta_1} \cdot \frac{\sqrt{Y_2}}{\beta_2} \cdots \frac{\sqrt{Y_n}}{\beta_n} \qquad \text{with } \beta_i = E_P[\sqrt{Y_i}] = \int \sqrt{\frac{d\nu_i}{d\mu_i}}\, d\mu_i$$

under the probability measure P. Note that for $n \in \mathbb{N}$,

$$E[M_n^2] = \prod_{i=1}^n E[Y_i]/\beta_i^2 = 1 \Big/ \left(\prod_{i=1}^n \beta_i\right)^2.$$

If (3) holds then (M_n) is bounded in $L^2(\Omega, \mathcal{A}, P)$. Therefore, by Doob's L^2 inequality, the supremum of M_n is in $\mathcal{L}^2(\Omega, \mathcal{A}, P)$, i.e.,

$$E[\sup |Z_n|\,] = E[\sup M_n^2] < \infty.$$

Thus (Z_n) is uniformly integrable and $Q \ll P$ on $\mathcal{F}_\infty$.

Conversely, if (3) does not hold then

$$Z_n = M_N^2 \cdot \prod_{i=1}^n \beta_i \longrightarrow 0 \qquad P\text{-almost surely,}$$

since M_n converges to a finite limit by the martingale convergence theorem. Therefore, the absolute continuous part Q_a vanishes by Theorem 4.19 (3), i.e., Q is singular w.r.t. P.

(3) $\iff$ (4): For reals $\beta_i \in (0,1)$, the condition $\prod_{i=1}^{\infty} \beta_i > 0$ is equivalent to $\sum_{i=1}^{\infty}(1-\beta_i) < \infty$. For β_i as above, we have

$$1 - \beta_i \;=\; 1 - \int \sqrt{\frac{d\nu_i}{d\mu_i}}\, d\mu_i \;=\; d_H^2(\nu_i, \mu_i).$$

(2) $\Rightarrow$ (1) is obvious.

(4) $\Rightarrow$ (2): Condition (4) is symmetric in ν_i and μ_i. Hence, if (4) holds then both $Q \ll P$ and $P \ll Q$. □

Example (Gaussian products). Let $P = \bigotimes_{i=1}^{\infty} N(0,1)$ and $Q = \bigotimes_{i=1}^{\infty} N(a_i,1)$ where $(a_i)_{i\in\mathbb{N}}$ is a sequence of reals. The relative density of the normal distributions $\nu_i := N(a_i,1)$ and $\mu := N(0,1)$ is

$$\frac{d\nu_i}{d\mu}(x) \;=\; \frac{\exp(-(x-a_i)^2)/2}{\exp(-x^2/2)} \;=\; \exp(a_i x - a_i^2/2),$$

and

$$\int \sqrt{\frac{d\nu_i}{d\mu}}\, d\mu \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x^2 - a_i x + a_i^2/2)\right) dx \;=\; \exp(-a_i^2/8).$$

Therefore, by condition (3) in Theorem 4.21,

$$Q \ll P \iff Q \approx P \iff \sum_{i=1}^{\infty} a_i^2 < \infty.$$

Hence mutual absolute continuity holds for the infinite products if and only if the sequence (a_i) is contained in ℓ^2, and otherwise Q and P are singular.

Remark (Relative entropy). (1). In the singular case, the exponential rate if degeneration of the relative densities on the σ-algebras $\mathcal{F}_n$ is related to the relative entropies

$$H(\nu_i \mid \mu_i) \;=\; \int \frac{d\nu_i}{d\mu_i} \log \frac{d\nu_i}{d\mu_i}\, d\mu_i \;=\; \int \log \frac{d\nu_i}{d\mu_i}\, d\nu_i.$$

For example in the i.i.d. case $\mu_i \equiv \mu$ and $\nu_i \equiv \nu$, we have

$$\frac{1}{n}\log Z_n \;=\; \frac{1}{n}\sum_{i=1}^{n}\log\frac{d\nu}{d\mu}(X_i) \;\longrightarrow\; H(\nu\,|\,\mu) \qquad Q\text{-a.s., and}$$

$$-\frac{1}{n}\log Z_n \;=\; \frac{1}{n}\log Z^{-1} \;\longrightarrow\; H(\mu\,|\,\nu) \qquad P\text{-a.s.}$$

as $n \to \infty$ by the Law of Large Numbers.
In general, $\log Z_n - \sum\limits_{i=1}^{n} H(\nu_i|\mu_i)$ is a martingale w.r.t. Q, and $\log Z_n + \sum\limits_{i=1}^{n} H(\nu_i|\mu_i)$ is a martingale w.r.t. P.

(2). The relative entropy is related to the squared Hellinger distance by the inequality

$$\frac{1}{2}H(\nu\,|\,\mu) \;\geq\; d_H^2(\nu\,|\,\mu),$$

which follows from the elementary inequality

$$\frac{1}{2}\log x^{-1} \;=\; -\log\sqrt{x} \;\geq\; 1-\sqrt{x} \qquad \text{for } x > 0.$$

4.5 Translations of Wiener measure

We now return to stochastic processes in continuous time. We endow the continuous path space $C([0,\infty),\mathbb{R}^d)$ with the σ-algebra generated by the evolution maps $X_t(\omega) = \omega(t)$, and with the filtration

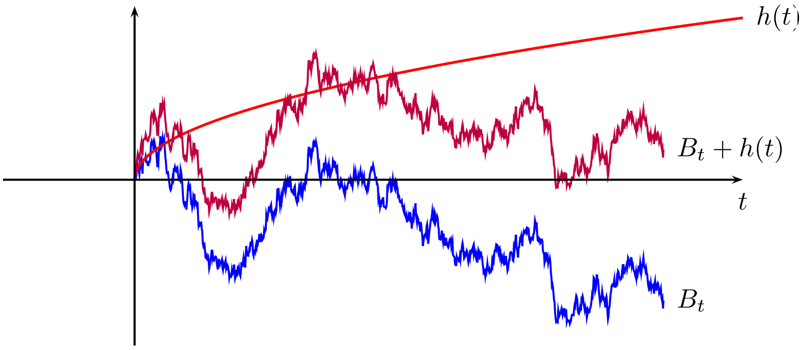
$$\mathcal{F}_t^X \;=\; \sigma(X_s \mid s \in [0,t]), \qquad t \geq 0.$$

Note that $\mathcal{F}_t^X$ consists of all sets of type

$$\left\{\omega \in C([0,\infty),\mathbb{R}^d) \;:\; \omega|_{[0,t]} \in \Gamma\right\} \qquad \text{with } \Gamma \in \mathcal{B}(C([0,t],\mathbb{R}^d)).$$

In many situations one is interested in the distribution on path space of a process

$$B_t^h \;=\; B_t + h(t)$$



obtained by translating a Brownian motion (B_t) by a deterministic function $h : [0, \infty) \to \mathbb{R}^d$. In particular, it is important to know if the distribution of (B_t^h) has a density w.r.t. the Wiener measure on the σ-algebras $\mathcal{F}_t^X$, and how to compute the densities if they exist.

Example. (1). Suppose we would like to evaluate the probability that $\sup\limits_{s\in[0,t]} |B_s - g(s)| < \varepsilon$ for a given $t > 0$ and a given function $g \in C([0, \infty), \mathbb{R}^d)$ asymptotically as $\varepsilon \searrow 0$. One approach is to study the distribution of the translated process $B_t - g(t)$ near 0.

(2). Similarly, computing the passage probability $P[B_s \geq a+bs \quad \text{for some } s \in [0, t]]$ to a line $s \mapsto a + bs$ for a one-dimensional Brownian motion is equivalent to computing the passage probability to the point a for the translated process $B_t - bt$.

(3). A solution to a stochastic differential equation

$$dY_t = dB_t + b(t, Y_t)dt$$

is a translation of the Brownian motion $(B_t - B_0)$ by the stochastic process $H_t = Y_0 + \int\limits_0^t b(s, Y_s)\, ds$, cf. below.
Again in the simplest case (when $b(t, y)$ only depends on t), H_t is a deterministic function.

The Cameron-Martin Theorem

Let (B_t) denote a continuous Brownian motion with $B_0 = 0$, and let $h \in C([0,\infty),\mathbb{R}^d)$. The distribution

$$\mu_h \ := \ P \circ (B+h)^{-1}$$

of the translated process $B_t^h = B_t + h(t)$ is the image of Wiener measure μ_0 under the translation map

$$\tau_h \ : \ C([0,\infty),\mathbb{R}^d) \ \longrightarrow \ C([0,\infty),\mathbb{R}^d), \quad \tau_h(x) \ = \ x+h.$$

Recall that Wiener measure is a Gaussian measure on the infinite dimensional space $C([0,\infty),\mathbb{R}^d)$. The next exercise discusses translations of Gaussian measures in $\mathbb{R}^n$:

Exercise (Translations of normal distributions). Let $C \in \mathbb{R}^{n\times n}$ be a symmetric non-negative definite matrix, and let $h \in \mathbb{R}^n$. the image of the normal distribution $N(0,C)$ under the translation map $x \mapsto x+h$ on $\mathbb{R}^n$ is the normal distribution $N(h,C)$.

(1). Show that if C is non-degenerate then $N(h,C) \approx N(0,C)$ with relative density

$$\frac{dN(h,C)}{dN(0,C)}(x) \ = \ e^{(h,x)-\frac{1}{2}(h,h)} \qquad \text{for } x \in \mathbb{R}^n, \tag{4.5.1}$$

where $(g,h) = (g, C^{-1}, h)$ for $g,h \in \mathbb{R}^n$.

(2). Prove that in general, $N(h,C)$ is absolutely continuous w.r.t. $N(0,C)$ if and only if h is orthogonal to the kernel of C w.r.t. the Euclidean inner product.

On $C([0,\infty),\mathbb{R}^d)$, we can usually not expect the existence of a global density of the translated measures μ_h w.r.t. μ_0. The Cameron-Martin Theorem states that for $t \geq 0$, a relative density on $\mathcal{F}_t^X$ exists if and only if h is contained in the corresponding Cameron-Martin space:

Theorem 4.22 (Cameron, Martin). *For $h \in C([0,\infty),\mathbb{R}^d)$ and $t \geq 0$ the translated measure $\mu_h = \mu \circ \tau_h^{-1}$ is absolutely continuous w.r.t. Wiener measure μ_0 on $\mathcal{F}_t^X$ if and only if h is an absolutely continuous function on $[0,t]$ with $h(0) = 0$ and $\int_0^t |h'(s)|^2\,ds <$*

∞.

In this case, the relative density is given by

$$\left.\frac{d\mu_h}{d\mu_0}\right|_{\mathcal{F}_t^X} = \exp\left(\int_0^t h'(s)\,dX_s - \frac{1}{2}\int_0^t |h'(s)|^2\,ds\right), \tag{4.5.2}$$

where the stochastic integral $\int_0^t h'(s)\,dX_s$ *is defined by*

$$\int_0^t h'(s)\,dX_s = \lim_{n\to\infty}\sum_{k=0}^{2^n-1}\frac{h((k+1)t/2^n)-h(kt/2^n)}{t/2^n}\cdot\left(X_{(k+1)t/2^n}-X_{kt/2^n}\right)$$

with convergence in $L^2(C([0,\infty),\mathbb{R}^d),\mathcal{F}_t^X,\mu_0)$ *and* μ_0*-almost surely.*

Before giving a rigorous proof let us explain heuristically why the result should be true. Clearly, absolute continuity does not holds if $h(0)\neq 0$, since then the translated paths do not start at 0. Now suppose $h(0)=0$, and fix $t\in(0,\infty)$. Absolute continuity on $\mathcal{F}_t^X$ means that the distribution μ_h^t of $(B_s^h)_{0\le s\le t}$ on $C([0,\infty),\mathbb{R}^d)$ is absolutely continuous w.r.t. Wiener measure μ_0^t on this space. The measure μ_0^t, however, is a kind of infinite dimensional standard normal distribution w.r.t.

$$(x,y)_H = \int_0^t x'(s)\cdot y'(s)\,ds$$

on function $x,y:[0,t]\to\mathbb{R}^d$ vanishing at 0, and the translated measure μ_h^t is a Gaussian measure with mean h and the same covariances.

Choosing an orthonormal basis $(e_i)_{i\in\mathbb{N}}$ w.r.t. the H-inner product (e.g. Schauder functions), we can identify μ_0^t and μ_h^t with the product measures $\bigotimes_{i=1}^{\infty}N(0,1)$ and $\bigotimes_{i=1}^{\infty}N(\alpha_i,1)$ respectively where $\alpha_i=(h,e_i)_H$ is the i-th coefficient of h in the basis expansion. Therefore, μ_h^t should be absolutely continuous w.r.t. μ_0^t if and only if

$$(h,h)_H = \sum_{i=1}^{\infty}\alpha_i^2 < \infty,$$

i.e., if and only if h is absolutely continuous with $h'\in\mathcal{L}^2(0,t)$.

Moreover, in analogy of the finite-dimensional case (4.5.1), we would expect informally a relative density of the type

$$\text{“}\ \frac{d\mu_h^t}{d\mu_0^t}(x) \;=\; e^{(h,x)_H - \frac{1}{2}(h,h)_H} \;=\; \exp\left(\int_0^t h'(s) - x'(s)\,ds - \frac{1}{2}\int_0^t |h'(s)|^2\,ds \right)\ \text{”}$$

Since μ_0^t-almost every path $x \in C([0,\infty),\mathbb{R}^d)$ is not absolutely continuous, this expression does not make sense. Nevertheless, using finite dimensional approximations and martingale methods, we can derive the rigorous expression (4.5.2) for the relative density where the integral $\int h'x'\,ds$ is replaced by the almost surely well-defined stochastic integral $\int_0^t h'\,dx$:

Proof of Theorem 4.22. We assume $t = 1$. The proof for other values of t is similar. Moreover, as explained above, it is enough to consider the case $h(0) = 0$.

(1). *Local densities:* We first compute the relative densities when the paths are only evaluated at dyadic time points. Fix $n \in \mathbb{N}$, let $t_i = i \cdot 2^{-n}$, and let

$$\Delta_i x \;=\; x_{t_{i+1}} - x_{t_i}$$

denote the i-th dyadic increment. Then the increments $\Delta_i B^h$ $(i = 0, 1, \ldots, 2^n - 1)$ of the translated Brownian motion are independent random variables with distributions

$$\Delta_i B^h \;=\; \Delta_i B + \Delta_i h \;\sim\; N(\Delta_i h, (\Delta t)\cdot I_d), \qquad \Delta t = 2^{-n}.$$

Consequently, the marginal distribution of $(B^h_{t_1}, B^h_{t_2}, \ldots, B^h_{t_{2^n}})$ is a normal distribution with density w.r.t. Lebesgue measure proportional to

$$\exp\left(-\sum_{i=0}^{2^n-1} \frac{|\Delta_i x - \Delta_i h|^2}{2\Delta t} \right), \qquad x = (x_{t_1}, x_{t_2}, \ldots, x_{t_{2^n}}) \in \mathbb{R}^{2^n d}.$$

Since the normalization constant does not depend on h, the joint distribution of $(B^h_{t_1}, B^h_{t_2}, \ldots, B^h_{t_{2^n}})$ is absolutely continuous w.r.t. that of $(B_{t_1}, B_{t_2}, \ldots, B_{t_{2^n}})$ with relative density

$$\exp\left(\sum \frac{\Delta_i h}{\Delta t} \cdot \Delta_i x - \frac{1}{2}\left|\frac{\Delta_i h}{\Delta t}\right|^2 \Delta t \right) \tag{4.5.3}$$

and consequently, μ_h is always absolutely continuous w.r.t. μ_0 on each of the σ-algebras

$$\mathcal{F}_n \;=\; \sigma(X_{i\cdot 2^{-n}} \mid i = 0, 1, \ldots, 2^n - 1), \qquad n \in \mathbb{N},$$

with relative densities

$$Z_n \;=\; \exp\left(\sum_{i=0}^{2^n-1} \frac{\Delta_i h}{\Delta t} \cdot \Delta_i X - \frac{1}{2}\sum_{i=0}^{2^n-1} \left|\frac{\Delta_i h}{\Delta t}\right|^2 \Delta t\right). \tag{4.5.4}$$

(2). *Limit of local densities:* Suppose that h is absolutely continuous with

$$\int_0^1 |h'(t)|^2 \, dt \;<\; \infty.$$

We now identify the limit of the relative densities Z_n as $n \to \infty$.
First, we note that

$$\sum_{i=0}^{2^n-1} \left|\frac{\Delta_i h}{\Delta t}\right|^2 \Delta t \quad\longrightarrow\quad \int_0^1 |h'(t)|^2 \, dt \qquad \text{as } n \to \infty.$$

In fact, the sum on the right hand side coincides with the squared L^2 norm

$$\int_0^1 |dh/dt|^2_{\sigma(\mathcal{D}_n)} \; dt$$

of the dyadic derivative

$$\left.\frac{dh}{dt}\right|_{\sigma(\mathcal{D}_n)} \;=\; \sum_{i=0}^{2^n-1} \frac{\Delta_i h}{\Delta t} \cdot I_{((i-1)\cdot 2^{-n}, i\cdot 2^{-n}]}$$

on the σ-algebra generated by the intervals $((i-1)\cdot 2^{-n}, i\cdot 2^{-n}]$ iff h is absolutely continuous with $h' \in L^2(0,1)$ then $\left.\frac{dh}{dt}\right|_{\sigma(\mathcal{D}_n)} \to h'(t)$ in $L^2(0,1)$ by the L^2 martingale convergence theorem. To study the convergence as $n \to \infty$ of the random sums

$$M_n = \sum_{i=0}^{2^n-1} \frac{\Delta_i h}{\Delta t} \cdot \Delta_i X \;=\; \sum_{i=0}^{2^n-1} 2^n \cdot (h((i+1)2^{-n}) - h(i2^{-n})) \cdot (X_{(i+1)2^{-n}} - X_{i2^{-n}}),$$

we note that under Wiener measure, (M_n) is a martingale w.r.t. the filtration $(\mathcal{F}_n)$. In fact, for $n \geq 0$, the conditional expectations w.r.t. Wiener measure of the $(n+1)$th dyadic increments are

$$\begin{aligned} E_{\mu_0}[X_{(i+1)2^{-n}} - X_{(i+\frac{1}{2})2^{-n}} \mid \mathcal{F}_n] &= E_{\mu_0}[X_{(i+\frac{1}{2})2^{-n}} - X_{i2^{-n}} \mid \mathcal{F}_n] \\ &= (X_{(i+1)2^{-n}} - X_{i2^{-n}})/2. \end{aligned}$$

Therefore,

$$\begin{aligned} E_{\mu_0}[M_{n+1} \mid \mathcal{F}_n] &= \sum_{i=0}^{2^n-1} 2^{n+1} \left(h((i+1)2^{-n}) - h((i+\frac{1}{2})2^{-n}) + h((i+\frac{1}{2})2^{-n}) - h(i2^{-n}) \right) \cdot \\ &\quad (X_{(i+1)2^{-n}} - X_{i2^{-n}})/2 \;=\; M_n. \end{aligned}$$

The martingale (M_n) is bounded in $L^2(\mu_0)$ because

$$E_{\mu_0}[M_n^2] = \sum_{i=0}^{2^n-1} \left| \frac{\Delta_i h}{\Delta t} \right|^2 \Delta t.$$

Therefore, the L^2 and almost sure martingale convergence theorems yield the existence of the limit

$$\int_0^1 h'(s)\, dX_s = \lim_{n\to\infty} M_n$$

in $L^2(\mu_0)$ and μ_0-almost surely. Summarizing, we have shown

$$\lim_{n\to\infty} Z_N = \exp\left(\int_0^1 h'(s)\, dX_s - \frac{1}{2} \int_0^1 |h'(s)|^2\, ds \right) \qquad \mu_0\text{-almost surely.} \tag{4.5.5}$$

(3). *Absolute continuity on* $\mathcal{F}_1^{\mathbf{X}}$*:* We still assume $h' \in L^2(0,1)$. Note that $\mathcal{F}_1^X = \sigma(\bigcup \mathcal{F}_n)$. Hence for proving that μ_h is absolutely continuous w.r.t. μ_0 on $\mathcal{F}_1^X$ with density given by (4.5.5), it suffices to show that $\limsup Z_n < \infty$ μ_h-almost surely (i.e., the singular part in the Lebesgue decomposition of μ_h w.r.t. μ_0 vanishes). Since $\mu_h = \mu_0 \circ \tau_h^{-1}$ the process

$$W_t = X_t - h(t) \qquad \text{is a Brownian motion w.r.t. } \mu_h,$$

and by (4.5.3) and (4.5.4),

$$Z_n \;=\; \exp\left(\sum_{i=0}^{2^n-1} \frac{\Delta_i h}{\Delta t} \cdot \Delta_i W + \frac{1}{2} \sum_{i=0}^{2^n-1} \left| \frac{\Delta_i h}{\Delta t} \right|^2 \Delta t \right).$$

Note that the minus sign in front of the second sum has turned into a plus by the translation! Arguing similarly as above, we see that (Z_n) converges μ_h-almost surely to a finite limit:

$$\lim Z_n \;=\; \exp\left(\int_0^1 h'(s)\, dW_s + \frac{1}{2} \int_0^1 |h'(s)|^2\, ds \right) \qquad \mu_h\text{-a.s.}$$

Hence $\mu_h \ll \mu_0$ with density $\lim Z_n$.

(4). *Singularity on $\mathcal{F}_1^X$:* Conversely, let us suppose now that h is not absolutely continuous or h' is not in $L^2(0,1)$. Then

$$\sum_{i=0}^{2^n-1} \left| \frac{\Delta_i h}{\Delta_i t} \right|^2 \Delta t \;=\; \int_0^1 \left| \frac{dh}{dt} \right|^2_{\sigma(\mathcal{D}_n)} dt \quad \longrightarrow \quad \infty \qquad \text{as } n \to \infty.$$

Since

$$\left\| \sum_{i=0}^{2^n-1} \frac{\Delta_i h}{\Delta_t} \cdot \Delta_i X \right\|_{L^2(\mu_0)} \;=\; E_{\mu_0}[M_n^2]^{1/2} \;=\; \left(\sum_{i=0}^{2^n-1} \left(\frac{\Delta_i h}{\Delta t} \right)^2 \Delta t \right)^{1/2},$$

we can conclude by (4.5.3) and (4.5.4) that

$$\lim Z_n \;=\; 0 \qquad \mu_0\text{-almost surely,}$$

i.e., μ_h is singular w.r.t. μ_0.

□

Passage times for Brownian motion with constant drift

We now consider a one-dimensional Brownian motion with constant drift β, i.e., a process

$$Y_t \;=\; B_t + \beta t, \qquad t \geq 0,$$

where B_t is a Brownian motion starting at 0 and $\beta \in \mathbb{R}$. We will apply the Cameron-Martin Theorem to compute the distributions of the first passage times

$$T_a^Y \;=\; \min\{t \geq 0 \;:\; Y_t = a\}, \qquad a > 0.$$

Note that T_a^Y is also the first passage time to the line $t \mapsto a - \beta t$ for the Brownian motion (B_t).

Theorem 4.23. *For $a > 0$ and $\beta \in \mathbb{R}$, the restriction of the distribution of T_a^Y to $(0, \infty)$ is absolutely continuous with density*

$$f_{a,\beta}(t) \;=\; \frac{a}{\sqrt{2\pi t^3}} \exp\left(-\frac{(a-\beta t)^2}{2t}\right).$$

In particular,

$$P[T_a^Y < \infty] \;=\; \int\limits_0^\infty f_{a,\beta}(s)\,ds.$$

Proof. Let $h(t) = \beta t$. By the Cameron-Martin Theorem, the distribution μ_h of (Y_t) is absolutely continuous w.r.t. Wiener measure on $\mathcal{F}_t^X$ with density

$$Z_t \;=\; \exp(\beta \cdot X_i - \beta^2 t/2).$$

Therefore, denoting by $T_0 = \min\{t \geq 0 \;:\; X_t = a\}$ the passage time of the canonical process, we obtain

$$\begin{aligned}
P[T_a^Y \leq t] \;&=\; \mu_h[T_a \leq t] \;=\; E_{\mu_0}[Z_t\,;\,T_a \leq t] \\
&=\; E_{\mu_0}[Z_{T_a}\,;\,T_a \leq t] \;=\; E_{\mu_0}[\exp(\beta a - \frac{1}{2}\beta^2 T_a)\,;\,T_a \leq t] \\
&=\; \int\limits_{0^t} \exp(\beta a - \beta^2 s/2) f_{T_a}(s)\,ds
\end{aligned}$$

by the optional sampling theorem. The claim follows by inserting the explicit expression for f_{T_a} derived in Corollary 1.25. □

Chapter 5

Stochastic Integral w.r.t. Brownian Motion

Suppose that we are interested in a continuous-time scaling limit of a stochastic dynamics of type $X_0^{(h)} = x_0$,

$$X_{k+1}^{(h)} - X_k^{(h)} \;=\; \sigma(X_k^{(h)}) \cdot \sqrt{h} \cdot \eta_{k+1}, \qquad k = 0, 1, 2, \ldots, \tag{5.0.1}$$

with i.i.d. random variables $\eta_i \in \mathcal{L}^2$ such that $E[\eta_i] = 0$ and $\mathrm{Var}[\eta_i] = 1$, a continuous function $\sigma : \mathbb{R} \to \mathbb{R}$, and a scale factor $h > 0$. Equivalently,

$$X_n^{(h)} \;=\; X_0^{(h)} + \sqrt{h} \cdot \sum_{k=0}^{n-1} \sigma(X_k^{(h)}) \cdot \eta_{k+1}, \qquad n = 0, 1, 2, \ldots. \tag{5.0.2}$$

If σ is constant then as $h \searrow 0$, the rescaled process $(X_{\lfloor t/h \rfloor}^{(h)})_{t \geq 0}$ converges in distribution to $(\sigma \cdot B_t)$ where (B_t) is a Brownian motion. We are interested in the scaling limit for general σ. One can prove that the rescaled process again converges in distribution, and the limit process is a solution of a stochastic integral equation

$$X_t \;=\; X_0 + \int_0^t \sigma(X_s)\, dB_s, \qquad t \geq 0. \tag{5.0.3}$$

Here the integral is an Itô stochastic integral w.r.t. a Brownian motion (B_t). Usually the equation (5.0.3) is written briefly as

$$dX_t \;=\; \sigma(X_t)\, dB_t, \tag{5.0.4}$$

and interpreted as a stochastic differential equation. Stochastic differential equations occur more generally when considering scaling limits of appropriately rescaled Markov chains on $\mathbb{R}^d$ with finite second moments. The goal of this section is to give a meaning to the stochastic integral, and hence to the equations (5.0.3), (5.0.4) respectively.

Example (**Stock prices, geometric Brownian motion**). A simple discrete time model for stock prices is given by

$$X_{k+1} - X_k \;=\; X_k \cdot \eta_{k+1}, \qquad \eta_i \text{ i.i.d.}$$

To set up a corresponding continuous time model we consider the rescaled equation (5.0.1) as $h \searrow 0$. The limit in distribution is a solution of a stochastic differential equation

$$dX_t \;=\; X_t \, dB_t \tag{5.0.5}$$

w.r.t. a Brownian motion (B_t). Although with probability one, the sample paths of Brownian motion are nowhere differentiable with probability one, we can give a meaning to this equation by rewriting it in the form (5.0.3) with an Itô stochastic integral.

A naive guess would be that the solution of (5.0.5) with initial condition $X_0 = 1$ is $X_t = \exp B_t$. However, more careful considerations show that this can not be true! In fact, the discrete time approximations satisfy

$$X_{k+1}^{(h)} \;=\; (1 + \sqrt{h}\eta_{k+1}) \cdot X_k^{(h)} \qquad \text{for } k \geq 0.$$

Hence $(X_k^{(h)})$ is a product martingale:

$$X_n^{(h)} \;=\; \prod_{k=1}^{n} (1 + \sqrt{h}\eta_k) \qquad \text{for any } n \geq 0.$$

In particular, $E[X_n^{(h)}] = 1$. We would expect similar properties for the scaling limit (X_t), but $\exp B_t$ is not a martingale and $E[\exp(B_t)] = \exp(t/2)$.

It turns out that in fact, the unique solution of (5.0.5) with $X_0 = 1$ is not $\exp(B_t)$ but the exponential martingale

$$X_t \;=\; \exp(B_t - t/2),$$

which is also called a geometric Brownian motion. The reason is that the irregularity of Brownian paths enforces a corrections term in the chain rule for stochastic differentials leading to Itô's famous formula, which is the fundament of stochastic calculus.

5.1 Defining stochastic integrals: A first attempt and a warning

Let us first fix some notation that will be used constantly below:

Basic notation

By a **partition** π of $\mathbb{R}_+$ we mean an increasing sequence $0 = t_0 < t_1 < t_2 < \ldots$ such that $\sup t_n = \infty$. The *mesh size* of the partition is

$$\mathrm{mesh}(\pi) \;=\; \sup\{|t_i - t_{i-1}| \;:\; i \in \mathbb{N}\}.$$

We are interested in defining the integrals of type

$$I_t \;=\; \int_0^t H_s \, dX_s, \qquad t \geq 0, \tag{5.1.1}$$

for continuous functions and, respectively, continuous adapted processes (H_s) and (X_s).

For a given $t \geq 0$ and a given partition π of $\mathbb{R}_+$, we define the increments of (X_s) up to time t by

$$\Delta X_s \;:=\; X_{s'\wedge t} - X_{s\wedge t} \qquad \text{for any } s \in \pi,$$

where $s' := \min\{u \in \pi \;:\; u > s\}$ denotes the next partition point after s. Note that the increments ΔX_s vanish for $s \geq t$. In particular, only finitely many of the increments are not equal to zero. A nearby approach for defining the integrals I_T in (5.1.1) would be Riemann sum approximations:

Riemann sum approximations

There are various possibilities to define approximating Riemann sums w.r.t. a given sequence (π_n) of partitions with $\mathrm{mesh}(\pi_n) \to 0$, for example:

Variant 1 (non-anticipative): $I_t^n \;=\; \sum_{s\in\pi_n} H_s \Delta X_s$,

Variant 2 (anticipative): $\hat{I}_t^n \;=\; \sum_{s\in\pi_n} H_{s'} \Delta X_s$, *Variant 3 (anticipative):* $\overset{\circ}{I}{}_t^n \;=\; \sum\limits_{s\in\pi_n} \frac{1}{2}(H_s + H_{s'})\Delta X_s$.

Note that for finite t, in each of the sums, only finitely many summands do not vanish. For example,

$$I_t^n \;=\; \sum_{\substack{s\in\pi_n\\ s<t}} H_s \Delta X_s \;=\; \sum_{\substack{s\in\pi_n\\ s<t}} H_s\cdot (X_{s'\wedge t} - X_s).$$

Now let us consider at first the case where $H_s = X_s$ and $t = 1$, i.e., we would like to define the integral $I = \int\limits_0^1 X_s\, dX_s$. Suppose first that $X : [0,1] \to \mathbb{R}$ is a continuous function of bounded variation, i.e.,

$$V^{(1)}(X) \;=\; \sup\left\{ \sum_{s\in\pi} |\Delta X_s| \;:\; \pi \text{ partition of } \mathbb{R}_+ \right\} < \infty.$$

Then for $H = X$ and $t = 1$ all the approximations above converge to the same limit as $n \to \infty$. For example,

$$\|\hat{I}_1^n - I_1^n\| \;=\; \sum_{s\in\pi_n} (\Delta X_s)^2 \;\le\; V^{(1)}(X)\cdot \sup_{s\in\pi_n} |\Delta X_s|,$$

and the right-hand side converges to 0 by uniform continuity of X on $[0,1]$. In this case the limit of the Riemann sums is a Riemann-Stieltjes integral

$$\lim_{n\to\infty} I_1^n \;=\; \lim_{n\to\infty} \hat{I}_1^n \;=\; \int\limits_0^1 X_s\, dX_s,$$

which is well-defined whenever the integrand is continuous and the integrator is of bounded variation or conversely. The sample paths of Brownian motion, however, are almost surely not of bounded variation. Therefore, the reasoning above does not apply, and in fact if $X_t = B_t$ is a one-dimensional Brownian motion and $H_t = X_t$ then

$$E[\,|\hat{I}_1^n - I_1^n|\,] \;=\; \sum_{s\in\pi_n} E[(\Delta B_s)^2] \;=\; \sum_{s\in\pi_n} \Delta s \;=\; 1,$$

i.e., the L^1-limits of the random sequence (I_1^n) and $(\hat{I}_1^n)$ are different if they exist. Below we will see that indeed the limits of the sequences (I_1^n), $(\hat{I}_1^n)$ and $(\overset{\circ}{I}{}_1^n)$ do exist in L^2, and all the limits are different. The limit of the non-anticipative Riemann sums I_1^n

is the *Itô stochastic integral* $\int_0^1 B_s\, dB_s$, the limit of $(\hat{I}_1^n)$ is the *backward Itô integral* $\int_0^1 B_s\, \hat{d}B_s$, and the limit of I_n° is the *Stratonovich integral* $\int_0^1 B_s \circ dB_s$. All three notions of stochastic integrals are relevant. The most important one is the Itô integral because the non-anticipating Riemann sum approximations imply that the Itô integral $\int_0^t H_s\, dB_s$ is a continuous time martingale transform of Brownian motion if the process (H_s) is adapted.

Itô integrals for continuous bounded integrands

We now give a first existence proof for Itô integrals w.r.t. Brownian motion. We start with a provisional definition that will ne made more precise later:

Definition. *For continuous functions or continuous stochastic processes (H_s) and (X_s) and a given sequence (π_n) of partitions with $\mathrm{mesh}(\pi_n) \to 0$, the* ***Itô integral of*** H ***w.r.t.*** X *is defined by*

$$\int_0^t H_s\, dX_s \;=\; \lim_{n\to\infty} \sum_{s\in\pi_n} H_s \Delta X_s$$

whenever the limit exists in a sense to be specified.

Note that the definition is vague since the mode of convergence is not specified. Moreover, the Itô integral might depend on the sequence (π_n), In the following sections we will see which kind of convergence holds in different circumstances, and in which sense the limit is independent of (π_n).

