A short introduction to stochastic differential equations

Tobias Jahnke

Karlsruher Institut für Technologie
Fakultät für Mathematik
Institut für Angewandte und Numerische Mathematik

tobias.jahnke@kit.edu

Karlsruhe Institute of Technology

Version: October 19, 2023

1 Motivation

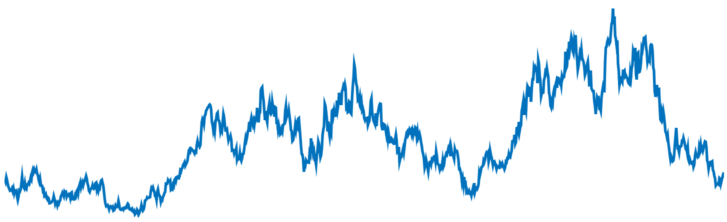
Financial markets trade investments into stocks of a company, commodities (e.g. oil, gold), bonds, or derivatives. Bonds evolve in a predictive way, but stocks and commodities do not. They are risky assets, because their value is affected by randomness. Our goal is to model the price of such a risky asset.
Many processes in science, technology, and engineering can be described very accurately with ordinary differential equations (ODEs)

$$\frac{dX}{dt} = f(t, X).$$

Here, $X = X(t)$ is the time-dependent solution, and $f = f(t, X)$ is a given function which depends on what is supposed to be modeled. But the solution of an ODE has no randomness – it is a purely deterministic object. If we know its value $X(t_0)$ at a given time t_0, then by solving the ODE we can compute its value $X(t)$ at any later time $t \geq 0$. This is certainly not true for stocks. In order to model risky assets, we have to put some randomness into the dynamics. A first and rather naïve approach to do so is to add a term which generates "random noise":

$$\underbrace{\frac{dX}{dt} = f(t, X)}_{\substack{\text{ordinary differential}\\ \text{equation}}} + \underbrace{g(t, X)Z(t)}_{\text{random noise}}. \tag{1}$$

The next questions are obviously how to choose $g(t, X)$, and how to define $Z(t)$ in a mathematically sound way. But even if these problems can be solved, Equation (1) is dubious. The solution of an ODE is, by definition, a *differentiable* function, whereas the chart of a stock typically looks like this:



It can be expected that such a function is continuous, but not differentiable. This raises the question if (1) makes sense at all. So what is the right way to define a *stochastic* differential equation?

The goal of these notes is to give an informal introduction to stochastic differential equation (SDEs) of Itô type. They are based on the *Itô integral*, which can be thought of as an extension of the classical Riemann-Stieltjes integral to cases where both the integrand and the integrator can be stochastic processes. Another important tool is the *Itô-Doeblin formula*, which is a stochastic counterpart of the classical chain rule.

The exposition in these notes is not mathematically rigorous. I have tried to be as precise as possible, but also to avoid technicalities whenever possible. Many proofs are omitted. For a comprehensive discussion of Itô SDEs the reader is referred to the books [Ste01, Shr04, Øks03, KP99, BK04].

2 Stochastic processes and filtrations

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space[1]: $\Omega \neq \emptyset$ is a set, $\mathcal{F}$ is a σ-algebra (or σ-field) on Ω, and $\mathbb{P} : \mathcal{F} \longrightarrow [0,1]$ is a probability measure. A probability space is complete if $\mathcal{F}$ contains all subsets G of Ω with $\mathbb{P}$-outer measure zero, i.e. with

$$\mathbb{P}^*(G) := \inf \big\{ \mathbb{P}(F) : F \in \mathcal{F} \text{ and } G \subset F \big\} = 0.$$

Any probability space can be completed. Hence, we can assume that every probability space in these notes is complete.

Definition 2.1 (Stochastic process) *Let $\mathcal{T}$ be an ordered set (e.g. $\mathcal{T} = [0, \infty)$ or $\mathcal{T} = \mathbb{N}$). A* ***stochastic process*** *is a family $X = \{X_t : t \in \mathcal{T}\}$ of random variables*

$$X_t : \Omega \longrightarrow \mathbb{R}^d.$$

Equivalent notations are $X(t,\omega)$, $X(t)$, $X_t(\omega)$, X_t, ... Below, we will often write X_t instead of $\{X_t : t \in \mathcal{T}\}$. For a fixed $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is called a realization (or path or trajectory) of X.

The path of a stochastic process is associated to some $\omega \in \Omega$. As time evolves, more information about ω becomes available.

Example (cf. chapter 2 in [Shr04]). If we toss a coin three times, then the possible results are:

ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8
HHH	HHT	HTH	HTT	THH	THT	TTH	TTT

(H = heads, T = tails).

- Before the first toss, we only know that $\omega \in \Omega = \{\omega_1, \dots, \omega_8\}$.
- After the first toss, we know if the final result will belong to

  $$\{HHH, HHT, HTH, HTT\} \text{ or to } \{THH, THT, TTH, TTT\}.$$

  These sets are "resolved by the information". Hence, we know in which of the sets

  $$\{w_1, w_2, w_3, w_4\}, \{w_5, w_6, w_7, w_8\}$$

  ω is.

[1] See "2.2.2 What is $(\Omega, \mathcal{F}, \mathbb{P})$ anyway?" in the book [CT04] for a nice discussion of this concept.

- After the second toss, the sets

$$\{HHH, HHT\}, \{HTH, HTT\}, \{THH, THT\}, \{TTH, TTT\}$$

are resolved, and we know in which of the sets

$$\{w_1, w_2\}, \{\omega_3, w_4\}, \{w_5, w_6\}, \{w_7, w_8\}$$

ω is.

This motivates the following definition.

Definition 2.2 (Filtration)

- *A **filtration** is a family $\{\mathcal{F}_t : t \geq 0\}$ of sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $t \geq s \geq 0$.*
Interpretation: A filtration models the fact that more and more information about the realization of a process is known as time evolves.
- *If $\{X_t : t \geq 0\}$ is a family of random variables and X_t is $\mathcal{F}_t$-measurable, then $\{X_t : t \geq 0\}$ is **adapted** to (or **nonanticipating** with respect to) $\{\mathcal{F}_t : t \geq 0\}$.*
Interpretation: At time t we know for each set $S \in \mathcal{F}_t$ if $\omega \in S$ or not. The value of X_t is revealed at time t.
- *For every $s \in [0, t]$ let $\sigma\{X_s\}$ be the σ-algebra generated by X_s, i.e. the collection of all sets*

$$X_s^{-1}(B) \text{ for all } B \in \mathcal{B}$$

where $\mathcal{B}$ denotes the Borel σ-algebra. By definition $\sigma\{X_s\}$ is the smallest σ-algebra where X_s is measurable.

3 The Wiener process

Robert Brown 1827, Louis Bachelier 1900, Albert Einstein 1905, Norbert Wiener 1923

Definition 3.1 (Normal distribution) *A random variable $X : \Omega \longrightarrow \mathbb{R}^d$ with $d \in \mathbb{N}$ is **normal** if it has a multivariate **normal (Gaussian) distribution** with mean $\mu \in \mathbb{R}^d$ and a symmetric, positive definite covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$, i.e.*

$$\mathbb{P}(X \in B) = \int_B \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\Big(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\Big)\, dx$$

for all Borel sets $B \subset \mathbb{R}^d$. Notation: $X \sim \mathcal{N}(\mu, \Sigma)$

Remarks:

1. If $X \sim \mathcal{N}(\mu, \Sigma)$, then $\mathbb{E}(X) = \mu$ and $\Sigma = (\sigma_{ij})$ with $\sigma_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$.
2. Standard normal distribution $\Leftrightarrow \mu = 0$, $\Sigma = I$ (identity matrix).
3. If $X \sim \mathcal{N}(\mu, \Sigma)$ and $Y = v + TX$ for some $v \in \mathbb{R}^d$ and a regular matrix $T \in \mathbb{R}^{d\times d}$, then
$$Y \sim \mathcal{N}(v + T\mu, T\Sigma T^T). \tag{2}$$
4. Warning: In one dimension, the covariance matrix is simply a number, namely the variance. Unfortunately, the variance is usually denoted by σ^2 instead of σ in the literature, which is somewhat confusing.

