Stochastic differential equations

Wojciech Dybalski

- There will be lecture notes and exercise sheets on https://wdybalski.faculty.wmi.amu.edu.pl. The exercise sheets will be posted about one week before the class at which they will be discussed.
- Office hours: Tuesdays 12-14 (Except Tuesday 7.10) or by appointment.
- Textbook: In general Øksendal [Ok]. But first lecture: Roepstorff [Ro, Ch 1].

Contents

Lecture 1

1 Introduction and motivation

1. ODE are prolific in physics. Imagine a particle moving under the influence of an external force, e.g. a pendulum

$$m\frac{d^2x(t)}{dt^2} = -kx(t), \quad k > 0. \tag{1.1}$$

A solution can be found given the initial conditions $x_0 = x(0), v_0 = \frac{dx(t)}{dt}|_{t=0}$:

$$x(t) = x_0 \cos(\omega t) + \frac{v_0}{\omega}\sin(\omega t), \quad \omega = \sqrt{\frac{k}{m}}. \tag{1.2}$$

The picture in phase space is an elipse.

2. In a real laboratory situations there are imperfections. For example, the air viscosity will modify the equation to

$$m\frac{d^2x(t)}{dt^2} = -kx(t) - \gamma\frac{dx(t)}{dt}, \quad \gamma > 0. \tag{1.3}$$

The picture in phase space is a spiral.

3. In addition there are perturbations which act like an external force: Trains passing by, seismic tremors, electromagnetic forces for nearby electric devices (acting on a metal pendulum):

$$m\frac{d^2x(t)}{dt^2} = -kx(t) - \gamma\frac{dx(t)}{dt} + \sigma\frac{dB_t}{dt}, \quad \sigma > 0. \tag{1.4}$$

The picture in phase space is a rough spiral. **The goal of the lecture is to understand the problem of existence and uniqueness of solutions for such equations.**

[It may seem hopeless to describe this influence of many factors which we do not control. What saves us is the Central Limit Theorem: If the particle feels the combined effect of a large number of independent kicks,

$$\int_0^t ds\big\{\text{independent random kicks at time } s\big\} \tag{1.5}$$

the net force looks approximately like a Gaussian random variable B_t called Brownian motion. $\frac{dB_t}{dt}$ is called the white noise.]

4. From the practitioner's point of view, random variable is something you can compute expectation E (or average) of and get numbers. For Brownian motion we have

- $E(B_t) = 0$,
- $E(B_t^2) = t$,
- $E((B_0 - B_t)(B_t - B_s)) = 0$ for $s > t > 0$, (independence of increments).
- $B_0 = 0$.

Dealing with $t \mapsto B_t$ naively as if it was a differentiable function is tricky: In fact,

$$1 = \frac{d}{dt}t = \frac{d}{dt}E(B_t^2) = 2E(B_t\frac{dB_t}{dt}) = \lim_{\Delta t \to 0} 2E\big(B_t\frac{B_{t+\Delta t} - B_t}{\Delta t}\big) = 0, \tag{1.6}$$

which is a contradiction. In reality, one can make sense of $\frac{dB_t^2}{dt} = 2B_t\frac{dB_t}{dt} + 1$, which saves the game. [This example shows, that closer mathematical scrutiny is needed to deal with such objects].

5. Equations involving such random noise are called stochastic differential equations (SDE).

   - Since B_t are random variables, also $x(t)$ are random variables. Concepts like existence and uniqueness of solutions have to be reconsidered.
   - Applications: climate fluctuations, turbulence, neural activity, option pricing.... [It is hard to escape SDE if you want to model real world phenomena].

6. A financial mathematics application:

   - Suppose a person has an asset or resource (e.g. house, stocks, oil...) that they want to sell.
   - To start with, suppose that the value of the asset grows with r (Δt - compounded, annual) interest rate:

$$\frac{\Delta X}{X} = r\Delta t. \tag{1.7}$$

   - If we take the idealization of a continuous interest rate, this gives

$$\frac{dX_t}{dt} = rX_t, \tag{1.8}$$

   with a solution $X_t = X_0 e^{rt}$ [t measured in years].

   - Since we are talking about a risky asset, the interest rate is fluctuating:

$$\frac{dX_t}{dt} = \left(r + \alpha\frac{dB_t}{dt}\right)X_t, \quad r, \alpha \geq 0. \tag{1.9}$$

   Naive solution has the form:

$$X_t = X_0 e^{rt + \alpha B_t} \tag{1.10}$$

   One can make a mathematical sense out of it. But one can also make mathematical sense of the solution:

$$X_t = X_0 e^{(r - \frac{1}{2}\alpha^2)t + \alpha B_t}, \tag{1.11}$$

   depending how exactly one defines stochastic integrals $\int f(t)dB_t$. [This will be important part of the course].