To get started let us consider the convergence of Riemann sum approximations for the Itô integrals $\int_0^t H_s\, dB_s$ of a bounded $(\mathcal{F}_s^B)$ adapted process $(H_s)_{s\geq 0}$ w.r.t. a Brownian motion (B_t). Let (π_n) be a fixed sequence of partitions with $\pi_n \subseteq \pi_{n+1}$ and $\mathrm{mesh}(\pi_n) \to 0$. Then for the Riemann-Itô sums

$$I_t^n \;=\; \sum_{s\in\pi_n} H_s \Delta B_s \;=\; \sum_{\substack{s\in\pi_n \\ s<t}} H_s (B_{s'\wedge t} - B_s)$$

we have

$$I_t^n - I_t^m = \sum_{\substack{s\in\pi_n\\ s<t}} (H_s - H_{\lfloor s \rfloor_m})\Delta B_s \qquad \text{for any } m \le n$$

where $\lfloor s \rfloor_m = \max\{r \in \pi_m : r \le s\}$ denotes the next partition point on π_m below s. Since Brownian motion is a martingale, we have $E[\Delta B_s \mid \mathcal{F}_s^B] = 0$ for any $s \in \pi_n$. Moreover, $E[(\Delta B_s)^2 \mid \mathcal{F}_s^B] = \Delta s$. Therefore, we obtain by conditioning on $\mathcal{F}_s^B, \mathcal{F}_r^B$ respectively:

$$\begin{aligned} E[(I_t^n - I_t^m)^2] &= \sum_{\substack{s\in\pi_n\\ s<t}} \sum_{\substack{r\in\pi_n\\ r<t}} E[(H_s - H_{\lfloor s \rfloor_m})(H_r - H_{\lfloor r \rfloor_m})\Delta B_s \Delta B_r] \\ &= \sum_{\substack{s\in\pi_n\\ s<t}} E[(H_s - H_{\lfloor s \rfloor_m})\Delta s] \\ &\le E[V_m] \cdot \sum_{\substack{s\in\pi_n\\ s<t}} \Delta s = E[V_m] \cdot t \end{aligned}$$

where

$$V_m := \sup_{|s-r|<\mathrm{mesh}(\pi_m)} (H_s - H_r)^2 \quad \longrightarrow \quad 0 \qquad \text{as } m \to \infty$$

by uniform continuity of (H_s) on $[0,t]$. Since H is bounded, $E[V_m] \to 0$ as $m \to \infty$, and hence (I_t^n) is a Cauchy sequence in $L^2(\Omega, \mathcal{A}, P)$ for any given $t \ge 0$. This proves that for any fixed $t \ge 0$, the Itô integral

$$\int_0^t H_s \, dB_s = \lim_{n\to\infty} I_t^n \tag{5.1.2}$$

exists as a limit in L^2. Arguing more carefully, one observes that the process $(I_t^n)_{t\ge 0}$ given by the n-th Riemann sum approximations is an L^2 bounded continuous martingale on any finite interval $[0,u], u \in (0,\infty)$. Therefore, the maximal inequality implies that the convergence in (5.1.2) even holds uniformly in t for $t \in [0,u]$ in the $L^2(P)$ sense. In particular, the Itô integral $t \mapsto \int_0^t H_s \, dB_s$ is again a continuous martingale.

A similar argument applies if Brownian motion is replaced by a bounded martingale with continuous sample paths, cf. Section **??** below. In the rest of this section we will work out the construction of the Itô integral w.r.t. Brownian motion more systematically and for a broader class of integrands.

5.2 Simple integrands and Itô isometry for BM

Let $(M_t)_{t\geq 0}$ be a continuous martingale w.r.t. a filtration $(\mathcal{F}_t)$ on a probability space $(\Omega, \mathcal{A}, P)$. Right now, we will mainly be interested in the case where (M_t) is a Brownian motion. We would like to define stochastic integrals $\int_0^t H_s\, dM_s$.

Predictable step functions

In a first step, we define the integrals for predictable step functions (H_t) of type

$$H_t(\omega) \;=\; \sum_{i=0}^{n-1} A_i(\omega) I_{(t_i,T_{i+1}]}(t)$$

with $n \in \mathbb{N}, 0 \leq t_0 < t_1 < t_2 < \ldots < t_n$, and bounded $\mathcal{F}_{t_i}$-measurable random variables $A_i, i = 0, 1, \ldots, n-1$. Let $\mathscr{E}$ denote the vector space consisting of all stochastic processes of this form.

Definition (Itô integral for predictable step functions). *For stochastic processes $H \in \mathscr{E}$ and $t \geq 0$ we define*

$$\int_0^t H_s\, dM_s \;:=\; \sum_{i=0}^{n-1} A_i \cdot (M_{t_{i+1}\wedge t} - M_{t_i \wedge t}) \;=\; \sum_{i\,:\,t_i<t} A_i \cdot (M_{t_{i+1}\wedge t} - M_{t_i}).$$

The stochastic processes $H_\bullet M$ given by

$$(H_\bullet M)_t \;:=\; \int_0^t H_s\, dM_s \qquad \textit{for } t \in [0,\infty]$$

*is called the **martingale transform** of M w.r.t. H.*

Note that the map $H \mapsto H_\bullet M$ is linear. The process $H_\bullet M$ models for example the net gain up to time t if we hold A_i units of an asset with price process (M_t) during each of the time intervals $(t_i, t_{i+1}]$.

Lemma 5.1. *For any $H \in \mathscr{E}$, the process $H_\bullet M$ is a continuous $(\mathcal{F}_t)$ martingale up to time $t = \infty$.*

Similarly to the discrete time case, the fact that A_i is predictable, i.e., $\mathcal{F}_{t_i}$-measurable, is essential for the martingale property:

Proof. By definition, $H_\bullet M$ is continuous and $(\mathcal{F}_t)$ adapted. It remains to verify that

$$E[(H_\bullet M)_t \mid \mathcal{F}_s] \;=\; (H_\bullet M)_s \qquad \text{for any } 0 \le s \le t. \tag{5.2.1}$$

We do this in three steps:

(1). At first we note that (5.2.1) holds for $s, t \in \{t_0, t_1, \ldots, t_n\}$.
Indeed, since A_i is $\mathcal{F}_{t_i}$-measurable, the process

$$(H_\bullet M)_{t_j} \;=\; \sum_{i=0}^{j-1} A_i \cdot (M_{t_{i+1}} - M_{t_i}), \qquad j = 0, 1, \ldots, n$$

is a martingale transform of the discrete time martingale (M_{t_i}), and hence again a martingale.

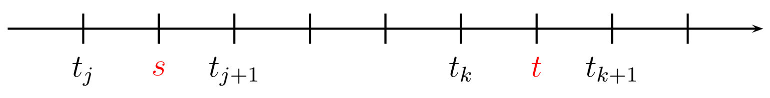
(2). Secondly, suppose $s, t \in [t_j, t_{j+1}]$ for some $j \in \{0, 1, 2, \ldots, n-1\}$. Then

$$E[(H_\bullet M)_t - (H_\bullet M)_s | \mathcal{F}_s] \;=\; E[A_j \cdot (M_t - M_s) | \mathcal{F}_s] \;=\; A_j \cdot E[M_t - M_s | \mathcal{F}_s] \;=\; 0$$

because A_j is $\mathcal{F}_{t_j}$-measurable and hence is $\mathcal{F}_s$-measurable, and (M_t) is a martingale.

(3). Finally, suppose that $s \in [t_j, t_{j+1}]$ and $t \in [t_k, t_{k+1}]$ with $j < k$. Then by the tower property for conditional expectations and by (1) and (2),

$$\begin{aligned} E[(H_\bullet M)_t \mid \mathcal{F}_s] &= E[E[E[(H_\bullet M)_t \mid \mathcal{F}_{t_k}] \mid \mathcal{F}_{t_{j+1}}] \mid \mathcal{F}_s] \\ &\overset{(2)}{=} E[E[(H_\bullet M)_{t_k} \mid \mathcal{F}_{t_{j+1}}] \mid \mathcal{F}_s] \overset{(1)}{=} E[(H_\bullet M)_{t_{j+1}} \mid \mathcal{F}_s] \\ &\overset{(2)}{=} (H_\bullet M)_s. \end{aligned}$$



□

Remark (Riemann sum approximations). Non-anticipative Riemann sum approximations of stochastic integrals are Itô integrals of predictable step functions: If (H_t) is an adapted stochastic process and $\pi = \{t_0, t_1, \dots, t_n\}$ is a partition then

$$\sum_{i=0}^{n-1} H_{t_i} \cdot (M_{t_{i+1}\wedge t} - M_{t_i \wedge t}) \;=\; \int_0^t H_s^\pi \, dM_s \tag{5.2.2}$$

where $H^\pi := \sum_{i=0}^{n-1} H_{t_i} \cdot I_{(t_i, t_{i+1}]}$ is a process in $\mathscr{E}$.

Itô isometry; Variant 1

Our goal is to prove that non-anticipative Riemann sum approximations for a stochastic integral converge.

Let (π_n) be a sequence of partitions of $[0,t]$ with $\mathrm{mesh}(\pi_n) \to 0$. By the remark above, the corresponding Riemann-Itô sums I_n defined by (5.2.2) are integrals of predictable step functions H^{π_n}:

$$I_n \;=\; \int_0^t H^{\pi_n} \, dM.$$

Hence in order to prove that the sequence (I_n) converges in $L^2(\Omega, \mathcal{A}, P)$ it suffices to show that

(1). (H^{π_n}) is a *Cauchy sequence w.r.t. an appropriate norm* on the vector space $\mathscr{E}$., and

(2). the "**Itô map**" $\mathcal{J} : \mathscr{E} \to L^2(\Omega, \mathcal{A}, P)$ defined by

$$\mathcal{J}(H) \;=\; \int_0^t H_s \, dM_s$$

is *continuous w.r.t. this norm*.

It turns out that we can even identify explicitly a simple norm on $\mathscr{E}$ such that the Itô map is an isometry. We first consider the case where (M_t) is a Brownian motion:

Theorem 5.2 (Itô isometry for Brownian motion, Variant 1). *If (B_t) is an $(\mathcal{F}_t)$ Brownian motion on $(\Omega, \mathcal{A}, P)$ then*

$$E\left[\left(\int_0^t H_s\, dB_s\right)^2\right] \;=\; E\left[\int_0^t H_s^2\, ds\right] \tag{5.2.3}$$

for any process $H \in \mathscr{E}$ and $t \in [0, \infty]$.

Proof. Suppose that $H = \sum_{i=1}^{n} A_i \cdot I_{(t_i, t_{i+1}]}$ with $n \in \mathbb{N}$, $0 \leq t_0 < t_1 < \ldots < t_n$ and A_i $\mathcal{F}_{t_i}$-measurable. With the notation

$$\Delta_i B \;:=\; B_{t_{i+1}\wedge t} - B_{t_i \wedge t}$$

we obtain for $t \geq 0$:

$$E\left[\left(\int_0^t H_s\, dB_s\right)^2\right] \;=\; E\left[\left(\sum_{i=0}^{n-1} A_i \Delta_i B\right)^2\right] \;=\; \sum_{i,k} E[A_i A_k \cdot \Delta_i B \Delta_k B]. \tag{5.2.4}$$

The summand on the right hand side vanishes for $i \neq k$, since

$$E[A_i A_k \Delta_i B \Delta_k B] \;=\; E[A_i A_k \Delta_i B \cdot E[\Delta_k B \mid \mathcal{F}_{t_k}]] \;=\; 0 \quad \text{ if } i < k.$$

Here we have used in an essential way, that A_k is $\mathcal{F}_{t_k}$-measurable. Similarly,

$$E[A_i^2 \cdot (\Delta_i B)^2] \;=\; E[A_i^2 E[(\Delta_i B)^2 \mid \mathcal{F}_{t_i}]] \;=\; E[A_i^2 \cdot \Delta_i t]$$

by the independence of the increments of Brownian motion. therefore, by (5.2.4) we obtain

$$E\left[\left(\int_0^t H_s\, dB_s\right)^2\right] \;=\; \sum_{i=0}^{n-1} E[A_i^2 \cdot (t_{i+1}\wedge t - t_i \wedge t)] \;=\; E\left[\int_0^t H_s^2\, ds\right].$$

□

Remark (Itô isometry w.r.t. continuous martingales). An Itô isometry also holds if Brownian motion is replaced by a continuous square-integrable martingale (M_t). Suppose that there exists a non-decreasing adapted continuous process $t \mapsto \langle M \rangle_t$ such that

$\langle M\rangle_0 = 0$ and $M_t^2 - \langle M\rangle_t$ s a martingale. The existence of a corresponding "variance process" in continuous time can always be shown | in fact, for martingales with continuous sample paths, $\langle M\rangle_t$ coincides with the quadratic variation process of M_t, cf. **??** below. Analogue arguments as in the proof above then yield the Itô isometry

$$E\left[\left(\int_0^t H_s\,dM_s\right)^2\right] \;=\; E\left[\int_0^t H_s^2\,d\langle M\rangle_s\right] \qquad \text{for any } H\in\mathscr{E} \text{ and } t\geq 0, \quad (5.2.5)$$

where $d\langle M\rangle$ denotes integration w.r.t. the measure with distribution function $F(t) = \langle M\rangle_t$. For Brownian motion $\langle B\rangle_t = t$, so (5.2.5) reduces to (5.2.3).

Theorem 5.2 shows that the linear map

$$\mathcal{J} : \mathscr{E} \to \mathcal{L}^2(\Omega,\mathcal{A},P), \quad \mathcal{J}(H) \;=\; \int_0^t H_s\,dB_s,$$

is an isometry of the space $\mathscr{E}$ of simple predictable processes $(s,\omega)\mapsto H_s(\omega)$ is endowed with the L^2 norm

$$\|H\|_{L^2(P\otimes\lambda_{[0,t]})} \;=\; E\left[\int_0^t H_s^2\,ds\right]^{1/2}$$

on the product space $\Omega\times[0,t]$. In particular, $\mathcal{J}$ respects $P\otimes\lambda$ classes, i.e., if $H_s(\omega) = \widetilde{H}_s(\omega)$ for $P\otimes\lambda$-almost every (ω,s) then $\int_0^t H\,dB = \int_0^t \widetilde{H}\,dB$ P-almost surely. Hence $\mathcal{J}$ also induces a linear map between the corresponding spaces of equivalence classes w.r.t. $P\otimes\lambda$, P respectively.
As usual, we do not always differentiate between equivalence classes and functions, and denote the linear map on equivalence classes again by $\mathcal{J}$:

$$\begin{aligned} \mathcal{J} \;:\; \mathscr{E}\subset L^2(P\otimes\lambda_{[0,t]}) \;&\to\; L^2(P) \\ \|\mathcal{J}(H)\|_{L^2(P)} \;&=\; \|H\|_{L^2(P\otimes\lambda_{[0,t]})}. \end{aligned} \qquad (5.2.6)$$

Defining stochastic integrals: A second attempt

Let $\overline{\mathscr{E}_t}$ denote the closure of the simple space $\mathscr{E}$ of elementary previsible processes in $L^2(\Omega \times [0,t], P \otimes \lambda)$. Since the Itô map $\mathcal{J}$ is an isometry, and hence a continuous linear map, it has a unique extension to a continuous linear map

$$\overline{\mathcal{J}} \,:\, \overline{\mathscr{E}_t} \subseteq L^2(P \otimes \lambda_[0,t]) \;\longrightarrow\; L^2(P),$$

and $\overline{\mathcal{J}}$ is again an isometry w.r.t. the corresponding L^2 norms. This can be used to define the Itô integral for any process in $\overline{\mathscr{E}_t}$, i.e., for any process that can be approximated by predictable step functions w.r.t. the $L^2(P \otimes \lambda_{[0,t]})$ norm. Explicitly, this leads to the following definition:

Definition. *For a given $t \geq 0$ and $H \in \overline{\mathscr{E}_t}$ we define*

$$\int_0^t H_s\, dB_s \;:=\; \lim_{n\to\infty} \int_0^t H_s^n\, dB_s \qquad \textit{in } L^2(\Omega, \mathcal{A}, P)$$

where (H^n) is an arbitrary sequence of simple predictable processes such that

$$E\left[\int_0^t (H_s - H_s^n)^2\, ds\right] \;\longrightarrow\; 0 \qquad \textit{as } n \to \infty.$$

The isometry (5.2.6) ensures that for a given $t \geq 0$, the stochastic integral is well-defined, i.e., the definition does not depend on the choice of the approximating sequence (H^n). Moreover, we will show in Section **??** that the space $\overline{\mathscr{E}_t}$ contains all square integrable (w.r.t. $P \otimes \lambda$) adapted processes, and is hence sufficiently large. As already remarked above, the map $H \mapsto \int_0^t H_s\, dB_s$ is again an isometry from $\overline{\mathscr{E}_t} \subseteq L^2(P \otimes \lambda_{[0,t]})$ to $L^2(P)$. Nevertheless, the definition above has two obvious drawbacks:

Drawback 1: For general $H \in \overline{\mathscr{E}_t}$ the Itô integral $\int_0^t H_s\, dB_s$ is only defined as an equivalence class in $L^2(\Omega, \mathcal{A}, P)$, i.e., uniquely up to modification on P-measure zero sets. In particular, we do not have a *pathwise definition* of $\int_0^t H_s(\omega)\, dB_s(\omega)$ for a given Brownian sample path $s \mapsto B_s(\omega)$.

Drawback 2: Even worse, the construction above works only for a fixed integration interval $[0,t)$. The exceptional sets may depend on t and therefore, the process $t \mapsto \int_0^t H_s \, dB_s$ does not have a meaning yet. In particular, we do not know yet if there exists a version of this process that is almost surely continuous.

The first drawback is essential: In certain cases it is indeed possible to define stochastic integrals pathwise, cf. Chapter **??** below. In general, however, pathwise stochastic integrals cannot be defined. The extra impact needed is the Lévy area process, cf. the rough path theory developed by T. Lyons and others [Lyons: "St. Flow", Friz and Victoir].

Fortunately, the second drawback can be overcome easily. By extending the Itô isometry to an isometry into the space M_c^2 of continuous L^2 bounded martingales, we can construct the complete process $t \mapsto \int_0^t H_s \, dB_s$ simultaneously as a continuous martingale. The key observation is that by the maximal inequality, continuous L^2 bounded martingales can be controlled uniformly in t by the L^2 norm of their final value.

The Hilbert space $\mathbf{M_c^2}$

Fix $u \in (0,\infty]$ and suppose that (H^n) is a sequence of elementary previsible processes converging in $L^2(P \otimes \lambda_{[0,u]})$. Our aim is to prove convergence of the continuous martingales $(H^n_\bullet B)_t = \int_0^t H^n \, dB_s$ to a further continuous martingale. Since the convergence holds only almost surely, the limit process will not necessarily be $(\mathcal{F}_t)$ adapted in general. To ensure adaptedness, we have to consider the **completed filtration**

$$\mathcal{F}_t^P \;=\; \{A \in \mathcal{A} \mid P[A \,\triangle\, B] = 0 \quad \text{for some } B \in \mathcal{F}_t\}, \qquad t \geq 0,$$

where $A \triangle B = (A \setminus B) \cup (B \setminus A)$ denotes the symmetric difference of the set A and B.

Note that the conditional expectations given $\mathcal{F}_t$ and $\mathcal{F}_t^P$ agree P-almost surely. Hence, if (B_t) is a Brownian motion resp. a martingale w.r.t. the filtration $(\mathcal{F}_t)$ then it is also a Brownian motion or a martingale w.r.t. $(\mathcal{F}_t^P)$.

Let $\mathcal{M}^2([0,u])$ denote the space of all L^2-bounded $(\mathcal{F}_t^P)$ martingales $(M_t)_{0\leq t\leq u}$ on

$(\Omega, \mathcal{A}, P)$. By $\mathcal{M}_c^2([0,u])$ we denote the subspace consisting of all continuous martingales $M \in \mathcal{M}^2([0,u])$. Recall that by the L^2 martingale convergence theorem, any right-continuous L^2-bounded martingale (M_t) defined for $t \in [0,u]$ can be extended to a martingale in $\mathcal{M}^2([0,u])$.

Two martingales $M, \widetilde{M} \in \mathcal{M}^2([0,u])$ are called **modifications** of each other if

$$P[M_t = \widetilde{M}_t] \;=\; 1 \qquad \text{for any } t \in [0,u].$$

If the martingales are right-continuous then two modifications agree almost surely, i.e.,

$$P[M_t = \widetilde{M}_t \quad \forall t \in [0,u]] \;=\; 1.$$

In order to obtain norms and not just semi-norms, we consider the spaces

$$M^2([0,u]) \;:=\; \mathcal{M}^2([0,u])/\sim \qquad \text{and} \qquad M_c^2([0,u]) \;:=\; \mathcal{M}_c^2([0,u])/\sim$$

of equivalence classes of martingales that are modifications of each other. We will frequently identify equivalence classes and their representatives.

We endow the space $M^2([0,u])$ with the inner product

$$(M,N)_{M^2([0,u])} \;=\; (M_u, N_u)_{L^2} \;=\; E[M_u N_u].$$

Since $M \in M^2([0,u])$, the process (M_t^2) is a submartingale, the norm corresponding to the inner product is given by

$$\|M\|^2_{M^2([0,u])} \;=\; E[M_u^2] \;=\; \sup_{0 \le t \le u} E[M_t^2].$$

Moreover, if (M_t) is right-continuous then by **Doob's $\mathbf{L^2}$-maximal inequality**,

$$\left\| \sup_{0 \le t \le u} |M_t| \right\|_{L^2(\Omega,\mathcal{A},P)} \;\le\; 2 \cdot \sup_{0 \le t \le u} \|M_t\|_{L^2(\Omega,\mathcal{A},P)} \;=\; 2\|M\|_{M^2([0,u])}. \tag{5.2.7}$$

This crucial estimate shows that on the subspace M_c^2, the M^2 norm is equivalent to the L^2 norm of the supremum of the martingale. Therefore, *the M^2 norm can be used to control (right-)continuous martingales uniformly in t!*

Lemma 5.3. *(1). The space $M^2([0,u])$ is a Hilbert space, and the linear map $M \mapsto M_u$ from $M^2([0,u])$ to $L^2(\Omega, \mathcal{F}_u, P)$ is onto and isometric.*

(2). *The space* $M_c^2([0,u])$ *is a closed subspace of* $M^2([0,u])$*, i.e., if* (M^n) *is a Cauchy sequence in* $M_c^2([0,u])$ *then there exists a continuous martingale* $M \in M_c^2([0,u])$ *such that*

$$\sup_{t\in[0,u]} |M_t^n - M_t| \longrightarrow 0 \qquad \textit{in } L^2(\Omega, \mathcal{A}, P).$$

Proof. (1). The map $M \mapsto M_u$ is an isometry by definition of the inner product on $M^2([0,u])$. Moreover, for any $X \in^2 (\Omega, \mathcal{F}_u, P)$, the process $M_t = E[X \mid \mathcal{F}_u]$ is in $M^2([0,u])$ with $M_u = X$. Hence, the image of the isometry is the whole space $L^2(\Omega, \mathcal{F}_u, P)$. Since $L^2(\Omega, \mathcal{F}_u, P)$ is complete w.r.t. the L^2 norm, the space $M^2([0,u])$ is complete w.r.t. the M^2 norm.

(2). If (M^n) is a Cauchy sequence in $M_c^2([0,u])$ then by (5.2.7),

$$\|M^n - M^m\|_{\sup} = \sup_{0\le t\le u} |M_t^n - M_t^m| \longrightarrow 0 \qquad \text{in } L^2(\Omega, \mathcal{A}, P).$$

In particular, we can choose a subsequence (M^{n_k}) such that

$$P[\, \|M^{n_{k+1}} - M^{n_k}\|_{\sup} \ge 2^{-k} \,] \le 2^{-k} \qquad \text{for all } k \in \mathbb{N}.$$

Hence, by the Borel-Cantelli Lemma,

$$P[\, \|M^{n_{k+1}} - M^{n_k}\|_{\sup} < 2^{-k} \quad \text{eventually} \,] = 1,$$

and therefore $M_t^{n_k}$ converges almost surely uniformly in t as $k \to \infty$. The limit of the sequence (M^n) in $M^2([0,u])$ exists by (1), and the process M defined by

$$M_t := \begin{cases} \lim M_t^{n_k} & \text{if } (M^{n_k}) \text{ converges uniformly}, \\ 0 & \text{otherwise} \end{cases} \tag{5.2.8}$$

is a continuous representative of the limit. Indeed, by Fatou's Lemma,

$$\begin{aligned} \|M^{n_k} - M\|^2_{M^2([0,u])} &\le E[\, \|M^{n_k} - M\|^2_{\sup} \,] = E[\lim_{l\to\infty} \|M^{n_k} - M^{n_l}\|^2_{\sup} \,] \\ &\le \liminf_{l\to\infty} E[\, \|M^{n_k} - M^{n_l}\|^2_{\sup} \,], \end{aligned}$$

and the right hand side converges to 0 as $k \to \infty$. Finally, M is a martingale w.r.t. $(\mathcal{F}_t^P)$, and hence an element in $M_c^2([0,u])$.

□

Remark. We point out that the continuous representative (M_t) defined by (5.2.8) is a martingale w.r.t. the complete filtration $(\mathcal{F}_t^P)$, but it is not necessarily adapted w.r.t. $(\mathcal{F}_t)$.

Itô isometry into M_c^2

For any simple predictable process H and any continuous martingale $M \in M_c^2([0,u])$, the process

$$(H_\bullet M)_t = \int_0^t H_s \, dM_s, \qquad t \in [0,u],$$

is again a continuous L^2 bounded martingale on $[0,u]$ by Lemma 5.1. We can therefore restate the Itô isometry in the following way:

Corollary 5.4 (Itô isometry for Brownian motion, Variant 2). *If (B_t) is a $(\mathcal{F}_t)$ Brownian motion on $(\Omega, \mathcal{A}, P)$ then*

$$\|H_\bullet B\|^2_{M^2([0,u])} = E\left[\int_0^u H_s^2 \, ds\right]$$

for any process $H \in \mathscr{E}$ and $u \in [0,\infty]$.

Proof. The assertion is an immediate consequence of the definition of the M^2 norm and Theorem 5.2. □

5.3 Itô integrals for square-integrable integrands

Let (B_t) be a Brownian motion w.r.t. a filtration $(\mathcal{F}_t)$ on $(\Omega, \mathcal{A}, P)$, and fix $u \in [0,\infty]$. the linear map

$$\mathcal{J} : \begin{array}{ccccc} \mathscr{E} & \subseteq & L^2(P \otimes \lambda_{[0,u]}) & \to & M_c^2([0,u]) \\ & & H & \mapsto & H_\bullet B \end{array}$$

mapping a simple predictable process H to the continuous martingale

$$(H_\bullet B)_t = \int_0^t H_s \, dB_s$$

is called **Itô map**. More precisely, we consider the induced map between equivalence classes.

Definition of Itô integral

By Corollary 5.4, the Itô map is an isometry. Therefore, there is a unique continuous (and even isometric) extension

$$\overline{\mathcal{J}} \;:\; \overline{\mathscr{E}_u} \subseteq L^2(P \otimes \lambda_{[0,u]}) \quad \to \quad M_c^2([0,u])$$

to the closure $\overline{\mathscr{E}_u}$ of the space $\mathscr{E}$ in $L^2(P \otimes \lambda_{[0,u]})$. this allows us to define the martingale transform $H_\bullet B$ and the stochastic integrals for any process $H \in \overline{\mathscr{E}_u}$ by

$$H_\bullet B \;:=\; \overline{\mathcal{J}}(H), \qquad \int_0^t H_s \, dB_s \;:=\; (H_\bullet B)_t.$$

We hence obtain the following definition of stochastic integrals for integrands in $\overline{\mathscr{E}_u}$:

Definition. *For $H \in \overline{\mathscr{E}_u}$ the process $H_\bullet B = \int_0^\bullet H_s \, dB_s$ is the up to modifications unique continuous martingale in $[0,u]$ such that*

$$(H^n_\bullet B)_t \;\to\; (H_\bullet B)_t \quad \text{in } L^2(P)$$

for any $t \in [0,u]$ and for any sequence (H^n) of simple predictable processes with $H^n \to H$ in $L^2(P \otimes \lambda_{[0,u]})$.

Remark. (1). By construction, the map $H \mapsto H_\bullet B$ is an isometry from $\overline{\mathscr{E}_u} \subseteq L^2(P \otimes \lambda_{[0,u]})$ to $M_c^2([0,u])$. We will prove below that the closure $\overline{\mathscr{E}_u}$ of the simple processes actually contains any $(\mathcal{F}_t^P)$ adapted process $(\omega, t) \mapsto H_t(\omega)$ that is square-integrable w.r.t. $P \otimes \lambda_{[0,u]}$.

(2). The definition above is consistent in the following sense: If $H_\bullet B$ is the stochastic integral defined on the time interval $[0,v]$ and $u \leq v$, then the restriction of $H_\bullet B$ to $[0,u]$ coincides with the stochastic integral on $[0,u]$.

For $0 \leq s \leq t$ we define

$$\int_s^t H_r \, dB_r \; := \; (H_\bullet B)_t - (H_\bullet B)_s.$$

Exercise. Verify that for any $H \in \overline{\mathscr{E}_t}$,

$$\int_s^t H_r \, dB_r \; = \; \int_0^t H_r \, dB_r - \int_0^t I_{(0,s)}(r) H_r \, dB_r \; = \; \int_0^t I_{(s,t)}(r) H_r \, dB_r.$$

Approximation by Riemann-Itô sums

We now show that bounded adapted processes with continuous sample paths are contained in the closure of the simple predictable processes, and the corresponding stochastic integrals are limits of predictable Riemann sum approximations. we consider partitions of $\mathbb{R}_+$ that are given by increasing sequences (t_n) of partition points with $t_0 = 0$ and $\lim\limits_{n\to\infty} t_n = \infty$:

$$\pi \; := \; \{t_0, t_1, t_2, \ldots\}.$$

For a point $s \in \pi$ we denote by s' the next largest partition point. the mesh size of the partition is defined by

$$\mathrm{mesh}(\pi) \; = \; \sup_{s\in\pi} |s' - s|.$$

now fix $u \in (0, \infty)$ and a sequence (π_n) of partitions of $\mathbb{R}_+$ such that $\mathrm{mesh}(\pi_n) \to 0$.

Theorem 5.5. *Suppose that $(H_t)_{t\in[0,u)}$ is a $(\mathcal{F}_t^P)$ adapted stochastic process on $(\Omega, \mathcal{A}, P)$ such that $(t, \omega) \mapsto H_t(\omega)$ is product-measurable and bounded. If $t \mapsto H_t$ is P-almost surely left continuous then H is in $\overline{\mathscr{E}_u}$, and*

$$\int_0^t H_s \, dB_s \; = \; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} H_s (B_{s'\wedge t} - B_s), \qquad t \in [0, u], \tag{5.3.1}$$

w.r.t. convergence uniformly in t in the $L^2(P)$ sense.

Remark. (1). In particular, a subsequence of the predictable Riemann sum approximations converges uniformly in t with probability one.

(2). The assertion also holds if H is unbounded with $\sup\limits_{0\leq s\leq u} |H_s| \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$.

Proof. For $t \in [0, u]$ the Riemann sums on the right hand side of (5.3.1) are the stochastic integrals $\int_0^t H_s^n \, dB_s$ of the predictable step functions

$$H_t^n \;:=\; \sum_{s<u} s \in \pi_n H_s \cdot I_{(s,s']}(t), \qquad n \in \mathbb{N}.$$

As $n \to \infty$, $H_t^n \to H_t$ for any $t \in [0, u]$ almost surely by left-continuity. Therefore, by dominated convergence,

$$H^n \;\to\; H \quad \text{ in } L^2(P \otimes \lambda_{[0,u)}),$$

because the sequence (H^n) is uniformly bounded by boundedness of H. Hence, by the Itô isometry,

$$\int_0^{\bullet} H_s \, dB_s \;=\; \lim_{n\to\infty} \int_0^{\bullet} H_s^n \, dB_s \qquad \text{in } M_c^2([0, u]).$$

□

Identification of admissible integrands

Let $u \in (0, \infty]$. We have already shown that if $u < \infty$ then any product-measurable adapted bounded process with left-continuous sample paths is in $\overline{\mathscr{E}_u}$. More generally, we will prove now that any adapted process in $\mathcal{L}^2(P \otimes \lambda_{[0,u)})$ is contained in $\overline{\mathscr{E}_u}$, and hence "integrable" w.r.t. Brownian motion.

Let $\mathcal{L}_a^2([0, u))$ denote the vector space of all product-measurable, $(\mathcal{F}_t^P)$ adapted stochastic processes $(\omega, t) \mapsto H_t(\omega)$ defined on $\Omega \times [0, u)$ such that

$$E\left[\int_0^u H_t^2 \, dt\right] \;<\; \infty.$$

The corresponding space of equivalence classes of $P\otimes\lambda$ versions is denoted by $L_a^2([0, u))$.

Lemma 5.6. $L_a^2([0, u))$ *is a closed subset of* $L^2(P \otimes \lambda_{[0,u)})$.

Proof. It only remains to show that an $L^2(P \otimes \lambda)$ limit of $(\mathcal{F}_t^P)$ adapted processes again has a $(\mathcal{F}_t^P)$ adapted $P \otimes \lambda$-version. Hence consider a sequence $H^n \in \mathcal{L}_a^2([0,u))$ with $H^n \to H$ in $L^2(P \otimes \lambda)$. Then there exists a subsequence (H^{n_k}) such that $H_t^{n_k}(\omega) \to H_t(\omega)$ for $P \otimes \lambda$-almost every $(\omega, t) \in \Omega \times [0,u)$, the process $\widetilde{H}$ defined by $\widetilde{H}_t(\omega) := \lim H_t^{n_k}(\omega)$ if the limit exists, $\widetilde{H}_t(\omega) := 0$ otherwise, is then a $(\mathcal{F}_t^P)$ adapted version of H. □

We can now identify the class of integrands H for which the stochastic integral $H_\bullet B$ is well-defined as a limit of integrals of predictable step functions in the space $M_c^2([0,u])$:

Theorem 5.7. *For any* $u \in (0, \infty]$,

$$\overline{\mathscr{E}_u} \;=\; L_a^2(P \otimes \lambda_{[0,u)}).$$

Proof. Since $\mathscr{E} \subseteq \mathcal{L}_a^2(P \otimes \lambda_{[0,u)})$ it only remains to show the inclusion "$\supseteq$". Hence fix a process $H \in \mathcal{L}_a^2(P \otimes \lambda_{[0,u)})$. We will prove in several steps that H can be approximated in $L^2(P \otimes \lambda_{[0,u)})$ by simple predictable processes:

(1). Suppose first that H is bounded and has almost surely continuous trajectories. Then for $u < \infty$, H is in $\overline{\mathscr{E}_u}$ by Theorem 5.5. For $u = \infty$, H is still in $\overline{\mathscr{E}_u}$ provided there exists $t_0 \in (0, \infty)$ such that H_t vanishes for $t \geq t_0$.

(2). Now suppose that (H_t) is bounded and, if $u = \infty$, vanishes for $t \geq t_0$. To prove $H \in \overline{\mathscr{E}_u}$ we approximate H by continuous adapted processes: Let $\psi_n : \mathbb{R} \to [0, \infty)$, $n \in \mathbb{N}$, be continuous functions such that $\psi(s) = 0$ for $s \notin (0, 1/n)$ and $\int_{-\infty}^{\infty} \psi_n(s)\, ds = 1$, and let $H^n := H * \psi_n$, i.e.,

$$H_t^n(\omega) \;=\; \int_0^{1/n} H_{t-\varepsilon}(\omega) \psi_n(\varepsilon)\, d\varepsilon, \tag{5.3.2}$$

where we set $H_t := 0$ for $t \leq 0$. We prove

(a) $H^n \to H$ in $L^2(P \otimes \lambda_{[0,u)})$, and

(b) $H^n \in \overline{\mathscr{E}_u}$ for any $n \in \mathbb{N}$.

Combining (a) and (b), we see that H is in $\overline{\mathscr{E}_u}$ as well.

(a) Since H is in $\mathcal{L}^2(P \otimes \lambda_{[0,u)})$, we have

$$\int_0^u H_t(\omega)^2 \, dt \; < \; \infty \tag{5.3.3}$$

for P-almost every ω. It is a standard fact from analysis that (5.3.3) implies

$$\int_0^u |H_t^n(\omega) - H_t(\omega)|^2 \, dt \quad \longrightarrow \quad 0 \qquad \text{as } n \to \infty.$$

By dominated convergence, we obtain

$$E\left[\int_0^u |H_t^n - H_t|^2 \, dt\right] \quad \longrightarrow \quad 0 \qquad \text{as } n \to \infty \tag{5.3.4}$$

because H is bounded, the sequence (H_n) is uniformly bounded, and H and H^n vanish for $t \geq t_0 + 1$.