Definition 3.2 (Wiener process, Brownian motion)
(a) A stochastic process $\{W_t : t \in [0, T)\}$ is called a ***standard Brownian motion*** *or* ***standard Wiener process*** *if it has the following properties:*

1. *$W_0 = 0$ (with probability one)*
2. *Independent increments: For all $0 \le t_1 < t_2 < \ldots < t_n < T$ the random variables*
$$W_{t_2} - W_{t_1}, \quad W_{t_3} - W_{t_2}, \quad \ldots \quad , \quad W_{t_n} - W_{t_{n-1}}$$
are independent.
3. *$W_t - W_s \sim \mathcal{N}(0, t - s)$ for any $0 \le s < t < T$.*
4. *There is a $\tilde{\Omega} \subset \Omega$ with $\mathbb{P}(\tilde{\Omega}) = 1$ such that $t \mapsto W_t(\omega)$ is continuous for all $\omega \in \tilde{\Omega}$.*

(b) If $W_t^{(1)}, \ldots, W_t^{(d)}$ are independent one-dimensional Wiener processes, then $W_t = \left(W_t^{(1)}, \ldots, W_t^{(d)}\right)$ is called a d-dimensional Wiener process, and
$$W_t - W_s \sim \mathcal{N}(0, (t - s)I).$$

The existence of Brownian motion was first proved in a mathematically rigorous way by Norbert Wiener in 1923.
The Wiener process will serve as the "source of randomness" in our model of the financial market.

Notation: $W_t = W_t(\omega) = W(t, \omega) = W(t)$

Numerical simulation of a Wiener process (d=1). In order to get an idea of what a Wiener process is, we sketch how a realisation of a Wiener process can be simulated numerically. All one has to do is to choose a step-size $\tau > 0$, set $t_n = n\tau$ and $\tilde{W}_0 = 0$, and to repeat the following steps:

for $n = 0, 1, 2, 3, \ldots$
    Generate random a number $Z_n \sim \mathcal{N}(0, 1)$
    $\tilde{W}_{n+1} = \tilde{W}_n + \sqrt{\tau} Z_n$
end for

For $\tau \longrightarrow 0$ the interpolation of $\tilde{W}_0, \tilde{W}_1, \tilde{W}_2, \ldots$ approximates a path of the Wiener process ($\tilde{W}_N \approx W_{n\tau}$).

How smooth is a path of a Wiener process? For simplicity, we consider only the case $d = 1$.

Hölder continuity and non-differentiability

Definition 3.3 (Hölder continuity) *A function $f : (a, b) \longrightarrow \mathbb{R}$ is **Hölder continuous of order α** for some $\alpha \in [0, 1]$ if there is a constant C such that*

$$|f(t) - f(s)| \leq C|t - s|^{\alpha} \qquad \text{for all } s, t \in (a, b).$$

If $\alpha = 0$, then f is bounded.
If $\alpha > 0$, then f is uniformly continuous.
If $\alpha = 1$, then f is Lipschitz continuous.

A path of the Wiener process on a bounded interval is

- Hölder continuous of order $\alpha \in [0, \frac{1}{2})$ with probability one, but
- not Hölder continuous of order $\alpha \geq \frac{1}{2}$ with probability one.

A path of the Wiener process is nowhere differentiable with probability one.

Proofs: [Ste01], chapter 5

Unbounded total variation

Definition 3.4 (Total variation) *Let $[a, b]$ be an interval and let*

$$P = \{t_0, t_1, \ldots, t_{N(P)}\}, \qquad a = t_0 < t_1 < \ldots < t_{N(P)} = b, \qquad N(P) \in \mathbb{N}$$

*be a partition of this interval. Let $\mathcal{P}$ be the set of all such partitions. The **total variation** of a function $f : [a, b] \longrightarrow \mathbb{R}$ is*

$$TV_{a,b}(f) = \sup_{P \in \mathcal{P}} \sum_{n=1}^{N(P)} |f(t_n) - f(t_{n-1})|. \tag{3}$$

If f is differentiable and f' is integrable, then it can be shown that

$$TV_{a,b}(f) = \int_a^b |f'(t)| \, dt.$$

Conversely, if a function f has bounded total variation, then its derivative exists for almost all $t \in [a, b]$.
Consequence: A path of the Wiener process has unbounded total variation with probability one.

Quadratic variation

The quadratic variation of a function $f : (a, b) \longrightarrow \mathbb{R}$ is

$$QV_{a,b}(f) = \lim_{\substack{N \longrightarrow \infty \\ |P_N| \to 0}} \sum_{n=1}^{N} \big(f(t_n) - f(t_{n-1})\big)^2, \qquad |P_N| = \max_{n=1,\dots,N} |t_n - t_{n-1}|$$

If f is continuously differentiable, then one can show that

$$QV_{a,b}(f) = 0.$$

For a path $t \mapsto W_t(\omega)$ with $t \in [0, T]$, however, it can be shown that

$$\lim_{\substack{N \longrightarrow \infty \\ |P_N| \longrightarrow 0}} \left\| \sum_{n=1}^{N} \big(W_{t_n}(\omega) - W_{t_{n-1}}(\omega)\big)^2 - T \right\|_{L^2(d\mathbb{P})} = 0,$$

where
$$\|X\|_{L^2(d\mathbb{P})} = \sqrt{\mathbb{E}(X^2)} = \left(\int_{\omega \in \Omega} X^2(\omega)\, d\mathbb{P}(\omega) \right)^{\frac{1}{2}}.$$

By choosing a suitable subsequence, it can be concluded that $QV_{0,t}\big(t \mapsto W_t(\omega)\big) = t$ with probability one.

Filtration of the Wiener process

The natural filtration of the Wiener process on $[0, T]$ is given by

$$\{\mathcal{F}_t : t \in [0, T]\}, \qquad \mathcal{F}_t = \sigma\{W_s,\ s \in [0, t]\}$$

(cf. Definition 2.2). $\mathcal{F}_t$ contains all information (but not more) which can be obtained by observing W_s on the interval $[0, t]$. For technical reasons, however, it is more advantageous to use an **augmented** filtration called the **standard Brownian filtration**. See pp. 50-51 in [Ste01] for details.

4 Construction of the Itô integral (step 1 and 2)

References: [KP99, Øks03, Shr04, Ste01]

Now we return to the naïve ansatz (1) for defining an SDE. We assume that $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are sufficiently regular functions. Although we do not yet know how to choose the random noise $Z(t)$, we formally apply the explicit Euler method:
Choose $t \geq 0$ and $N \in \mathbb{N}$, let $\tau = t/N$, $t_n = n\tau$ and define approximations $X_n \approx X(t_n)$ by

$$X_{n+1} = X_n + \tau f(t_n, X_n) + \tau g(t_n, X_n) Z(t_n) \qquad (n = 0, 1, 2, \dots)$$

with initial data $X_0 = X(0)$. In the special case $f(t,X) = 0$, $g(t,X) = 1$ and $X(0) = 0$, we want that $X_n = W(t_n)$ is the Wiener process, i.e. we postulate that

$$W(t_{n+1}) \stackrel{!}{=} W(t_n) + \tau Z(t_n).$$

This yields

$$X_{n+1} = X_n + \tau f(t_n, X_n) + g(t_n, X_n)\Big(W(t_{n+1}) - W(t_n)\Big)$$

and after N steps

$$X_N = X_0 + \tau \sum_{n=0}^{N-1} f(t_n, X_n) + \sum_{n=0}^{N-1} g(t_n, X_n)\Big(W(t_{n+1}) - W(t_n)\Big). \tag{4}$$

Now we keep t fixed and let $N \longrightarrow \infty$, which means that $\tau = t/N \longrightarrow 0$. Then, (4) should somehow converge to

$$X(t) = X(0) + \int_0^t f(s, X(s))\, ds + \underbrace{\int_0^t g(s, X(s)) dW(s)}_{(\star)}. \tag{5}$$

Problem: We cannot define $(\star)$ as a pathwise Riemann-Stieltjes integral (cf. Appendix B). When $N \longrightarrow \infty$, the sum

$$\sum_{n=0}^{N-1} g(t_n, X_n(\omega))\Big(W(t_{n+1},\omega) - W(t_n,\omega)\Big)$$

diverges with probability one, because a path of the Wiener process has unbounded total variation with probability one.