   - **Optimal stopping problem**: The person knows $\{X_s\}_{0 \leq s \leq t}$ and wants to find an optimal time to sale the asset. [If you sell too early, you may loose future gains. If you sell to late the pay-off may shrink].

- **An optimal portfolio problem**: Suppose a person has two possible investments: a risky investment (1.9) (a stock) and a safe investment (1.8) (a bond). How to divide the asset $X_t = u_t X_t + (1 - u_t) X_t$ into a risky investment and a safe investment, given the selling time T?
- **Pricing of options**: Suppose that at time $t = 0$ the person of the previous question is offered the right (but no obligation) to buy one unit of the risky asset at a specified price K and at a specified future time $t = T$. [European call option]. How much should the person be willing to pay for such an option? Problem solved by Black and Scholes in (1973), related to Nobel Prize in Economics in 1997 (Scholes and Merton).

2 Brownian motion

2.1 General probability

I will explain how B_t emerges from a random walk. We need some basic probabilistic notions needed to describe the random walk and state the Central Limit Theorem (CLT). For more information on basic probability see [BT, JP].

1. Def: If Ω is a given set then a σ-algebra $\mathcal{F}$ on Ω is a family of subsets on Ω with the following properties:

   (i) $\emptyset \in \mathcal{F}$.

   (ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$, where A^c is the complement of A in Ω.

   (iii) $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

   - Example: For $\Omega = \mathbb{R}^d$ the smallest σ-algebra containing all open sets is called the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$.
   - If (iii) holds only for finite sums, then $\mathcal{F}$ is called an algebra.

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

3. Def: A probability measure on a measurable space $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to [0, 1]$ s.t.

   (a) $P(\emptyset) = 0$, $P(\Omega) = 1$.

   (b) If $A_1, A_2, \ldots \in \mathcal{F}$ and $\{A_i\}_{i=1}^{\infty}$ disjoint (i.e., $A_i \cap A_j = \emptyset$ is $i \neq j$) then
   $$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \tag{2.1}$$

   [Question: Why do we need a sigma field and cannot just define the measure on all subsets? Banach-Tarski paradox.]

4. Def: The triple $(\Omega, \mathcal{F}, P)$ is called a probability space.

5. Def. We say that the events A, B, are independent if $P(A \cap B) = P(A)P(B)$. [Note that disjoint sets are dependent. Examples: flipping a coin several times, rolling a die several times]. For a family of events $A_1, \ldots A_n$, we require for any non-empty subset $I \subset \{1, \ldots n\}$
   $$P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i). \tag{2.2}$$

6. Def. Conditional probability:
   $$P(A|B) := \frac{P(A \cap B)}{P(B)} \text{ for } P(B) \neq 0, \tag{2.3}$$
   That is, probability that A occurs provided that B is known to occur.

7. Random variable is a functions $X : \Omega \to \mathbb{R}^d$ which is $\mathcal{F}$-measurable. That is, $X^{-1}(O) \in \mathcal{F}$ for any open $O \subset \mathbb{R}^d$.

8. Def. Every random variable induces a probability measure μ_X on $\mathbb{R}^d$, given by
$$\mu_X(B) = P(X^{-1}(B)) = P(X \in B). \tag{2.4}$$
It is called the distribution (or law) of X. One also writes $\mathcal{L}(X) := \mu_X$ or $X \sim$ [Name of $\mathcal{L}(X)$].
For example, if $X : \Omega \to \mathbb{R}$ satisfies
$$\mu_X(A) = \int_A \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}} dx, \tag{2.5}$$
we say that X is a Gaussian variable with law $N(m, \sigma^2)$ (normal disribution) or $X \sim N(m, \sigma^2)$.

9. Def. For a finite family of random variables $X_1, \ldots, X_n$ the joint law is the measure on $\mathbb{R}^{dn}$ given by
$$\mu_{X_1,\ldots,X_n}(A) = P((X_1, \ldots, X_n) \in A), \quad A \in \mathcal{B}(\mathbb{R}^{dn}). \tag{2.6}$$

10. Def. If $\int |X(\omega)| dP(\omega) < \infty$ then the number
$$E(X) := \int_\Omega X(\omega) dP(\omega) = \int_{\mathbb{R}} x \, d\mu_X(x) \tag{2.7}$$
is called the expectation of X (w.r.t. P). Change of variables formula for measures.
If you are not comfortable with integration w.r.t. an arbitrary measure you may want to have a look at [Ru, Chapter 1].

11. Def. If $\int |X(\omega)|^2 dP(\omega) < \infty$ then the number
$$\mathrm{var}(X) = E(X^2) - E(X)^2 = E[(X - E(X))^2] \tag{2.8}$$
is called the variance.