(b) This is essentially a consequence of part (1) of the proof. We sketch how to verify that H^n satisfies the assumptions made there:

- The sample paths $t \mapsto H_t^n(\omega)$ are continuous for all ω,
- $|H_t^n|$ is bounded by $\sup |H|$
- The map $(\omega, t) \mapsto H_t^n(\omega)$ is product measurable by (5.3.2) and Fubini's Theorem, because the map $(\omega, t, \varepsilon) \mapsto H_{t-\varepsilon}(\omega)\psi_\varepsilon(\omega)$ is product measurable.
- If the process (H_t) is *progressively measurable*, i.e., if the map $(s, \omega) \mapsto H_s(\omega) \quad (0 \leq s \leq t, \omega \in \Omega)$ is measurable w.r.t. the product σ-algebra $\mathcal{B}([0,t]) \otimes \mathcal{F}_t^P$ for any $t \geq 0$, then (H_t^n) is $(\mathcal{F}_t^P)$ adapted by (5.3.2) and Fubini's Theorem. This i for example the case if (H_t) is right continuous or left continuous.
- In general, one can prove that (H_t) has a progressively measurable modification, where (H_t^n) has a $(\mathcal{F}_t^P)$ adapted modification. We omit the details.

(3). We finally prove that general $H \in \mathcal{L}_a^2(P \otimes \lambda_{[0,u)})$ are contained in $\overline{\mathscr{E}_u}$. This is a consequence of (2), because we can approximate H by the processes

$$H_t^n := ((H_t \wedge n) \vee (-n)) \cdot I_{[0,n]}(t), \qquad n \in \mathbb{N}.$$

These processes are bounded, they vanish for $t \geq n$, and $H^n \to H$ in $L^2(P \otimes \lambda_{[0,u)})$. By (2) H^n is contained in $\overline{\mathscr{E}_u}$ for any n, so H is in $\overline{\mathscr{E}_u}$ as well.

□

Remark (Riemann sum approximations). For discontinuous integrands, the predictable Riemann sum approximations considered above do not converge to the stochastic integral in general. However, one can prove that for $u < \infty$ any process $H \in L_a^2(P \otimes \lambda_{[0,u)})$ is the limit of the simple predictable processes

$$H_t^n = \sum_{i=1}^{2^n-1} 2^n \int_{(i-1)2^{-n}u}^{i2^{-n}u} H_s \, ds \cdot I_{(i2^{-n}u,(i-1)2^{-n}u]}(t)$$

w.r.t. the $L^2(P \otimes \lambda_{[0,u)})$ norm, cf. [Steele: "Stochastic calculus and financial applications", Sect 6.6]. Therefore, the stochastic integral $\int_0^t H \, dB$ can be approximated for $t \leq u$ by the correspondingly modified Riemann sums.

Local dependence on the integrand

We conclude this section by pointing out that the approximations considered above imply that the stochastic integral depends locally on the integrand in the following sense:

Corollary 5.8. *Suppose that $T : \Omega \to [0, \infty]$ is a random variable, and $H, \widetilde{H}$ are processes in $\mathcal{L}_a^2([0, \infty))$ such that $H_t = \widetilde{H}_t$ for any $t \in [0, T)$ holds P-almost surely. Then, P-almost surely,*

$$\int_0^t H_s \, dB_s = \int_0^t \widetilde{H}_s \, dB_s \qquad \textit{for any } t \in [0, T].$$

Proof. W.l.o.g. we may assume $\widetilde{H} = 0$ and

$$H_t \;=\; 0 \qquad \text{for } t < T. \tag{5.3.5}$$

We then have to prove that $\int_0^t H\,dB = 0$ for $t < T$. For this purpose we go through the same approximations as in the proof of Theorem 5.7 above:

(1). If H_t is almost surely continuous and bounded, and $H_t = 0$ for $t \geq t_0$ then by Theorem 5.5, for any sequence (π_n) of partitions with $\mathrm{mesh}(\pi_n) \to 0$, we have

$$\int_0^t H\,dB \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} H_s \cdot (B_{s'\wedge t} - B_s)$$

with convergence uniformly in t, P-almost surely along a subsequence. For $t \leq T$ the right-hand side vanishes by (5.3.5).

(2). If H is bounded and $H_t = 0$ for $t \geq t_0$ then the approximations

$$H_t^n \;=\; \int_0^{1/n} H_{t-\varepsilon}\psi(\varepsilon)\,d\varepsilon$$

(with ψ_n defined as in the proof of Theorem 5.7 and $H_t := 0$ for $t < 0$) vanish for $t \leq T$. Hence by (1) and (5.3.4),

$$\int_0^t H\,dB \;=\; \lim \int_0^t H^n\,dB \;=\; 0 \qquad \text{for } t \leq T$$

where the convergence holds again almost surely uniformly in t along a subsequence.

(3). Finally, in the general case the assertion follows by approximating H by the bounded processes

$$H_t^n \;=\; ((H_t \wedge n) \vee (-n)) \cdot I_{[0,n]}(t).$$

□

5.4 Localization

Square-integrability of the integrand is still an assumption that we would like to avoid, since it is not always easy to verify or may even fail to hold. The key to extending the class of admissible integrands further is localization, which enables us to define a stochastic integral w.r..t Brownian motion for any continuous adapted process. The price we have to pay is that for integrands that are not square integrable, the Itô integral is in general not a martingale, but only a local martingale.

Itô integrals for locally square-integrable integrands

Let $T : \Omega \to [0,\infty]$ be an $(\mathcal{F}_t^P)$ stopping time. We will also be interested in the case where $T = \infty$. By $\mathcal{L}^2_{a,\mathrm{loc}}([0,T))$ we denote the vector space of all stochastic processes $(t,\omega) \mapsto H_t(\omega)$ defined for $t \in [0, T(\omega))$ such that the trivially extended process

$$\widetilde{H}_t := \begin{cases} H_t & \text{for } t < T, \\ 0 & \text{for } t \geq T, \end{cases}$$

is product measurable in (t,ω), adapted w.r.t. the filtration $(\mathcal{F}_t^P)$, and

$$t \mapsto H_t(\omega) \quad \text{is in } \mathcal{L}^2_{\mathrm{loc}}([0,T(\omega)), dt) \quad \text{for } P\text{-a.e. } \omega. \tag{5.4.1}$$

Here for $u \in (0,\infty]$, the space $\mathcal{L}^2_{\mathrm{loc}}([0,T(\omega)), dt)$ consists of all functions $f : [0,u) \to [-\infty,\infty]$ such that $\int_0^s f(t)^2\, dt < \infty$ for any $s \in (0,u)$. From now on, we use the notation $H_t \cdot I_{\{t<T\}}$ for the trivial extension $(\widetilde{H}_t)_{0 \leq t < \infty}$ of a process $(H_t)_{0 \leq t < T}$ beyond the stopping time T. Locally square integrable adapted processes allow for a localization by stopping times:

Lemma 5.9. *If $(H_t)_{0 \leq t < T}$ is a process in $\mathcal{L}^2_{a,loc}([0,T))$ then there exists an increasing sequence (T_n) of $(\mathcal{F}_t^P)$ stopping times such that $T = \sup T_n$ almost surely and*

$$H_t \cdot I_{\{t<T_n\}} \in \mathcal{L}^2_a([0,T)) \qquad \textit{for any } n \in \mathbb{N}.$$

Proof. One easily verifies that the random variables T_n defined by

$$T_n := \int_{0 \leq t < T} \left\{ \int_0^t H_s^2\, ds \geq n \right\} \wedge T, \qquad n \in \mathbb{N}, \tag{5.4.2}$$

are $(\mathcal{F}_t^P)$ stopping times. Moreover, for almost every ω, $t \mapsto H_t(\omega)$ is in $\mathcal{L}^2_{\text{loc}}([0,T))$. Hence the function $t \mapsto \int_0^t H_s(\omega)^2\, ds$ is increasing and finite on $[0, T(\omega))$, and therefore $T_n(\omega) \nearrow T(\omega)$ as $n \to \infty$. Since T_n is an $(\mathcal{F}_t^P)$ stopping time, the process $H_t \cdot I_{\{t<T_n\}}$ is $(\mathcal{F}_t^P)$-adapted, and

$$E\left[\int_0^\infty (H_s \cdot I_{\{s<T_n\}})^2\, ds\right] \;=\; E\left[\int_0^{T_n} H_s^2\, ds\right] \;=\; n \qquad \text{for any } n$$

by (5.4.2). □

A sequence of stopping times as in the lemma will also be called a **localizing sequence**. We can now extend the definition of the Itô integral to locally square-integrable adapted integrands:

Definition. *For a process $h \in \mathcal{L}^2_{a,loc}([0,T))$ the Itô stochastic integral w.r.t. Brownian motion is defined for $t \in [0,T)$ by*

$$\int_0^t H_s\, dB_s \;:=\; \int_0^t H_s \cdot I_{\{s<\hat{T}\}}\, dB_s \qquad \textit{for } t \in [0, \hat{T}] \tag{5.4.3}$$

whenever $\hat{T}$ is a $(\mathcal{F}_t^P)$ stopping time with $H_t \cdot I_{\{t<\hat{T}\}} \in \mathcal{L}^2_a([0,\infty))$

Theorem 5.10. *For $H \in \mathcal{L}^2_{a,loc}([0,T))$ the Itô integral $t \mapsto \int_0^t H_s\, dB_s$ is almost surely well defined by (5.4.3) as a continuous process on $[0,T)$.*

Proof. We have to verify that the definition does not depend on the choice of the localizing stopping times. This is a direct consequence of Corollary 5.8: Suppose that $\hat{T}$ and $\widetilde{T}$ are $(\mathcal{F}_t)$ stopping times such that $H_t \cdot I_{t<\hat{T}}$ and $H_t \cdot I_{t<\widetilde{T}}$ are both in $\mathcal{L}^2_a([0,T))$. Since the two trivially extended processes agree on $[0, \hat{T} \wedge \widetilde{T})$, Corollary 5.8 implies that almost surely,

$$\int_0^t H_s \cdot I_{\{s<\hat{T}\}}\, dB_s \;=\; \int_0^t H_s \cdot I_{\{s<\widetilde{T}\}}\, dB_s \qquad \text{for any } t \in [0, \hat{T} \wedge \widetilde{T}).$$

Hence, by Lemma 5.9, the stochastic integral is well defined on $[0,T)$. □

Stochastic integrals as local martingales

Itô integrals w.r.t. Brownian motion are not necessarily martingales if the integrands are not square integrable. However, they are still local martingales in the sense of the definition stated below.

Definition. *An $(\mathcal{F} - t^P)$ stopping time T is called **predictable** iff there exists an increasing sequence of $(\mathcal{F}_t^P)$ stopping times $(T_k)_{k\in\mathbb{N}}$ such that $T_k < T$ on $\{T \neq 0\}$ for any k, and $T = \sup T_k$.*

Example. The hitting time T_A of a closed set A by a continuous adapted process is predictable, as it can be approximated from below by the hitting times T_{A_k} of the neighbourhoods $A_k = \{x \,:\, \mathrm{dist}(x, A) < 1/k\}$ of the set A. On the other hand, the hitting time of an open set is usually not predictable.

Definition. *Suppose that $T : \Omega \to [0, \infty]$ is a predictable stopping time. A stochastic process $M_t(\omega)$ defined for $0 \leq t < T(\omega)$ is called a **local martingale up to time** T, if and only if there exists an increasing sequence (T_k) of stopping times with $T = \sup T_k$ such that for any $k \in \mathbb{N}$, $T_k < T$ on $\{T > 0\}$, and the stopped process $(M_{t\wedge T_k})$ is a martingale for $t \in [0, \infty)$.*

Recall that by the Optional Stopping Theorem, a continuous martingale stopped at a stopping time is again a martingale. Therefore, any continuous martingale $(M_t)_{t\geq 0}$ is a local martingale up to $T = \infty$. Even if (M_t) is assumed to be uniformly integrable, the converse implication fails to hold:

Exercise (A uniformly integrable local martingale that is not a martingale). Let $x \in \mathbb{R}^3$ with $x \neq 0$, and suppose that (B_t) is a three-dimensional Brownian motion with initial value $B_0 = x$. Prove that the process $M_t = 1/|B_t|$ is a uniformly integrable local martingale up to $T = \infty$, but (M_t) is not a martingale.

On the other hand, note that if (M_t) is a continuous local martingale up to $T = \infty$, and the family $\{M_{t\wedge T_k} \mid k \in \mathbb{N}\}$ is uniformly integrable for each *fixed* $t \geq 0$, then (M_t) is a martingale, because for $0 \leq s \leq t$

$$E[M_t \mid \mathcal{F}_s] \;=\; \lim_{k\to\infty} E[M_{t\wedge T_k} \mid \mathcal{F}_s] \;=\; \lim_{k\to\infty} M_{s\wedge T_k} \;=\; M_s$$

with convergence in L^1.

As a consequence of the definition of the Itô integral by localization, we immediately obtain:

Theorem 5.11. *Suppose that T is a predictable stopping time w.r.t. $(\mathcal{F}_t^P)$. Then for any $H \in \mathcal{L}^2_{a,loc}([0,T))$, the Itô integral process $t \mapsto \int_0^t H_s\, dB_s$ is a continuous local martingale up to time T.*

Proof. We can choose an increasing sequence (T_k) of stopping times such that $T_k < T$ on $\{T > 0\}$ and $H_t \cdot I_{t<T_k} \in \mathcal{L}^2_a([0,\infty))$ for any k. Then, by definition of the stochastic integral,

$$\int_0^{t\wedge T_k} H_s\, dB_s \;=\; \int_0^{t\wedge T_k} H_s \cdot I_{\{s<T_k\}}\, dB_s \qquad \text{for any } k \in \mathbb{N},$$

and the right-hand side is a continuous martingale in $M_c^2([0,\infty))$. □

The theorem shows that for a predictable $(\mathcal{F}_t^P)$ stopping time T, the Itô map $H \mapsto \int_0^\bullet H\, dB$ extends to a linear map

$$\mathcal{J} \;:\; L^2_{\text{loc}}([0,T)) \;\longrightarrow M_{c,\text{loc}}([0,T)),$$

where $M_{c,\text{loc}}([0,T))$ denotes the space of equivalence classes of local $(\mathcal{F}_t^P)$ martingales up to time T.
We will finally note that continuous local martingales (and hence stochastic integrals w.r.t. Brownian motion) can always be localized by a sequence of *bounded* martingales:

Exercise. Suppose that (M_t) is a continuous local martingale up to time T, and (T_k) is a localizing sequence of stopping times.

(1). Show that

$$\widetilde{T}_k \;=\; T_k \wedge \inf\{t \geq :\; |M_t| > k\}$$

is another localizing sequence, and the stopped process $(M_{t\wedge\widetilde{T}_k})_{t\geq 0}$ are bounded martingales for all k.

(2). If $T = \infty$ then $\hat{T}_k := \inf\{t \geq :\; |M_t| > k\}$ is a localizing sequence.

Approximation by Riemann-Itô sums

If the integrand (H_t) of a stochastic integral $\int H\, dB$ has continuous sample paths then local square integrability always holds, and the stochastic integral is a limit of Riemann-Itô sums: Let (π_n) be a sequence of partition of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$.

Theorem 5.12. *Suppose that T is a predictable stopping time, and $(H_t)_{0\le t<T}$ is a stochastic process defined for $t < T$. If the sample paths $t \mapsto H_t(\omega)$ are continuous on $[0, T(\omega))$ for any ω, and the trivially extended process $H_t \cdot I_{\{t<T\}}$ is $(\mathcal{F}_t^P)$ adapted, then H is in $\mathcal{L}^2_{a,loc}([0, T))$, and for any $t \ge 0$,*

$$\int_0^t H_s\, dB_s \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} H_s \cdot (B_{s'\wedge t} - B_s) \qquad \text{on } \{t < T\} \tag{5.4.4}$$

with convergence in probability.

Proof. Let $\lfloor t\rfloor_n = \max\{s \in \pi_n \;:\; s \le t\}$ denote the next partition point below t. By continuity,

$$H_t \cdot I_{\{t<T\}} \;=\; \lim_{n\to\infty} H_{\lfloor t\rfloor_n} \cdot I_{\{t<T\}}.$$

Hence $(H_t \cdot I_{\{t<T\}})$ is $(\mathcal{F}_t^P)$ adapted. It is also product-measurable, because

$$H_{\lfloor t\rfloor_n} \cdot I_{\{t<T\}} \;=\; \sum_s \in \pi_n H_s \cdot I_{\{s<T\}} \cdot I_{(s,s')}(t) \cdot I_{(0,\infty)}(T - t).$$

Thus $H \in \mathcal{L}^2_{a,\text{loc}}([0, T))$. Moreover, suppose that (T_k) is a sequence of stopping times approaching T from below in the sense of the definition of a predictable stopping time given above. Then

$$\widetilde{T}_k \;:=\; T_k \wedge \inf\{t \ge 0 \;:\; |H_t| \ge k\}, \qquad k \in \mathbb{N},$$

is a localizing sequence of stopping times with $H_t \cdot I_{\{t<T_k\}}$ in $\mathcal{L}^2_a([0, T))$ for any k, and $\widetilde{T}_k \nearrow T$. Therefore, by definition of the Itô integral and by Theorem 5.5,

$$\begin{aligned}\int_0^t H_s\, dB_s \;&=\; \int_0^t H_s \cdot I_{\{s<\widetilde{T}_k\}}\, dB_s \;=\; \int_0^t H_s \cdot I_{\{s\le\widetilde{T}_k\}}\, dB_s \\ &=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} H_s \cdot (B_{s'\wedge t} - B_s) \qquad \text{on } \{t \le \widetilde{T}_k\}\end{aligned}$$

w.r.t. convergence in probability. Since

$$P\left[\{t < T\} \setminus \bigcup_k \{t \leq \widetilde{T}_k\}\right] \;=\; 0,$$

we obtain (5.4.4). □

Chapter 6

Itô's formula and pathwise integrals

Our approach to Itô's formula in this chapter follows that of [Föllmer: Stochastic Analysis, Vorlesungsskript Uni Bonn WS91/92]. We start with a heuristic derivation of the formula that will be the central topic of this chapter.

Suppose that $s \mapsto X_s$ is a function from $[0,t]$ to $\mathbb{R}$, and F is a smooth function on $\mathbb{R}$. If (π_n) is a sequence of partitions of the interval $[0,t]$ with $\operatorname{mesh}(\pi_n) \to 0$ then by Taylor's theorem

$$F(X_{s'}) - F(X_s) \;=\; F'(X_s)\cdot(X_{s'}-X_s) + \frac{1}{2}F''(X_s)\cdot(X_{s'}-X_s)^2 + \text{higher order terms.}$$

Summing over $s \in \pi_n$ we obtain

$$F(X_t) - F(X_0) \;=\; \sum_{s\in\pi_n} F'(X_s)\cdot(X_{s'}-X_s) + \frac{1}{2}F''(X_s)\cdot(X_{s'}-X_s)^2 + \ldots \tag{6.0.1}$$

We are interested in the limit of this formula as $n \to \infty$.

(a) Classical case, e.g. X continuously differentiable For $X \in C^1$ we have

$$\begin{aligned} X_{s'} - X_s &= \frac{dX_s}{ds}(s'-s) + O(|s-s'|^2), \qquad \text{and} \\ (X_{s'}-X_s)^2 &= O(|s-s'|^2). \end{aligned}$$

Therefore, the second order terms can be neglected in the limit of (6.0.1) as $\operatorname{mesh}(\pi_n) \to 0$. Similarly, the higher order terms can be neglected, and we obtain the limit equation

$$F(X_t) - F(X_0) \;=\; \int_0^t F'(X_s)\, dX_s, \tag{6.0.2}$$

or, in differential notation,

$$dF(X_t) \;=\; F'(X_t)\,dX_t, \tag{6.0.3}$$

Of course, (6.0.3) is just the chain rule of classical analysis, and (6.0.2) is the equivalent chain rule for Stieltjes integrals, cf. Section 6.1 below.

(b) X_t **Brownian motion** If (X_t) is a Brownian motion then

$$E[(X_{s'} - X_s)^2] \;=\; s' - s.$$

Summing these expectations over $s \in \pi_n$, we obtain the value t independently of n. This shows that the sum of the second order terms in (6.0.1) can not be neglected anymore. Indeed, as $n \to \infty$, a law of large numbers type result implies that we can almost surely replace the squared increments $(X_{s'}-X_s)^2$ in (6.0.1) asymptotically by their expectation values. The higher order terms are on average $O(|s'-s|^{3/2})$ whence their sum can be neglected. Therefore, in the limit of (6.0.1) as $n \to \infty$ we obtain the modified chain rule

$$F(X_t) - F(X_0) \;=\; \int_0^t F'(X_s)\,dX_s + \frac{1}{2}\int_0^t F''(X_s)\,ds \tag{6.0.4}$$

with probability one. The equation (6.0.4) is the basic version of Itô's celebrated formula, which, as turned out recently, has been independently discovered by W. Doeblin.

In this chapter, we will first introduce Stieltjes integrals and the chain rule from Stieltjes calculus systematically. After computing the quadratic variation of Brownian motion in Section **??**, we will prove in Section **??** a general version of Itô's formula in dimension one. As an aside we obtain a pathwise definition for stochastic integrals involving only a single one-dimensional process due to Föllmer. The subsequent sections contain extensions to the multivariate and time-dependent case, as well as first applications.

6.1 Stieltjes integrals and chain rule

In this section, we define Lebesgue-Stieltjes integrals w.r.t. deterministic functions of bounded variation, and prove a corresponding chain rule. The resulting calculus can then be applied path by path to stochastic processes with sample paths of bounded variation.

Lebesgue-Stieltjes integrals

Fix $u \in (0, \infty]$, and suppose that $t \mapsto A_t$ is a right-continuous and non-decreasing function on $[0, u)$. Then $A_t - A_0$ is the distribution function of the positive measure μ on $(0, u)$ determined uniquely by

$$\mu_A[(s,t]] \;=\; A_t - A_s \qquad \text{for any } 0 \leq s \leq t < u.$$

Therefore, we can define integrals of type $\int_s^t H_s \, dA_s$ as Lebesgue integrals w.r.t. the measure μ_A. We extend μ trivially to the interval $[0, u)$, so $\mathcal{L}^1_{\text{loc}}([0, u), \mu_A)$ is the space of all functions $H : [0, u) \to \mathbb{R}$ that are integrable w.r.t. μ_A on any interval $(0, t)$ with $t < u$. Then for any $u \in [0, \infty]$ and any function $H \in \mathcal{L}^1_{\text{loc}}([0, u), \mu_A)$, the **Lebesgue-Stieltjes integral of H w.r.t. A** is defined by

$$\int_s^t H_r \, dA_r \;:=\; \int H_r \cdot I_{(s,t]}(r) \mu_A(dr) \qquad \text{for } 0 \leq s \leq t < u.$$

It is easy to verify that the definition is consistent, i.e., varying u does not change the definition of the integrals, and that $t \mapsto \int_0^t H_r \, dA_r$ is again a right-continuous function.

For an arbitrary right-continuous function $A : [0, u) \to \mathbb{R}$, the (first order) variation of A on an interval $[0, t)$ is defined by

$$V^{(1)}_t(A) \;:=\; \sup_\pi \sum_{\substack{s \in \pi \\ s < t}} |A_{s' \wedge t} - A_{s \wedge t}| \qquad \text{for } 0 \leq t < u,$$

where the supremum is over all partitions π of $\mathbb{R}_+$. The function $t \mapsto A_t$ is said to be **(locally) of bounded variation** on the interval $[0, u)$ iff $V^{(1)}_t(A) < \infty$ for any $t \in [0, u)$. Any right-continuous function of bounded variation can be written as the difference of two non-decreasing right-continuous functions. In fact, we have

$$A_t \;=\; A^{\nearrow}_t - A^{\searrow}_t \tag{6.1.1}$$

with

$$A_t^{\nearrow} \;=\; \sup_{\pi} \sum_{s\in\pi} (A_{s'\wedge t} - A_{s\wedge t})^+ \;=\; \frac{1}{2}(V_t^{(1)}(A) + A_t), \tag{6.1.2}$$

$$A_t^{\searrow} \;=\; \sup_{\pi} \sum_{s\in\pi} (A_{s'\wedge t} - A_{s\wedge t})^- \;=\; \frac{1}{2}(V_t^{(1)}(A) - A_t). \tag{6.1.3}$$

Exercise. Prove that if A_t is right-continuous and is locally of bounded variation on $[0,u)$ then the functions $V_t^{(1)}(A)$, $A_t^{\nearrow}$ and $A_t^{\searrow}$ are all right-continuous and non-decreasing for $t < u$.

Remark (Hahn-Jordan decomposition). The functions $A_t^{\nearrow} - A_0^{\nearrow}$ and $A_t^{\searrow} - A_0^{\searrow}$ are again distribution functions of positive measures μ_A^+ and μ_A^- on $(0,u)$. Correspondingly, $A_t - A_0$ is the distribution function of the signed measure

$$\mu_A[B] \;:=\; \mu_A^+[B] - \mu_A^-[B], \qquad B \in \mathcal{B}(0,u), \tag{6.1.4}$$

and $V_t^{(1)}$ is the distribution of the measure $|\mu_A| = \mu_A^+ - \mu_A^-$. It is a consequence of (6.1.5) and (6.1.6) that the measures μ_A^+ and μ_A^- are singular, i.e., the mass is concentrated on disjoint sets S^+ and S^-. The decomposition (6.1.7) is hence a particular case of the Hahn-Jordan decomposition of a signed measure μ of bounded variation into a positive and a negative part, and the measure $|\mu|$ is the total variation measure of μ, cf. e.g. [Alt: Lineare Funktionalanalysis].

We can now apply (6.1.1) to define Lebesgue-Stieltjes integrals w.r.t. functions of bounded variation. A function is integrable w.r.t. a signed measure μ if and only if it is integrable w.r.t. both the positive part μ^+ and the negative part μ^-. The Lebesgue integral w.r.t. μ is then defined as the difference of the Lebesgue integrals w.r.t. μ^+ and μ^-. Correspondingly, we define the Lebesgue-Stieltjes integral w.r.t. a function A_t of bounded variation as the Lebesgue integral w.r.t. the associated signed measure μ_A:

Definition. *Suppose that $t \mapsto A_t$ is right-continuous and locally of bounded variation on $[0,u)$. Then the* ***Lebesgue-Stieltjes integral w.r.t.*** *A is defined by*

$$\int_s^t H_r\, dA_r \;:=\; \int H_r \cdot I_{(s,t]}(r)\, dA_r^{\nearrow} - \int H_r \cdot I_{(s,t)}(r)\, dA_r^{\searrow}, \qquad 0 \le s \le t < u,$$

for any function $H \in \mathcal{L}^1_{loc}([0,u),|dA|)$ where

$$\mathcal{L}^1_{loc}([0,u),|dA|) := \mathcal{L}^1_{loc}([0,u),dA^{\nearrow}) \cap \mathcal{L}^1_{loc}([0,u),dA^{\searrow})$$

is the intersection of the local $\mathcal{L}^1$ spaces w.r.t. the positive measures $dA^{\nearrow} = \mu_A^+$ and $dA^{\searrow} = \mu_A^-$ on $[0,u)$, or, equivalently, the local $\mathcal{L}^1$ space w.r.t. the total variation measure $|dA| = |\mu_A|$.

Remark. (1). *Simple integrands:* If $H_t = \sum_{i=0}^{n-1} c_i \cdot I_{(t_i,t_{i+1}]}$ is a step function with $0 \le t_0 < t_1 < \ldots < t_n < u$ and $c_0, c_1, \ldots, c_{n-1} \in \mathbb{R}$ then

$$\int_0^t H_s\, dA_s \;=\; \sum_{i=0}^{n-1} c_i \cdot (A_{t_{i+1}\wedge t} - A_{t_i \wedge t}).$$

(2). *Continuous integrands; Riemann-Stieltjes integral:* If H is a continuous function then the Stieltjes integral can be approximated by Riemann sums:

$$\int_0^t H_s\, dA_s \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} H_s \cdot (A_{s'\wedge t} - A_s), \qquad t \in [0,u),$$

for any sequence (π_n) of partitions of $\mathbb{R}_+$ such that $\mathrm{mesh}(\pi_n) \to 0$. For the proof note that the step functions

$$H_r^n \;=\; \sum_{\substack{s\in\pi_n \\ s<t}} H_s \cdot I_{(s,s']}(r), \qquad r \in [0,u),$$

converge to H_r pointwise on $(0,u)$ by continuity. Moreover, again by continuity, H_r is locally bounded on $[0,u)$,and hence the sequence H_r^n is locally uniformly bounded. Therefore,

$$\int H_r^n I_{(0,t]}(r)\, dA_r \;\longrightarrow\; \int H_r \dot{I}_{(0,t]}(r)\, dA_r$$

for any $t < n$ by the dominated convergence theorem.

(3). *Absolutely continuous integrators:* If A_t is an absolutely continuous function on $[0,u)$ then A_t has locally bounded variation

$$V_t^{(1)}(A) \;=\; \int_0^t |A_s'|\, ds \;<\; \infty \qquad \text{for } t \in [0,u).$$

The signed measure μ_A with distribution function $A_t - A_0$ is then absolutely continuous w.r.t. Lebesgue measure with Radon-Nikodym density

$$\frac{d\mu_A}{dt}(t) = A'_t \qquad \text{for almost every } t \in [0,u).$$

Therefore,

$$\mathcal{L}^1_{\text{loc}}([0,u), |dA|) = \mathcal{L}^1_{\text{loc}}([0,u), |A'|dt),$$

and the Lebesgue-Stieltjes integral of a locally integrable function H is given by

$$\int_0^t H_s\, dA_s = \int_0^t H_s A'_s\, ds \qquad \text{for } t \in [0,u).$$

In the applications that we are interested in, the integrand will mostly be continuous, and the integrator absolutely continuous. Hence Remarks (2) and (3) above apply.

The chain rule in Stieltjes calculus

We are now able to prove Itô's formula in the special situation where the integrator has bounded variation. In this case, the second order correction disappears, and Itô's formula reduces to the classical chain rule from Stieltjes calculus:

Theorem 6.1 (Fundamental Theorem of Stieltjes Calculus). *Suppose that $A : [0,u) \to \mathbb{R}$ is a continuous function of locally bounded variation. Then for any $F \in C^2(\mathbb{R})$,*

$$F(A_t) - F(A_0) = \int_0^t F'(A_s)\, dA_s \qquad \forall t \in [0,u). \tag{6.1.5}$$

Proof. Let $t \in [0,u)$ be given. Choose a sequence of partitions (π_n) of $\mathbb{R}_+$ with $\text{mesh}(\pi_n) \to 0$, and let

$$\Delta A_s := A_{s'\wedge t} - A_{s\wedge t} \qquad \text{for } s \in \pi_n,$$

where, as usual, s' denotes the next partition point. By Taylor's formula, for $s \in \pi_n$ with $s < t$ we have

$$F(A_{s'\wedge t}) - F(A_s) = F'(A_s)\Delta A_s + \frac{1}{2}F''(Z_s)\cdot(\Delta A_s)^2,$$

where Z_s is an intermediate value between A_s and $A_{s'\wedge t}$. Summing over $s \in \pi_n$, we obtain

$$F(A_t) - F(A_0) \;=\; \sum_{\substack{s\in\pi_n \\ s<t}} F'(A_s)\Delta A_s + \frac{1}{2}\sum_{\substack{s\in\pi_n \\ s<t}} F''(Z_s)(\Delta A_s)^2. \tag{6.1.6}$$

As $n \to \infty$, the first (Riemann) sum converges to the Stieltjes integral $\int_0^t F'(A_s)\,dA_s$ by continuity of $F'(A_s)$, cf. Remark (2) above.

To see that the second sum converges to zero, note that the range of the continuous function A restricted to $[0,t]$ is a bounded interval. Since F'' is continuous by assumption, F'' is bounded on this range by a finite constant c. As Z_s is an intermediate value between A_s and $A_{s'\wedge t}$, we obtain

$$\left|\sum_{\substack{s\in\pi_n \\ s<t}} F''(Z_s)(\Delta A_s)^2\right| \;\le\; c\cdot \sum_{\substack{s\in\pi_n \\ s<t}} (\Delta A_s)^2 \;\le\; c\cdot V_t^{(1)}(A)\cdot \sup_{\substack{s\in\pi_n \\ s<t}} |\Delta A_s|.$$

Since $V_t^{(1)}(A) < \infty$, and A is a uniformly continuous function on $[0,t]$, the right hand side converges to 0 as $n \to \infty$. Hence we obtain (6.1.5) in the limit of (6.1.6) as $n \to \infty$. □

To see that (6.1.5) can be interpreted as a chain rule, we write the equation in differential form:

$$dF(A) \;=\; F'(A)dA. \tag{6.1.7}$$

In general, the equation (6.1.7) is to be understood mathematically only as an abbreviation for the integral equation (6.1.5). For intuitive arguments, the differential notation is obviously much more attractive than the integral form of the equation. However, for the differential form to be useful at all, we should be able to multiply the equation (6.1.7) by another function, and still obtain a valid equation. This is indeed possible due to the next result, which states briefly that if $dI = H\,dA$ then also $G\,dI = GH\,dA$:

Theorem 6.2 (Stieltjes integrals w.r.t. Stieltjes integrals). *Suppose that $I_s = \int_0^s H_r dA_r$ where $A : [0,u) \to \mathbb{R}$ is a function of locally bounded variation, and $H \in \mathcal{L}^1_{loc}([0,u), |dA|)$.*

Then the function $s \mapsto I_s$ *is again right continuous with locally bounded variation* $V_t^{(1)}(I) \leq \int\limits_0^t |H|\,|dA| < \infty$*, and, for any function* $G \in \mathcal{L}^1_{loc}([0,u),|dI|)$*,*

$$\int\limits_0^t G_s\,dI_s \;=\; \int\limits_0^t G_s H_s\,dA_s \qquad \textit{for } t \in [0,u). \tag{6.1.8}$$

Proof. Right continuity of I_t and the upper bound for the variation are left as an exercise. We now use Riemann sum approximations to prove (6.1.8) if G is continuous. For a partition $0 = t_0 < t_1 < \ldots < t_k = t$, we have

$$\sum_{i=0}^{n-1} G_{t_i}(I_{t_{i+1}} - I_{t_i}) \;=\; \sum_{i=0}^{n-1} G_{t_i} \cdot \int\limits_{t_i}^{t_{i+1}} H_s\,dA_s \;=\; \int\limits_0^t G_{\lfloor s \rfloor} H_s\,dA_s$$

where $\lfloor s \rfloor$ denotes the largest partition point $\leq s$. Choosing a sequence (π_n) of partitions with $\mathrm{mesh}(\pi_n) \to 0$, the integral on the right hand side converges to the Lebesgue-Stieltjes integral $\int\limits_0^t G_s H_s\,dA_s$ by continuity of G and the dominated convergence theorem, whereas the Riemann sum on the left hand side converges to $\int\limits_0^t G_s\,dI_s$. Hence (6.1.8) holds for continuous G. The equation for general $G \in \mathcal{L}^1_{\mathrm{loc}}([0,u),|dI|)$ follows then by standard arguments. □

6.2 Quadratic variation, Itô's formula and pathwise Itô integrals

Our next goal is to derive a generalization of the chain rule from Stieltjes calculus to continuous functions that are not of bounded variation. Examples of such functions are typical sample paths of Brownian motion. As pointed out above, an additional term will appear in the chain rule in this case.