New goal: Define the integral

$$\mathcal{I}_t[u](\omega) = \int_0^t u(s,\omega)\, dW_s(\omega)$$

in a "reasonable" way for the following class of functions.

Definition 4.1 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{\mathcal{F}_t : t \in [0,T]\}$ be the standard Brownian filtration. Then, we define $\mathcal{H}^2[0,T]$ to be the class of functions*

$$u = u(t,\omega), \qquad u : [0,T] \times \Omega \longrightarrow \mathbb{R}$$

with the following properties:

- $(t,\omega) \mapsto u(t,\omega)$ *is* $(\mathcal{B}\times\mathcal{F})$*-measurable.*
- u *is adapted to* $\{\mathcal{F}_t : t \in [0,T]\}$*, i.e.* $u(t,\cdot)$ *is* $\mathcal{F}_t$*-measurable.*
- $\mathbb{E}\left(\int_0^T u^2(t,\cdot)\,dt\right) < \infty$

Remark: $\mathcal{B}\times\mathcal{F}$ is the σ-algebra generated by all sets of the form $B \times F$ with $B \in \mathcal{B}$ and $F \in \mathcal{F}$. The product measure satisfies $(\mu \times \mathbb{P})(B \times F) = \mu(B)\mathbb{P}(F)$.

Step 1: Itô integral for elementary functions

Definition 4.2 (Elementary functions) *A function* $\phi \in \mathcal{H}^2[0,T]$ *is called* ***elementary*** *if it is a stochastic step function of the form*

$$\begin{aligned}\phi(t,\omega) &= a_0(\omega)\mathbf{1}_{[0,0]}(t) + \sum_{n=0}^{N-1} a_n(\omega)\mathbf{1}_{(t_n,t_{n+1}]}(t)\\ &= a_0(\omega)\mathbf{1}_{[0,t_1]}(t) + \sum_{n=1}^{N-1} a_n(\omega)\mathbf{1}_{(t_n,t_{n+1}]}(t)\end{aligned}$$

with a partition $0 = t_0 < t_1 < \ldots < t_{N-1} < t_N = T$*. The random variables* a_n *must be* $\mathcal{F}_{t_n}$*-measurable with* $\mathbb{E}(a_n^2) < \infty$*. Here and below,*

$$\mathbf{1}_{[c,d]}(t) = \begin{cases} 1 & \text{if } t \in [c,d] \\ 0 & \text{else} \end{cases} \tag{6}$$

is the indicator function of an interval $[c,d]$*.*

For $0 \le c < d \le T$, the only reasonable way to define the Itô integral of an indicator function $\mathbf{1}_{(c,d]}$ is

$$\mathcal{I}_T[\mathbf{1}_{(c,d]}](\omega) = \int_0^T \mathbf{1}_{(c,d]}(s)\,dW(s,\omega) = \int_c^d dW(s,\omega) = W(d,\omega) - W(c,\omega).$$

Hence, by linearity, we define the Itô integral of an elementary function by

$$\mathcal{I}_T[\phi](\omega) = \sum_{n=0}^{N-1} a_n(\omega)\big(W(t_{n+1},\omega) - W(t_n,\omega)\big).$$

Lemma 4.3 (Itô isometry for elementary functions) *For all elementary functions we have*

$$\mathbb{E}\left(\mathcal{I}_T[\phi]^2\right) = \mathbb{E}\left(\int_0^T \phi^2(t,\cdot)\,dt\right)$$

or equivalently

$$\|\mathcal{I}_T[\phi]\|_{L^2(d\mathbb{P})} = \|\phi\|_{L^2(dt\times d\mathbb{P})}$$

with

$$\|\phi\|_{L^2(dt\times d\mathbb{P})} = \left(\int_\Omega \int_0^T \phi^2(t,\omega)\ dt\ d\mathbb{P}\right)^{\frac{1}{2}} = \left(\mathbb{E}\left(\int_0^T \phi^2(t,\cdot)\ dt\right)\right)^{\frac{1}{2}}.$$

Proof. Since

$$\phi^2(t,\omega) = a_0^2(\omega)\mathbf{1}_{[0,0]}(t) + \sum_{n=0}^{N-1} a_n^2(\omega)\mathbf{1}_{(t_n,t_{n+1}]}(t)$$

we obtain

$$\mathbb{E}\left(\int_0^T \phi^2(t,\cdot)\ dt\right) = \sum_{n=0}^{N-1} \mathbb{E}\left(a_n^2\right)(t_{n+1}-t_n) \tag{7}$$

for the right-hand side. If we let $\Delta W_n = W(t_{n+1}) - W(t_n)$, then

$$\mathbb{E}\left(\mathcal{I}_T[\phi]^2\right) = \mathbb{E}\left(\left(\sum_{n=0}^{N-1} a_n \Delta W_n\right)^2\right) = \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \mathbb{E}(a_m a_n \Delta W_m \Delta W_n). \tag{8}$$

- If $n > m$, then $a_m a_n \Delta W_m$ is $\mathcal{F}_{t_n}$-measurable. ΔW_n is independent of $\mathcal{F}_{t_n}$ because the Wiener process has independent increments. Hence, $a_m a_n \Delta W_m$ and ΔW_n are independent. It follows that

$$\mathbb{E}\left(a_n a_m \Delta W_n \Delta W_m\right) = \mathbb{E}\left(a_n a_m \Delta W_m\right)\mathbb{E}\left(\Delta W_n\right) = 0 \qquad \text{for } n > m$$

  because $\Delta W_n \sim \mathcal{N}(0, t_{n+1}-t_n)$ by definition.
- By the same argument a_n^2 and ΔW_n^2 are independent. Hence, we obtain

$$\mathbb{E}\left(a_n^2 \Delta W_n^2\right) = \mathbb{E}\left(a_n^2\right)(t_{n+1}-t_n)$$

  because $\mathbb{E}(\Delta W_n^2) = \mathbb{V}(\Delta W_n) = t_{n+1} - t_n$.

Hence, (8) simplifies to

$$\mathbb{E}\left(\mathcal{I}_T[\phi]^2\right) = \sum_{n=0}^{N-1} \mathbb{E}\left(a_n^2\right)(t_{n+1}-t_n). \tag{9}$$

Comparing (7) and (9) yields the assertion. ■

Step 2: Itô integral on $\mathcal{H}^2[0,T]$

Lemma 4.4 *For any $u \in \mathcal{H}^2[0,T]$ there is a sequence $(\phi_k)_{k\in\mathbb{N}}$ of elementary functions $\phi_k \in \mathcal{H}^2[0,T]$ such that*

$$\lim_{k\to\infty} \|u - \phi_k\|_{L^2(dt\times d\mathbb{P})} = 0$$

Proof: Section 6.6 in [Ste01].

Let $u \in \mathcal{H}^2[0,T]$ and let $(\phi_k)_{k\in\mathbb{N}}$ be elementary functions such that

$$u = \lim_{k\to\infty} \phi_k \qquad \text{in } L^2(dt\times d\mathbb{P})$$

as in Lemma 4.4. The linearity of $\mathcal{I}_T[\cdot]$ and Lemma 4.3 yield

$$\|\mathcal{I}_T[\phi_j] - \mathcal{I}_T[\phi_k]\|_{L^2(d\mathbb{P})} = \|\mathcal{I}_T[\phi_j - \phi_k]\|_{L^2(d\mathbb{P})} = \|\phi_j - \phi_k\|_{L^2(dt\times d\mathbb{P})} \longrightarrow 0$$

for $j,k \longrightarrow \infty$. Hence, $\left(\mathcal{I}_T[\phi_k]\right)_k$ is a Cauchy sequence in the Hilbert space $L^2(d\mathbb{P})$. Thus, $\left(\mathcal{I}_T[\phi_k]\right)_k$ converges in $L^2(d\mathbb{P})$, and we can define

$$\mathcal{I}_T[u] = \lim_{k\to\infty} \mathcal{I}_T[\phi_k].$$

The choice of the sequence does not matter: If $(\psi_k)_{k\in\mathbb{N}}$ is another sequence of elementary functions with $u = \lim_{k\to\infty}\psi_k$ in $L^2(dt\times d\mathbb{P})$, then by Lemma 4.3 we obtain for $k \longrightarrow \infty$

$$\begin{aligned}
\|\mathcal{I}_T[\phi_k] - \mathcal{I}_T[\psi_k]\|_{L^2(d\mathbb{P})} &= \|\mathcal{I}_T[\phi_k - \psi_k]\|_{L^2(d\mathbb{P})} \\
&= \|\phi_k - \psi_k\|_{L^2(dt\times d\mathbb{P})} \\
&\le \|\phi_k - u\|_{L^2(dt\times d\mathbb{P})} + \|u - \psi_k\|_{L^2(dt\times d\mathbb{P})} \longrightarrow 0.
\end{aligned}$$

Theorem 4.5 (Itô isometry) *For all $u \in \mathcal{H}^2[0,T]$ we have*

$$\|\mathcal{I}_T[u]\|_{L^2(d\mathbb{P})} = \|u\|_{L^2(dt\times d\mathbb{P})}.$$

Proof: Let $(\phi_k)_{k\in\mathbb{N}}$ again be elementary functions such that $u = \lim_{k\to\infty}\phi_k$ in $L^2(dt\times d\mathbb{P})$; cf. Lemma 4.4. Then

$$\lim_{k\to\infty} \|\phi_k\|_{L^2(dt\times d\mathbb{P})} = \|u\|_{L^2(dt\times d\mathbb{P})},$$

because the reverse triangle inequality yields

$$\left| \|\phi_k\|_{L^2(dt\times d\mathbb{P})} - \|u\|_{L^2(dt\times d\mathbb{P})} \right| \le \|\phi_k - u\|_{L^2(dt\times d\mathbb{P})} \to 0.$$

By the same argument, we obtain

$$\lim_{k\to\infty} \|\mathcal{I}_T[\phi_k]\|_{L^2(d\mathbb{P})} = \|\mathcal{I}_T[u]\|_{L^2(d\mathbb{P})}$$

Now the assertion follows from Lemma 4.3 by taking the limit. ∎

5 Martingales

Definition 5.1 (conditional expectation) *Let X be an integrable random variable, and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then, Y is a* ***conditional expectation*** *of X with respect to $\mathcal{G}$ if Y is $\mathcal{G}$-measurable and if*

$$\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(Y\mathbf{1}_A) \qquad \text{for all } A \in \mathcal{G},$$

$$\Leftrightarrow \qquad \int_A X(\omega)\, d\mathbb{P}(\omega) = \int_A Y(\omega)\, d\mathbb{P}(\omega) \qquad \text{for all } A \in \mathcal{G}.$$

In this case, we write $Y = \mathbb{E}(X \mid \mathcal{G})$.

"This definition is not easy to love. Fortunately, love is not required."
J.M. Steele in [Ste01], p. 45.

Interpretation. $\mathbb{E}(X \mid \mathcal{G})$ is a random variable on $(\Omega, \mathcal{G}, \mathbb{P})$ and hence on $(\Omega, \mathcal{F}, \mathbb{P})$, too. Roughly speaking, $\mathbb{E}(X \mid \mathcal{G})$ is the best approximation of X detectable by the events in $\mathcal{G}$. The more $\mathcal{G}$ is refined, the better $\mathbb{E}(X \mid \mathcal{G})$ approximates X.

Examples.

1. If $\mathcal{G} = \{\Omega, \emptyset\}$, then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X) = \int_\Omega X(\omega)\, d\mathbb{P}(\omega)$.
2. If $\mathcal{G} = \mathcal{F}$, then $\mathbb{E}(X \mid \mathcal{G}) = X$.
3. If $F \in \mathcal{F}$ with $\mathbb{P}(F) > 0$ and

$$\mathcal{G} = \{\emptyset, F, \Omega \setminus F, \Omega\}$$

then it can be shown that

$$\mathbb{E}(X \mid \mathcal{G})(\omega) = \begin{cases} \dfrac{1}{\mathbb{P}(F)} \displaystyle\int_F X d\mathbb{P} & \text{if } \omega \in F \\ \dfrac{1}{\mathbb{P}(\Omega \setminus F)} \displaystyle\int_{\Omega \setminus F} X d\mathbb{P} & \text{if } \omega \in \Omega \setminus F. \end{cases}$$

4. If X is independent of $\mathcal{G}$, then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X)$.

Proof: Exercise.

Lemma 5.2 (Properties of the conditional expectation) *For all integrable random variables X and Y and all sub-σ-algebras $\mathcal{G} \subset \mathcal{F}$, the conditional expectation has the following properties:*

- *Linearity:* $\mathbb{E}(X + Y \mid \mathcal{G}) = \mathbb{E}(X \mid \mathcal{G}) + \mathbb{E}(Y \mid \mathcal{G})$
- *Positivity: If $X \geq 0$, then $\mathbb{E}(X \mid \mathcal{G}) \geq 0$.*
- *Tower property: If $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ are sub-σ-algebras, then*

$$\mathbb{E}\Big(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}\Big) = \mathbb{E}(X \mid \mathcal{H})$$

- $\mathbb{E}\Big(\mathbb{E}(X \mid \mathcal{G})\Big) = \mathbb{E}(X)$
- *Factorization property: If Y is $\mathcal{G}$-measurable and $|XY|$ and $|Y|$ are integrable, then*

$$\mathbb{E}(XY \mid \mathcal{G}) = Y\mathbb{E}(X \mid \mathcal{G})$$

Proof: Exercise.

Definition 5.3 (martingale) *Let X_t be a stochastic process which is adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$ of $\mathcal{F}$. If*

1. $\mathbb{E}(|X_t|) < \infty$ *for all* $0 \leq t < \infty$, *and*
2. $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$ *for all* $0 \leq s \leq t < \infty$,

*then X_t is called a **martingale**. A martingale X_t is called continuous if there is a set $\Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0) = 1$ such that the path $t \mapsto X_t(\omega)$ is continuous for all $\omega \in \Omega_0$.*

Interpretation: A martingale models a fair game. Observing the game up to time s does not give any advantage for future times.

Examples. It can be shown that each of the following processes is a continuous martingale with respect to the standard Brownian filtration:

$$W_t, \qquad W_t^2 - t, \qquad \exp\left(\alpha W_t - \frac{\alpha^2}{2}t\right) \quad \text{with } \alpha \in \mathbb{R}$$

Proof: Exercise.

6 Construction of the Itô integral (step 3 and 4)

Step 3: The Itô integral as a process

So far we have defined the Itô integral $\mathcal{I}_T[u](\omega)$ over the interval $[0,T]$ for **fixed** T. For applications in mathematical finance, however, we want to consider $\big\{\mathcal{I}_t[u](\omega) : t \in [0,T]\big\}$ as a stochastic **process**.

If $u(s,\omega) \in \mathcal{H}^2[0,T]$, then $\mathbf{1}_{[0,t]}(s)u(s,\omega) \in \mathcal{H}^2[0,T]$. Can we define $\mathcal{I}_t[u](\omega)$ by $\mathcal{I}_T[\mathbf{1}_{[0,t]}u](\omega)$?

Problem: The integral $\mathcal{I}_T[\mathbf{1}_{[0,t]}u](\omega)$ is only defined in $L^2(d\mathbb{P})$. Hence, the value $\mathcal{I}_T[\mathbf{1}_{[0,t]}u](\omega)$ is arbitrary on sets $Z \in \mathcal{Z}_t := \{Z \in \mathcal{F}_t : \mathbb{P}(Z) = 0\}$. Since the set $[0,T]$ is uncountable[2], the union

$$\bigcup_{t\in[0,T]} \mathcal{Z}_t$$

(i.e. the set where the process is not well-defined) could be "very large"! Fortunately, this can be fixed:

[2] We only know that countable unions of null sets have measure zero, but this is not true for uncountabel unions.

Theorem 6.1 *For any $u \in \mathcal{H}^2[0,T]$ there is a process $\{X_t : t \in [0,T]\}$ that is a continuous martingale with respect to the standard Brownian filtration $\mathcal{F}_t$ such that*

$$\mathbb{P}\Big(\big\{\omega \in \Omega : X_t(\omega) = \mathcal{I}_T[\mathbf{1}_{[0,t]}u](\omega)\big\}\Big) = 1$$

for each $t \in [0,T]$.

A proof can be found in [Ste01], Theorem 6.2, pages 83-84.