Quadratic variation

Consider once more the approximation (6.1.6) that we have used to prove the fundamental theorem of Stieltjes calculus. We would like to identify the limit of the last sum

$\sum_{s\in\pi_n} F''(Z_s)(\Delta A_s)^2$ when A has unbounded variation on finite intervals. For $F'' = 1$ this limit is called the quadratic variation of A if it exists:

Definition. *Let $u \in (0,\infty]$ and let (π_n) be a sequence of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$. The* ***quadratic variation*** *$[X]_t$ of a continuous function $X : [0,u) \to \mathbb{R}$ w.r.t. the sequence (π_n) is defined by*

$$[X]_t \;=\; \lim_{n\to\infty} \sum_{s\in\pi_n} (X_{s'\wedge t} - X_{s\wedge t})^2 \qquad \textit{for } t \in [0,u)$$

whenever the limit exists.

WARNINGS (Dependence on partition, classical 2-variation)

(1). The quadratic variation should not be confused with the classical 2-variation defined by

$$V_t^{(2)}(X) \;:=\; \sup_{\pi} \sum_{s\in\pi} |X_{s'\wedge t} - X_{s\wedge t}|^2$$

where the supremum is over all partitions π. The classical 2-variation $V_t^{(2)}(X)$ is strictly positive for any function X that is not constant on $[0,t]$ whereas $[X]_t$ vanishes in many cases, cf. Example (1) below.

(2). In general, the quadratic variation may depend on the sequence of partitions considered. See however Examples (1) and (3) below.

Example. (1). *Functions of bounded variation:* For any continuous function $A : [0,u) \to \mathbb{R}$ of locally bounded variation, the quadratic variation along (π_n) vanishes:

$$[A]_t \;=\; 0 \qquad \text{for any } t \in [0,u).$$

In fact, for $\Delta A_s = A_{s'\wedge t} - A_{s\wedge t}$ we have

$$\sum_{s\in\pi_n} |\Delta A_s|^2 \;\le\; V_t^{(1)}(A) \cdot \sup_{\substack{s\in\pi_n \\ s<t}} |\Delta A_s| \;\to\; 0 \qquad \text{as } n\to\infty$$

by uniform continuity and since $V_t^{(1)}(A) < \infty$.

(2). *Perturbations by functions of bounded variation:* If the quadratic variation $[X]_t$ of X w.r.t. (π_n) exists, then $[X + A_t]$ also exists, and

$$[X + A]_t = [X]_t.$$

This holds since

$$\sum |\Delta(X + A)|^2 = \sum (\Delta X)^2 + 2 \sum \Delta X \Delta A + \sum (\Delta A)^2,$$

and the last two sums converge to 0 as $\mathrm{mesh}(\pi_n) \to 0$ by Example (1) and the Cauchy-Schwarz inequality.

(3). *Brownian motion:* If $(B_t)_{t \geq 0}$ is a one-dimensional Brownian motion then P-almost surely,

$$[B]_t = t \qquad \text{for all } t \geq 0$$

w.r.t. any *fixed* sequence (π_n) of partitions such that $\mathrm{mesh}(\pi_n) \to 0$, cf. Theorem **??** below.

(4). *Itô processes:* If $I_t = \int_0^t H_s \, dB_s$ is the stochastic integral of a process $H \in \mathcal{L}^2_{a,\mathrm{loc}}(0, \infty)$ w.r.t. Brownian motion then almost surely, the quadratic variation w.r.t. a fixed sequence of partitions is

$$[I]_t = \int_0^t H_s^2 \, ds \qquad \text{for all } t \geq 0.$$

Note that the exceptional sets in Example (3) and (4) may depend on the sequence (π_n)

If it exists, the quadratic variation $[X]_t$ is a non-decreasing function in t.

Lemma 6.3. *Suppose that $X : [0, u) \to \mathbb{R}$ is a continuous function. If the quadratic variation $[X]_t$ along (π_n) exists for $t \in [0, u)$, and $t \mapsto [X]_t$ is continuous then*

$$\sum_{\substack{s \in \pi_n \\ s < t}} H_s \cdot (X_{s' \wedge t} - X_s)^2 \quad \longrightarrow \quad \int_0^t H_s \, d[X]_s \qquad \text{as } n \to \infty \tag{6.2.1}$$

for any continuous function $H : [0, u) \to \mathbb{R}$ and any $t \geq 0$.

Remark. Heuristically, the assertion of the lemma tells us

$$\text{“} \int H \, d[X] \; = \; \int H \, (dX)^2 \text{”},$$

i.e., the infinitesimal increments of the quadratic variation are something like squared infinitesimal increments of X. This observation is crucial for controlling the second order terms in the Taylor expansion used for proving the Itô-Doeblin formula.

Proof. The sum on the left-hand side of (6.2.1) is the integral of H w.r.t. the finite positive measure

$$\mu_n \; := \; \sum_{\substack{s \in \pi_n \\ s < t}} (X_{s' \wedge t} - X_s)^2 \cdot \delta_s$$

on the interval $[0, t)$. The distribution function of μ_n is

$$F_n(u) \; =: \sum_{\substack{s \in \pi_n \\ s \leq t}} (X_{s' \wedge t} - X_s)^2, \qquad u \in [0, t].$$

As $n \to \infty$, $F_n(u) \to [X]_u$ for any $u \in [0, t]$ by continuity of X. Since $[X]_u$ is a continuous function of u, convergence of the distribution functions implies weak convergence of the measures μ_n to the measure $d[X]$ on $[0, t)$ with distribution function $[X]$. Hence,

$$\int H_s \mu_n(ds) \; \longrightarrow \; \int H_s \, d[X]_s \qquad \text{as } n \to \infty$$

for any continuous function $H : [0, t] \to \mathbb{R}$. □

Itô's formula and pathwise integrals in $\mathbb{R}^1$

We are now able to complete the proof of the following purely deterministic (pathwise) version of the one-dimensional Itô formula going back to [Föllmer: Calcul d'Itô sans probabilités, Sém. Prob XV, LNM850]:

Theorem 6.4 (Itô's formula without probability). *Suppose that $X : [0, u) \to \mathbb{R}$ is a continuous function with continuous quadratic variation $[X]$ w.r.t. (π_n). Then for any function F that is C^2 in a neighbourhood of $X([0, u))$, and for any $t \in [0, u)$, the Itô integral*

$$\int_0^t F'(X_s) \, dX_s \; = \; \lim_{n \to \infty} \sum_{\substack{s \in \pi_n \\ s < t}} F'(X_s) \cdot (X_{s' \wedge t} - X_s) \tag{6.2.2}$$

exists, and Itô’s formula

$$F(X_t) - F(X_0) \;=\; \int_0^t F'(X_s)\,dX_s + \frac{1}{2}\int_0^t F''(X_s)\,d[X]_s \tag{6.2.3}$$

holds. In particular, if the quadratic variation $[X]$ *does not depend on* (π_n) *then the Itô integral (6.2.2) does not depend on* (π_n) *either.*

Note that the theorem **implies the existence** of $\int_0^t f(X_S)\,dX_s$ for any function $f \in C^1(\mathbb{R})$! Hence this type of Itô integrals can be defined in a purely deterministic way without relying on the Itô isometry. Unfortunately, the situation is more complicated in higher dimensions, cf. **??** below.

Proof. Fix $t \in [0,u)$ and $n \in \mathbb{N}$. As before, for $s \in \pi_n$ we set $\Delta X_s = X_{s'\wedge t} - X_{s\wedge t}$ where s' denotes the next partition point. Then as above we have

$$F(X_t) - F(X_0) \;=\; \sum_{\substack{s\in\pi_n\\ s<t}} F'(X_s)\Delta X_s + \frac{1}{2}\sum_{\substack{s\in\pi_n\\ s<t}} F''(Z_s^{(n)})(\Delta X_s)^2 \tag{6.2.4}$$

$$\;=\; \sum_{\substack{s\in\pi_n\\ s<t}} F'(X_s)\Delta X_s + \frac{1}{2}\sum_{\substack{s\in\pi_n\\ s<t}} F''(X_s)(\Delta X_s)^2 + \sum_{\substack{s\in\pi_n\\ s<t}} R_s^{(n)}, \tag{6.2.5}$$

where $Z_s^{(n)}$ is an intermediate point between X_s and $X_{s'\wedge t}$, and $R_s^{(n)} := \frac{1}{2}(F''(Z_s^{(n)}) - F''(X_s))\cdot(\Delta X_s)^2$. As $n \to \infty$, the second sum on the right hand side of (6.2.4) converges to $\int_0^t F''(X_s)\,d[X]_s$ by Lemma 6.3. We claim that the sum of the remainders $R_s^{(n)}$ converges to 0. To see this note that $Z_s^{(n)} = X_r$ for some $r \in [s, s'\wedge t]$, whence

$$|R_s^{(n)}| \;=\; |F''(Z_s^{(n)}) - F''(X_s)|\cdot(\Delta X_s)^2 \;\le\; \frac{1}{2}\varepsilon_n(\Delta X_s)^2,$$

where

$$\varepsilon_n \;:=\; \sup_{\substack{a,b\in[0,t]\\ |a-b|\le \mathrm{mesh}(\pi_n)}} |F''(X_a) - F''(X_b)|.$$

As $n \to \infty$, ε_n converges to 0 by uniform continuity of $F'' \circ X$ on the interval $[0, t]$. Thus

$$\sum |R_s^{(n)}| \;\leq\; \frac{1}{2}\varepsilon_n \sum_{\substack{s\in\pi_n \\ s<t}} (\Delta X_s)^2 \;\to\; 0 \qquad \text{as well,}$$

because the sum of the squared increments converges to the finite quadratic variation $[X]_t$.

We have shown that all the terms on the right hand side of (6.2.4) except the first Riemann-Itô sum converge as $n \to \infty$. Hence, by (6.2.4), the limit $\int\limits_0^t F'(X_s)\, dX_s$ of the Riemann Itô sums also exists, and the limit equation (6.2.2) holds. □

Remark. (1). In differential notation, we obtain the Itô chain rule

$$dF(X) \;=\; F'(X)\, dX + \frac{1}{2}F''(X)\, d[X]$$

which includes a second order correction term due to the quadratic variation. A justification for the differential notation is given in Section **??**.

(2). For functions X with $[X] = 0$ we recover the classical chain rule $dF(X) = F'(X)\, dX$ from Stieltjes calculus as a particular case of Itô's formula.

Example. (1). *Exponentials:* Choosing $F(x) = e^x$ in Itô's formula, we obtain

$$e^{X_t} - e^{X_0} \;=\; \int\limits_0^t e^{X_s}\, dX_s + \frac{1}{2}\int\limits_0^t e^{X_s}\, d[X]_s,$$

or, in differential notation,

$$de^X \;=\; e^X\, dX + \frac{1}{2}e^X\, d[X].$$

Thus e^X does *not* solve the Itô differential equation

$$dZ \;=\; Z\, dX \tag{6.2.6}$$

if $[X] \neq 0$. An appropriate renormalization is required instead. We will see below that the correct solution of (6.2.6) is given by

$$Z_t \;=\; \exp\left(X_t - [X]/2\right),$$

cf. Theorem **??**.

(2). *Polynomials:* Similarly, choosing $F(x) = x^n$ for some $n \in \mathbb{N}$, we obtain

$$dX^n \;=\; nX^{n-1}\,dX + \frac{n(n-1)}{2}X^{n-2}\,[X].$$

Again, X^n does not solve the equation $dX^n = nX^{n-1}\,dX$. Here, the appropriate renormalization leads to the Hermite polynomials $:X:^n$, cf. **??** below.

The chain rule for anticipative integrals

The for of the second order correction term appearing in Itô's formula depends crucially on choosing non-anticipative Riemann sum approximations, we obtain different correction terms, and hence also different notions of integrals.

Theorem 6.5. *Suppose that $X : [0,u) \to \mathbb{R}$ is continuous with continuous quadratic variation $[X]$ along (π_n). Then for any function F that is X^2 in a neighbourhood of $X([0,u))$ and for any $t \geq 0$, the* ***backward Itô integral***

$$\int_0^t F'(X_s)\,\hat{d}X_s \;:=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} F'(X_{s'\wedge t}) \cdot (X_{s'\wedge t} - X_s),$$

and the ***Stratonovich integral***

$$\int_0^t F'(X_s)\,\circ dX_s \;:=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} \frac{1}{2}(F'(X_s) + F'(X_{s'\wedge t})) \cdot (X_{s'\wedge t} - X_s)$$

exist, and

$$F(X_t) - F(X_0) \;=\; \int_0^t F'(X_s)\,\hat{d}X_s - \frac{1}{2}\int_0^t F''(X_s)\,d[X]_s \tag{6.2.7}$$

$$\;=\; \int_0^t F'(X_s)\,\circ dX_s. \tag{6.2.8}$$

Proof. The proof of the backward Itô formula (6.2.7) is completely analogous to that of Itô's formula. Th Stratonovich formula (6.2.8) follows by averaging the Riemann sum approximations to the forward and backward Itô rule. □

Note that Stratonovich integrals satisfy the classical chain rule

$$\circ dF(X) \;=\; F'(X) \,\circ dX.$$

This makes them very attractive for various applications. For example, in stochastic differential geometry, the chain rule is of fundamental importance to construct stochastic processes that stay on a given manifold. Therefore, it is common to use Stratonovich instead of Itô calculus in this context, cf. **??** also the example in the next section.
On the other hand, Stratonovich calculus has a significant disadvantage against Itô calculus: The Stratonovich integrals

$$\int_0^t H_s \,\circ dB_s \;=\; \lim_{n\to\infty} \sum \frac{1}{2}(H_s + H_{s'\wedge t})(B_{s'\wedge t} - B_s)$$

w.r.t. Brownian motion typically are not martingales, because the coefficients $\frac{1}{2}(H_s + H_{s'\wedge t})$ are not predictable.

6.3 First applications to Brownian motion and martingales

Our next aim is to compute the quadratic variation and to state Itô's formula for typical sample paths of Brownian motion. Let (π_n) be a sequence of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$. We note first that for any function $t \mapsto X_t$ the identity

$$X_t^2 - X_0^2 \;=\; \sum_{\substack{s\in\pi_n \\ s<t}} (X_{s'\wedge t}^2 - X_s^2) \;=\; V_t^n + 2I_t^n \tag{6.3.1}$$

with

$$V_t^n \;=\; \sum_{\substack{s\in\pi_n \\ s<t}} (X_{s'\wedge t} - X_s)^2 \quad \text{and} \quad I_t^n \;=\; \sum_{\substack{s\in\pi_n \\ s<t}} X_s \cdot (X_{s'\wedge t} - X_s)$$

holds. The equation (6.3.1) is a discrete approximation of Itô's formula for the function $F(x) = x^2$. The remainder terms in the approximation vanish in this particular case.

Note that by (6.3.1), the quadratic variation $[X]_t = \lim_{n\to\infty} V_t^n$ exists if and only if the Riemann sum approximations I_t^n to the Itô integral $\int_0^t X_s\, dX_s$ converge:

$$\exists\, [X]_t = \lim_{n\to\infty} V_t^n \quad \iff \quad \exists \int_0^t X_s\, dX_s = \lim_{n\to\infty} I_t^n.$$

Now suppose that (X_t) is a continuous martingale with $E[X_t^2] < \infty$ for any $t \geq 0$. Then the Riemann sum approximations (I_t^n) are continuous martingales for any $n \in \mathbb{N}$. Therefore, by the maximal inequality, for a given $u > 0$, the processes (I_t^n) and (V_t^n) converge uniformly for $t \in [0, u]$ in $L^2(P)$ if and only if the random variables I_u^n or I_u^n respectively converge in $L^2(P)$.

Quadratic variation of Brownian motion

For the sample paths of a Brownian motion B, the quadratic variation $[B]$ exists almost surely along any *fixed* sequence of partitions (π_n) with $\mathrm{mesh}(\pi_n) \to 0$, and $[B]_t = t$. In particular, $[B]$ is a *deterministic* function that does not depend on (π_n). The reason is a law of large numbers type effect when taking the limit of the sum of squared increments as $n \to \infty$.

Theorem 6.6 (P. Lévy). *If (B_t) is a one-dimensional Brownian motion on $(\Omega, \mathcal{A}, P)$ then as $n \to \infty$*

$$\sup_{t\in[0,u]} \left| \sum_{\substack{s\in\pi_n \\ s<t}} (B_{s'\wedge t} - B_s)^2 - t \right| \quad \longrightarrow \quad 0 \qquad P\text{-a.s. and in } L^2(\Omega, \mathcal{A}, P) \tag{6.3.2}$$

for any $u \in (0, \infty)$, and for each sequence (π_n) of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$.

Warning. (1). Although the almost sure limit in (6.3.2) does not depend on the sequence (π_n), the exceptional set may depend on the chosen sequence!

(2). The classical quadratic variation $V_t^{(2)}(B) = \sup_\pi \sum_{s\in\pi} (\Delta B_s)^2$ is almost surely infinite for any $t \geq 0$. The classical p-Variation is almost surely finite if and only if $p > 2$.

Proof. (1). *L^2-convergence for fixed t:* As usual, the proof of L^2 convergence is comparatively simple. For $V_t^n = \sum_{s\in\pi_n} (\Delta B_s)^2$ with $\Delta B_s = B_{s'\wedge t} - B_{s\wedge t}$, we have

$$\begin{aligned} E[V_t^n] &= \sum_{s\in\pi_n} E[(\Delta B_s)^2] = \sum_{s\in\pi_n} \Delta s = t, \qquad \text{and} \\ \mathrm{Var}[V_t^n] &= \sum_{s\in\pi_n} \mathrm{Var}[(\Delta B_s)^2] = \sum_{s\in\pi_n} E[((\Delta B_s)^2 - \Delta s)^2] \\ &= E[(Z^2-1)^2] \cdot \sum_{s\in\pi_n} (\Delta s)^2 \le \text{const.} \cdot t \cdot \mathrm{mesh}(\pi_n) \end{aligned}$$

where Z is a standard normal random variable. Hence, as $n \to \infty$,

$$V_t^n - t = V_t^n - E[V_t^n] \to 0 \qquad \text{in } L^2(\Omega, \mathcal{A}, P).$$

Moreover, by (6.3.1), $V_t^n - V_t^m = I_t^n - I_t^m$ is a continuous martingale for any $n, m \in \mathbb{N}$. Therefore, the maximal inequality yields uniform convergence of V_t^n to t for t in a finite interval in the $L^2(P)$ sense.

(2). *Almost sure convergence if $\sum \mathrm{mesh}(\pi_n) < \infty$:* Similarly, by applying the maximal inequality to the process $V_t^n - V_t^m$ and taking the limit as $m \to \infty$, we obtain

$$P\left[\sup_{t\in[0,u]} |V_t^n - t| > \varepsilon\right] \le \frac{2}{\varepsilon^2} E[(V_t^n - t)^2] \le \text{const.} \cdot t \cdot \mathrm{mesh}(\pi_n)$$

for any given $\varepsilon > 0$ and $u \in (0,\infty)$. If $\sum \mathrm{mesh}(\pi_n) < \infty$ then the sum of the probabilities is finite, and hence $\sup_{t\in[0,u]} |V_t^n - t| \to 0$ almost surely by the Borel-Cantelli Lemma.

(3). *Almost sure convergence if $\sum \mathrm{mesh}(\pi_n) = \infty$:* In this case, almost sure convergence can be shown by the backward martingale convergence theorem. We refer to Proposition 2.12 in [Revuz, Yor], because for our purposes almost sure convergence w.r.t arbitrary sequences of partitions is not essential.

□

Itô's formula for Brownian motion

By Theorem 6.6, we can apply Theorem 6.4 to almost every sample path of a one-dimensional Brownian motion (B_t). If $I \subseteq \mathbb{R}$ is an open interval then for any $F \in C^2(I)$, we obtain almost surely the identity

$$F(B_t) - F(B_0) \;=\; \int_0^t F'(B_s)\,dB_s + \frac{1}{2}\int_0^t F''(B_s)\,ds \qquad \text{for all } t < T, \tag{6.3.3}$$

where $T = \min\{t \geq 0 \,:\, B_t \notin I\}$ is the first exit time from I.

Note that the pathwise integral and the Itô integral as defined in Section 5 coincide almost surely since both are limits of Riemann-Itô sums w.r.t. uniform convergence for t in a finite interval, almost surely along a common (sub)sequence of partitions.

Consequences

(1). *Doob decomposition in continuous time:* The Itô integral $M_t^F = \int_0^t F'(B_s)\,dB_s$ is a local martingale up to T, and M_t^F is a square integrable martingale if $I = \mathbb{R}$ and F' is bounded. Therefore, (6.3.3) can be interpreted as a *continuous time Doob decomposition* of the process $(F(B_t))$ into the (local) martingale part M^F and an adapted process of bounded variation. This process takes over the role of the predictable part in discrete time.

In particular, we obtain

Corollary 6.7 (Martingale problem for Brownian motion). *Brownian motion is a solution of the martingale problem for the operator* $\mathscr{L} = \frac{1}{2}\frac{d^2}{dx^2}$ *with domain* $\mathrm{Dom}(\mathscr{L}) = \{F \in C^2(\mathbb{R}) \,:\, \frac{dF}{dx}$ *is bounded*$\}$*, i.e., the process*

$$M_t^F \;=\; F(B_t) - F(B_0) - \int_0^t (\mathscr{L}f)(B_s)\,ds$$

is a martingale for any $F \in \mathrm{Dom}(\mathscr{L})$.

The corollary demonstrates how Itô's formula can be applied to obtain solutions of martingale problems, cf. ??? and **??** below for generalizations.

(2). *Kolmogorov's backward equation:* Taking expectation values in (6.3.3), we recover Kolmogorov's equation

$$E[F(B_t)] \;=\; E[F(B_0)] + \int_0^t E[(\mathscr{L}F)(B_s)]\,ds \qquad \forall\, t \geq 0$$

for any $F \in C_b^2(\mathbb{R})$. In differential form,

$$\frac{d}{dt}E[F(B_t)] \;=\; E[(\mathscr{L}F)(B_t)].$$

(3). *Computation of expectation values:* The Itô formula can be applied in many ways to compute expectation values:

Example. (a) For any $n \in \mathbb{N}$, the process

$$B_t^n - \frac{n(n-1)}{2}\int_0^t B_s^{n-2}\,ds \;=\; n \cdot \int_0^t B_s^{n-1}\,dB_s$$

is a martingale. By taking expectation values for $t = 1$ we obtain the recursion

$$\begin{aligned} E[B_1^n] \;&=\; \frac{n(n-1)}{2}\int_0^1 E[B_s^{n-2}]\,ds \;=\; \frac{n(n-1)}{2}\int_0^1 s^{n-2/2}\,ds \cdot E[B_1^{n-2}] \\ &=\; (n-1)\cdot E[B_1^{n-2}] \end{aligned}$$

for the moments of the standard normally distributed random variable B_1. Of course this identity can be obtained directly by integration by parts in the Gaussian integral $\int x^n \cdot e^{-x^2/2}\,dx$.

(b) For $\alpha \in \mathbb{R}$, the process

$$\exp(\alpha B_t) - \frac{\alpha^2}{2}\int_0^t \exp(\alpha B_s)\,ds \;=\; \alpha\int_0^t \exp(\alpha B_s)\,dB_s$$

is a martingale because $E[\int_0^t \exp(2\alpha B_s)\, ds] < \infty$. Denoting by $T_b = \min\{t \geq 0 : B_t = b\}$ the first passage time to a level $b > 0$, we obtain the identity

$$E\left[\int_0^{T_b} \exp(\alpha B_s)\, ds\right] = \frac{2}{\alpha^2}(e^{\alpha b} - 1) \qquad \text{for any } \alpha > 0$$

by optional stopping and dominated convergence.

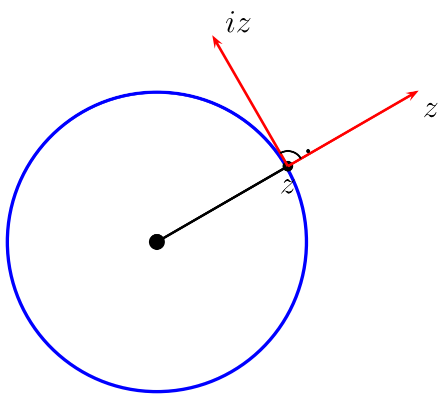
Itô's formula is also the key tool to derive or solve stochastic differential equations for various stochastic processes of interest:

Example (Brownian motion on $\mathbf{S^1}$). Brownian motion on the unit circle $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ is the process given by

$$Z_t = \exp(iB_t) = \cos B_t + i \cdot \sin B_t$$

where (B_t) is a standard Brownian motion on $\mathbb{R}^1$. Itô's formula yields the stochastic differential equation

$$dZ_t = A(Z_t)\, dB_t - \frac{1}{2} n(Z_t)\, dt, \tag{6.3.4}$$



where $A(z) = iz$ is the unit tangent vector to S^1 at the point z, and $n(z) = z$ is the outer normal vector. If we would omit the correction term $-\frac{1}{2}n(Z_t)\, dt$ in (6.3.4), the solution to the s.d.e. would not stay on the circle. This is contrary to classical o.d.e. where the correction term is not required. For Stratonovich integrals, we obtain the modified equation

$$\circ dZ_t \;=\; A(Z_t) \circ dB_t,$$

which does not involve a correction term!

We conclude this section with a first remark on applications of Itô calculus to continuous martingales, see Section **??** for details.

Remark (Itô calculus for continuous martingales).

(1). *Quadratic variation and Doob decomposition:* If (M_t) is a continuous square-integrable martingale then one can show that the Itô integral $\int_0^t M_s\,dM_s$ and, correspondingly, the quadratic variation $[M]_t$ exists w.r.t. uniform convergence of the Riemann sum approximations for t in a finite interval in mean square w.r.t. P, and the identity

$$M_t^2 - M_0^2 \;=\; 2\int_0^t M_s\,dM_s + [M]_t \qquad \text{for any } t \geq 0 \tag{6.3.5}$$

holds P-almost surely, cf. Section **??** below. The Itô integral is a continuous martingale, and thus (6.3.5) yields a continuous time Doob decomposition of the submartingale M_t^2 into a martingale and the increasing adapted process $[M]_t$. In particular, we can interpret $[M]_t$ as a continuous replacement for the conditional variance process $\langle M\rangle_t$. By localization, the identity (6.3.5) extends to continuous local martingales.

(2). *Non-constant martingales have non-trivial quadratic variation:* A first remarkable consequence of (6.3.5) is that if $[M]_u$ vanishes for some $u > 0$ then by the maximal inequality,

$$\begin{aligned} E\left[\sup_{t\in[0,u]} |M_t - M_0|^2\right] &\leq 2\cdot E[(M_u - M_0)^2] \;=\; E[M_u^2] - E[M_0^2] \\ &= E[[M]_u] \;=\; 0, \end{aligned}$$

and hence the martingale (M_t) is almost surely constant on the interval $[0,u]$. Thus in particular, *any bounded martingale with continuous sample paths of bounded variation (or, more generally, of vanishing quadratic variation) is almost surely constant!* Again, this statement extends to continuous local martingales. As a consequence, the Doob type decomposition of a stochastic process into a continuous local martingale and a continuous process of bounded variation is unique up to equivalence.

(3). *Continuous local martingales with $[M]_t = t$ are Brownian motions:* A second remarkable consequence of Itô's formula for martingales is that any continuous local martingale (M_t) (up to $T = \infty$) with quadratic variation given by $[M]_t = t$ for any $t \geq 0$ is a Brownian motion. In fact, for $0 \leq s \leq t$ and $p \in \mathbb{R}$, Itô's formula yields

$$e^{ipM_t} - e^{ipM_s} \;=\; ip \int_s^t e^{ipM_r}\, dM_r - \frac{p^2}{2} \int_s^t e^{ipM_r}\, dr$$

where the stochastic integral can be identified as a local martingale. From this identity it is not difficult to conclude that the increment $M_t - M_s$ is conditionally independent of $\mathcal{F}_s^M$ with characteristic function

$$E[e^{ip(M_t - M_s)}] \;=\; e^{-p^2(t-s)/2} \qquad \text{for any } p \in \mathbb{R},$$

i.e., (M_t) has independent increments with distribution $M_t - M_s \sim N(0, t-s)$. A detailed proof and an extension to the multi-dimensional case are given in Theorem **??** below.

(4). *Continuous local martingales as time-changed Brownian motion:* More generally, it can be shown that any continuous local martingale (M_t) is a time-changed Brownian motion:

$$M_t \;=\; B_{[M]_t}, \qquad \text{cf. Section ?? below.}$$

Independently of K. Itô, W. Doeblin has developed during the Second World War an alternative approach to stochastic calculus where stochastic integrals are defined as time changes of Brownian motion. Doeblin died at the front, and his results have been published only recently, more than fifty years later, cf. [Doeblin, Sur l'équation de Kolmogoroff, C.R.A.S. 1940], [Yor: Présentation du duplicadeté, C.R.A.S. 2000].

6.4 Multivariate and time-dependent Itô formula

We now extend Itô's formula to $\mathbb{R}^d$-valued functions and stochastic processes. Let $u \in (0, \infty]$ and suppose that $X : [0, u) \to D, X_t = (X_t^{(1)}, \ldots, X_t^{(d)})$, is a continuous function taking values in an open set $D \subseteq \mathbb{R}^d$. As before, we fix a sequence (π_n) of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$. For a function $F \in C^2(D)$, we have similarly as in the one-dimensional case:

$$\begin{aligned} F(X_{s'\wedge t}) - F(X_s) \;&=\; \nabla F(X_s) \cdot (X_{s'\wedge t} - X_s) + \\ &\frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 F}{\partial x_i \partial x_j}(X_s)(X_{s'\wedge t}^{(i)} - X_s^{(i)})(X_{s'\wedge t}^{(j)} - X_s^{(j)}) + R_s^{(n)} \end{aligned} \tag{6.4.1}$$

for any $s \in \pi_n$ with $s < t$ where the dot denotes the Euclidean inner product $R_s^{(n)}$ is the remainder term in Taylor's formula. We would like to obtain a multivariate Itô formula by summing over $s \in \pi_n$ with $s < t$ and taking the limit as $n \to \infty$. A first problem that arises in this context is the identification of the limit of the sums

$$\sum_{\substack{s \in \pi_n \\ s<t}} g(X_s) \Delta X_s^{(i)} \Delta X_s^{(j)}$$

for a continuous function $g : D \to \mathbb{R}$ as $n \to \infty$.

Covariation

Suppose that $X, Y : [0, u) \to \mathbb{R}$ are continuous functions with continuous quadratic variations $[X]_t$ and $[Y]_t$ w.r.t. (π_n).

Definition. *The function*

$$[X, Y]_t \;=\; \lim_{n\to\infty} \sum_{s \in \pi_n} (X_{s'\wedge t} - X_{s\wedge t})(Y_{s'\wedge t} - Y_{s\wedge t}), \qquad t \in [0, u),$$

is called the ***covariation of*** X ***and*** Y ***w.r.t.*** (π_n) *if the limit exists.*

The covariation $[X, Y]_t$ is the bilinear form corresponding to the quadratic form $[X]_t$. In particular, $[X, X] = [X]$. Furthermore:

Lemma 6.8 (Polarization identity). *The covariation $[X, Y]_t$ exists and is a continuous function in t if and only if the quadratic variation $[X + Y]_t$ exists and is continuous respectively. In this case,*

$$[X, Y]_t \;=\; \frac{1}{2}([X+Y]_t - [X]_t - [Y]_t).$$

Proof. For $n \in \mathbb{N}$ we have

$$2\sum_{s\in\pi_n} \Delta_s X \Delta_s Y \;=\; \sum_{s\in\pi_n}(\Delta_s X + \Delta_s Y)^2 - \sum_{s\in\pi_n}(\Delta_s X)^2 - \sum_{s\in\pi_n}(\Delta_s Y)^2.$$

The assertion follows as $n \to \infty$ because the limits $[X]_t$ and $[Y]_t$ of the last two terms are continuous functions by assumption. □

Remark. Note that by the polarization identity, the covariation $[X, Y]_t$ is the difference of two increasing functions, i.e., $t \mapsto [X, Y]_t$ has bounded variation.

Example. (1). *Functions and processes of bounded variation:* If Y has bounded variation then $[X, Y]_t = 0$ for any $t \geq 0$. Indeed,

$$\left|\sum_{s\in\pi_n} \Delta X_s \Delta Y_s\right| \;\leq\; \sup_{s\in\pi_n} |\Delta X_s| \cdot \sum_{s\in\pi_n} |\Delta Y_s|$$

and the right hand side converges to 0 by uniform continuity of X on $[0, t]$. In particular, we obtain again

$$[X + Y] \;=\; [X] + [Y] + 2[X, Y] \;=\; [X].$$

(2). *Independent Brownian motions:* If (B_t) and $(\widetilde{B}_t)$ are independent Brownian motions on a probability space $(\Omega, \mathcal{A}, P)$ then for any given sequence (π_n),

$$[B, \widetilde{B}]_t \;=\; \lim_{n\to\infty} \sum_{s\in\pi_n} \Delta B_s \Delta \widetilde{B}_s \;=\; 0 \qquad \text{for any } t \geq 0$$

P-almost surely. For the proof note that $(B_t + \widetilde{B}_t)/\sqrt{2}$ is again a Brownian motion, whence

$$[B, \widetilde{B}]_t \;=\; [(B + \widetilde{B})/\sqrt{2}]_t - \frac{1}{2}[B]_t - \frac{1}{2}[\widetilde{B}]_t \;=\; t - \frac{t}{2} - \frac{t}{2} \;=\; 0 \qquad \text{almost surely.}$$

(3). *Itô processes:* If $I_t = \int_0^t G_s\, dB_s$ and $\mathcal{F}_t = \int_0^t H_s\, dB_s$ with continuous adapted processes (G_t) and (H_t) and Brownian motions (B_t) and $(\widetilde{B}_t)$ then

$$\langle I, J\rangle_t = 0 \qquad \text{if } B \text{ and } \widetilde{B} \text{ are independent, and} \tag{6.4.2}$$

$$\langle I, J\rangle_t = \int_0^t G_s H_s\, ds \qquad \text{if } B = \widetilde{B}, \tag{6.4.3}$$

cf. Theorem **??** below.

More generally, under appropriate assumptions on G, H, X and Y, the identity

$$\langle I, J\rangle_t = \int_0^t G_s H_s\, d\langle X, Y\rangle_s$$

holds for Itô integrals $I_t = \int_0^t G_s\, dX_s$ and $J_t = \int_0^t H_s\, dY_s$, cf. e.g. Corollary **??**.

Itô to Stratonovich conversion

The covariation also occurs as the correction term in Itô compared to Stratonovich integrals:

Theorem 6.9. *If the Itô integral*

$$\int_0^t X_s\, Y_s = \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} X_s \Delta Y_s$$

and the covariation $[X, Y]_t$ *exists along a sequence* (π_n) *of partitions with* $\mathrm{mesh}(\pi_n) \to 0$ *then the corresponding backward Itô integral* $\int_0^t X_s\, \hat{d}Y_s$ *and the Stratonovich integral* $\int_0^t X_s \circ dY_s$ *also exist, and*

$$\int_0^t X_s\, \hat{d}Y_s = \int_0^t X_s\, Y_s + [X, Y]_t, \qquad \textit{and}$$

$$\int_0^t X_s \circ dY_s = \int_0^t X_s\, Y_s + \frac{1}{2}[X, Y]_t.$$

Proof. This follows from the identities

$$\begin{aligned} \sum X_{S'\wedge t}\Delta Y_s &= \sum X_s \Delta Y_s + \sum \Delta X_s \Delta Y_s, \qquad \text{and} \\ \sum \frac{1}{2}(X_s + X_{s'\wedge t})\Delta Y_s &= \sum X_s \Delta Y_s + \frac{1}{2}\sum \Delta X_s \Delta Y_s. \end{aligned}$$

□

Itô's formula in $\mathbb{R}^d$

By the polarization identity, if $[X]_t, [Y]_t$ and $[X+Y]_t$ exist and are continuous then $[X,Y]_t$ is a continuous function of bounded variation.