Step 4: The Itô integral on $\mathcal{L}^2_{loc}[0,T]$

So far we have defined the Itô integral for functions $u \in \mathcal{H}^2[0,T]$; cf. Definition 4.1. Such functions must satisfy

$$\mathbb{E}\left(\int_0^T u^2(t,\cdot)\, dt\right) < \infty, \tag{10}$$

and this condition is sometimes too restrictive.
Example: If $y(x) = \exp(x^4)$, then $u(t,\omega) = y(W_t(\omega)) \notin \mathcal{H}^2[0,T]$, because

$$\begin{aligned}\mathbb{E}\left(\int_0^T u^2(t,\cdot)\, dt\right) &= \int_\Omega \int_0^T \exp(2W_t^4(\omega))\, dt d\omega \\ &= \int_{-\infty}^{\infty} \int_0^T \exp(2x^4) \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right)\, dt\, dx = \infty\end{aligned}$$

since the first exp term is stronger than the second.

With some more work, the Itô integral can be extended to the class $\mathcal{L}^2_{loc}[0,T]$, i.e. to all functions

$$u = u(t,\omega), \qquad u : [0,T] \times \Omega \longrightarrow \mathbb{R}$$

with the following properties:

- $(t,\omega) \mapsto u(t,\omega)$ is $(\mathcal{B} \times \mathcal{F})$-measurable.
- u is adapted to $\{\mathcal{F}_t : t \in [0,T]\}$.
- $\mathbb{P}\left(\int_0^T u^2(t,\omega)\, dt < \infty\right) = 1.$

The first two conditions are the same as for $\mathcal{H}^2[0,T]$, but the third condition is weaker than (10). If $y : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous, then $u(t,\omega) = y\big(W(t,\omega)\big) \in \mathcal{L}^2_{loc}[0,T]$, because $t \mapsto y\big(W(t,\omega)\big)$ is continuous with probability one and hence bounded on $[0,T]$ with probability one.

Details: Chapter 7 in [Ste01].

Notation

The process X constructed above is called the **Itô integral** (Itô Kiyoshi 1944) of $u \in \mathcal{L}^2_{loc}[0,T]$ and is denoted by

$$X(t,\omega) = \int_0^t u(s,\omega)\, dW(s,\omega).$$

The Itô integral over an arbitrary interval $[a,b] \subset [0,T]$ is defined by

$$\int_a^b u(s,\omega)\, dW(s,\omega) = \int_0^b u(s,\omega)\, dW(s,\omega) - \int_0^a u(s,\omega)\, dW(s,\omega).$$

Alternative notations:

$$\int_a^b u(s,\omega)\, dW(s,\omega) = \int_a^b u(s,\omega)\, dW_s(\omega) = \int_a^b u_s(\omega)\, dW_s(\omega) = \int_a^b u_s\, dW_s$$

Properties of the Itô integral

Lemma 6.2 *Let $c \in \mathbb{R}$ and $u, v \in \mathcal{L}^2_{loc}[0,T]$. The Itô integral on $[a,b] \subset [0,T]$ has the following properties:*

1. *Linearity:*

$$\int_a^b \Big(cu(s,\omega) + v(s,\omega)\Big)\, dW_s(\omega) = c\int_a^b u(s,\omega)\, dW_s(\omega) + \int_a^b v(s,\omega)\, dW_s(\omega)$$

   with probability one.

2. $\mathbb{E}\left(\int_a^b u(s,\cdot)\, dW_s\right) = 0$

3. $\int_a^t u(s,\omega)\, dW_s(\omega)$ *is $\mathcal{F}_t$-measurable for $t \geq a$.*

4. *Itô isometry on $[a,b]$:*

$$\mathbb{E}\left(\left(\int_a^b u(s,\cdot)\, dW_s\right)^2\right) = \mathbb{E}\left(\int_a^b u^2(s,\cdot)\, ds\right)$$

   (cf. Theorem 4.5).

5. *Martingale property: The Itô integral*

$$X(t,\omega) = \int_0^t u(s,\omega)\, dW(s,\omega).$$

of a function $u \in \mathcal{H}^2[0,T]$ *is a continuous martingale with respect to the standard Brownian filtration; cf. Theorem 6.1. If* $u \in \mathcal{L}^2_{loc}[0,T]$*, then the Itô integral is only a local martingale; cf. Proposition 7.7 in [Ste01].*

The first four properties can be shown by considering elementary functions and passing to the limit.

7 Stochastic differential equations and the Itô-Doeblin formula

Definition 7.1 (SDE) *A* ***stochastic differential equation (SDE)*** *is an equation of the form*

$$X(t) = X(0) + \int_0^t f\big(s, X(s)\big)\, ds + \int_0^t g\big(s, X(s)\big)\, dW(s). \tag{11}$$

The solution $X(t)$ *of (11) is called an* ***Itô process****.*

The last term is an Itô integral, with $W(t)$ denoting the Wiener process. The functions $f : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$ and $g : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$ are called drift and diffusion coefficients, respectively. These functions are typically given while $X(t) = X(t,\omega)$ is unknown.
This equation is actually not a **differential** equation, but an **integral** equation! Often people write

$$dX_t = f(t, X_t)dt + g(t, X_t)dW_t$$

as a shorthand notation for (11). Some people even "divide by dt" in order to make the equation look like a differential equation, but this is more than audacious since "dW_t/dt" does not make sense.

Two special cases:

- If $g\big(t, X(t)\big) \equiv 0$, then (11) is reduced to

$$X(t) = X(0) + \int_0^t f\big(s, X(s)\big)\, ds.$$

If $X(t)$ is differentiable, this is equivalent to the ordinary differential equation

$$\frac{dX(t)}{dt} = f\big(t, X(t)\big)$$

with initial data $X(0)$.

- For $f\big(t, X(t)\big) \equiv 0$, $g\big(t, X(t)\big) \equiv 1$ and $X(0) = 0$, (11) turns into

$$X(t) = \underbrace{X(0)}_{=0} + \underbrace{\int_0^t f\big(s, X(s)\big)\, ds}_{=0} + \int_0^t \underbrace{g\big(s, X(s)\big)}_{=1}\, dW(s) = W(t) - W(0) = W(t).$$

Computing Riemann integrals via the basic definition is usually very tedious. The fundamental theorem of calculus provides an alternative which is more convenient in most cases. For Itô integrals, the situation is similar: The approximation via elementary functions which is used to *define* the Itô integral is rarely used to *compute* the integral. What is the counterpart of the fundamental theorem of calculus for the Itô integral?

Theorem 7.2 (Itô-Doeblin formula) *Let X_t be the solution of the SDE*

$$dX_t = f(t, X_t)dt + g(t, X_t)dW_t$$

and let $F(t,x)$ be a function with continuous partial derivatives $\partial_t F = \frac{\partial F}{\partial t}$, $\partial_x F = \frac{\partial F}{\partial x}$, and $\partial_x^2 F = \frac{\partial^2 F}{\partial x^2}$. Then, we have for $Y_t := F(t, X_t)$ that

$$\begin{aligned} dY_t &= \partial_t F\, dt + \partial_x F\, dX_t + \frac{1}{2}(\partial_x^2 F) g^2\, dt \\ &= \left(\partial_t F + (\partial_x F) f + \frac{1}{2}(\partial_x^2 F) g^2\right) dt + (\partial_x F) g\, dW_t. \end{aligned} \tag{12}$$

with $f = f(t, X_t)$, $g = g(t, X_t)$, $\partial_x F = \partial_x F(t, X_t)$, and so on.

Notation. Evaluations of the derivatives of F are to be understood in the sense of, e.g.,

$$\partial_x F(s, X_s) := \partial_x F(t, x)\big|_{(t,x)=(s,X_s)}$$

and so on.

Remarks:

1. If $y(t)$ is a smooth deterministic fuctions, then according to the chain rule the derivative of $t \mapsto F\big(t, y(t)\big)$ is
$$\frac{d}{dt} F\big(t, y(t)\big) = \partial_t F\big(t, y(t)\big) + \partial_x F\big(t, y(t)\big) \cdot \frac{dy(t)}{dt}$$
and in shorthand notation
$$dF = \partial_t F dt + \partial_x F dy.$$
The Itô-Doeblin formula can be considered as a stochastic version of the chain rule, but the term $\frac{1}{2}(\partial_x^2 F) \cdot g^2 dt$ is surprising since such a term does not appear in the deterministic chain rule.

2. Let $f(t, X_t) = 0$, $g(t, X_t) = 1$, $X_t = W_t$ and suppose that $F(t, x) = F(x)$ does not depend on t. Then, the Itô-Doeblin formula yields for $Y_t := F(W_t)$ that
$$dY_t = F'(W_t) dW_t + \frac{1}{2} F''(W_t) dt$$
which is the shorthand notation for
$$F(W_t) = F(W_0) + \int_0^t F'(W_s) dW_s + \frac{1}{2} \int_0^t F''(W_s) ds.$$
This can be seen as a counterpart of the fundamental theorem of calculus. Again, the last term is surprising, because for a suitable deterministic function $v(t) = v_t$ we obtain
$$F(v_t) = F(v_0) + \int_0^t F'(v_s) dv_s.$$

Sketch of the proof of Theorem 7.2.

(i) Equation (12) is the shorthand notation for
$$Y_t = Y_0 + \int_0^t \Big(\partial_t F(s, X_s) + \partial_x F(s, X_s) \cdot f(s, X_s) + \frac{1}{2} \partial_x^2 F(s, X_s) \cdot g^2(s, X_s) \Big) ds$$
$$+ \int_0^t \partial_x F(s, X_s) \cdot g(s, X_s) dW_s$$

Assume that F is twice continuously differentiable with bounded partial derivatives. (Otherwise F can be approximated by such functions with uniform convergence on compact subsets of $[0,\infty) \times \mathbb{R}$.) Moreover, assume that $(t,\omega) \mapsto f(t, X_t(\omega))$ and $(t,\omega) \mapsto g(t, X_t(\omega))$ are elementary functions. (Otherwise approximate by elementary functions.) Hence, there is a partition $0 = t_0 < t_1 < \ldots < t_N = t$ such that

$$f(t, X_t(\omega)) = f(0, X_0(\omega))\mathbf{1}_{[0,t_1]}(t) + \sum_{n=1}^{N-1} f(t_n, X_{t_n}(\omega))\mathbf{1}_{(t_n,t_{n+1}]}(t)$$

and the same equation with f replaced by g.

(ii) For the rest of the proof, we will use the short-hand notation

$$f^{(n)} := f(t_n, X_{t_n}), \qquad F^{(n)} := F(t_n, X_{t_n}),$$
$$g^{(n)} := g(t_n, X_{t_n}), \qquad \partial_t F^{(n)} := \partial_t F(t_n, X_{t_n})$$

and so on, and

$$\Delta t_n = t_{n+1} - t_n, \qquad \Delta X_n = X_{t_{n+1}} - X_{t_n}, \qquad \Delta W_n = W_{t_{n+1}} - W_{t_n}.$$

Since f and g are elementary functions, we have

$$\begin{aligned} X_{t_n} &= X_0 + \int_0^{t_n} f(s, X_s)\,ds + \int_0^{t_n} g(s, X_s)\,dW_s \\ &= X_0 + \sum_{k=0}^{n-1} \underbrace{f(t_k, X_{t_k})}_{f^{(k)}} \Delta t_k + \sum_{k=0}^{n-1} \underbrace{g(t_k, X_{t_k})}_{g^{(k)}} \Delta W_k. \end{aligned}$$

and hence

$$\Delta X_n = X_{t_{n+1}} - X_{t_n} = f^{(n)}\Delta t_n + g^{(n)}\Delta W_n.$$

(iii) Now Y_t can be expressed by the telescoping sum

$$Y_t = Y_{t_N} = Y_0 + \sum_{n=0}^{N-1} (Y_{t_{n+1}} - Y_{t_n}) = Y_0 + \sum_{n=0}^{N-1} \big(F^{(n+1)} - F^{(n)}\big).$$

Applying Taylor's theorem yields

$$\begin{aligned} &F^{(n+1)} - F^{(n)} \\ &= \partial_t F^{(n)} \cdot \Delta t_n + \partial_x F^{(n)} \cdot \Delta X_n + \frac{1}{2}\partial_t^2 F^{(n)} \cdot (\Delta t_n)^2 + \partial_t \partial_x F^{(n)} \cdot \Delta t_n \Delta X_n \\ &\quad + \frac{1}{2}\partial_x^2 F^{(n)} \cdot (\Delta X_n)^2 + R_n(\Delta t_n, \Delta X_n) \end{aligned}$$

with a remainder term R_n. This identity is inserted into the telescoping sum.

(iv) Consider the limit $N \longrightarrow \infty$, $\Delta t_n \longrightarrow 0$ with respect to $\|\cdot\|_{L^2(d\mathbb{P})}$. For the first two terms, this yields

$$\lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_t F^{(n)} \cdot \Delta t_n = \lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_t F(t_n, X_{t_n}) \cdot \Delta t_n = \int_0^t \partial_t F(s, X_s)\, ds$$

and

$$\begin{aligned} &\lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_x F^{(n)} \cdot \Delta X_n \\ &= \lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_x F^{(n)} \cdot f^{(n)} \Delta t_n + \lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_x F^{(n)} \cdot g^{(n)} \Delta W_n \\ &= \int_0^t \partial_x F(s, X_s) \cdot f(s, X_s)\, ds + \int_0^t \partial_x F(s, X_s) \cdot g(s, X_s)\, dW_s. \end{aligned}$$

(v) Next, we investigate the "$\partial_x^2 F^{(n)}$ term". Since

$$(\Delta X_n)^2 = \left(f^{(n)} \Delta t_n + g^{(n)} \Delta W_n \right)^2$$

we have

$$\frac{1}{2} \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot (\Delta X_n)^2 = \frac{1}{2} \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot \left(f^{(n)}\right)^2 (\Delta t_n)^2 \tag{13}$$

$$+ \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot f^{(n)} g^{(n)} \Delta t_n \Delta W_n \tag{14}$$

$$+ \frac{1}{2} \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot \left(g^{(n)}\right)^2 (\Delta W_n)^2. \tag{15}$$

For the right-hand side of (13), we obtain

$$\left\| \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot \left(f^{(n)}\right)^2 (\Delta t_n)^2 \right\|_{L^2(d\mathbb{P})}^2 = \mathbb{E}\left[\left(\sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot \left(f^{(n)}\right)^2 (\Delta t_n)^2 \right)^2 \right] \longrightarrow 0.$$

With the abbreviation $\alpha^{(n)} := \partial_x^2 F^{(n)} \cdot f^{(n)} g^{(n)}$ we obtain for the right-hand side of (14) that

$$\begin{aligned} \left\| \sum_{n=0}^{N-1} \alpha^{(n)} \Delta t_n \Delta W_n \right\|_{L^2(d\mathbb{P})}^2 &= \mathbb{E}\left[\left(\sum_{n=0}^{N-1} \alpha^{(n)} \Delta t_n \Delta W_n \right)^2 \right] \\ &= \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \mathbb{E}\left(\alpha^{(n)} \alpha^{(m)} \Delta W_n \Delta W_m \right) \Delta t_n \Delta t_m. \end{aligned}$$

For $n < m$ we have

$$\mathbb{E}\left(\alpha^{(n)}\alpha^{(m)}\Delta W_n \Delta W_m\right) = \mathbb{E}\left(\alpha^{(n)}\alpha^{(m)}\Delta W_n\right) \underbrace{\mathbb{E}\left(\Delta W_m\right)}_{=0} = 0$$

and similar for $m < n$. Hence, only the terms with $n = m$ have to be considered, which yields

$$\left\| \sum_{n=0}^{N-1} \alpha^{(n)} \Delta t_n \Delta W_n \right\|_{L^2(d\mathbb{P})}^2 = \sum_{n=0}^{N-1} \mathbb{E}\left((\alpha^{(n)})^2\right) (\Delta t_n)^2 \underbrace{\mathbb{E}\left[(\Delta W_n)^2\right]}_{=\Delta t_n} \longrightarrow 0.$$