Lemma 6.10. *Suppose that X, Y and $X+Y$ are continuous function on $[0,u)$ with continuous quadratic variations w.r.t. (π_n). Then*

$$\sum_{\substack{s\in\pi_n \\ s<t}} H_s (X_{s'\wedge t} - X_s)(Y_{s'\wedge t} - Y_s) \quad \longrightarrow \quad \int_0^t H_s \, d[X,Y]_s \qquad \text{as } n\to\infty$$

for any continuous function $H : [0,u) \to \mathbb{R}$ and any $t \geq 0$.

Proof. The assertion follows from Lemma 6.3 by polarization. □

By Lemma 6.10, we can take the limit as $\mathrm{mesh}(\pi_n) \to 0$ in the equation derived by summing (6.4.2) over all $s \in \pi_n$ with $s < t$. In analogy to the one-dimensional case, this yields the following multivariate version of the pathwise Itô formula:

Theorem 6.11 (Multivariate Itô formula without probability). *Suppose that $X : [0,u) \to D \subseteq \mathbb{R}^d$ is a continuous function with continuous covariations $[X^{(i)}, X^{(j)}]_t, 1 \leq i,j \leq d$, w.r.t. (π_n). Then for any $F \in C^2(D)$ and $t \in [0,u)$,*

$$F(X_t) \;=\; F(X_0) + \int_0^t \nabla F(X_s)\cdot dX_s + \frac{1}{2}\sum_{i,j=1}^d \int_0^t \frac{\partial^2 F}{\partial x_i \partial x_j}(X_s)\, d[X^{(i)}, X^{(j)}]_s,$$

where the Itô integral is the limit of Riemann sums along (π_n):

$$\int_0^t \nabla F(X_s)\cdot dX_s \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} \nabla F(X_s)\cdot (X_{s'\wedge t} - X_s). \tag{6.4.4}$$

The details of the proof are similar to the one-dimensional case and left as an exercise to the reader. Note that the theorem shows in particular that the Itô integral in (6.4.4) is independent of the sequence (π_n) if the same holds for the covariations $[X^{(i)}, X^{(j)}]$.

Remark (Existence of pathwise Itô integrals). The theorem implies the existence of the Itô integral $\int_0^t b(X_s) \cdot dX_s$ if $b = \nabla F$ is the gradient of a C^2 function $F : D \subseteq \mathbb{R}^d \to \mathbb{R}$. In contrast to the one-dimensional case, not every C^1 vector field $b : D \to \mathbb{R}^d$ is a gradient. Therefore, for $d \geq 2$ we do **not** obtain existence of $\int_0^t b(X_s) \cdot dX_s$ for any $b \in C^1(D, \mathbb{R}^d)$ from Itô's formula. In particular, *we do not know in general* if the integrals $\int_0^t \frac{\partial F}{\partial x_i}(X_s)\, dX_s^{(i)}, 1 \leq i \leq d$, exists and if

$$\int_0^t \nabla F(X_s) \cdot dX_s \;=\; \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial x_i}(X_s)\, dX_s^{(i)}.$$

If (X_t) is a Brownian motion this is almost surely the case by the existence proof for Itô integrals w.r.t. Brownian motion from Section 5.

Example (Itô's formula for Brownian motion in $\mathbb{R}^d$). Suppose that $B_t = (B_t^{(1)}, \ldots, B_t^{(d)})$ is a d-dimensional Brownian motion defined on a probability space $(\Omega, \mathcal{A}, P)$. Then the component processes $B_t^{(1)}, \ldots, B_t^{(d)}$ are independent one-dimensional Brownian motions. Hence for a given sequence of partitions (π_n) with $\mathrm{mesh}(\pi_n) \to 0$, the covariations $[B^{(i)}, B^{(j)}], 1 \leq i, j \leq d$, exists almost surely by Theorem 6.6 and the example above, and

$$[B^{(i)}, B^{(j)}] \;=\; t \cdot \delta_{ij} \qquad \forall t \geq 0$$

P-almost surely. Therefore, we can apply Itô's formula to almost every trajectory. For an open subset $D \subseteq \mathbb{R}^d$ and a function $F \in C^2(D)$ we obtain:

$$F(B_t) \;=\; F(B_0) + \int_0^t \nabla F(B_s) \cdot dB_s + \frac{1}{2} \int_0^t \Delta F(B_s)\, ds \qquad \forall t < T_{D^C} \quad P\text{-a.s.} \tag{6.4.5}$$

where $T_{D^C} := \inf\{t \geq 0 \;:\; B_t \notin D\}$ denotes the first exit time from D. As in the one-dimensional case, (6.4.5) yields a decomposition of the process $F(B_t)$ into a continuous local martingale and a continuous process of bounded variation, cf. Section **??** for applications.

Product rule, integration by parts

As a special case of the multivariate Itô formula, we obtain the following integration by parts identity for Itô integrals:

Corollary 6.12. *Suppose that $X, Y : [0, u) \to \mathbb{R}$ are continuous functions with continuous quadratic variations $[X]$ and $[Y]$, and continuous covariation $[X, Y]$. Then*

$$X_tY_t - X_0Y_0 \;=\; \int_0^t \begin{pmatrix} Y_s \\ X_s \end{pmatrix} \cdot d\left(X_sY_s\right) + [X,Y]_t \qquad \textit{for any } t \in [0,u). \tag{6.4.6}$$

If one, or, equivalently, both of the Itô integrals $\int_0^t Y_s\,dX_s$ and $\int_0^t X_s\,dY_s$ exist then (6.4.6) yields

$$X_tY_t - X_0Y_0 \;=\; \int_0^t Y_s\,dX_s + \int_0^t X_s\,dY_s + [X,Y]_t. \tag{6.4.7}$$

Proof. The identity (6.4.6) follows by applying Itô's formula in $\mathbb{R}^2$ to the process (X_t, Y_t) and the function $F(x, y) = xy$. If one of the integrals $\int_0^t Y\,dX$ or $\int_0^t X\,dY$ exists, then the other exists as well, and

$$\int_0^t \begin{pmatrix} Y_s \\ X_s \end{pmatrix} \cdot d\begin{pmatrix} X_s \\ Y_s \end{pmatrix} \quad = \quad \int_0^t Y_s\,dX_s + \int_0^t X_s\,dY_s.$$

□

As it stands, (6.4.7) is an integration by parts formula for Itô integrals which involves the correction term $[X, Y]_t$. In differential notation, it is a product rule for Itô differentials:

$$d(XY) \;=\; X\,dY + Y\,dX + [X,Y].$$

Again, in Stratonovich calculus a corresponding product rule holds without the correction term $[X, Y]$:

$$\circ d(XY) \;=\; X\,\circ dY + Y\,\circ dX.$$

Remark / Warning (Existence of $\int$ X dY, Lévy area). Under the conditions of the theorem, the Itô integrals $\int_0^t X\,dY$ and $\int_0^t Y\,dX$ do not necessarily exist! The following statements are equivalent:

(1). The Itô integral $\int_0^t Y_s\, dX_s$ exists (along (π_n)).

(2). The Itô integral $\int_0^t X_s\, dY_s$ exists.

(3). The **Lévy area** $A_t(X,Y)$ defined by

$$A_t(X,Y) \;=\; \int_0^t (Y\, dX - X\, dY) \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} (Y_s \Delta X_s - X_s \Delta Y_s)$$

exists.

Hence, if the Lévy area $A_t(X,Y)$ is given, the stochastic integrals $\int X\, dY$ and $\int Y\, dX$ can be constructed pathwise. Pushing these ideas further leads to the rough paths theory developed by T. Lyons and others, cf. [Lyons, St. Flour], [Friz: Rough paths theory].

Example (Integrating bounded variation processes w.r.t. Brownian motion). If (H_t) is an adapted process with continuous sample paths of bounded variation and (B_t) is a one-dimensional Brownian motion then $[H,B]=0$, and hence

$$H_t B_t - H_0 B_0 \;=\; \int_0^t H_s\, dB_s + \int_0^t B_s\, dH_s.$$

This integration by parts identity can be used as an alternative definition of the stochastic integral $\int_0^t H\, dB$ for integrands of bounded variation, which can then again be extended to general integrands in $\mathcal{L}^2_a(0,t)$ by the Itô isometry.

Time-dependent Itô formula

The multi-dimensional Itô formula can be applied to functions that depend explicitly on the time variable t or on the quadratic variation $[X]_t$. For this purpose we simply add t or $[X]_t$ respectively as an additional component to the function, i.e., we apply the multi-dimensional Itô formula to $Y_t = (t, X_t)$ or $Y_t = (t, [X]_t)$ respectively.

Theorem 6.13. *Suppose that $X : [0,u) \to \mathbb{R}^d$ is a continuous function with continuous covariations $[X^{(i)}, X^{(j)}]_t$, along (π_n), and let $F \in C^2(A([0,u))\times\mathbb{R}^d)$. If $A : [0,u) \to \mathbb{R}$ is a continuous function of bounded variation then the integral*

$$\int_0^t \nabla_x F(A_s, X_s) \cdot dX_s \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} \nabla_x F(A_s, X_s) \cdot (X_{s'\wedge t} - X_s)$$

exists, and the Itô formula

$$F(A_t, X_t) \;=\; F(0, X_0) + \int_0^t \nabla_x F(A_s, X_s) \cdot dX_s + \int_0^t \frac{\partial F}{\partial a}(A_s, X_s)\, dA_s + \tag{6.4.8}$$

$$\frac{1}{2}\sum_{i,j=1}^d \int_0^t \frac{\partial^2 F}{\partial x_i \partial x_j}(A_s, X_s)\, d[X^{(i)}, X^{(j)}]_s \tag{6.4.9}$$

holds for any $t \geq 0$. Here $\partial F/\partial a$ denotes the derivative of $F(a,x)$ w.r.t. the first component, and $\nabla_x F$ and $\partial^2 F/\partial x_i \partial x_j$ are the gradient and the second partial derivatives w.r.t. the other components. The most important application of the theorem is for $A_t = t$. Here we obtain the time-dependent Itô formula

$$dF(t, X_t) \;=\; \nabla_x F(t, X_t) \cdot dX_t + \frac{\partial F}{\partial t}(t, X_t)\, dt + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 F}{\partial x_i \partial x_j}(t, X_t)\, d[X^{(i)}, X^{(j)}]_t. \tag{6.4.10}$$

Similarly, if $d = 1$ and $A_t = [X]_t$ then we obtain

$$dF([X]_t, X_t) \;=\; \frac{\partial F}{\partial t}([X]_t, X_t)\, dt + \left(\frac{\partial F}{\partial a} + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}\right)([X]_t, X_t)\, d[X]_t. \tag{6.4.11}$$

If $(X)_t$ is a Brownian motion and $d = 1$ then both formulas coincide.

Proof. Let $Y_t = (Y_t^{(0)}, Y_t^{(1)}, \ldots, Y_t^{(d)}) := (A_t, X_t)$. Then $[Y^{(0)}, Y^{(i)}]_t = 0$ for any $t \geq 0$ and $0 \leq i \leq d$ because $Y_t^{(0)} = A_t$ has bounded variation. Therefore, by Itô's formula in $\mathbb{R}^{d+1}$,

$$F(A_t, X_t) \;=\; F(A_0, X_0) + I_t + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 F}{\partial x_i \partial x_j}(A_s, X_s)\, d[X^{(i)}, X^{(j)}]_s$$

where

$$\begin{aligned} I_t &= \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} \nabla^{\mathbb{R}^{d+1}} F(A_s, X_s) \cdot \begin{pmatrix} A_{s'\wedge t} - A_s \\ X_{s'\wedge t} - X_s \end{pmatrix} \\ &= \lim_{n\to\infty} \left(\sum \frac{\partial F}{\partial a}(A_s, X_s)(A_{s'\wedge t} - A_s) + \sum \nabla_x F(A_s, X_s) \cdot (X_{s'\wedge t} - X_s) \right). \end{aligned}$$

The first sum on the right hand side converges to the Stieltjes integral $\int_0^t \frac{\partial F}{\partial a}(A_s, X_s)\, dA_s$ as $n \to \infty$. Hence, the second sum also converges, and we obtain (6.4.7) in the limit as $n \to \infty$. □

Note that if $h(t, x)$ is a solution of the dual heat equation

$$\frac{\partial h}{\partial t} + \frac{1}{2}\frac{\partial^2 h}{\partial x^2} = 0 \qquad \text{for } t \geq 0, x \in \mathbb{R}, \tag{6.4.12}$$

then by (6.4.11),

$$h([X]_t, X_t) = h(0, X_0) + \int_0^t \frac{\partial h}{\partial x}([X]_s, X_s)\, dX_s.$$

In particular, if (X_t) is a Brownian motion, or more generally a local martingale, then $h([X]_t, X_t)$ is also a local martingale. The next example considers two situations where this is particular interesting:

Example. (1). *Itô exponentials:* For any $\alpha \in \mathbb{R}$, the function

$$h(t, x) = \exp(\alpha x - \alpha^2 t/2)$$

satisfies (6.4.12) and $\partial h/\partial x = \alpha h$. Hence the function

$$Z_t^{(\alpha)} := \exp\left(\alpha X_t - \frac{1}{2}\alpha^2 [X]_t\right)$$

is a solution of the Itô differential equation

$$dZ_t^{(\alpha)} = \alpha Z_t^{(\alpha)}\, dX_t$$

with initial condition $Z_0^{(\alpha)} = 1$. This shows that in Itô calculus, the functions $Z_t^{(\alpha)}$ are the correct replacements for the exponential functions. The additional factor

$\exp(-\alpha^2[X]_t/2)$ should be thought of as an appropriate renormalization in the continuous time limit.

For a Brownian motion (X_t), we obtain the exponential martingales as generalized exponentials.

(2). *Hermite polynomials:* For $n = 0, 1, 2, \ldots$, the Hermite polynomials

$$h_n(t, x) \;=\; \frac{\partial^n}{\partial \alpha^n} \exp(\alpha x - \frac{1}{2}\alpha^2 t)|_{\alpha=0}$$

also satisfy (6.4.12). The first Hermite polynomials are $1, x, x^2 - t, x^3 - 3tx, \ldots$. Note also that

$$\exp(\alpha x - \alpha^2 t/2) \;=\; \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} h_n(t, x)$$

by Taylor's theorem. Moreover, the following properties can be easily verified:

$$h_n(1, x) \;=\; e^{x^2/2}(-1)^n \frac{d^n}{dx^n} e^{-x^2/2} \qquad \text{for any } x \in \mathbb{R}, \tag{6.4.13}$$

$$h_n(t, x) \;=\; t^{n/2} h_n(1, x/\sqrt{t}) \qquad \text{for any } t \geq 0, x \in \mathbb{R}, \tag{6.4.14}$$

$$\frac{\partial h_n}{\partial x} \;=\; n h_{n-1}, \qquad \frac{\partial h_n}{\partial t} + \frac{1}{2}\frac{\partial^2 h_n}{\partial x^2} \;=\; 0. \tag{6.4.15}$$

For example, (6.4.13) holds since

$$\exp(\alpha x - \alpha^2/2) \;=\; \exp(-(x-a)^2/2)\exp(x^2/2)$$

yields

$$h_N(1, x) \;=\; \exp(x^2/2)(-1)^n \, \frac{d^n}{d\beta^n} \exp(-\beta^2/2)\bigg|_{\beta=x},$$

and (6.4.14) follows from

$$\begin{aligned} \exp(\alpha x - \alpha^2 t/2) \;&=\; \exp(\alpha\sqrt{t} \cdot (x/\sqrt{t}) - (\alpha\sqrt{t})^2/2) \\ &=\; \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} t^{n/2} h_n(1, x/\sqrt{t}). \end{aligned}$$

By (6.4.13) and (6.4.14), h_n is a polynomial of degree n. For any $n \geq 0$, the function

$$H_t^{(n)} \;:=\; h_n([X]_t, X_t)$$

is a solution of the Itô equation

$$dH_t^{(n)} \;=\; nH_t^{(n-1)}\, dX_t. \tag{6.4.16}$$

Therefore, the Hermite polynomials are appropriate replacements for the ordinary monomials x^n in Itô calculus. If $X_0 = 0$ then $H_0^{(n)} = 0$ for $n \geq 1$, and we obtain inductively

$$H_t^{(0)} = 1, \quad H_t^{(1)} \;=\; \int_0^t dX_s, \quad H_t^{(2)} \;=\; \int H_s^{(1)}\, dX_s \;=\; \int_0^t \int_0^s dX_r\, dX_s,$$

and so on.

Corollary 6.14 (Itô 1951). *If $X : [0, u) \to \mathbb{R}$ is continuous with continuous variation then for $t \in [0, u)$,*

$$\int_0^t \int_0^{s_n} \cdots \int_0^{s_2} dX_{s_1} \cdots\, dX_{s_{n-1}}\, dX_{s_n} \;=\; \frac{1}{n!} h_n([X]_t, X_t).$$

Proof. The equation follows from (6.4.16) by induction on n. □

Iterated Itô integrals occur naturally in Taylor expansions of Itô calculus. Therefore, the explicit expression from the corollary is valuable for numerical methods for stochastic differential equations, cf. Section **??** below.

Chapter 7

Brownian Motion and Partial Differential Equations

The stationary and time-dependent Itô formula enable us to work out the connection of Brownian motion to several partial differential equations involving the Laplace operator in detail. One of the many consequences is the evaluation of probabilities and expectation values for Brownian motion by p.d.e. methods. More generally, Itô's formula establishes a link between stochastic processes and analysis that is extremely fruitful in both directions.

Suppose that (B_t) is a d-dimensional Brownian motion defined on a probability space $(\Omega, \mathcal{A}, P)$ such that *every* sample path $t \mapsto B_t(\omega)$ is continuous. We first note that Itô's formula shows that Brownian motion solves the martingale problem for the operator $\mathscr{L} = \frac{1}{2}\Delta$ in the following sense:

Corollary 7.1 (Time-dependent martingale problem). *The process*

$$M_t^F \;=\; F(t, B_t) - F(0, B_0) - \int_0^t \left(\frac{\partial F}{\partial s} + \frac{1}{2}\Delta F \right)(s, B_s)\, ds$$

is a continuous $(\mathcal{F}_t^B)$ martingale for any C^2 function $F : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ with bounded first derivatives. Moreover, M^F is a continuous local martingale up to $T_{D^C} = \inf\{t \geq 0 \,:\, B_t \notin D\}$ for any $F \in C^2([0,\infty) \times D), D \subseteq \mathbb{R}^d$ open.

Proof. By the continuity assumptions one easily verifies that M^F is $(\mathcal{F}_t^B)$ adapted. Moreover, by the time-dependent Itô formula (6.4.10),

$$M_t^F \;=\; \int_0^t \nabla_x F(s, B_s) \cdot dB_s \qquad \text{for } t < T_{D^C},$$

which implies the claim. □

Choosing a function F that does not explicitly depend on t, we obtain in particular that

$$M_t^F \;=\; F(B_t) - F(B_0) - \int_0^t \frac{1}{2}\Delta F(B_s)\, ds$$

is a martingale for any $f \in C_b^2(\mathbb{R}^d)$, and a local martingale up to T_{D^C} for any $F \in C^2(D)$.

7.1 Recurrence and transience of Brownian motion in $\mathbb{R}^{\mathbf{d}}$

As a first consequence of Corollary 7.1 we can now complete the proof of the stochastic representation for solutions of the Dirichlet problem, cf. Theorem 3.6 above. By solving the Dirichlet problem for balls explicitly, we will then study recurrence, transience and polar sets for multi-dimensional Brownian motion.

The Dirichlet problem revisited

Suppose that $h \in C^2(D) \cap C(\overline{D})$ is a solution of the Dirichlet problem

$$\Delta h \;=\; 0 \quad \text{on } D, \quad h = f \quad \text{on } \partial D,$$

for a bounded open set $D \subset \mathbb{R}^d$ and a continuous function $f : \partial D \to \mathbb{R}$. If (B_t) is under P_x a continuous Brownian motion with $B_0 = x$ P_x-almost surely, then by Corollary 7.1, the process $h(B_t)$ is a local $(\mathcal{F}_t^B)$ martingale up to T_{D^C}. By applying the optional stopping theorem with a localizing sequence of bounded stopping times $S_n \nearrow T_{D^C}$, we obtain

$$h(x) \;=\; E_x[h(B_0)] \;=\; E_x[h(B_{S_n})] \qquad \text{for any } n \in \mathbb{N}.$$

Since $P_x[T_{D^C} < \infty] = 1$ and h is bounded on $\overline{D}$, dominated convergence then yields the stochastic representation

$$h(x) \;=\; E_x[h(B_{T_{D^C}})] \;=\; E_x[f(B_{T_{D^C}})] \qquad \text{for any } x \in \mathbb{R}^d.$$

We will generalize this result substantially in Theorem **??** below. Before, we apply the Dirichlet problem to study recurrence and transience of Brownian motions:

Recurrence and transience of Brownian motion in $\mathbb{R}^d$

Let (B_t) be a d-dimensional Brownian motion on $(\Omega, \mathcal{A}, P)$ with initial value $B_0 = x_0, x_0 \neq 0$. For $r \geq 0$ let

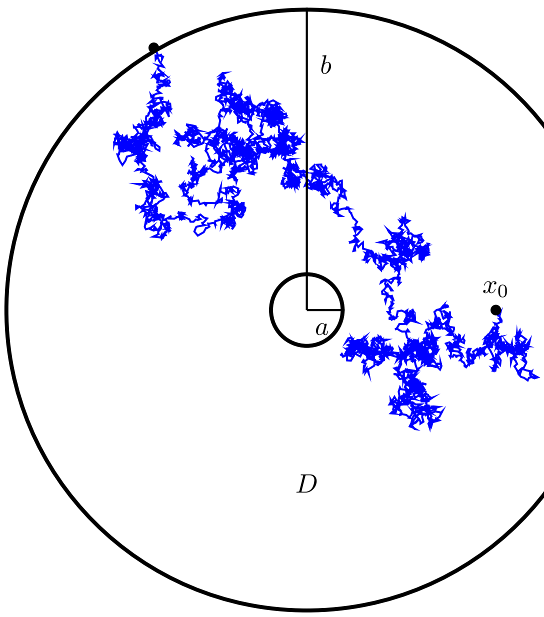
$$T_r \;=\; \inf\{t > 0 \;:\; |B_t| = r\}.$$

We now compute the probabilities $P[T_a < T_b]$ for $a < |x_0| < b$. Note that this is a multi-dimensional analogue of the *classical ruin problem*. To compute the probability for given a, b we consider the domain

$$D \;=\; \{x \in \mathbb{R}^d \;:\; a < |x| < b\}.$$

For $b < \infty$, the first exit time T_{D^C} is almost surely finite,

$$T_{D^C} \;=\; \min(T_a, T_b), \qquad \text{and} \qquad P[T_a < T_b] \;=\; P[|B_{T_{D^C}}| = a].$$



Suppose that $h \in C(\overline{U}) \cap C^2(U)$ is a solution of the Dirichlet problem

$$\Delta h(x) \;=\; 0 \qquad \text{for all } x \in D, \qquad h(x) \;=\; \begin{cases} 1 & \text{if } |x| = a, \\ 0 & \text{if } |x| = b. \end{cases} \tag{7.1.1}$$

Then $h(B_t)$ is a bounded local martingale up to T_{D^C} and optional stopping yields

$$P[T_a < T_b] \;=\; E[h(B_{T_{D^C}})] \;=\; h(x_0). \tag{7.1.2}$$

By rotational symmetry, the solution of the Dirichlet problem (7.1.1) can be computed explicitly. The Ansatz $h(x) = f(|x|)$ leads us to the boundary value problem

$$\frac{d^2 f}{dr^2}(|x|) + \frac{d-1}{|x|}\frac{df}{dr}(|x|) \;=\; 0, \qquad f(a) = 1, f(b) = 0,$$

for a second order ordinary differential equation. Solutions of the o.d.e. are linear combinations of the constant function 1 and the function

$$\phi(s) \;:=\; \begin{cases} s & \text{for } d = 1, \\ \log s & \text{for } d = 2, \\ s^{2-d} & \text{for } d \geq 3. \end{cases}$$

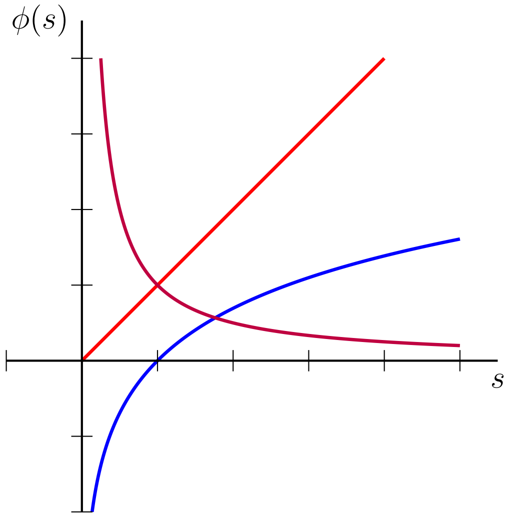


Figure 7.1: The function $\phi(s)$ for different values of d: red ($d = 1$), blue ($d = 2$) and purple ($d = 3$)

Hence, the unique solution f with boundary conditions $f(a) = 1$ and $f(b) = 0$ is

$$f(r) = \frac{\phi(b) - \phi(r)}{\phi(b) - \phi(a)}.$$

Summarizing, we have shown:

Theorem 7.2 (Ruin problem in $\mathbb{R}^d$). *For $a, b > 0$ with $a < |x_0| < b$,*

$$P[T_a < T_b] = \frac{\phi(b) - \phi(|x_0|)}{\phi(b) - \phi(a)}, \qquad \textit{and}$$

$$P[T_b < \infty] = \begin{cases} 1 & \textit{for } d \leq 2 \\ (a/|x_0|)^{d-2} & \textit{for } d > 2. \end{cases}$$

Proof. The first equation follows by 6.4.12. Moreover,

$$P[T_a < \infty] = \lim_{b \to \infty} P[T_a < T_b] = \begin{cases} 1 & \text{for } d \leq 2 \\ \phi(|x_0|)/\phi(a) & \text{for } d \geq 3. \end{cases}$$

□

Corollary 7.3. *For a Brownian motion in $\mathbb{R}^d$ the following statements hold for any initial value $x_0 \in \mathbb{R}^d$:*

(1). If $d \leq 2$ then every non-empty ball $D \subseteq \mathbb{R}^d$ is ***recurrent****, i.e., the last visit time of D is almost surely infinite:*

$$L_d = \sup\{t \geq 0 : B_t \in D\} = \infty \qquad P\text{-a.s.}$$

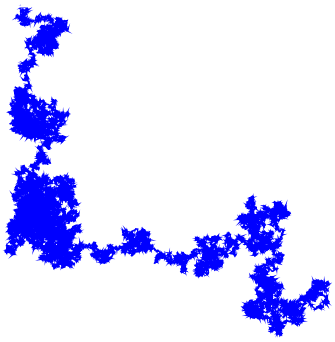
(2). If $d \geq 3$ then every ball D is ***transient****, i.e.,*

$$L_d < \infty \qquad P\text{-a.s.}$$

(3). If $d \geq 2$ then every point $x \in \mathbb{R}^d$ is ***polar****, i.e.,*

$$P[\,\exists\, t > 0 : B_t = x] = 0.$$

We point out that the last statement holds even if the starting point x_0 coincides with x. the first statement implies that a typical Brownian sample path is dense in $\mathbb{R}^2$, whereas by the second statement, $\lim_{t\to\infty} |B_t| = \infty$ almost surely for $d \geq 3$.



Proof.

(1),(2) The first two statements follow from Theorem 7.2 and the Markov property.

(3). For the third statement we assume w.l.o.g. $x = 0$. If $x_0 \neq 0$ then

$$P[T_0 < \infty] = \lim_{b\to\infty} P[T_0 < T_b]$$

for any $a > 0$. By Theorem 7.2,

$$P[T_0 < T_b] \leq \inf_{a>0} P[T_a < T_b] = 0 \qquad \text{for } d \geq 2,$$

whence $T_0 = \infty$ almost surely. If $x_0 = 0$ then by the Markov property,

$$P[\,\exists\, t > \varepsilon \,:\, B_t = 0] \;=\; E[P_{B_\varepsilon}[T_0 < \infty]] \;=\; 0$$

for any $\varepsilon > 0$. thus we again obtain

$$P[T_0 < \infty] \;=\; \lim_{\varepsilon \searrow 0} P[\,\exists\, t > \varepsilon \,:\, B_t = 0] \;=\; 0.$$

□

Remark (Polarity of linear subspaces). For $d \geq 2$, any $(d-2)$ dimensional subspace $V \subseteq \mathbb{R}^d$ is polar for Brownian motion. For the proof note that the orthogonal projection of a one-dimensional Brownian motion onto the orthogonal complement $V^\perp$ is a 2-dimensional Brownian motion.

7.2 Boundary value problems, exit and occupation times

The connection of Brownian motion to boundary value problems for partial differential equations involving the Laplace operator can be extended substantially:

The stationary Feynman-Kac-Poisson formula

Suppose that $f : \partial D \to \mathbb{R}, V : D \to \mathbb{R}$ and $g : D \to [0, \infty)$ are continuous functions defined on an open bounded domain $D \subset \mathbb{R}^d$, or on its boundary respectively. We assume that under P_x, (B_t) is Brownian motion with $P_x[B_0 = x] = 1$, and that

$$E_x\left[\exp \int_0^T V^-(B_s)\, ds\right] \;<\; \infty \qquad \text{for any } x \in D, \tag{7.2.1}$$

where $T = T_{D^C}$ is the first exit time from D.
Note that (7.2.1) always holds if V is non-negative.

Theorem 7.4. *If $u \in C^2(D) \cap C(\overline{D})$ is a solution of the boundary problem*

$$\frac{1}{2}\Delta u(x) \;=\; V(x)u(x) - g(x) \qquad \text{for } x \in D \tag{7.2.2}$$

$$u(x) \;=\; f(x) \qquad \text{for } x \in \partial D, \tag{7.2.3}$$

and (7.2.1) holds then

$$u(x) \;=\; E_x\left[\exp\left(-\int_0^T V(B_s)\,ds\right)\cdot f(B_T)\right] + E_x\left[\int_0^T \exp\left(-\int_0^t V(B_s)\,ds\right)\cdot g(B_t)\,dt\right] \tag{7.2.4}$$

for any $x \in D$.

Remark. Note that we *assume* the existence of a smooth solution of the boundary value problem (7.2.2). Proving that the function u defined by (7.2.4) is a solution of the b.v.p. without assuming existence is much more demanding.

Proof. By continuity of V and (B_s), the sample paths of the process

$$A_t \;=\; \int_0^t V(B_s)\,ds$$

are C^1 and hence of bounded variation for $t < T$. Let

$$X_t \;=\; e^{-A_t}u(B_t), \qquad t < T.$$

Applying Itô's formula with $F(a,b) = e^{-a}u(b)$ yields the decomposition

$$\begin{aligned} dX_t \;&=\; e^{-A_t}\nabla u(B_t)\cdot dB_t - e^{-A_t}u(B_t)\,dA_t + \frac{1}{2}e^{-A_t}\Delta u(B_t)\,dt \\ &=\; e^{-A_t}\nabla u(B_t)\cdot dB_t + e^{-A_t}\left(\frac{1}{2}\Delta u - V\cdot u\right)(B_t)\,dt \end{aligned}$$

of X_t into a local martingale up to time T and an absolutely continuous part. Since u is a solution of (7.2.2), we have $\frac{1}{2}\Delta u - Vu = -g$ on D. By applying the optional stopping theorem with a localizing sequence $T_n \nearrow T$ of stopping times, we obtain the representation

$$\begin{aligned} u(x) \;=\; E_x[X_0] \;&=\; E_x[X_{T_n}] + E_x\left[\int_0^{T_n} e^{-A_t}g(B_t)\,dt\right] \\ &=\; E_x[e^{-A_{T_n}}u(B_{T_n})] + E_x\left[\int_0^{T_n} e^{-A_t}g(B_t)\,dt\right] \end{aligned}$$

for $x \in D$. The assertion (7.2.4) now follows provided we can interchange the limit as $n \to \infty$ and the expectation values. For the second expectation on the right hand side this is possible by the monotone convergence theorem, because $g \geq 0$. For the first expectation value, we can apply the dominated convergence theorem, because

$$\left|e^{-A_{T_n}} u(B_{T_n})\right| \;\leq\; \exp\left(\int_0^T V^-(B_s)\, ds\right) \cdot \sup_{y \in \overline{D}} |u(y)| \qquad \forall n \in \mathbb{N},$$

and the majorant is integrable w.r.t. each P_x by Assumption 7.2.1. □

Remark (Extension to diffusion processes). A corresponding result holds under appropriate assumptions if the Brownian motion (B_t) is replaced by a diffusion process (X_t) solving a stochastic differential equation of the type $dX_t = \sigma(X_t)\, dB_t + b(X_t)\, dt$, and the operator $\frac{1}{2}\Delta$ in (7.2.2) is replaced by the generator

$$\mathscr{L} \;=\; \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j} + b(x)\cdot\nabla, \quad a(x) = \sigma(x)\sigma(x)^\top,$$

of the diffusion process, cf. **??**. The theorem hence establishes a general connection between Itô diffusions and boundary value problems for linear second order elliptic partial differential equations.

By Theorem 7.4 we can compute many interesting expectation values for Brownian motion by solving appropriate p.d.e. We now consider various corresponding applications.

Let us first recall the Dirichlet problem where $V \equiv 0$ and $g \equiv 0$. In this case, $u(x) = E_x[f(B_t)]$. We have already pointed out in the last section that this can be used to compute exit distributions and to study recurrence, transience and polarity of linear subspaces for Brownian motion in $\mathbb{R}^d$. A second interesting case of Theorem 7.4 is the stochastic representation for solutions of the Poisson equation:

Poisson problem and mean exit time

If V and f vanish in Theorem 7.2, the boundary value problem (7.2.2) reduces to the boundary value problem

$$\frac{1}{2}\Delta u \;=\; -g \quad \text{on } D, \quad u \;=\; 0 \quad \text{on } D,$$

for the Poisson equation. The solution has the stochastic representation

$$u(x) \;=\; E_x\left[\int_0^T g(B_t)\,dt\right], \qquad x \in D, \tag{7.2.5}$$

which can be interpreted as an average cost accumulated by the Brownian path before exit from the domain D. In particular, choosing $g \equiv 1$, we can compute the mean exit time

$$u(x) \;=\; E_x[T]$$

from D for Brownian motion starting at x by solving the corresponding Poisson problem.

Example. If $D = \{x \in \mathbb{R}^d \;:\; |x| < r\}$ is a ball around 0 of radius $r > 0$, then the solution $u(x)$ of the Poisson problem

$$\frac{1}{2}\Delta u(x) \;=\; \begin{cases} -1 & \text{for } |x| < r \\ 0 & \text{for } |x| = r \end{cases}$$

can be computed explicitly. We obtain

$$E_x[T] \;=\; u(x) \;=\; \frac{r^2 - |x|^2}{d} \qquad \text{for any } x \in D.$$

Occupation time density and Green function

If (B_t) is a Brownian motion in $\mathbb{R}^d$ then the corresponding Brownian motion with absorption at the first exit time from the domain D is the Markov process (X_t) with state space $D \cup \{\Delta\}$ defined by

$$X_t \;=\; \begin{cases} B_t & \text{for } t < T \\ \Delta & \text{for } t \geq T \end{cases},$$

where Δ is an extra state added to the state space. By setting $g(\Delta) \;=\; 0$, the stochastic representation (7.2.5) for a solution of the Poisson problem can be written in the form

$$u(x) \;=\; E_x\left[\int_0^\infty g(X_t)\,dt\right] \;=\; \int_0^\infty (p_t^D g)(x)\,dt, \tag{7.2.6}$$

where

$$p_t^D(x,A) \;=\; P_x[X_t \in A], \qquad A \subseteq \mathbb{R}^d measurable,$$

is the transition function for the absorbed process (X_t). Note that for $A \subset \mathbb{R}^d$,

$$p_t^D(x,A) \;=\; P_x[B_t \in A \text{ and } t < T] \;\le\; p_t(x,A) \tag{7.2.7}$$

where p_t is the transition function of Brownian motion on $\mathbb{R}^d$. For $t > 0$ and $x \in \mathbb{R}^d$, the transition function $p_t(x, \bullet)$ of Brownian motion is absolutely continuous. Therefore, by (7.2.7), the sub-probability measure $p_t^D(x, \bullet)$ restricted to $\mathbb{R}^d$ is also absolutely continuous with non-negative density

$$p_t^D(x,y) \;\le\; p_t(x,y) \;=\; (2\pi t)^{-d/2} \exp\left(-\frac{|x-y|^2}{2t}\right).$$

The function p_t^D is called the **heat kernel on the domain** D w.r.t. absorption on the boundary. Note that

$$G^D(x,y) \;=\; \int_0^\infty p_t^D(x,y)\,dt$$

is an **occupation time density**, i.e., it measures the average time time a Brownian motion started in x spends in a small neighbourhood of y before it exits from the Domain D. By (7.2.6), a solution u of the Poisson problem $\frac{1}{2}\Delta u = -g$ on D, $u = 0$ on ∂D, can be represented as

$$u(x) \;=\; \int_D G^D(x,y) g(y)\,dy \qquad \text{for } x \in D.$$

This shows that the occupation time density $G^D(x,y)$ is the **Green function** (i.e., the fundamental solution of the Poisson equation) **for the operator $\frac{1}{2}$ with Dirichlet boundary conditions on the domain D.**

Note that although for domains with irregular boundary, the Green's function might not exist in the classical sense, the function $G^D(x,y)$ is always well-defined!

Stationary Feynman-Kac formula and exit time distributions

Next, we consider the case where g vanishes and $f \equiv 1$ in Theorem 7.4. Then the boundary value problem (7.2.4) takes the form

$$\frac{1}{2}\Delta u \;=\; Vu \quad \text{on } D, \quad u \;=\; 1 \quad \text{on } \partial D. \tag{7.2.8}$$

The p.d.e. $\frac{1}{2}\Delta u = Vu$ is a stationary Schrödinger equation. We will comment on the relation between the Feynman-Kac formula and Feynman's path integral formulation of quantum mechanics below. For the moment, we only note that for the solution of (**??**), the stochastic representation

$$u(x) \;=\; E_x\left[\exp\left(-\int_0^T V(B_t)\,dt\right)\right]$$

holds for $x \in D$.

As an application, we can, at least in principle, compute the full distribution of the exit time T. In fact, choosing $V \equiv \alpha$ for some constant $\alpha > 0$, the corresponding solution u_α of (7.2.8) yields the Laplace transform

$$u_\alpha(x) \;=\; E_x[e^{-\alpha T}] \;=\; \int_0^\infty e^{-\alpha t}\mu_x(dt) \tag{7.2.9}$$

of $\mu_x = P_x \circ T^{-1}$.

Example (Exit times in $\mathbb{R}^1$). Suppose $d = 1$ and $D = (-1, 1)$. Then (7.2.8) with $V = \alpha$ reads

$$\frac{1}{2}u_\alpha''(x) \;=\; \alpha u_\alpha(x) \quad \text{for } x \in (-1,1), \quad u_\alpha(1) \;=\; u_\alpha(-1) \;=\; 1.$$

This boundary value problem has the unique solution

$$u_\alpha(x) \;=\; \frac{\cosh(x\cdot\sqrt{2\alpha})}{\cosh(\sqrt{2\alpha})} \qquad \text{for } x \in [-1,1].$$

By inverting the Laplace transform (7.2.9), one can now compute the distribution μ_x of the first exit time T from $(-1, 1)$. It turns out that μ_x is absolutely continuous with density

$$f_T(t) = \frac{1}{\sqrt{2\pi t^3}}\sum_{n=-\infty}^{\infty}\left((4n+1+x)e^{-\frac{(4n+1+x)^2}{2t}} + (4n+1-x)e^{-\frac{(4n+1-x)^2}{2t}}\right), \quad t \geq 0.$$

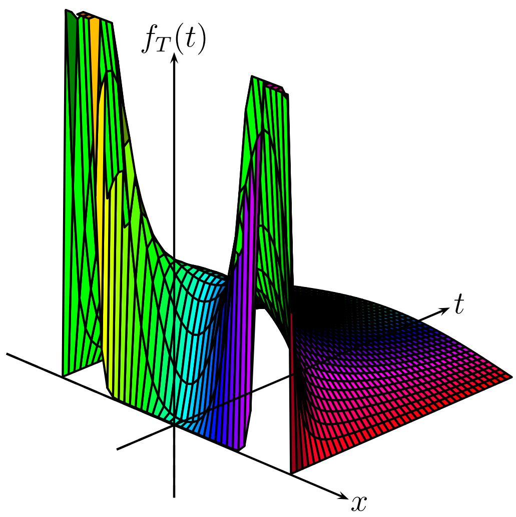


Figure 7.2: The density of the first exit time T depending on the starting point $x \in [-1, 1]$ and the time $t \in (0, 2]$.

Boundary value problems in $\mathbb{R}^d$ and total occupation time

Suppose we would like to compute the distribution of the total occupation time

$$\int_0^\infty I_A(B_s)\, ds$$

of a bounded domain $A \subset \mathbb{R}^d$ for Brownian motion. This only makes sense for $d \geq 3$, since for $d \leq 2$, the total occupation time of any non-empty open set is almost surely infinite by recurrence of Brownian motion in $\mathbb{R}^1$ and $\mathbb{R}^2$. The total occupation time is of the form $\int_0^\infty V(B_s)\, ds$ with $V = I_A$. Therefore, we should in principle be able to apply Theorem 7.2, but we have to replace the exit time T by $+\infty$ and hence the underlying bounded domain D by $\mathbb{R}^d$.

Corollary 7.5. *Suppose $d \geq 3$ and let $V : \mathbb{R}^d \to [0, \infty)$ be continuous. If $u \in C^2(\mathbb{R}^d)$ is a solution of the boundary value problem*

$$\frac{1}{2}\Delta u = Vu \quad \text{on } \mathbb{R}^d, \qquad \lim_{|x|\to\infty} u(x) = 1 \tag{7.2.10}$$

then

$$u(x) \;=\; E_x\left[\exp\left(-\int_0^\infty V(B_t)\,dt\right)\right] \qquad \textit{for any } x\in\mathbb{R}^d.$$

Proof. Applying the stationary Feynman-Kac formula on an open bounded subset $D\subset\mathbb{R}^d$, we obtain the representation

$$u(x) \;=\; E_x\left[u(B_{T_{D^C}})\exp\left(-\int_0^{T_{D^C}} V(B_t)\,dt\right)\right] \tag{7.2.11}$$

by Theorem 7.2. Now let $D_n = \{x\in\mathbb{R}^d : |x|<n\}$. Then $T_{D_n^C}\nearrow\infty$ as $n\to\infty$. Since $d\geq 3$, Brownian motion is transient, i.e., $\lim\limits_{t\to\infty}|B_t| = \infty$, and therefore by (7.2.10)

$$\lim_{n\to\infty} u(B_{T_{D_n^C}}) \;=\; 1 \qquad P_x\text{-almost surely for any } x.$$

Since u is bounded and V is non-negative, we can apply dominated convergence in (7.2.11) to conclude

$$u(x) \;=\; E_x\left[\exp\left(-\int_0^\infty V(B_t)\,dt\right)\right].$$

□

Let us now return to the computation of occupation time distributions. consider a bounded subset $A\subset\mathbb{R}^d, d\geq 3$, and let

$$v_\alpha(x) \;=\; E_x\left[\exp\left(-\alpha\int_0^\infty I_A(B_s)\,ds\right)\right], \qquad \alpha>0,$$

denote the Laplace transform of the total occupation time of A. Although $V=\alpha I_A$ is not a continuous function, a representation of v_α as a solution of a boundary problem holds:

Exercise. Prove that if $A\subset\mathbb{R}^d$ is a bounded domain with smooth boundary ∂A and $u_\alpha\in C^1(\mathbb{R}^d)\cap C^2(\mathbb{R}^d\setminus\partial A)$ satisfies

$$\frac{1}{2}\Delta u_\alpha \;=\; \alpha I_A u_\alpha \quad \text{on } \mathbb{R}^d\setminus\partial A, \qquad \lim_{|x|\to\infty} u_\alpha(x) \;=\; 1, \tag{7.2.12}$$

then $v_\alpha = u_\alpha$.

Remark. The condition $u_\alpha \in C^1(\mathbb{R}^d)$ guarantees that u_α is a weak solution of the p.d.e. (7.2.10) on all of $\mathbb{R}^d$ including the boundary ∂U.

Example (Total occupation time of the unit ball in $\mathbb{R}^3$). Suppose $A = \{x \in \mathbb{R}^3 : |x| < 1\}$. In this case the boundary value problem (7.2.10) is rotationally symmetric. The ansatz $u_\alpha(x) = f_\alpha(|x|)$, yields a Bessel equation for f_α on each of the intervals $(0,1)$ and $(1,\infty)$:

$$\frac{1}{2}f_\alpha''(r) + r^{-1}f_\alpha'(r) \;=\; \alpha f_\alpha(r) \quad \text{for } r<1, \quad \frac{1}{2}f_\alpha''(r) + r^{-1}f_\alpha(r) \;=\; 0 \quad \text{for } r>1.$$

Taking into account the boundary condition and the condition $u_\alpha \in C^1(\mathbb{R}^d)$, one obtains the rotationally symmetric solution

$$u_\alpha(x) \;=\; \begin{cases} 1 + \left(\frac{\tanh(\sqrt{2\alpha})}{\sqrt{2\alpha}} - 1\right)\cdot r^{-1} & \text{for } r \in [1,\infty), \\ \frac{\sinh(\sqrt{2\alpha}r)}{\sqrt{2\alpha}\cosh\sqrt{2\alpha}}\cdot r^{-1} & \text{for } r\in(0,1) \\ \frac{1}{\cosh(\sqrt{2\alpha})} & \text{for } r=0 \end{cases}.$$

of (7.2.10), and hence an explicit formula for v_α. In particular, for $x=0$ we obtain the simple formula

$$E_0\left[\exp\left(-\alpha\int_0^\infty I_A(B_t)\,dt\right)\right] \;=\; u_\alpha(0) \;=\; \frac{1}{\cosh(\sqrt{2\alpha})}.$$

The right hand side has already appeared in the example above as the Laplace transform of the exit time distribution of a one-dimensional Brownian motion starting at 0 from the interval $(-1,1)$. Since the distribution is uniquely determined by its Laplace transform, we have proven the remarkable fact that the total occupation time of the unit ball for a standard Brownian motion in $\mathbb{R}^3$ has the same distribution as the first exit time from the unit ball for a standard one-dimensional Brownian motion:

$$\int_0^\infty I_{\{|B_t^{\mathbb{R}^3}|<1\}}\,dt \;\sim\; \inf\{t>0 \,:\, |B_t^{\mathbb{R}^3}|>1\}.$$

This is a particular case of a theorem of Ciesielski and Taylor who proved a corresponding relation between Brownian motion in $\mathbb{R}^{d+2}$ and $\mathbb{R}^d$ for arbitrary d.

7.3 Heat Equation and Time-Dependent Feynman-Kac Formula

Itô's formula also yields a connection between Brownian motion (or, more generally, solutions of stochastic differential equations) and parabolic partial differential equations. The parabolic p.d.e. are Kolmogorov forward or backward equations for the corresponding Markov processes. In particular, the time-dependent Feynman-Kac formula shows that the backward equation for Brownian motion with absorption is a heat equation with dissipation.

Brownian Motion with Absorption

Suppose we would like to describe the evolution of a Brownian motion that is absorbed during the evolution of a Brownian motion that is absorbed during an infinitesimal time interval $[t, t+dt]$ with probability $V(t,x)dt$ where x is the current position of the process. We assume that the *absorption rate* $V(t,x)$ is given by a measurable locally-bounded function

$$V : [0,\infty) \times \mathbb{R}^d \;\to\; [0,\infty).$$

Then the accumulated absorption rate up to time t is given by the increasing process

$$A_t \;=\; \int_0^t V(s, B_s)\, ds, \qquad t \geq 0.$$

We can think of the process A_t as an internal clock for the Brownian motion determining the absorption time. More precisely, we define:

Definition. *Suppose that $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion and T is a with parameter 1 exponentially distributed random variable independent of (B_t). Let Δ be a separate state added to the state space $\mathbb{R}^d$. Then the process (X_t) defined by*

$$X_t \;:=\; \begin{cases} B_t & \text{for } A_t < T, \\ \Delta & \text{for } A_t \geq T, \end{cases}$$

is called a ***Brownian motion with absorption rate*** $V(t,x)$*, and the random variable*

$$\zeta := \inf\{t \geq 0 \,:\, A_t \geq T\}$$

is called the ***absorption time****.*

A justification for the construction is given by the following informal computation: For an infinitesimal time interval $[t, t+dt]$ and almost every ω,

$$\begin{aligned} P[\zeta \leq t+dt \mid (B_s)_{s\geq 0}, \zeta > t](\omega) &= P[A_{t+dt}(\omega) \geq T \mid A_t(\omega) < T] \\ &= P[A_{t+dt}(\omega) - A_t(\omega) \geq T] \\ &= P[V(t, B_t(\omega))dt \geq T] \\ &= V(t, B_t(\omega))dt \end{aligned}$$

by the memoryless property of the exponential distribution, i.e., $V(t,x)$ is indeed the infinitesimal absorption rate.

Rigorously, it is not difficult to verify that (X_t) is a Markov process with state space $\mathbb{R}^d \cup \{\Delta\}$ where Δ is an absorbing state. The Markov process is time-homogeneous if $V(t,x)$ is independent of t.

For a measurable subset $D \subseteq \mathbb{R}^d$ and $t \geq 0$ the distribution μ_t of X_t is given by

$$\begin{aligned} \mu_t[D] &= P[X_t \in D] = P[B_t \in D \text{ and } A_t < T] \\ &= E[P[A_t < T \mid (B_t)] \,;\, B_t \in D] \\ &= E\left[\exp\left(-\int_0^t V(s, B_s)\, ds\right) \,;\, B_t \in D\right]. \end{aligned} \tag{7.3.1}$$

Itô's formula can be used to prove a Kolmogorov type forward equation:

Theorem 7.6 (Forward equation for Brownian motion with absorption). *The sub-probability measures* μ_t *on* $\mathbb{R}^d$ *solve the heat equation*

$$\frac{\partial \mu_t}{\partial t} = \frac{1}{2}\Delta\mu_t - V(t, \bullet)\mu_t \tag{7.3.2}$$

in the following distributional sense:

$$\int f(x)\mu_t(dx) - \int f(x)\mu_0(dx) = \int_0^t \int (\frac{1}{2}\Delta f(x) - V(s,x)f(x))\mu_s(dx)\, ds$$

for any function $f \in C_0^2(\mathbb{R}^d)$.

Here $C_0^2(\mathbb{R}^d)$ denotes the space of C^2-functions with compact support. Under additional regularity assumptions it can be shown that μ_t has a smooth density that solves (7.3.1) in the classical sense. The equation (7.3.1) describes heat flow with cooling when the heat at x at time t dissipates with rate $V(t,x)$.

Proof. By (7.3.1),

$$\int f \, d\mu_t \;=\; E[\exp(-A_t)\,;\; f(B_t)] \tag{7.3.3}$$

for any bounded measurable function $f : \mathbb{R}^d \to \mathbb{R}$. For $f \in C_0^2(\mathbb{R}^d)$, an application of Itô's formula yields

$$e^{-A_t} f(B_t) \;=\; f(B_0) + M_t + \int_0^t e^{-A_s} f(B_s) V(s, B_s)\, ds + \frac{1}{2}\int_0^t e^{-A_s} \Delta f(B_s)\, ds,$$

for $t \geq 0$, where (M_t) is a local martingale. Taking expectation values for a localizing sequence of stopping times and applying the dominated convergence theorem subsequently, we obtain

$$E[e^{-A_t} f(B_t)] \;=\; E[f(B_0)] + \int_0^t E[e^{-A_s}(\frac{1}{2}\Delta f - V(s,\bullet)f)(B_s)]\, ds.$$

Here we have used that $\frac{1}{2}\Delta f(x) - V(s,x) f(x)$ is uniformly bounded for $(s,x) \in [0,t] \times \mathbb{R}^d$, because f has compact support and V is locally bounded. The assertion now follows by (7.3.3). □

Exercise (Heat kernel and Green's function). The transition kernel for Brownian motion with time-homogeneous absorption rate $V(x)$ restricted to $\mathbb{R}^d$ is given by

$$p_t^V(x, D) \;=\; E_x\left[\exp\left(-\int_0^t V(B_s)\, ds\right)\;;\; B_t \in D\right].$$

(1). Prove that for any $t > 0$ and $x \in \mathbb{R}^d$, the sub-probability measure $p_t^V(x, \bullet)$ is absolutely continuous on $\mathbb{R}^d$ with density satisfying

$$0 \;\leq\; p_t^V(x,y) \;\leq\; (2\pi t)^{-d/2} \exp(-|x-y|^2/(2t)).$$

(2). Identify the occupation time density

$$G^V(x,y) \;=\; \int_0^\infty p_t^V(x,y)\,dt$$

as a fundamental solution of an appropriate boundary value problem. Adequate regularity may be assumed.

Time-dependent Feynman-Kac formula

In Theorem 7.6 we have applied Itô's formula to prove a Kolmogorov type forward equation for Brownian motion with absorption. To obtain a corresponding backward equation, we have to reverse time:

Theorem 7.7 (Feynman-Kac). *Fix $t > 0$, and let $f : \mathbb{R}^d \to \mathbb{R}$ and $V, g : [0,t] \times \mathbb{R}^d \to \mathbb{R}$ be continuous functions. Suppose that f is bounded, g is non-negative, and V satisfies*

$$E_x\left[\exp\int_0^t V(t-s,B_s)^-\,ds\right] \;<\; \infty \qquad \text{for all } x \in \mathbb{R}^d. \tag{7.3.4}$$

If $u \in C^{1,2}((0,t] \times \mathbb{R}^d) \cap C([0,t] \times \mathbb{R}^d)$ is a bounded solution of the heat equation

$$\begin{aligned} \frac{\partial u}{\partial s}(s,x) \;&=\; \frac{1}{2}\Delta u(s,x) - V(s,x)u(s,x) + g(s,x) \qquad \text{for } s \in (0,t], x \in \mathbb{R}^d, \\ u(0,x) \;&=\; f(x), \end{aligned} \tag{7.3.5}$$

then u has the stochastic representation

$$\begin{aligned} u(t,x) \;&=\; E_x\left[f(B_t)\exp\left(-\int_0^t V(t-s,B_s)\,ds\right)\right] + \\ &\quad E_x\left[\int_0^t g(t-r,B_r)\exp\left(-\int_0^r V(t-s,B_s)\,ds\right)\,dr\right]. \end{aligned}$$

Remark. The equation (7.3.5) describes heat flow with sinks and dissipation.

Proof. We first reverse time on the interval $[0,t]$. The function

$$\hat{u}(s,x) \;=\; u(t-s,x)$$

solves the p.d.e.

$$\begin{aligned}\frac{\partial \hat{u}}{\partial s}(s,x) \;&=\; -\frac{\partial u}{\partial t}(t-s,x) \;=\; -\left(\frac{1}{2}\Delta u - Vu + g\right)(t-s,x)\\ &=\; -\left(\frac{1}{2}\Delta \hat{u} - \hat{V}\hat{u} + \hat{g}\right)(s,x)\end{aligned}$$

on $[0,t]$ with terminal condition $\hat{u}(t,x) \;=\; f(x)$. Now let $X_r = \exp(-A_r)\hat{u}(r,B_r)$ for $r \in [0,t]$, where

$$A_r \;:=\; \int_0^r \hat{V}(s,B_s)\,ds \;=\; \int_0^r V(t-s,B_s)\,ds.$$

By Itô's formula, we obtain for $\tau \in [0,t]$,

$$\begin{aligned}X_\tau - X_0 \;&=\; M_\tau - \int_0^\tau e^{-A_r}\hat{u}(r,B_r)\,dA_r + \int_0^\tau e^{-A_r}\left(\frac{\partial \hat{u}}{\partial s} + \frac{1}{2}\Delta\hat{u}\right)(r,B_r)\,dr\\ &=\; M_\tau + \int_0^\tau e^{-A_r}\left(\frac{\partial \hat{u}}{\partial s} + \frac{1}{2}\Delta\hat{u} - \hat{V}\hat{u}\right)(r,B_r)\,dr\\ &=\; M_\tau - \int_0^\tau e^{-A_r}\hat{g}(r,B_r)\,dr\end{aligned}$$

with a local martingale $(M_\tau)_{\tau\in[0,t]}$ vanishing at 0. Choosing a corresponding localizing sequence of stopping times T_n with $T_n \nearrow t$, we obtain by the optional stopping theorem and dominated convergence

$$\begin{aligned}u(t,x) \;&=\; \hat{u}(0,x) \;=\; E_x[X_0]\\ &=\; E_x[X_t] + E_x\left[\int_0^t e^{-A_r}\hat{g}(r,B_r)\,dr\right]\\ &=\; E_x[e^{-A_t}u(0,B_t)] + E_x\left[\int_0^t e^{-A_r}g(t-r,B_r)\,dr\right].\end{aligned}$$

□

Remark (Extension to diffusion processes). Again a similar result holds under a appropriate regularity assumptions for Brownian motion replaced by a solution of a s.d.e. $dX_t = \sigma(X_t)dB_t + b(X_t)dt$ and $\frac{1}{2}\Delta$ replaced by the corresponding generator, cf. **??**.

Occupation times and arc-sine law

The Feynman-Kac formula can be used to study the distribution of occupation times of Brownian motion. We consider an example where the distribution can be computed explicitly: The proportion of time during the interval $[0, t]$ spent by a one-dimensional standard Brownian motion (B_t) in the interval $(0, \infty)$. Let

$$A_t \;=\; \lambda(\{s \in [0, t] \;:\; B_s > 0\}) \;=\; \int_0^t I_{(0,\infty)}(B_s)\, ds.$$

Theorem 7.8 (Arc-sine law of P.Lévy). *For any* $t > 0$ *and* $\theta \in [0, 1]$,

$$P_0[A_t/t \le \theta] \;=\; \frac{2}{\pi} \arcsin \sqrt{\theta} \;=\; \frac{1}{\pi} \int_0^\theta \frac{ds}{\sqrt{s(1-s)}}.$$

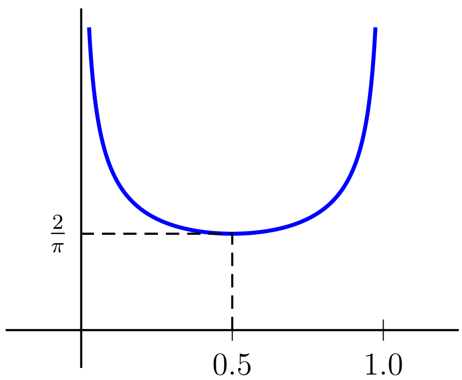


Figure 7.3: Density of A_t/t.

Note that the theorem shows in particular that a law of large numbers does **not** hold! Indeed, for each $\varepsilon > 0$,

$$P_0\left[\left|\frac{1}{t}\int_0^t I_{(0,\infty)}(B_s)\, ds - \frac{1}{2}\right| > \varepsilon\right] \quad \not\to \quad 0 \qquad \text{as } t \to \infty.$$

Even for large times, values of A_t/t close to 0 or 1 are the most probable. By the functional central limit theorem, the proportion of time that one player is ahead in a long coin tossing game or a counting of election results is also close to the arcsine law. In particular, it is more then 20 times more likely that one player is ahead for more than 98% of the time than it is that each player is ahead between 49% and 51% of the time [Steele].

Before proving the arc-sine law, we give an informal derivation based on the time-dependent Feynman-Kac formula.
The idea for determining the distribution of A_t is again to consider the Laplace transforms

$$u(t,x) \;=\; E_x[\exp(-\beta A_t)], \qquad \beta > 0.$$

By the Feynman-Kac formula, we could expect that u solves the equation

$$\frac{\partial u}{\partial t} \;=\; \frac{1}{2}\frac{\partial^2 u}{\partial x^2} \tag{7.3.6}$$

with initial condition $u(0,x) = 1$. To solve the parabolic p.d.e. (7.3.6), we consider another Laplace transform: The Laplace transform

$$v_\alpha(x) \;=\; \int_0^\infty e^{-\alpha t} u(t,x)\,dt \;=\; E_x\left[\int_0^\infty e^{-\alpha t - \beta A_t}\,dt\right], \qquad \alpha > 0,$$

of a solution $u(t,x)$ of (7.3.6) w.r.t. t. An informal computation shows that v_α should satisfy the o.d.e.

$$\begin{aligned}
\frac{1}{2}v_\alpha'' - \beta I_{(0,\infty)} v_\alpha \;&=\; \int_0^\infty e^{-\alpha t}\left(\frac{1}{2}\frac{\partial^2 u}{\partial x^2} - \beta I_{(0,\infty)} u\right)(t,\bullet)\,dt \\
&=\; \int_0^\infty e^{-\alpha t}\frac{\partial u}{\partial t}(t,\bullet)\,dt \;=\; e^{-\alpha t}u(t,\bullet)|_0^\infty - \alpha\int_0^\infty e^{-\alpha t}u(t,\bullet)\,dt \\
&=\; 1 - \alpha v_\alpha,
\end{aligned}$$

i.e., v_α should be a bounded solution of

$$\alpha v_\alpha - \frac{1}{2}v_\alpha'' + \beta I_{0,\infty} v_\alpha \;=\; g \tag{7.3.7}$$

where $g(x) = 1$ for all x. The solution of (7.3.7) can then be computed explicitly, and yield the arc-sine law by Laplace inversion.

Remark. The method of transforming a parabolic p.d.e. by the Laplace transform into an elliptic equation is standard and used frequently. In particular, the Laplace transform of a transition semigroup $(p_t)_{t\geq 0}$ is the corresponding resolvent $(g_\alpha)_{\alpha\geq 0}, g_\alpha = \int_0^\infty e^{-\alpha t} p_t \, dt$, which is crucial for potential theory.

Instead of trying to make the informal argument above rigorous, one can directly prove the arc-sine law by applying the stationary Feynman-Kac formula:

Exercise. Prove Lévy's arc-sine law by proceeding in the following way:

(1). Let $g \in C_b(\mathbb{R})$. Show that if v_α is a bounded solution of (7.3.7) on $\mathbb{R} \setminus \{0\}$ with $v_\alpha \in C^1(\mathbb{R}) \cap C^2(\mathbb{R} \setminus \{0\})$ then

$$v_\alpha(x) \;=\; E_x\left[\int_0^\infty g(B_t) e^{-\alpha t - \beta A_t} \, dt\right] \qquad \text{for any } x \in \mathbb{R}.$$

(2). Compute a corresponding solution v_α for $g \equiv 1$, and conclude that

$$\int_0^\infty e^{-\alpha t} E_0[e^{-\beta A_t}] \, dt \;=\; \frac{1}{\sqrt{\alpha(\alpha+\beta)}}.$$

(3). Now use the uniqueness of the Laplace inversion to show that the distribution μ_t of A_t/t under $P_\bullet$ is absolutely continuous with density

$$f_{A_t/t}(s) \;=\; \frac{1}{\pi\sqrt{s\cdot(1-s)}}.$$

Chapter 8

Stochastic Differential Equations: Explicit Computations

Suppose that $(B_t)_{t\geq 0}$ is a given Brownian motion defined on a probability space $(\Omega, \mathcal{A}, P)$. We will now study solutions of stochastic differential equations (SDE) of type

$$dX_t \;=\; b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t \tag{8.0.1}$$

where b and σ are continuous functions defined on $\mathbb{R}_+ \times \mathbb{R}^d$ or an appropriate subset.

Recall that $\mathcal{F}_t^{B,P}$ denotes the completion of the filtration $\mathcal{F}_t^B = \sigma(B_s \mid 0 \leq s \leq t)$ generated by the Brownian motion. Let T be an $(\mathcal{F}_t^{B,P})$ stopping time. We call a process $(t, \omega) \mapsto X_t(\omega)$ defined for $t < T(\omega)$ **adapted w.r.t.** $\mathbf{\mathcal{F}_t^{B,P}}$, if the trivially extended process $\widetilde{X}_t = X_t \cdot I_{\{t<T\}}$ defined by

$$\widetilde{X}_t \;:=\; \begin{cases} X_t & \text{for } t < T \\ 0 & \text{for } t \geq T \end{cases},$$

is $(\mathcal{F}_t^{B,P})$-adapted.

Definition. *An almost surely continuous stochastic process* $(t, \omega) \mapsto X_t(\omega)$ *defined for* $t \in [0, T(\omega))$ *is called a* ***strong solution*** *of the stochastic differential equation (8.0.1) if it is* $(\mathcal{F}_t^{B,P})$*-adapted, and the equation*

$$X_t \;=\; X_0 + \int_0^t b(s, X_s)\,ds + \int_0^t \sigma(s, X_s)\,dB_s \quad \textit{for } t \in [0, T) \tag{8.0.2}$$

holds P-almost surely.

The terminology "strong" solution will be explained later when we introduce "weak" solutions. The point is that a strong solution is adapted w.r.t. the filtration $(\mathcal{F}_t^{B,P})$ generated by the Brownian motion. Therefore, a strong solution is essentially (up to modification on measure zero sets) a *measurable function of the given Brownian motion*! The concept of strong and weak solutions of SDE is not related to the analytic definition of strong and weak solutions for partial differential equations.

In this section we study properties of solutions and explicit solutions for one-dimensional SDE. We start with an example:

Example (Asset price model in continuous time). A nearby model for an asset price process $(S_n)_{n=0,1,2,...}$ in discrete time is to define S_n recursively by

$$S_{n+1} - S_n \;=\; \alpha_n(S_0, \ldots, S_n)S_n + \sigma_n(S_0, \ldots, S_n)S_n\eta_{n+1}$$

with i.i.d. random variables $\eta_i, i \in \mathbb{N}$, and measurable functions $\alpha_n, \sigma_n : \mathbb{R}^n \to \mathbb{R}$. Trying to set up a corresponding model in continuous time, we arrive at the stochastic differential equation

$$dS_t \;=\; \alpha_t S_t\, dt + \sigma_t S_t\, dB_t \tag{8.0.3}$$

with an $(\mathcal{F}_t)$-Brownian motion (B_t) and $(\mathcal{F}_t^P)$ adapted continuous stochastic processes $(\alpha_t)_{t\geq 0}$ and $(\sigma_t)_{t\geq 0}$, where $(\mathcal{F}_t)$ is a given filtration on a probability space $(\Omega, \mathcal{A}, P)$. The processes α_t and σ_t describe the *instantaneous mean rate of return* and the *volatility*. Both are allowed to be time.dependent and random.
In order to compute a solution of (8.0.3), we assume $S_t > 0$ for any $t \geq 0$, and derive the equation by S_t:

$$\frac{1}{S_t}\, dS_t \;=\; \alpha_t\, dt + \sigma_t\, dB_t. \tag{8.0.4}$$

We will prove in Section **??** that if an s.d.e. holds then the s.d.e. multiplied by a continuous adapted process also holds, cf. Theorem 8.1. Hence (8.0.4) is equivalent to (8.0.3) if $S_t > 0$. If (8.0.4) would be a classical o.d.e. then we could use the identity $d\,\log S_t = \frac{1}{S_t}\, dS_t$ to solve the equation (8.0.4). In Itô calculus, however, the classical chain rule is violated. Nevertheless, it is still useful to compute $d\log S_t$ by Itô's formula. The process (S_t) has quadratic variation

$$[S]_t \;=\; \left[\int_0^{\bullet} \sigma_r S_r\, dB_r\right]_t \;=\; \int_0^t \sigma_r^2 S_r^2\, dr \qquad \text{for any } t \geq 0,$$

almost surely along any appropriate sequence (π_n) of partitions with $\mathrm{mesh}(\pi_n) \to 0$. The first equation holds by (8.0.3), since $t \mapsto \int_0^t \alpha_r S_r\, dr$ has bounded variation, and the second identity is proved in Theorem 8.1 below. Therefore, Itô's formula implies:

$$\begin{aligned} d\,\log S_t &= \frac{1}{S_t}\, dS_t - \frac{1}{2S_t^2}\, d[S]_t \\ &= \alpha_t\, dt + \sigma_t\, dB_t - \frac{1}{2}\sigma_t^2\, dt \\ &= \mu_t\, dt + \sigma_t\, dB_t, \end{aligned}$$

where $\mu_t := \alpha_t - \sigma_t^2/2$, i.e.,

$$\log S_t - \log S_0 \;=\; \int_0^t \mu_s\, ds + \int_0^t \sigma_s\, dB_s,$$

or, equivalently,

$$S_t \;=\; S_0 \cdot \exp\left(\int_0^t \mu_s\, ds + \int_0^t \sigma_s\, dB_s\right). \tag{8.0.5}$$

Conversely, one can verify by Itô's formula that (S_t) defined by (8.0.5) is indeed a solution of (8.0.3). Thus we have proven existence, uniqueness and an explicit representation for a strong solution of (8.0.3). In the special case when $\alpha_t \equiv \alpha$ and $\sigma_t \equiv \sigma$ are constants in t and ω, the solution process

$$S_t \;=\; S_0 \exp\big(\sigma B_t + (\alpha - \sigma^2/2)t\big)$$

is called a **geometric Brownian motion with parameters** α **and** σ.

8.1 Stochastic Calculus for Itô processes

By definition, any solution of an SDE of the form (8.0.1) is the sum of an absolutely continuous adapted process and an Itô stochastic integral w.r.t. the underlying Brownian motion, i.e.,

$$X_t \;=\; A_t + I_t \qquad \text{for } t < T \tag{8.1.1}$$

where

$$A_t \;=\; \int_0^t K_s\,ds \quad \text{and} \quad I_t \;=\; \int_0^t H_s\,dB_s \qquad \text{for } t < T \tag{8.1.2}$$

with $(H_t)_{t<T}$ and $(K_t)_{t<T}$ almost surely continuous and $(\mathcal{F}_t^{B,P})$-adapted. A stochastic process of type (8.1.1) is called an **Itô process**. In order to compute and analyze solutions of SDE we will apply Itô's formula to Itô processes. Since the absolutely continuous process A_t has bounded variation, classical Stieltjes calculus applies to this part of an Itô process. It remains to consider the stochastic integral part $(I_t)_{t<T}$:

Stochastic integrals w.r.t. Itô processes

Let (π_n) be a sequence of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$. Recall that for $t \geq 0$,

$$I_t \;=\; \int_0^t H_s\,dB_s \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} H_s \cdot (B_{s'\wedge t} - B_s)$$

w.r.t. convergence in probability on $\{t < T\}$, cf. Theorem 5.12.