The third term (15), however, has a non-zero limit: We show that

$$\lim_{N\to\infty} \frac{1}{2} \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot \left(g^{(n)}\right)^2 (\Delta W_n)^2 = \frac{1}{2} \int_0^t \partial_x^2 F(s, X_s) \cdot \left(g(s, X_s)\right)^2 ds$$

which yields the strange additional term in the Itô-Doeblin formula. With the abbreviation $\beta^{(n)} = \frac{1}{2}\partial_x^2 F^{(n)} \cdot \left(g^{(n)}\right)^2$ we have

$$\begin{aligned}
&\left\| \sum_{n=0}^{N-1} \beta^{(n)} \left((\Delta W_n)^2 - \Delta t_n\right) \right\|_{L^2(d\mathbb{P})}^2 \\
&= \mathbb{E}\left[\left(\sum_{n=0}^{N-1} \beta^{(n)} \left((\Delta W_n)^2 - \Delta t_n\right) \right)^2 \right] \\
&= \mathbb{E}\left[\sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \beta^{(n)}\beta^{(m)} \left((\Delta W_n)^2 - \Delta t_n\right) \left((\Delta W_m)^2 - \Delta t_m\right) \right].
\end{aligned}$$

For $n < m$ we have

$$\begin{aligned}
&\mathbb{E}\left[\beta^{(n)}\beta^{(m)} \left((\Delta W_n)^2 - \Delta t_n\right) \left((\Delta W_m)^2 - \Delta t_m\right)\right] \\
&= \mathbb{E}\left[\beta^{(n)}\beta^{(m)} \left((\Delta W_n)^2 - \Delta t_n\right)\right] \underbrace{\mathbb{E}\left[\left((\Delta W_m)^2 - \Delta t_m\right)\right]}_{=0} = 0
\end{aligned}$$

and vice versa for $n > m$. Hence, only the terms with $n = m$ have to be considered, and we obtain

$$\begin{aligned}
\left\| \sum_{n=0}^{N-1} \beta^{(n)} \left((\Delta W_n)^2 - \Delta t_n\right) \right\|_{L^2(d\mathbb{P})}^2 &= \mathbb{E}\left[\sum_{n=0}^{N-1} \left(\beta^{(n)}\right)^2 \left((\Delta W_n)^2 - \Delta t_n\right)^2 \right] \\
&= \sum_{n=0}^{N-1} \mathbb{E}\left[\left(\beta^{(n)}\right)^2\right] \mathbb{E}\left[\left((\Delta W_n)^2 - \Delta t_n\right)^2\right] \to 0
\end{aligned}$$

because it can be shown that $\mathbb{E}\left[\left((\Delta W_n)^2 - \Delta t_n\right)^2\right] = 2\Delta t_n^2$.

(vi) With essentially the same arguments, it can be shown that

$$\lim_{N\to\infty} \frac{1}{2}\sum_{n=0}^{N-1} \partial_t^2 F^{(n)} \cdot (\Delta t_n)^2 = 0$$

$$\lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_t\partial_x F^{(n)} \cdot \Delta t_n \Delta X_n = 0$$

and that the remainder term from the Taylor expansion can be neglected when the limit is taken. ■

Example 1. Consider the integral

$$\int_0^t W_s \, dW_s.$$

$X_t := W_t$ solves the SDE with $f(t, X_t) \equiv 0$ and $g(t, X_t) \equiv 1$. For

$$F(t, x) = x^2, \qquad Y_t = F(t, X_t) = X_t^2 = W_t^2$$

the Itô-Doeblin formula

$$dY_t = \left(\partial_t F + (\partial_x F)f + \frac{1}{2}(\partial_x^2 F)g^2\right) dt + (\partial_x F)g \, dW_t$$

yields

$$d(W_t^2) = 0 + 0 + \frac{1}{2} \cdot 2 \cdot 1^2 \, dt + 2W_t \cdot 1 \, dW_t = dt + 2W_t \, dW_t$$

$$\Longrightarrow \qquad W_t \, dW_t = \frac{1}{2}\big(d(W_t^2) - dt\big)$$

This means that

$$\int_0^t W_s \, dW_s = \frac{1}{2}\int_0^t 1 \, d(W_s^2) - \frac{1}{2}\int_0^t 1 \, ds = \frac{1}{2}W_t^2 - \frac{1}{2}t.$$

Example 2. The solution of the SDE

$$dY_t = \mu Y_t \, dt + \sigma Y_t \, dW_t$$

with constants $\mu, \sigma \in \mathbb{R}$ and deterministic initial value $Y_0 \in \mathbb{R}$ is given by

$$Y_t = \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right) Y_0.$$

This process is called a **geometric Brownian motion** and is often used in mathematical finance to model stock prices (see below).

The proof is left as an exercise.

Ordinary differential equations can have multiple solutions with the same initial value, and solutions do not necessarily exist for all times. Hence, we cannot expect that every SDE has a unique solution. As in the ODE case, however, existence and uniqueness can be shown under certain assumptions concerning the coefficients f and g:

Theorem 7.3 (existence and uniqueness)
Let $f : \mathbb{R}_+ \times \mathbb{R} \longrightarrow \mathbb{R}$ and $g : \mathbb{R}_+ \times \mathbb{R} \longrightarrow \mathbb{R}$ be functions with the following properties:

- ***Lipschitz condition:*** *There is a constant $L \geq 0$ such that*

$$|f(t,x) - f(t,y)| \leq L|x-y|, \qquad |g(t,x) - g(t,y)| \leq L|x-y| \tag{16}$$

  for all $x, y \in \mathbb{R}$ and $t \geq 0$.

- ***Linear growth condition:*** *There is a constant $K \geq 0$ such that*

$$|f(t,x)|^2 \leq K\left(1 + |x|^2\right), \qquad |g(t,x)|^2 \leq K\left(1 + |x|^2\right) \tag{17}$$

  for all $x \in \mathbb{R}$ and $t \geq 0$.

Then, the SDE

$$dX(t) = f\big(t, X(t)\big)dt + g\big(t, X(t)\big)dW(t), \qquad t \in [0,T]$$

with deterministic initial value $X(0) = X_0$ has a continuous adapted solution and

$$\sup_{t \in [0,T]} \mathbb{E}\left(X^2(t)\right) < \infty.$$

If both $X(t)$ and $\widetilde{X}(t)$ are such solutions, then

$$\mathbb{P}\Big(X(t) = \widetilde{X}(t) \text{ for all } t \in [0,T]\Big) = 1.$$

Proof: Theorem 9.1 in [Ste01] or Theorem 4.5.3 in [KP99].

Remark: The assumptions can be weakened.

8 Extension to higher dimensions

In order to model options on several underlying assets (e.g. basket options), we have to consider vector-valued Itô integrals and SDEs. A d-dimensional SDE takes the form

$$X_j(t) = X_j(0) + \int_0^t f_j(s, X(s))\, ds + \sum_{k=1}^m \int_0^t g_{jk}(s, X(s))\, dW_k(s) \qquad (j = 1, \ldots, d) \tag{18}$$

for $d, m \in \mathbb{N}$ and suitable functions

$$f_j : \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}, \qquad g_{jk} : \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}.$$

$W_1(s), \dots, W_m(s)$ are one-dimensional scalar Wiener processes which are pairwise independent. (18) is equivalent to

$$X(t) = X(0) + \int_0^t f(s, X(s))\, ds + \int_0^t g(s, X(s))\, dW(s) \tag{19}$$

with vectors

$$W(t) = \big(W_1(t), \dots, W_m(t)\big)^T \in \mathbb{R}^m$$
$$f(t,x) = \big(f_1(t,x), \dots, f_d(t,x)\big)^T \in \mathbb{R}^d$$

and a matrix

$$g(t,x) = \begin{pmatrix} g_{11}(t,x) & \cdots & g_{1m}(t,x) \\ \vdots & & \vdots \\ g_{d1}(t,x) & \cdots & g_{dm}(t,x) \end{pmatrix} \in \mathbb{R}^{d\times m}$$

Theorem 8.1 (Multi-dimensional Itô-Doeblin formula) *Let X_t be the solution of the SDE (19) and let $F : [0,\infty) \times \mathbb{R}^d \longrightarrow \mathbb{R}^n$ be a function with continuous partial derivatives $\partial_t F$, $\partial_{x_j} F$, and $\partial_{x_j}\partial_{x_k} F$. Then, the process $Y(t) := F(t, X_t)$ satisfies*

$$\begin{aligned} dY_\ell(t) &= \partial_t F_\ell(t, X_t)\, dt \\ &\quad + \sum_{i=1}^d \partial_{x_i} F_\ell(t, X_t) \cdot f_i(t, X_t)\, dt \\ &\quad + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \partial_{x_i}\partial_{x_j} F_\ell(t, X_t) \cdot \left(\sum_{k=1}^m g_{ik}(t, X_t) g_{jk}(t, X_t) \right) dt \\ &\quad + \sum_{i=1}^d \partial_{x_i} F_\ell(t, X_t) \cdot \sum_{k=1}^m g_{ik}(t, X_t)\, dW_k \end{aligned}$$

or equivalently

$$dY_\ell = \left\{ \partial_t F_\ell + f^T \nabla F_\ell + \frac{1}{2} \mathrm{tr}\Big(g^T (\nabla^2 F_\ell) g \Big) \right\} dt + (\nabla F_\ell)^T g\, dW(t)$$

where ∇F_ℓ is the gradient and $\nabla^2 F_\ell$ is the Hessian of F_ℓ, and where $\mathrm{tr}(A) = \sum_{j=1}^m a_{jj}$ is the trace of a matrix $A = (a_{ij})_{i,j} \in \mathbb{R}^{m\times m}$.