Theorem 8.1 (Composition rule and quadratic variation). *Suppose that T is a predictable stopping time and $(H_t)_{t<T}$ is almost surely continuous and adapted.*

(1). For any almost surely continuous, adapted process $(G_t)_{0\leq t<T}$, and for any $t \geq 0$,

$$\lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} G_s (I_{s'\wedge t} - I_s) \;=\; \int_0^t G_s H_s\,dB_s \tag{8.1.3}$$

with convergence in probability on $\{t < T\}$. Moreover, if H is in $\mathcal{L}_a^2([0,a])$ and G is bounded on $[0,a]\times\Omega$ for some $a > 0$, then the convergence holds in $M_c^2([0,a])$ and thus uniformly for $t \in [0,a]$ in the $L^2(P)$ sense.

(2). *For any* $t \geq 0$*, the quadratic variation* $[I]_t$ *is given by*

$$[I]_t \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} (I_{s'\wedge t} - I_s)^2 \;=\; \int\limits_0^t H_s^2\, ds \tag{8.1.4}$$

w.r.t. convergence in probability on $\{t < T\}$*.*

Remark (Uniform convergence). Similarly to the proof of Theorem 5.12 one can show that there is a sequence of bounded stopping times $T_k \nearrow T$ such that almost surely along a subsequence, the convergence in (8.1.3) and (8.1.4) holds uniformly on $[0, T_k]$ for any k.

Proof. (1). We first fix $a > 0$ and assume that H is in $\mathcal{L}_a^2([0,a))$ and G is bounded, left-continuous and adapted on $[0,\infty)\times\Omega$. Since $I_{s'\wedge t} - I_s \;=\; \int\limits_s^{s'\wedge t} H_r\, dB_r$, we obtain

$$\sum_{\substack{s\in\pi_n\\ s<t}} G_s(I_{s'\wedge t} - I_s) \;=\; \int\limits_0^t G_{\lfloor r\rfloor} H_r\, dB_r$$

where $\lfloor r\rfloor_n \;=\; \max\{s\in\pi_n \;:\; s\leq r\}$ is the next partition point below r. As $n\to\infty$, the right-hand side converges to $\int\limits_0^t G_r H_r\, dB_r$ in $M_c^2([0,a])$ because $G_{\lfloor r\rfloor_n} H_r \to G_r H_r$ in $L^2(P\otimes\lambda_{[0,a)})$ by continuity of G and dominated convergence.

The assertion in the general case now follows by localization: Suppose (S_k) and (T_k) are increasing sequences of stopping times with $T_k \nearrow T$ and $H_t I_{\{t\leq S_k\}} \in \mathcal{L}_a^2([0,\infty))$, and let

$$\widetilde{T}_k \;=\; S_k \wedge T_k \wedge \inf\{t\geq 0 \;:\; |G_t| > k\} \wedge k.$$

Then $\widetilde{T}_k \nearrow T$, the process $H_t^{(k)} := H_t I_{\{t\leq T_k\}}$ is in $\mathcal{L}_a^2([0,\infty))$ the process $G_t^{(k)} := G_t I_{\{t\leq T_k\}}$ is bounded, left-continuous and adapted, and

$$I_s \;=\; \int\limits_0^s H_r^{(k)}\, dB_r, \quad G_s \;=\; G_s^{(k)} \qquad \text{for any } s\in[0,t]$$

holds almost surely on $\{t \le \widetilde{T}_k\}$. Therefore,

$$\begin{aligned}\sum_{\substack{s\in\pi_n\\ s<t}} G_s(I_{s'\wedge t}) &= \sum_{\substack{s\in\pi_n\\ s<t}} G_s^{(k)}(I_{s'\wedge t}) \\ &\to \int_0^t G_r^{(k)} H_r^{(k)}\, dB_r \;=\; \int_0^t G_r H_r\, dB_r\end{aligned}$$

uniformly for $t \le \widetilde{T}_k$ in $L^2(P)$. The claim follows, since

$$P\left[\{t < T\} \setminus \bigcup_k \{t \le \widetilde{T}_k\}\right] \;=\; 0.$$

(2). We first assume H is in $\mathcal{L}_a^2([0,\infty))$, continuous and bounded. Then for $s \in \pi_n$,

$$\Delta I_s \;=\; I_{s'\wedge t} - I_s \;=\; \int_s^{s'\wedge t} H_r\, dB_r \;=\; H_s \Delta B_s + R_s^{(n)}$$

where $R_s^{(n)} \;:=\; \int_s^{s'\wedge t} (H_r - H_{\lfloor r\rfloor})\, dB_r$. Therefore,

$$\sum_{\substack{s\in\pi_n\\ s<t}} (\Delta I_s)^2 \;=\; \sum_{\substack{s\in\pi_n\\ s<t}} H_s^2 (\Delta B_s)^2 + 2\sum_{\substack{s\in\pi_n\\ s<t}} R_s^{(n)} H_s \Delta B_s + \sum_{\substack{s\in\pi_n\\ s<t}} (R_s^{(n)})^2.$$

Since $[B]_t \;=\; t$ almost surely, the first term on the right-hand side converges to $\int_0^t H_s^2\, ds$ with probability one. It remains to show that the remainder terms converge to 0 in probability as $n \to \infty$. This is the case, since

$$\begin{aligned}E\left[\sum (R_s^{(n)})^2\right] &= \sum E[(R_s^{(n)})^2] \;=\; \sum \int_s^{s'\wedge t} E[(H_r - H_{\lfloor r\rfloor_n})^2]\, dr \\ &= \int_0^t E[(H_r - H_{\lfloor r\rfloor_n})^2]\, dr \;\longrightarrow\; 0\end{aligned}$$

by the Itô isometry and continuity and boundedness of H, whence $\sum (R_s^{(n)})^2 \to 0$ in $\mathcal{L}^1$ and in probability, and $\sum R_s^{(n)} H_s \Delta B_s \;\to\; 0$ in the same sense by the

Schwarz inequality.

For H defined up to a stopping time T, the assertion follows by a localization procedure similar to the one applied above.

□

The theorem and the corresponding composition rule for Stieltjes integrals suggest that we may define stochastic integrals w.r.t. an Itô process

$$X_t \;=\; X_0 + \int_0^t H_s \, dB_s + \int_0^t K_s \, ds, \qquad t < T,$$

in the following way:

Definition. *Suppose that (B_t) is a Brownian motion on $(\Omega, \mathcal{A}, P)$ a filtration $(\mathcal{F}_t^P)$, X_0 is an $(\mathcal{F}_0^P)$-measurable random variable, T is a predictable $(\mathcal{F}_t^P)$-stopping time, and $(G_t), (H_t)$ and (K_t) are almost surely continuous, $(\mathcal{F}_t^P)$ adapted processes defined for $t < T$. Then the stochastic integral of (G_t) w.r.t. (X_t) is the Itô process defined by*

$$\int_0^t G_s \, dX_s \;=\; \int_0^t G_s H_s \, dB_s + \int_0^t G_s K_s \, ds, \qquad t < T.$$

By Theorem 8.1, this definition is consistent with a definition by Riemann sum approximations. Moreover, the definition shows that the class of almost surely $(\mathcal{F}_t^P)$ adapted Itô process w.r.t. a given Brownian motion is *closed under taking stochastic integrals*! In particular, strong solutions of SDE w.r.t. Itô processes are again Itô processes.

Calculus for Itô processes

We summarize calculus rules for Itô processes that are immediate consequences of the definition above and Theorem 8.1: Suppose that (X_t) and (Y_t) are Itô processes, and (G_t), $(\widetilde{G}_t)$ and (H_t) are adapted continuous process that are all defined up to a stopping time T. Then the following calculus rules hold for Itô stochastic differentials:

Linearity:

$$d(X + cY) \;=\; dX + c\, dY \qquad \text{for any } c \in \mathbb{R},$$
$$(G + cH)\, dX \;=\; G\, dX + cH\, dX \qquad \text{for any } c \in \mathbb{R}.$$

Composition rule:

$$dY = G\,dX \quad \Rightarrow \quad \widetilde{G}\,dY = \widetilde{G}G\,dX,$$

Quadratic variation:

$$dY = G\,dX \quad \Rightarrow \quad d[Y] = G^2\,d[X],$$

Itô rule:

$$dF(t,X) = \frac{\partial F}{\partial x}(t,X)\,dX + \frac{\partial F}{\partial t}F(t,X)\,dt + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t,X)\,d[X]$$

for any function $F \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R})$.

All equations are to be understood in the sense that the corresponding stochastic integrals over any interval $[0,t], t < T$, coincide almost surely. The proofs are straightforward: For example, if

$$Y_t = Y_0 + \int_0^t G_s\,dX_s$$

and

$$X_t = X_0 + \int_0^t K_s\,ds + \int_0^t H_s\,dB_s$$

then, by the definition above, for $t < T$,

$$Y_t = Y_0 + \int_0^t G_sK_s\,ds + \int_0^t G_sH_s\,dB_s,$$

and hence

$$\int_0^t \widetilde{G}_s\,dY_s = \int_0^t \widetilde{G}_sG_sK_s\,ds + \int_0^t \widetilde{G}_sG_sH_s\,dB_s = \int_0^t \widetilde{G}_sG_s\,dX_s,$$

and

$$[Y]_t = \left[\int_0^{\bullet} G_sH_s\,dB_s\right]_t = \int_0^t G_s^2H_s^2\,ds = \int_0^t G_s^2\,d[X]_s.$$

Moreover, Theorem 8.1 guarantees that the stochastic integrals in Itô's formula, which are limits of Riemann-Itô sums coincide with the stochastic integrals for Itô processes defined above.

Example (Option Pricing in continuous time I). We again consider the continuous time asset price model introduced in the beginning of Section **??**. Suppose an agent is holding ϕ_t units of a single asset with price process (S_t) at time t, and he invests the remainder $V_t - \phi_t S_t$ of his wealth V_t in the money market with interest rate R_t. We assume that (ϕ_t) and (R_t) are continuous adapted processes. Then the change of wealth in a small time unit should be described by the Itô equation

$$dV_t \;=\; \phi_t\, dS_t + R_t(V_t - \phi_t S_t)\, dt.$$

Similarly to the discrete time case, we consider the discounted wealth process

$$\widetilde{V}_t \;:=\; \exp\left(-\int_0^t R_s\, ds\right) V_t.$$

Since $t \mapsto \int_0^t R_s\, ds$ has bounded variation, the Itô rule and the composition rule for stochastic integrals imply:

$$\begin{aligned}
d\widetilde{V}_t &= \exp\left(-\int_0^t R_s\, ds\right)\, dV_t - \exp\left(-\int_0^t R_s\, ds\right) R_t V_t\, dt \\
&= \exp\left(-\int_0^t R_s\, ds\right) \phi_t\, dS_t - \exp\left(-\int_0^t R_s\, ds\right) R_t \phi_t S_t\, dt \\
&= \phi_t \cdot \left(\exp\left(-\int_0^t R_s\, ds\right)\, dS_t - \exp\left(-\int_0^t R_s\, ds\right) R_t S_t\, dt\right) \\
&= \phi_t\, d\widetilde{S}_t,
\end{aligned}$$

where $\widetilde{S}_t$ is the discounted asset price process. Therefore,

$$\widetilde{V}_t - \widetilde{V}_0 \;=\; \int_0^t \phi_s\, d\widetilde{S}_s \qquad \forall t \geq 0 \quad P\text{-almost surely}.$$

As a consequence, we observe that if $(\widetilde{S}_t)$ is a (local) martingale under a probability measure P_* that is equivalent to P then the discounted wealth process $\widetilde{V}_t$ is also a local martingale under P_*. A corresponding probability measure P_* is called an *equivalent* *martingale* measure or *risk neutral measure*, and can be identified by Girsanov's theorem, cf. **??** below. Once we have found P_*, option prices can be computed similarly as in discrete time under the additional assumption that the true measure P for the asset price process is equivalent to P_*.

The Itô-Doeblin formula in $\mathbb{R}^1$

We can now apply Itô's formula to solutions of stochastic differential equations. Let $b, \sigma \in C(\mathbb{R}_+ \times I)$ where $I \subseteq \mathbb{R}$ is an open interval. Suppose that (B_t) is an $(\mathcal{F}_t)$-Brownian motion on $(\Omega, \mathcal{A}, P)$, and $(X_t)_{0 \le t < T}$ is an $(\mathcal{F}_t^P)$-adapted process with values in I and defined up to an $(\mathcal{F}_t^P)$ stopping time T such that the s.d.e.

$$X_t - X_0 = \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \qquad \text{for any } t < T \tag{8.1.5}$$

holds almost surely.

Corollary 8.2 (Doeblin 1941, Itô 1944). *Let $F \in C^{1,2}(R_+ \times I)$. Then almost surely,*

$$\begin{aligned} F(t, X_t) - F(0, X_0) &= \int_0^t (\sigma F')(s, X_s)\, dB_s \\ &\quad + \int_0^t \left(\frac{\partial F}{\partial t} + \frac{1}{2}\sigma^2 F'' + bF' \right)(s, X_s)\, ds \qquad \text{for any } t < T, \end{aligned} \tag{8.1.6}$$

where $F' = \partial F / \partial x$ denotes the partial derivative w.r.t. x.

Proof. Let (π_n) be a sequence of partitions with $\mathrm{mesh}(\pi_n) \to 0$. Since the process $t \mapsto X_0 + \int_0^t b(s, X_s)\, ds$ has sample paths of locally bounded variation, the quadratic variation of (X_t) is given by

$$[X]_t = \left[\int_0^\bullet \sigma(s, X_s)\, dB_s \right]_t = \int_0^t \sigma(s, X_s)^2\, ds \qquad \forall t < T$$

w.r.t. almost sure convergence along a subsequence of (π_n). Hence Itô's formula can be applied to almost every sample path of (X_t), and we obtain

$$F(t,X_t) - F(0,X_0) \;=\; \int_0^t F'(s,X_s)\,dX_s + \int_0^t \frac{\partial F}{\partial t}(s,X_s)\,ds + \frac{1}{2}\int_0^t F''(s,X_s)\,d[X]_s$$

$$= \int_0^t (\sigma F')(s,X_s)\,dB_s + \int_0^t (bF')(s,X_s)\,ds + \int_0^t \frac{\partial F}{\partial t}(s,X_s)\,ds + \frac{1}{2}\int_0^t (\sigma^2 F'')(s,X_s)\,ds$$

for all $t < T$, P-almost surely. Here we have used (8.1.5) and the fact that the Itô integral w.r.t. X is an almost sure limit of Riemann-Itô sums after passing once more to an appropriate subsequence of (π_n). □

Exercise (Black Scholes partial differential equation). A stock price is modeled by a geometric Brownian Motion (S_t) with parameters $\alpha, \sigma > 0$. We assume that the interest rate is equal to a real constant r for all times. Let $c(t,x)$ be the value of an option at time t if the stock price at that time is $S_t = x$. Suppose that $c(t,S_t)$ is replicated by a hedging portfolio, i.e., there is a trading strategy holding ϕ_t shares of stock at time t and putting the remaining portfolio value $V_t - \phi_t S_t$ in the money market account with fixed interest rate r so that the total portfolio value V_t at each time t agrees with $c(t,S_t)$.

"Derive" the *Black-Scholes partial differential equation*

$$\frac{\partial c}{\partial t}(t,x) + rx\frac{\partial c}{\partial x}(t,x) + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 c}{\partial x^2}(t,x) \;=\; rc(t,x) \tag{8.1.7}$$

and the *delta-hedging rule*

$$\phi_t \;=\; \frac{\partial c}{\partial x}(t,S_t) \qquad (=:\ \mathrm{Delta}). \tag{8.1.8}$$

Hint: Consider the discounted portfolio value $\widetilde{V}_t = e^{-rt}V_t$ *and, correspondingly,* $e^{-rt}c(t,S_t)$. *Compute the Ito differentials, and conclude that both processes coincide if* c *is a solution to (8.1.7) and* ϕ_t *is given by (8.1.8).*

Martingale problem for solutions of SDE

The Itô-Doeblin formula shows that

$$M_t^F \;=\; F(t,X_t) - F(0,X_0) - \int_0^t (\mathscr{L}_s F)(s,X_s)\,ds$$

is a local martingale up to T for any $F \in C^{1,2}(\mathbb{R}_+ \times I)$ and

$$\mathscr{L}_t F \;=\; \frac{1}{2}\sigma(t,\bullet)^2 F'' + b(t,\bullet)F'.$$

In particular, in the time-homogeneous case and for $T = \infty$, any solution of (8.1.5) solves the martingale problem for the Operator $\mathscr{L}F = \frac{1}{2}\sigma^2 F'' + bF'$ with domain $C_0^2(I)$.

Similar as for Brownian motion, the martingales identified by the Itô-Doeblin formula can be used to compute various expectation values for the Itô diffusion (X_t). In the next section we will look at first examples.

Remark (Uniqueness and Markov property of strong solutions). If the coefficients are, for example, Lipschitz continuous, then the strong solution of the s.d.e. (8.1.5) is unique, and it has the strong Markov property, i.e., it is a diffusion process in the classical sense (a strong Markov process with continuous sample paths). By the Itô-Doeblin formula, the generator of this Markov process is an extension of the operator $(\mathscr{L}, C_0^2(I))$.

Although in general, uniqueness and the Markov property may not hold for solutions of the s.d.e. (8.1.5), we call any solution of this equation an **Itô diffusion**.

8.2 Stochastic growth

In this section we consider time-homogeneous Itô diffusions taking values in the interval $I = (0,\infty)$. They provide natural models for stochastic growth processes, e.g. in mathematical biology, financial mathematics and many other application fields. Analogue results also hold if I is replaced by an arbitrary non-empty open interval.

Suppose that $(X_t)_{0\le t<T}$ is a strong solution of the s.d.e.

$$\begin{aligned} dX_t &= b(X_t)\,dt + \sigma(X_t)\,dB_t \qquad \text{for } t \in [0,T), \\ X_0 &= x_0, \end{aligned}$$

with a given Brownian motion (B_t), $x_0 \in (0,\infty)$, and continuous time-homogeneous coefficients $b, \sigma : (0,\infty) \to \mathbb{R}$ such that the solution is defined up to the explosion time

$$T \;=\; \sup_{\varepsilon, r>0} T_{\varepsilon,r}, \qquad T_{\varepsilon,r} \;=\; \inf\{t \ge 0 \mid X_t \notin (\varepsilon, r)\}.$$

The corresponding generator is

$$\mathscr{L}F \;=\; bF' + \frac{1}{2}\sigma^2 F''.$$

Before studying some concrete models, we show in the general case how harmonic functions can be used to compute exit distributions (e.g. ruin probabilities) to analyze the asymptotic behaviour of X_t as $t \nearrow T$.

Scale function and exit distributions

To determine the exit distribution from a finite subinterval $(\varepsilon, r) \subset (0,\infty)$ we compute the harmonic functions of $\mathscr{L}$. For $h \in C^2(0,\infty)$ with $h' > 0$ we obtain:

$$\mathscr{L}h \;=\; 0 \quad \iff \quad h'' \;=\; -\frac{2b}{\sigma^2}h' \quad \iff \quad (\log h')' \;=\; -\frac{2b}{\sigma^2}.$$

Therefore, the two-dimensional vector space of harmonic functions is spanned by the constant function 1 and by

$$s(x) \;=\; \int_{x_0}^{x} \exp\left(-\int_{x_0}^{z} \frac{2b(y)}{\sigma(y)^2}\, dy \right)\, dz.$$

The function s is called the **scale function** of the process (X_t). It is strictly increasing and harmonic on $(0,\infty)$. Hence we can think of $s : (0,\infty) \to (s(0), s(\infty))$ as a coordinate transformation, and the transformed process $s(X_t)$ is a local martingale up to the explosion time T.

Applying the martingale convergence theorem and the optional stopping theorem to $s(X_t)$ one obtains:

Theorem 8.3. *For any $\varepsilon, r \in (0,\infty)$ with $\varepsilon < x_0 < r$ we have:*

(1). The exit time $T_{\varepsilon,r} \;=\; \inf\{t \in [0,T) \;:\; X_t \notin (\varepsilon, r)\}$ is almost surely smaller than T.

(2). $P[T_\varepsilon < T_r] \;=\; P[X_{T_{\varepsilon,r}} = \varepsilon] \;=\; \dfrac{s(r) - s(x)}{s(r) - s(\varepsilon)}.$

Remark. (1). Note that any affine transformation $\widetilde{s}(x) = cs(x) + d$ with constants $c > 0$ and $d \in \mathbb{R}$ is also harmonic and strictly increasing, and hence a scale function. The ratio $(s(r) - s(x))/(s(r) - s(\varepsilon))$ is invariant under non-degenerate affine transformations of s.

(2). The scale function and the ruin probabilities depend only on the ratio $b(x)/\sigma(x)^2$.

The proof of Theorem 8.3 is left as an exercise.

Recurrence and asymptotics

As a consequence of the computation of exit distributions we can study the asymptotics of one-dimensional non-degenerate Itô diffusions as $t \nearrow T$. For example, for $\varepsilon \in (0, x_0)$ we obtain

$$\begin{aligned} P[T_\varepsilon < T] &= P[T_\varepsilon < T_r \quad \text{for some } r \in (x_0, \infty)] \\ &= \lim_{r\to\infty} P[T_\varepsilon < T_r] \;=\; \lim_{r\to\infty} \frac{s(r) - s(x_0)}{s(r) - s(\varepsilon)}. \end{aligned}$$

In particular,

$$\begin{aligned} P[X_t = \varepsilon \quad \text{for some } t \in [0, T)] \quad &= \quad P[T_\varepsilon < T] \;=\; 1 \\ \iff \quad & s(\infty) \;=\; \lim_{r\nearrow\infty} s(r) \;=\; \infty. \end{aligned}$$

Similarly, one obtains for $r \in (x_0, \infty)$:

$$\begin{aligned} P[X_t = \varepsilon \quad \text{for some } t \in [0, T)] \quad &= \quad P[T_r < T] \;=\; 1 \\ \iff \quad & s(0) \;=\; \lim_{\varepsilon\searrow 0} s(\varepsilon) \;=\; -\infty. \end{aligned}$$

Moreover,

$$P[X_t \to \infty \text{ as } t \nearrow T] \;=\; P\left[\bigcup_{\varepsilon>0} \bigcap_{r<\infty} \{T_r < T_\varepsilon\}\right] \;=\; \lim_{\varepsilon\searrow 0} \lim_{r\nearrow\infty} \frac{s(x_0) - s(\varepsilon)}{s(r) - s(\varepsilon)},$$

and

$$P[X_t \to 0 \text{ as } t \nearrow T] \;=\; P\left[\bigcup_{r<\infty} \bigcap_{\varepsilon>0} \{T_\varepsilon < T_r\}\right] \;=\; \lim_{r\nearrow\infty} \lim_{\varepsilon\searrow 0} \frac{s(x_0) - s(\varepsilon)}{s(r) - s(\varepsilon)}.$$

Summarizing, we have shown:

Corollary 8.4 (Asymptotics of one-dimensional Itô diffusions). *(1). If $s(0) = -\infty$ and $s(\infty) = \infty$, then the process (X_t) is recurrent, i.e.,*

$$P[X_t = y \quad \text{for some } t \in [0,T)] \;=\; 1 \qquad \text{for any } x_0, y \in (0,\infty).$$

(2). If $s(0) > -\infty$ and $s(\infty) = \infty$ then $\lim_{t \nearrow T} X_t = 0$ almost surely.

(3). If $s(0) = -\infty$ and $s(\infty) < \infty$ then $\lim_{t \nearrow T} X_t = \infty$ almost surely.

(4). If $s(0) > -\infty$ and $s(\infty) < \infty$ then

$$P\left[\lim_{t \nearrow T} X_t = 0\right] \;=\; \frac{s(\infty) - s(x_0)}{s(\infty) - s(0)}$$

and

$$P\left[\lim_{t \nearrow T} X_t = \infty\right] \;=\; \frac{s(x_0) - s(0)}{s(\infty) - s(0)}$$

Intuitively, if $s(0) = -\infty$, in the natural scale the boundary is transformed to $-\infty$, which is not a possible limit for the local martingale $s(X_t)$, whereas otherwise $s(0)$ is finite and approached by $s(X_t)$ with strictly positive probability.

Example. Suppose that $b(x)/\sigma(x)^2 \approx \gamma x^{-1}$ as $x \nearrow \infty$ and $b(x)/\sigma(x)^2 \approx \delta x^{-1}$ as $x \searrow 0$ holds for $\gamma, \delta \in \mathbb{R}$ in the sense that $b(x)/\sigma(x)^2 - \gamma x^{-1}$ is integrable at ∞ and $b(x)/\sigma(x)^2 - \delta x^{-1}$ is integrable at 0. Then $s'(x)$ is of order $x^{-2\gamma}$ as $x \nearrow \infty$ and of order $x^{-2\delta}$ as $x \searrow 0$. Hence

$$s(\infty) \;=\; \infty \quad \iff \quad \gamma \leq \frac{1}{2}, \qquad s(0) \;=\; -\infty \quad \iff \quad \delta \geq \frac{1}{2}.$$

In particular, recurrence holds if and only if $\gamma \leq \frac{1}{2}$ and $\delta \geq \frac{1}{2}$.

More concrete examples will be studied below.

Remark (Explosion in finite time, Feller's test). Corollary 8.4 does not tell us whether the explosion time T is infinite with probability one. It can be shown that this is always the case if (X_t) is recurrent. In general, *Feller's test for explosions* provides a necessary and sufficient condition for the absence of explosion in finite time. The idea is to compute a function $g \in C(0,\infty)$ such that $e^{-t}g(X_t)$ is a local martingale and to apply the optional stopping theorem. The details are more involved than in the proof of corollary above, cf. e.g. Section 6.2 in [Durrett: Stochastic calculus].

Geometric Brownian motion

A geometric Brownian motion with parameters $\alpha \in \mathbb{R}$ and $\sigma > 0$ is a solution of the s.d.e.

$$dS_t \;=\; \alpha S_t\,dt + \sigma S_t\,dB_t. \tag{8.2.1}$$

We have already shown in the beginning of Section **??** that for $B_0 = 0$, the unique strong solution of (8.2.1) with initial condition $S_0 = x_0$ is

$$S_t \;=\; x_0 \cdot \exp\left(\sigma B_t + (\alpha - \sigma^2/2)t\right).$$

The distribution of S_t at time t is a **lognormal distribution**, i.e., the distribution of $c \cdot e^Y$ where c is a constant and Y is normally distributed. Moreover, one easily verifies that (S_t) is a time-homogeneous Markov process with log-normal transition densities

$$p_t(x,y) \;=\; \frac{1}{\sqrt{2\pi t\sigma^2}}\exp\left(-\frac{(\log(y/x)-\mu t)^2}{2t\sigma^2}\right), \qquad t,x,y>0,$$

where $\mu = \alpha - \sigma^2/2$. By the Law of Large Numbers for Brownian motion,

$$\lim_{t\to\infty} S_t \;=\; \begin{cases} +\infty & \text{if } \mu > 0 \\ 0 & \text{if } \mu < 0 \end{cases}.$$

If $\mu = 0$ then (S_t) is recurrent since the same holds for (B_t).
We now convince ourselves that we obtain the same results via the scale function:
The ratio of the drift and diffusion coefficient is

$$\frac{b(x)}{\sigma(x)^2} \;=\; \frac{\alpha x}{(\sigma x)^2} \;=\; \frac{\alpha}{\sigma^2 x},$$

and hence

$$s'(x) \;=\; \text{const.}\cdot\exp\left(-\int_{x_0}^{x}\frac{2\alpha}{\sigma^2 y}\,dy\right) \;=\; \text{const.}\cdot x^{-2\alpha/\sigma^2}.$$

Therefore,

$$s(\infty) \;=\; \infty \quad\iff\quad 2\alpha/\sigma^2 \le 1, \qquad s(0) \;=\; \infty \quad\iff\quad 2\alpha/\sigma^2 \ge 1,$$

which again shows that $S_t \to \infty$ for $\alpha > \sigma^2/2$, $S_t \to 0$ for $\alpha < \sigma^2/2$, and S_t is recurrent for $\alpha = \sigma^2/2$.

Feller's branching diffusion

Our second growth model is described by the stochastic differential equation

$$dX_t \;=\; \beta X_t\,dt + \sigma\sqrt{X_t}\,dB_t, \qquad X_0 = x_0, \tag{8.2.2}$$

with given constants $\beta \in \mathbb{R}, \sigma > 0$, and values in $\mathbb{R}_+$. Note that in contrast to the equation of geometric Brownian motion, the multiplicative factor $\sqrt{X_t}$ in the noise term is not a linear function of X_t. As a consequence, there is no explicit formula for a solution of (8.2.2). Nevertheless, a general existence result guarantees the existence of a strong solution defined up to the explosion time

$$T \;=\; \sup_{\varepsilon,r>0} T_{\mathbb{R}\setminus(\varepsilon,r)},$$

cf. **??**. SDEs similar to (8.2.2) appear in various applications.

Example (Diffusion limits of branching processes). We consider a Galton-Watson branching process Z_t^h with time steps $t = 0, h, 2h, 3h, \ldots$ of size $h > 0$, i.e., Z_0^h is a given initial population size, and

$$Z_{t+h}^h \;=\; \sum_{i=1}^{Z_t^h} N_i, t/h \quad \text{for } t = k\cdot h, k = 0, 1, 2, \ldots,$$

with independent identically distributed random variables $N_{i,k}, i \geq 1, k \geq 0$. The random variable Z_{kh}^h describes the size of a population in the k-th generation when $N_{i,l}$ is the number of offspring of the i-th individual in the l-th generation. We assume that the mean and the variance of the offspring distribution are given by

$$E[N_{i,l}] \;=\; 1 + \beta h \quad \text{and} \quad \mathrm{Var}[N_{i,l}] \;=\; \sigma^2$$

for finite constants $\beta, \sigma \in \mathbb{R}$.

We are interested in a scaling limit of the model as the size h of time steps goes to 0. To establish convergence to a limit process as $h \searrow 0$ we rescale the population size by h, i.e., we consider the process

$$X_t^h \;:=\; h \cdot Z_{\lfloor t \rfloor}^h, \qquad t \in [0, \infty).$$

The mean growth ("drift") of this process in one time step is

$$E[X_{t+h}^h - X_t^h \mid \mathcal{F}_t^h] \;=\; h \cdot E[Z_{t+h}^h - Z_t^h \mid \mathcal{F}_t^h] \;=\; h\eta h Z_t^h \;=\; h\beta X_t^h,$$

and the corresponding condition variance is

$$\mathrm{Var}[X_{t+h}^h - X_t^h \mid \mathcal{F}_t^h] \;=\; h^2 \cdot \mathrm{Var}[Z_{t+h}^h - Z_t^h \mid \mathcal{F}_t^h] \;=\; h^2\sigma^2 Z_t^h \;=\; h\sigma^2 X_t^h,$$

where $\mathcal{F}_t^h \;=\; \sigma(N_{i,l} \mid i \geq 1, 0 \leq l \leq k)$ for $t = k \cdot h$. Since both quantities are of order $O(h)$, we can expect a limit process (X_t) as $h \searrow 0$ with drift coefficient $\beta \cdot X_t$ and diffusion coefficient $\sqrt{\sigma^2 X_t}$, i.e., the scaling limit should be a diffusion process solving a s.d.e. of type (8.2.2). A rigorous derivation of this diffusion limit can be found e.g. in Section 8 of [Durrett: Stochastic Calculus].

We now analyze the asymptotics of solutions of (8.2.2). The ratio of drift and diffusion coefficient is $\beta x/(\sigma\sqrt{x})^2 = \beta/\sigma$, and hence the derivative of a scale function is

$$s'(x) \;=\; \text{const.} \cdot \exp(-2\beta x/\sigma).$$

Thus $s(0)$ is always finite, and $s(\infty) = \infty$ if and only if $\beta \leq 1$. Therefore, by Corollary 8.4, in the subcritical and critical case $\beta \leq 1$, we obtain

$$\lim_{t \nearrow T} X_t \;=\; 0 \qquad \text{almost surely},$$

whereas in the supercritical case $\beta > 1$,

$$P\left[\lim_{t \nearrow T} X_t = 0\right] > 0 \quad \text{and} \quad P\left[\lim_{t \nearrow T} X_t = \infty\right] > 0.$$

This corresponds to the behaviour of Galton-Watson processes in discrete time. It can be shown by Feller's boundary classification for one-dimensional diffusion processes that if $X_t \to 0$ then the process actually dies out almost surely in finite time, cf. e.g. Section 6.5 in [Durrett: Stochastic Calculus]. On the other hand, for trajectories with $X_t \to \infty$, the explosion time T is almost surely infinite and X_t grows exponentially as $t \to \infty$.

Cox-Ingersoll-Ross model

The CIR model is a model for the stochastic evolution of interest rates or volatilities. The equation is

$$dR_t \;=\; (\alpha - \beta R_t)\,dt + \sigma\sqrt{R_t}\,dB_t \qquad R_0 = x_0, \tag{8.2.3}$$

with a one-dimensional Brownian motion (B_t) and positive constants $\alpha, \beta, \sigma > 0$. Although the s.d.e. looks similar to the equation for Feller's branching diffusion, the behaviour of the drift coefficient near 0 is completely different. In fact, the idea is that the positive drift α pushes the process away from 0 so that a recurrent process on $(0,\infty)$ is obtained. We will see that this intuition is true for $\alpha \geq \sigma^2/2$ but not for $\alpha < \sigma^2/2$. Again, there is no explicit solution for the s.d.e. (8.13), but existence of a strong solution holds. The ratio of the drift and diffusion coefficient is $(\alpha-\beta x)/\sigma^2 x$, which yields

$$s'(x) \;=\; \text{const.}\cdot x^{-2\alpha/\sigma^2} e^{2\beta x/\sigma^2}.$$

Hence $s(\infty) = \infty$ for any $\beta > 0$, and $s(0) = \infty$ if and only if $2\alpha \geq \sigma^2$. Therefore, the CIR process is recurrent if and only if $\alpha \geq \sigma^2/2$, whereas $X_t \to 0$ as $t \nearrow T$ almost surely otherwise.

By applying Itô's formula one can now prove that X_t has finite moments, and compute the expectation and variance explicitly. Indeed, taking expectation values in the s.d.e.

$$R_t \;=\; x_0 + \int_0^t (\alpha - \beta R_s)\,ds + \int_0^t \sigma\sqrt{R_s}\,dB_s,$$

we obtain informally

$$\frac{d}{dt}E[R_t] \;=\; \alpha - \beta E[R_t],$$

and hence by variation of constants,

$$E[R_t] \;=\; x_0\cdot e^{-\beta t} + \frac{\alpha}{\beta}(1 - e^{-\beta t}).$$

To make this argument rigorous requires proving that the local martingale $t \mapsto \int_0^t \sigma\sqrt{R_s}dB_s$ is indeed a martingale:

Exercise. Consider a strong solution $(R_t)_{t\geq 0}$ of (8.13) for $\alpha \geq \sigma^2/2$.

(1). Show by applying Itô's formula to $x \mapsto |x|^p$ that $E[|R_t|^p] < \infty$ for any $t \geq 0$ and $p \geq 1$.