Proof: Similar to the case $d = m = 1$.

Final remark: Itô vs. Stratonovich. The Itô integral is not the only stochastic integral, and the Stratonovich integral is a famous alternative. The Stratonovich integral has the advantage that the ordinary chain rule remains valid, i.e. the additional term in the Itô-Doeblin formula does not appear when the Stratonovich integral is used. However, a Stratonovich integral is not a martingale, whereas an Itô integral is, and this is the reason why typically the Itô integral is used to model risky assets in financial markets. However, Stratonovich integrals can be transformed into Itô integrals and vice versa. See 3.1, 3.3 in [Øks03] and 3.5, 4.9 in [KP99]. If the Itô integral in (11) or (18) is replaced by the Stratonovich integral, the properties of the SDE change. In order to distinguish between both concepts, one should distinguish between "Itô SDEs" and "Stratonovich SDEs".

References

[BK04] N. H. Bingham and Rüdiger Kiesel. *Risk-neutral valuation. Pricing and hedging of financial derivatives.* Springer Finance. Springer, London, 4nd ed. edition, 2004.

[CT04] Rama Cont and Peter Tankov. *Financial modelling with jump processes.* CRC Financial Mathematics Series. Chapman & Hall, Boca Raton, FL, 2004.

[KP99] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations.* Number 23 in Applications of Mathematics. Springer, Berlin, corr. 3rd printing edition, 1999.

[Øks03] Bernt Øksendal. *Stochastic differential equations. An introduction with applications.* Universitext. Springer, Berlin, 6th ed. edition, 2003.

[Shr04] Steven E. Shreve. *Stochastic calculus for finance. II: Continuous-time models.* Springer Finance. Springer, 2004.

[Ste01] J. Michael Steele. *Stochastic calculus and financial applications.* Number 45 in Applications of Mathematics. Springer, New York, NY, 2001.

A Some definitions from probability theory

Definition A.1 (Probability space) *The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a* ***probability space****, if the following holds:*

1. *$\Omega \neq \emptyset$ is a set, and $\mathcal{F}$ is a* ***σ-algebra*** *(or σ-field) on Ω, i.e. a family of subsets of Ω with the following properties:*
    - $\emptyset \in \mathcal{F}$
    - *If $F \in \mathcal{F}$, then $\Omega \setminus F \in \mathcal{F}$*
    - *If $F_i \in \mathcal{F}$ for all $i \in \mathbb{N}$, then $\bigcup_{i=1}^{\infty} F_i \in \mathcal{F}$*

*The pair $(\Omega, \mathcal{F})$ is called a **measurable space**.*

2. *$\mathbb{P} : \mathcal{F} \longrightarrow [0, 1]$ is a **probability measure**, i.e.*
   - *$\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$*
   - *If $F_i \in \mathcal{F}$ for all $i \in \mathbb{N}$ are pairwise disjoint (i.e. $F_i \cap F_j = \emptyset$ for $i \neq j$), then*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(F_i).$$

Definition A.2 (Borel σ-algebra) *If $\mathcal{U}$ is a family of subsets of Ω, then **the σ-algebra generated by $\mathcal{U}$** is*

$$\mathcal{F}_{\mathcal{U}} = \bigcap \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-algebra of } \Omega \text{ and } \mathcal{U} \subset \mathcal{F}\}.$$

*If $\mathcal{U}$ is the collection of all open subsets of a topological space Ω (e.g. $\Omega = \mathbb{R}^d$), then $\mathcal{B} = \mathcal{F}_{\mathcal{U}}$ is called the **Borel σ-algebra** on Ω. The elements $B \in \mathcal{B}$ are called Borel sets.*

For the rest of this section $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Definition A.3 (Measurable functions, random variables)

- *A function $X : \Omega \longrightarrow \mathbb{R}^d$ is called **$\mathcal{F}$-measurable** if*

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

  *for all Borel sets $B \in \mathcal{B}$. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, then every $\mathcal{F}$-measurable function is called a **random variable**.*
- *If $X : \Omega \longrightarrow \mathbb{R}^d$ is any function, then the **σ-algebra generated by X** is the collection of all subsets*

$$X^{-1}(B) \text{ for all } B \in \mathcal{B}.$$

  Notation: $\mathcal{F}^X = \sigma\{X\}$
  $\mathcal{F}^X$ is the smallest σ-algebra where X is measurable.

Definition A.4 (Independence)

- *Two sets $A \subset \Omega$ and $B \subset \Omega$ are called independent if*

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

- *For $n \in \mathbb{N}$ let $\mathcal{G}_1, \ldots, G_n \subset \mathcal{F}$ be a collection of sub-σ-algebras of $\mathcal{F}$. $\mathcal{G}_1, \ldots, G_n$ are independent if*

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \cdots \mathbb{P}(A_n) \quad \text{for all} \quad A_i \in \mathcal{G}_i.$$

- *Random variables $X_1, \dots, X_n$ are called* ***independent*** *if*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} X_i^{-1}(A_i)\right) = \prod_{i=1}^{n} \mathbb{P}\big(X_i^{-1}(A_i)\big)$$

  for all $A_1, \dots, A_n \in \mathcal{B}$. Equivalent: The random variables $X_1, \dots, X_n$ are independent if the σ-algebras generated by $X_1, \dots, X_n$ are independent.
- *If X and Y are independent random variables with $\mathbb{E}(|XY|) < \infty$, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.*
- *A random variable X is independent of a sub-σ-algebra $G \subset \mathcal{F}$ if the σ-algebra generated by X is independent of $\mathcal{G}$.*

B The Riemann-Stieltjes integral

Let $f : [a, b] \to \mathbb{R}$ be bounded and let $w : [a, b] \to \mathbb{R}$ be increasing, i.e. $w(t) \geq w(s)$ for all $t \geq s$. For a partition $a = t_0 < t_1 < \ldots < t_N = b$ we define the lower and upper sums

$$\overline{S_N} := \sum_{n=0}^{N-1} \sup\{f(t) : t \in [t_n, t_{n+1}]\}\big(w(t_{n+1}) - w(t_n)\big)$$
$$\underline{S_N} := \sum_{n=0}^{N-1} \inf\{f(t) : t \in [t_n, t_{n+1}]\}\big(w(t_{n+1}) - w(t_n)\big).$$

If $\overline{S_N}$ and $\underline{S_N}$ converge to the same value as the partition is refined, then the Riemann-Stieltjes integral is defined by

$$\int_a^b f(t)dw(t) := \lim_{N\to\infty} \overline{S_N} = \lim_{N\to\infty} \underline{S_N}.$$

For $w(t) = t$, this is the standard Riemann integral.

If $w : [a, b] \to \mathbb{R}$ is not increasing but has bounded variation, then there are increasing functions $w_1 : [a, b] \to \mathbb{R}$ and $w_2 : [a, b] \to \mathbb{R}$ such that $w(t) = w_1(t) - w_2(t)$, and the Riemann-Stieltjes integral can be defined by

$$\int_a^b f(t)dw(t) := \int_a^b f(t)dw_1(t) - \int_a^b f(t)dw_2(t).$$