(2). Compute the expectation of R_t, e.g. by applying Itô's formula to $e^{\beta t}x$.

(3). Proceed in a similar way to compute the variance of R_t. Find its asymptotic value $\lim\limits_{t\to\infty} \mathrm{Var}[R_t]$.

8.3 Linear SDE with additive noise

We now consider stochastic differential equations of the form

$$dX_t \;=\; \beta_t C_t\, dt + \sigma_t\, dB_t, \qquad X_0 = x, \tag{8.3.1}$$

where (B_t) is a Brownian motion, and the coefficients are *deterministic* continuous functions $\beta, \sigma : [0, \infty) \to \mathbb{R}$. Hence the drift term $\beta_t X_t$ is linear in X_t, and the diffusion coefficient does not depend on X_t, i.e., the noise increment $\sigma_t\, dB_t$ is proportional to white noise dB_t with a proportionality factor that does not depend on X_t.

Variation of constants

An explicit strong solution of the SDE (8.3.1) can be computed by a "variation of constants" Ansatz. We first note that the general solution in the deterministic case $\sigma_t \equiv 0$ is given by

$$X_t \;=\; \text{const.} \cdot \exp\left(\int_0^t \beta_s\, ds\right).$$

To solve the SDE in general we try the ansatz

$$X_t \;=\; C_t \cdot \exp\left(\int_0^t \beta_s\, ds\right)$$

with a continuous Itô process (C_t) driven by the Brownian motion (B_t). By the Itô product rule,

$$dX_t = \beta_t X_t \, dt + \exp\left(\int_0^t \beta_s \, ds\right) dC_t.$$

Hence (X_t) solves (8.3.1) if and only if

$$dC_t = \exp\left(-\int_0^t \beta_s \, ds\right) \sigma_t \, dB_t,$$

i.e.,

$$C_t = C_0 + \int_0^t \exp\left(-\int_0^r \beta_s \, ds\right) \sigma_r \, dB_r.$$

We thus obtain:

Theorem 8.5. *The almost surely unique strong solution of the SDE (8.3.1)with initial value x is given by*

$$X_t^x = x \cdot \exp\left(-\int_0^t \beta_s \, ds\right) + \int_0^t \exp\left(\int_r^t \beta_s \, ds\right) \sigma_r \, dB_r.$$

Note that the theorem not only yields an explicit solution but it also shows that the solution depends smoothly on the initial value x. The effect of the noise on the solution is additive and given by a Wiener-Itô integral, i.e., an Itô integral with deterministic integrand. The average value

$$E[X_t^x] = x \cdot \exp\left(\int_0^t B_s \, ds\right), \tag{8.3.2}$$

coincides with the solution in the absence of noise, and the mean-square deviation from this solution due to random perturbation of the equation is

$$\mathrm{Var}[X_t^x] = \mathrm{Var}\left[\int_0^t \exp\left(\int_r^t \beta_s \, ds\right) \sigma_r \, dB_r\right] = \int_0^t \exp\left(2\int_r^t \beta_s \, ds\right) \sigma_r^2 \, dr$$

by the Itô isometry.

Solutions as Gaussian processes

We now prove that the solution (X_t) of a linear s.d.e. with additive noise is a Gaussian process. We first observe that X_t is normally distributed for any $t \geq 0$.

Lemma 8.6. *For any deterministic function $h \in L^2(0,t)$, the Wiener-Itô integral $I_t = \int_0^t h_s\, dB_s$ is normally distributed with mean 0 and variance $\int_0^t h_s^2\, ds$.*

Proof. Suppose first that $h = \sum_{i=0}^{n-1} c_i \cdot I_{(t_i,t_{i+1}]}$ is a step function with $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{R}$, and $0 \leq t_0 < t_1 < \ldots < t_n$. Then $I_t = \sum_{i=0}^{n-1} c_i \cdot (B_{t_{i+1}} - B_{t_i})$ is normally distributed with mean zero and variance

$$\mathrm{Var}[I_t] \;=\; \sum_{i=0}^{n-1} c_i^2 (t_{i+1} - t_i) \;=\; \int_0^t h_s^2\, ds.$$

In general, there exists a sequence $(h^{(n)})_{n\in\mathbb{N}}$ of step functions such that $h^{(n)} \to h$ in $L^2(0,t)$, and

$$I_t \;=\; \int_0^t h\, dB \;=\; \lim_{n\to\infty} \int_0^t h^{(n)}\, dB \qquad \text{in } L^2(\Omega, \mathcal{A}, P).$$

Hence I_t is again normally distributed with mean zero and

$$\mathrm{Var}[I_t] \;=\; \lim_{n\to\infty} \mathrm{Var}\left[\int_0^t h^{(n)}\, dB\right] \;=\; \int_0^t h^2\, ds.$$

□

Theorem 8.7 (Wiener-Itô integrals are Gaussian processes). *Suppose that $h \in L^2_{loc}([0,\infty), \mathbb{R})$. Then $I_t = \int_0^t h_s\, dB_s$ is a continuous Gaussian process with*

$$E[I_t] \;=\; 0 \quad \textit{and} \quad \mathrm{Cov}[I_t, I_s] \;=\; \int_0^{t\wedge s} h_r^2\, ds \qquad \textit{for any } t, s \geq 0.$$

Proof. Let $0 \leq t_1 < \ldots < t_n$. To show that $(I_{t_1}, \ldots, I_{t_n})$ has a normal distribution it suffices to prove that any linear combination of the random variables $I_{t_1}, \ldots, I_{t_n}$ is normally distributed. This holds true since any linear combination is again an Itô integral with deterministic integrand:

$$\sum_{i=1}^{n} \lambda_i I_{t_i} = \int_0^{t_n} \sum_{i=1}^{n} \lambda_i \cdot I_{(0,t_i)}(s) h_s \, dB_s$$

for any $n \in \mathbb{N}$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. Hence (I_t) is a Gaussian process with $E[I_t] = 0$ and

$$\begin{aligned} \mathrm{Cov}[I_t, I_s] &= E[I_t I_s] \\ &= E\left[\int_0^\infty h_r \cdot I_{(0,t)}(r) \, dB_r \int_0^\infty h_r \cdot I_{(0,s)}(r) \, dB_r\right] \\ &= (h \cdot I_{(0,t)}, h \cdot I_{(0,s)})_{L^2(0,\infty)} \\ &= \int_0^{s \wedge t} h_r^2 \, dr. \end{aligned}$$

□

Example (Brownian motion). If $h \equiv 1$ then $I_t = B_t$. The Brownian motion (B_t) is a centered Gaussian process with $\mathrm{Cov}[B_t, B_s] = t \wedge s$.

More generally, by Theorem 8.7 and Theorem 8.5, any solution (X_t) of a linear SDE with additive noise and deterministic (or Gaussian) initial value is a continuous Gaussian process. In fact by (8.3.1), the marginals of (X_t) are affine functions of the corresponding marginals of a Wiener-Itô integral:

$$X_t^x = \frac{1}{h_t} \cdot \left(x + \int_0^t h_r \sigma_r \, dB_r \right) \quad \text{with} \quad h_r = \exp\left(- \int_0^r \beta_u \, du \right).$$

Hence all finite dimensional marginals of (X_t^x) are normally distributed with

$$E[X_t^x] = x/H_t \quad \text{and} \quad \mathrm{Cov}[X_t^x, X_s^x] = \frac{1}{h_t h_s} \cdot \int_0^{t \wedge s} h_r^2 \sigma_r^2 \, dr.$$

The Ornstein-Uhlenbeck process

In 1905, Einstein introduced a model for the movement of a "big" particle in a fluid. Suppose that V_t^{abs} is the absolute velocity of the particle, $\overline{V}_t$ is the mean velocity of the fluid molecules and $V_t = V_t^{\text{abs}} - \overline{V}_t$ is the velocity of the particle relative to the fluid. Then the velocity approximatively can be described as a solution to an s.d.e.

$$dV_t = -\gamma V_t \, dt + \sigma dB_t. \tag{8.3.3}$$

Here (B_t) is a Brownian motion in $\mathbb{R}^d, d = 3$, and γ, σ are strictly positive constants that describe the damping by the viscosity of the fluid and the magnitude of the random collisions. A solution to the s.d.e. (8.3.3) is called an **Ornstein-Uhlenbeck process**. Although it has first been introduced as a model for the velocity of physical Brownian motion, the Ornstein-Uhlenbeck process is a fundamental stochastic process that is almost as important as Brownian motion for mathematical theory and stochastic modeling. In particular, it is a continuous-time analogue of an AR(1) autoregressive process. Note that (8.3.3) is a system of d decoupled one-dimensional stochastic differential equations $dV_t^{(i)} = -\gamma V_t^{(i)} + \sigma dB_t^{(i)}$. Therefore, we will assume w.l.o.g. $d = 1$. By the considerations above, the one-dimensional Ornstein-Uhlenbeck process is a continuous Gaussian process. The unique strong solution of the s.d.e. (8.3.3) with initial condition x is given explicitly by

$$V_t^x = e^{-\gamma t} \left(x + \sigma \int_0^t e^{\gamma s} \, dB_s \right). \tag{8.3.4}$$

In particular,

$$E[V_t^x] = e^{-\gamma t} x,$$

and

$$\begin{aligned} \operatorname{Cov}[V_t^x, V_s^x] &= e^{-\gamma(t+s)} \sigma^2 \int_0^{t \wedge s} e^{2\gamma r} \, dr \\ &= \frac{\sigma^2}{2\gamma} (e^{-\gamma|t-s|} - e^{-\gamma(t+s)}) \qquad \text{for any } t, s \geq 0. \end{aligned}$$

Note that as $t \to \infty$, the effect of the initial condition decays exponentially fast with rate γ. Similarly, the correlations between V_t^x and V_s^x decay exponentially as $|t-s| \to \infty$. The distribution at time t is

$$V_t^x \quad \sim \quad N\left(e^{-\gamma t}x,\ \frac{\sigma^2}{2\gamma}(1-e^{-2\gamma t})\right). \tag{8.3.5}$$

In particular, as $t \to \infty$

$$V_t^x \quad \xrightarrow{\mathcal{D}} \quad N\left(0, \frac{\sigma^2}{2\gamma}\right).$$

One easily verifies that $N(0,\sigma^2/2\gamma)$ is an *equilibrium* for the process: If $V_0 \sim N(0,\sigma^2/2\gamma)$ and (B_t) is independent of V_0 then

$$\begin{aligned} V_t &= e^{-\gamma t}V_0 + \sigma \int_0^t e^{\gamma(s-t)}\, dB_s \\ &\sim N\left(0, \frac{\sigma^2}{2\gamma}e^{-2\gamma t} + \sigma^2 \int_0^t e^{2\gamma(s-t)}\, ds\right) \ = \ N(0,\sigma^2/2\gamma) \end{aligned}$$

for any $t \geq 0$.

Theorem 8.8. *The Ornstein-Uhlenbeck process (V_t^x) is a time-homogeneous Markov process w.r.t. the filtration $(\mathcal{F}_t^{B,P})$ with stationary distribution $N(0,\sigma^2/2\gamma)$ and transition probabilities*

$$p_t(x,A) \ = \ P\left[e^{-\gamma t}x + \frac{\sigma}{\sqrt{2\gamma}}\sqrt{1-e^{-2\gamma t}}Z \in A\right], \qquad Z \sim N(0,1).$$

Proof. We first note that by (8.3.5),

$$V_t^x \ \sim \ e^{-\gamma t}x + \frac{\sigma}{\sqrt{2\gamma}}\sqrt{1-e^{-2\gamma t}}Z \qquad \text{for any } t \geq 0$$

with $Z \sim N(0,1)$. Hence,

$$E[f(V_t^x)] \ = \ (p_t f)(x)$$

for any non-negative measurable function $f : \mathbb{R} \to \mathbb{R}$. We now prove a *pathwise counterpart to the Markov property*: For $t, r \geq 0$, by (8.3.4)

$$V_{t+r}^x = e^{-\gamma(t+r)} \left(x + \sigma \int_0^t e^{\gamma s} \, dB_s \right) + \sigma \int_0^{t+r} e^{\gamma(s-t-r)} \, dB_s$$

$$= e^{-\gamma r} V_t^x + \sigma \int_0^r e^{\gamma(u-r)} \, d\overline{B}_u, \tag{8.3.6}$$

where $\overline{B}_u := B_{t+u} - B_t$ is a Brownian motion that is independent of $\mathcal{F}_t^{B,P}$. Hence, the random variable $\sigma \cdot \int_0^r e^{\gamma(u-r)} \, d\overline{B}_u$ is also independent of $\mathcal{F}_t^{B,P}$ and, by (8.3.4), it has the same distribution as the Ornstein-Uhlenbeck process with initial condition 0:

$$\sigma \cdot \int_0^r e^{\gamma(u-r)} \, d\overline{B}_u \quad \sim \quad V_r^0.$$

Therefore, by (8.3.6), the conditional distribution of V_{t+r}^x given $\mathcal{F}_t^{B,P}$ coincides with the distribution of the process with initial V_t^x at time r:

$$E[f(V_{t+r}^x) \mid \mathcal{F}_t^{B,P}] = E[f(e^{-\gamma r} V_t^x(\omega) + V_r^0)]$$

$$= E[f(V_r^{V_t^x(\omega)})] = (p_r f)(V_t^x(\omega)) \qquad \text{for } P\text{-a.e. } \omega.$$

This proves that (V_t^x) is a Markov process with transition kernels $p_r, r \geq 0$. □

Remark. The pathwise counterpart of the Markov property used in the proof above is called **cocycle property** of the stochastic flow $x \mapsto V_t^x$.

The Itô-Doeblin formula can now be used to identify the generator of the Ornstein-Uhlenbeck process: Taking expectation values, we obtain the forward equation

$$E[F(V_t^x)] = F(x) + \int_0^t E[(\mathscr{L} F)(V_s^x)] \, ds$$

for any function $F \in C_0^2(\mathbb{R})$ and $t \geq 0$, where

$$(\mathscr{L} F)(x) = \frac{1}{2} \sigma^2 f''(x) - \gamma x f'(x).$$

For the transition function this yields

$$(p_t F)(x) \;=\; F(x) + \int_0^t (p_s \mathscr{L} F)(x) \qquad \text{for any } x \in \mathbb{R},$$

whence

$$\lim_{t \searrow 0} \frac{(p_t f)(x) - f(x)}{t} \;=\; \lim_{t \searrow 0} \frac{1}{t} \int_0^t E[(\mathscr{L} f)(V_s^x)]\, ds \;=\; (\mathscr{L} f)(x)$$

by continuity and dominated convergence. This shows that the infinitesimal generator of te Ornstein-Uhlenbeck process is an extension of the operator $(\mathscr{L}, C_0^2(\mathbb{R}))$.

Change of time-scale

We will now prove that Wiener-Itô integrals can also be represented as Brownian motion with a coordinate transformation on the time axis. Hence solutions of one-dimensional linear SDE with additive noise are affine functions of time changed Brownian motions. We first note that a Wiener-Itô integral $I_t = \int_0^t h_r \, dB_r$ with $h \in L^2_{\text{loc}}(0,\infty)$ is a continuous centered Gaussian process with covariance

$$\text{Cov}[I_t, I_s] \;=\; \int_0^{t \wedge s} h_r^2 \, dr \;=\; \tau(t) \wedge \tau(s)$$

where

$$\tau(t) \;:=\; \int_0^t h_r^2 \, dr \;=\; \text{Var}[I_t]$$

is the corresponding variance process. The variance process should be thought of as an "internal clock" for the process (I_t). Indeed, suppose $h > 0$ almost everywhere. Then τ is strictly increasing and continuous, and

$$\tau \;:\; [0,\infty) \quad \to \quad [0, \tau(\infty)) \qquad \text{is a homeomorphism.}$$

Transforming the time-coordinate by τ, we have

$$\text{Cov}[I_{\tau^{-1}(t)}, I_{\tau^{-1}(s)}] \;=\; t \wedge s \qquad \text{for any } t, s \in [0, \tau(\infty)].$$

These are exactly the covariance of a Brownian motion. Since a continuous Gaussian process is uniquely determined by its expectations and covariances, we can conclude:

Theorem 8.9 (Wiener-Itô integrals as time changed Brownian motions). *The process* $\widetilde{B}_s := I_{\tau^{-1}(s)}, \quad 0 \leq s < \tau(\infty)$, *is a Brownian motion, and*

$$I_t = \widetilde{B}_{\tau(t)} \qquad \textit{for any } t \geq 0, P\textit{-almost surely}.$$

Proof. Since $(\widetilde{B}_s)_{0 \leq s < \tau(\infty)}$ has the same marginal distributions as the Wiener-Itô integral $(I_t)_{t \geq 0}$ (but at different times), $(\widetilde{B}_s)$ is again a continuous centered Gaussian process. Moreover, $\mathrm{Cov}[B_t, B_s] = t \wedge s$, so that (B_s) is indeed a Brownian motion. □

Note that the argument above is different from previous considerations in the sense that the Brownian motion $(\widetilde{B}_s)$ is constructed from the process (I_t) and not vice versa.
This means that we can not represent (I_t) as a time-change of a given Brownian motion (e.g. (B_t)) but we can only show that there exists a Brownian motion $(\widetilde{B}_s)$ such that I is a time-change of $\widetilde{B}$. This way of representing stochastic processes w.r.t. Brownian motions that are constructed from the process corresponds to the concept of weak solutions of stochastic differential equations, where driving Brownian motion is not given a priori. We return to these ideas in Section 9, where we will also prove that continuous local martingales can be represented as time-changed Brownian motions.

Theorem 8.9 enables us to represent solution of linear SDE with additive noise by time-changed Brownian motions. We demonstrate this with an example: By the explicit formula (8.3.4) for the solution of the Ornstein-Uhlenbeck SDE, we obtain:

Corollary 8.10 (Mehler formula). *A one-dimensional Ornstein-Uhlenbeck process* V_t^x *with initial condition* x *can be represented as*

$$V_t^x = e^{-\gamma t}(x + \sigma \widetilde{B}_{\frac{1}{2\gamma}(e^{2\gamma t}-1)})$$

with a Brownian motion $(\widetilde{B}_t)_{t \geq 0}$ *such that* $\widetilde{B}_0 = 0$.

Proof. The corresponding time change for the Wiener-Itô integral is given by

$$\tau(t) = \int_0^t \exp(2\gamma s)\, ds = (\exp(2\gamma t) - 1)/2\gamma.$$

□

8.4 Brownian bridge

In many circumstances one is interested in conditioning diffusion process on taking a given value at specified times. A basic example is the Brownian bridge which is Brownian motion conditioned to end at a given point x after time t_0. We now present several ways to describe and characterize Brownian bridges. The first is based on the Wiener-Lévy construction and specific to Brownian motion, the second extends to Gaussian processes, whereas the final characterization of the bridge process as the solution of a time-homogeneous SDE can be generalized to other diffusion processes. From now on, we consider a one-dimensional Brownian motion $(B_t)_{0\leq t\leq 1}$ with $B_0 = 0$ that we would like to condition on taking a given value y at time 1

Wiener-Lévy construction

Recall that the Brownian motion (B_t) has the Wiener-Lévy representation

$$B_t(\omega) \;=\; Y(\omega)t + \sum_{n=0}^{\infty}\sum_{k=0} 2^n - 1 Y_{n,k}(\omega) e_{n,k}(t) \qquad \text{for } t \in [0,1] \tag{8.4.1}$$

where $e_{n,k}$ are the Schauder functions, and Y and $Y_{n,k}$ $(n \geq 0, k = 0, 1, 2, \ldots, 2^n - 1)$ are independent and standard normally distributed. The series in (8.4.1) converges almost surely uniformly on $[0,1]$, and the approximating partial sums are piecewise linear approximations of B_t. The random variables $Y = B_1$ and

$$X_t \;:=\; \sum_{n=0}^{\infty}\sum_{k=0}^{2^n-1} Y_{n,k} e_{n,k}(t) \;=\; B_t - tB_1$$

are independent. This suggests that we can construct the bridge by replacing $Y(\omega)$ by the constant value y. Let

$$X_t^y \;:=\; yt + X_t \;=\; B_t + (y - B_1)\cdot t,$$

and let μ_y denote the distribution of the process $(X_t^y)_{0\leq t\leq 1}$ on $C([0,1])$. The next theorem shows that X_t^y is indeed a Brownian motion conditioned to end at y at time 1:

Theorem 8.11. *The map $y \mapsto \mu_y$ is a regular version of the conditional distribution of $(B_t)_{0\leq t\leq 1}$ given B_1, i.e.,* *(1).* μ_y *is a probability measure on* $C([0,1])$ *for any* $y \in \mathbb{R}$,

(2). $P[(B_t)_{0\le t\le 1} \in A \mid B_1] = \mu_{B_1}[A]$ *holds* P*-almost surely for any given Borel subset* $A \subseteq C([0,1])$.

(3). If $F : C([0,1]) \to \mathbb{R}$ *is a bounded and continuous function (w.r.t. the supremum norm on* $C([0,1])$*) then the map* $y \mapsto \int F\, d\mu_y$ *is continuous.*

The last statement says that $<\mapsto \mu_y$ is a continuous function w.r.t. the topology of weak convergence.

Proof. By definition, μ_y is a probability measure for any $y \in \mathbb{R}$. Moreover, for any Borel set $A \subseteq C([0,1])$,

$$\begin{aligned} P[(B_t)_{0\le t\le 1} \in A \mid B_1](\omega) &= P[(X_t + tB_1) \in A \mid B_1](\omega) \\ &= P[(X_t + tB_1(\omega)) \in A] \;=\; P[(X_t^{B_t(\omega)}) \in A] \;=\; \mu_{B_1(\omega)}[A] \end{aligned}$$

for P-almost every ω by independence of (X_T) and B_1. Finally, if $F : C([0,1]) \to \mathbb{R}$ is continuous and bounded then

$$\int F\, d\mu_y \;=\; E[F((y_t + X_t)_{0\le t\le 1})]$$

is continuous in y by dominated convergence. □

Finite-dimensional distributions

We now compute the marginals of the Brownian bridge X_t^y:

Corollary 8.12. *For any* $n \in \mathbb{N}$ *and* $0 < t_1 < \ldots < t_n < 1$*, the distribution of* $(X_{t_1}^y, \ldots, X_{t_n}^y)$ *on* $\mathbb{R}^n$ *is absolutely continuous with density*

$$f_y(x_1,\ldots,x_n) \;=\; \frac{p_{t_1}(0,x_1)p_{t_2-t_1}(x_1,x_2)\cdots p_{t_n-t_{n-1}}(x_{n-1},x_n)p_{1-t_n}(x_n,y)}{p_1(0,y)}. \quad (8.4.2)$$

Proof. The distribution of $(B_{t_1},\ldots,B_{t_n},B_1)$ is absolutely continuous with density

$$f_{B_{t_1},\ldots,B_{t_n},B_1}(x_1,\ldots,x_n,y) = p_{t_1}(0,x_0)p_{t_2-t_1}(x_1,x_2)\cdots p_{t_n-t_{n-1}}(x_{n-1},x_n)p_{1-t_n}(x_n,y).$$

Since the distribution of $(X^y_{t_1}, \dots, X^y_{t_n})$ is a regular version of the conditional distribution of $(B_{t_1}, \dots, B_{t_n})$ given B_1, it is absolutely continuous with the conditional density

$$\begin{aligned} f_{B_{t_1},\dots,B_{t_n}|B_1}(x_1,\dots,x_n|y) &= \frac{f_{B_{t_1},\dots,B_{t_n},B_1}(x_1,\dots,x_n,y)}{\int\cdots\int f_{B_{t_1},\dots,B_{t_n},B_1}(x_1,\dots,x_n,y)\,dx_1\cdots dx_n} \\ &= f_y(x_1,\dots,x_n). \end{aligned}$$

□

In general, any almost surely continuous process on $[0,1]$ with marginals given by (8.4.2) is called a **Brownian bridge from** 0 **to** y **in time** 1. A Brownian bridge from x to y in time t is defined correspondingly for any $x, y \in \mathbb{R}$ and any $t > 0$. In fact, this definition of the bridge process in terms of the marginal distributions carries over from Brownian motion to arbitrary Markov processes with strictly positive transition densities. In the case of the Brownian bridge, the marginals are again normally distributed:

Theorem 8.13 (**Brownian bridge as a Gaussian process**). *The Brownian bridge from* 0 *to* y *in time* 1 *is the (in distribution unique) continuous Gaussian process* $(X^y_t)_{t\in[0,1]}$ *with*

$$E[X^y_t] = ty \quad \textit{and} \quad \mathrm{Cov}[X^y_t, X^y_s] = t\wedge s - ts \qquad \textit{for any } s,t\in[0,1]. \tag{8.4.3}$$

Proof. A continuous Gaussian process is determined uniquely in distribution by its means and covariances. Therefore, it suffices to show that the bridge $X^y_t = B_t + (y - B_1)t$ defined above is a continuous Gaussian process such that (8.4.3) holds. This holds true: By (8.4.2), the marginals are normally distributed, and by definition, $t \mapsto X^y_t$ is almost surely continuous. Moreover,

$$\begin{aligned} E[X^y_t] &= E[B_t] + E[y - B_1]\cdot t \;=\; yt, \qquad \text{and} \\ \mathrm{Cov}[X^y_t, X^y_s] &= \mathrm{Cov}[B_t, B_s] - t\cdot\mathrm{Cov}[B_1, B_s] - s\cdot\mathrm{Cov}[B_t, B_1] + ts\,\mathrm{Var}[B_1] \\ &= t\wedge s - ts - st + ts \;=\; t\wedge s - ts. \end{aligned}$$

□

Remark (Covariance as Green function, Cameron-Martin space). The covariances of the Brownian bridge are given by

$$c(t,s) \;=\; \mathrm{Cov}[X_t^y, X_s^y] \;=\; \begin{cases} t\cdot(1-s) & \text{for } t\le s,\\ (1-t)\cdot s & \text{for } t\ge s.\end{cases}$$

The function $c(t,s)$ is the Green function of the operator d^2/dt^2 with Dirichlet boundary conditions on the interval $[0,1]$. This is related to the fact that the distribution of the Brownian bridge from 0 to 0 can be viewed as a standard normal distribution on the space of continuous paths $\omega : [0,1] \to \mathbb{R}$ with $\omega(0) = \omega(1) = 0$ w.r.t. the Cameron-Martin inner product

$$(g,h)_H \;=\; \int_0^1 g'(s)h'(s)\,ds.$$

The second derivative d^2/dt^2 is the linear operator associated with this quadratic from.

SDE for the Brownian bridge

Our construction of the Brownian bridge by an affine transformation of Brownian motion has two disadvantages:

- It can not be carried over to more general diffusion processes with possibly non-linear drift and diffusion coefficients.
- The bridge $X_t^y = B_t + t(y - B_1)$ does not depend on (B_t) in an adapted way, because the terminal value B_1 is required to define X_t^y for any $t > 0$.

We will now show how to construct a Brownian bridge from a Brownian motion in an adapted way. The idea is to consider an SDE w.r.t. the given Brownian motion with a drift term that forces the solution to end at a given point at time 1. The size of the drift term will be large if the process is still far away from the given terminal point at a time close to 1. For simplicity we consider a bridge (X_t) from 0 to 0 in time 1. Brownian bridges with other end points can be constructed similarly. Since the Brownian bridge is a Gaussian process, we may hope that there is a linear stochastic differential equation with additive noise that has a Brownian bridge as a solution. We therefore try the Ansatz

$$dX_t \;=\; -\beta_t X_t\,dt + dB_t, \qquad X_0 = 0 \tag{8.4.4}$$

with a given continuous deterministic function $\beta_t, 0 \le t < 1$. By variation of constants, the solution of (8.4.4) is the Gaussian process $X_t, 0 \le t < 1$, given by

$$X_t \;=\; \frac{1}{h_t}\int_0^t h_r\,dB_r \quad \text{where} \quad h_t \;=\; \exp\left(\int_0^t \beta_s\,ds\right).$$

The process (X_t) is centered and has covariances

$$\mathrm{Cov}[X_t, X_s] \;=\; \frac{1}{h_t h_s}\int_0^{t\wedge s} h_r^2\,dr.$$

Therefore, (X_t) is a Brownian bridge if and only if

$$\mathrm{Cov}[X_t, X_s] \;=\; t\cdot(1-s) \qquad \text{for any } t \le s,$$

i.e., if and only if

$$\frac{1}{th_t}\int_0^t h_r^2\,dr \;=\; h_s\cdot(1-s) \qquad \text{for any } 0 < t \le s. \tag{8.4.5}$$

The equation (8.4.5) holds if and only if h_t is a constant multiple of $1/1-t$, and in this case

$$\beta_t \;=\; \frac{d}{dt}\log h_t \;=\; \frac{h_t'}{h_t} \;=\; \frac{1}{1-t} \qquad \text{for } t \in [0,1].$$

Summarizing, we have shown:

Theorem 8.14. *If (B_t) is a Brownian motion then the process (X_t) defined by*

$$X_t \;=\; \int_0^t \frac{1-t}{1-r}\,dB_r \qquad \text{for } t \in [0,1], \quad X_1 = 0,$$

is a Brownian bridge from 0 to 0 in time 1. It is unique continuous process solving the SDE

$$dX_t \;=\; -\frac{X_t}{1-t}\,dt + \,dB_t \qquad \text{for } t \in [0,1). \tag{8.4.6}$$

Proof. As shown above, $(X_t)_{t\in[0,1)}$ is a continuous centered Gaussian process with the covariances of the Brownian bridge. Hence its distribution on $C([0,1))$ coincides with that of the Brownian bridge from 0 to 0. In particular, this implies $\lim_{t\nearrow 1} X_t = 0$ almost surely, so the trivial extension from $[0,1)$ to $[0,1]$ defined by $X_1 = 0$ is a Brownian bridge. □

If the Brownian bridge is replaced by a more general conditioned diffusion process, the Gaussian characterization does not apply. Nevertheless, it can still be shown by different means (the keyword is "h-transform") that the bridge process solves an SDE generalizing (8.4.6), cf. **??** below.

8.5 Stochastic differential equations in $\mathbb{R}^n$

We now explain how to generalize our considerations to systems of stochastic differential equations, or, equivalently, SDE in several dimensions. For the moment, we will not initiate a systematic study but rather consider some examples. Before, we extend the rules of Itô calculus to the multidimensional case. The setup is the following: We are given a d-dimensional Brownian motion $B_t = (B_t^1, \ldots, B_t^d)$. The component processes $B_t^k, 1 \leq k \leq d$, are independent one-dimensional Brownian motions that drive the stochastic dynamics. We are looking for a stochastic process $X_t : \Omega \to \mathbb{R}^n$ solving an SDE of the form

$$dX_t \;=\; b(t, X_t)\, dt + \sum_{k=1}^{d} \sigma_k(t, X_t)\, dB_t^k. \tag{8.5.1}$$

Here n and d may be different, and $b, \sigma_1, \ldots, \sigma_d : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ are time-dependent continuous vector fields on $\mathbb{R}^n$. In matrix notation,

$$dX_t \;=\; b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \tag{8.5.2}$$

where $\sigma(t,x) = (\sigma_1(t,x)\sigma_2(t,x)\cdots\sigma_d(t,x))$ is an $n \times d$-matrix.

Itô processes driven by several Brownian motions

Any solution to the SDE (8.5.1) is an Itô process pf type

$$X_t = \int_0^t G_s\, ds + \sum_{k=1}^{d} \int_0^t H_s^k\, dB_s^k \tag{8.5.3}$$

with continuous $(\mathcal{F}_t^{B,P})$ adapted stochastic processes $G_s, H_s^1, H_s^2, \ldots, H_s^d$. We now extend the stochastic calculus rules to such Itô processes that are driven by several independent Brownian motions. Let H_s and $\widetilde{H}_s$ be continuous $(\mathcal{F}_t^{B,P})$ adapted processes.

Lemma 8.15. *If (π_n) is a sequence of partitions of $\mathbb{R}_+$ with $\operatorname{mesh}(\pi_n) \to 0$ then for any $1 \le k, l \le d$ and $a \in \mathbb{R}_+$, the covariation of the Itô integrals $t \mapsto \int_0^t H_s\, dB_s^k$ and $t \mapsto \int_0^t \widetilde{H}_s\, dB_s^l$ exists almost surely uniformly for $t \in [0, a]$ along a subsequence of (π_n), and*

$$\left[\int_0^\bullet H\, dB^k, \int_0^\bullet \widetilde{H}\, dB^l\right]_t = \int_0^t H\widetilde{H}\, d[B^k, B^l] = \delta_{kl} \int_0^t H_s \widetilde{H}_s\, ds.$$

The proof is an extension of the proof of Theorem 8.1(ii), where the assertion has been derived for $k = l$ and $H = \widetilde{H}$. The details are left as an exercise.

Similarly to the one-dimensional case, the lemma can be used to compute the covariation of Itô integrals w.r.t. arbitrary Itô processes. If X_s and Y_s are Itô processes as in (8.5.1), and K_s and L_s are adapted and continuous then we obtain

$$\left[\int_0^\bullet K\, dX, \int_0^\bullet L\, dY\right]_t = \int_0^t K_s L_s\, d[X, Y]_s$$

almost surely uniformly for $t \in [0, u]$, along an appropriate subsequence of (π_n).

Multivariate Itô-Doeblin formula

We now assume again that $(X_t)_{t \ge 0}$ is a solution of a stochastic differential equation of the form (8.5.1). By Lemma 8.15, we can apply Itô's formula to almost every sample path $t \mapsto X_t(\omega)$:

Theorem 8.16 (Itô-Doeblin). *Let $F \in C^2(\mathbb{R}_+ \times \mathbb{R}^n)$. Then almost surely,*

$$\begin{aligned} F(t, X_t) &= F(0, X_0) + \int_0^t (\sigma^\top \nabla_x F)(s, X_s) \cdot dB_s \\ &\quad + \int_0^t \left(\frac{\partial F}{\partial t} + \mathscr{L}_t F \right)(s, X_s)\, ds \qquad \textit{for all } t \geq 0, \end{aligned}$$

where ∇_x denotes the gradient in the space variable, and

$$(\mathscr{L}_t F)(t, x) := \frac{1}{2} \sum_{i,j=1}^n a_{i,j}(t, x) \frac{\partial^2 F}{\partial x_i \partial x_j}(t, x) + \sum_{i=1}^n b_i(t, x) \frac{\partial F}{\partial x_i}(t, x)$$

with $a(t, x) := \sigma(t, x)\sigma(t, x)^\top \in \mathbb{R}^{n \times n}$.

The details of the proof are again left as an exercise. The Itô-Doeblin formula shows that for any $F \in C^2(\mathbb{R}_+ \times \mathbb{R}^n)$, the process

$$M_s^F = F(s, X_s) - F(0, X_0) - \int_0^s \left(\frac{\partial F}{\partial t} + \mathscr{L}_t F \right)(t, X_t)\, dt$$

is a local martingale. If $\sigma^\top \nabla_x F$ is bounded then M^F is a global martingale.

Exercise (Drift and diffusion coefficients). Show that the processes

$$M_s^i = X_s^i - X_0^i - \int_0^s b^i(s, X_s)\, ds, \qquad 1 \leq i \leq n,$$

are local martingales with covariations

$$[M^i, M^j]_s = a_{i,j}(s, X_s) \qquad \text{for any } s \geq 0,\ P\text{-almost surely.}$$

The vector field $b(s, x)$ is called the *drift vector field* of the SDE, and the coefficients $a_{i,j}(s, x)$ are called *diffusion coefficients*.

Index