12. Def. We say that real-valued random variables $X_1, \ldots, X_n$ are independent if
$$\mu_{X_1,\ldots,X_n} = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}. \tag{2.9}$$

13. Fact: If two random variables X, Y are independent, then they are uncorrelated
$$E[XY] = E[X]E[Y] \tag{2.10}$$
provided that $E[|X|] < \infty$ and $E[|Y|] < \infty$. (Converse not true in general. But true for Gaussian random variables as we will see).
Idea of proof: We compute for characteristic functions $X = \chi_{F_1}$, $Y = \chi_{F_2}$.
$$E[\chi_{F_1}\chi_{F_2}] = \int \chi_{F_1}(\omega)\chi_{F_2}(\omega) dP(\omega) = P(F_1 \cap F_2) = P(F_1)P(F_2). \tag{2.11}$$
Then, it follows for step functions. Recalling the definition of the Lebesgue integral, a measurable function is approximated pointwise by step functions.

14. Thm (Central Limit Theorem (CLT)): Let $(X_j)_{j\geq 1}$ be i.i.d. (independent identically distributed) $\mathbb{R}$-valued random variables with $E(X_j) = m$ and $\mathrm{Var}(X_j) = \sigma^2$ with $0 < \sigma^2 < \infty$. Let $S_n = \sum_{j=1}^n X_j$. Let $Y_n = \frac{S_n - nm}{\sigma\sqrt{n}}$. Then Y_n converges in distribution to Y, where $Y \sim N(0, 1)$. That is, for any bounded, continuous function f on $\mathbb{R}$
$$\int f(x) d\mu_{Y_n}(x) \to \int f(x) d\mu_Y(x). \tag{2.12}$$
(This is also called weak convergence of measures).

2.2 Example: Bernoulli trials and binomial probabilities

1. Consider one Bernoulli trial with a toss of a coin where we get heads (H) with probability p and tails (T) with probability $1-p$.
   - Sample space: $\Omega_0 = \{H, T\}$
   - σ-algebra: All subsets of Ω_0
   $$\mathcal{F}_0 = \{\emptyset, \{H\}, \{T\}, \{H, T\}\} \tag{2.13}$$
   - Probability measure: $P_0(\emptyset) = 0$, $P_0(\{H\}) = p$, $P_0(\{T\}) = 1-p$, $P_0(\{H, T\}) = 1$.
2. Consider n independent Bernoulli trials.
   - Sample space $\Omega = \Omega_0^n$. Any $\omega \in \Omega$ has the form $\omega = (\omega_1, \ldots, \omega_n)$, $\omega_i \in \{H, T\}$.
   - σ-algebra: All subsets of Ω.
   - Probability measure: By independence of trials,
   $$P(\{\omega\}) = \prod_{i=1}^{n} P_0(\{\omega_i\}) = p^{R_n(\omega)}(1-p)^{n-R_n(\omega)}, \tag{2.14}$$
   where $R_n(\omega)$ is the number of H in ω. Note that $R_n : \Omega \to \mathbb{R}$ is a random variable.
3. Define random variables $\xi_i : \Omega \to \{-1, 1\}$ s.t.
   $$\xi_i(\omega) = \begin{cases} +1 & \text{if } \omega_i = H, \\ -1 & \text{if } \omega_i = T. \end{cases} \tag{2.15}$$
4. Then $R_n = \#\{\, i : \xi_i = +1\}$ is the number of heads in n trials.
5. Fact: We have
   $$P(R_n = \ell) = P(\ell \text{ heads come up in an } n\text{-toss sequence}) = \binom{n}{\ell} p^\ell (1-p)^{n-\ell}. \tag{2.16}$$
   We will write $R_n \sim \text{Binomial}(n, p)$, i.e. the random variable R_n has a binomial probability distribution.

   Proof:
   - Suppose we fix a particular individual outcome, e.g. HTTHTT. There are $\ell = 2$ heads in $n = 6$ trials. By independence of the tosses,
   $$P(\text{HTTHTT}) = P(\text{H})P(\text{H})P(\text{T})P(\text{H})P(\text{T})P(\text{T}) = p^\ell (1-p)^{n-\ell}. \tag{2.17}$$
   - There are many events which have $\ell = 2$ heads in $n = 6$ trials. For example HHTTTT. As this is a disjoint event from HTTHTT we have
   $$P(\text{HHTHTT or HTTHTT}) = P(\text{HHTHTT}) + P(\text{HTTHTT}). \tag{2.18}$$
   - Now, how many different sequences of n tosses contains exactly ℓ heads? Clearly, $\binom{n}{\ell}$. $\square$

2.3 From random walks to Brownian motion

1. Particle moves on a lattice $h\mathbb{Z}$, $h > 0$.

2. There occurs one step, to the left or to the right, within the period of time $\tau > 0$.

3. Successive steps are independent: e.g. a fair coin ($p = \frac{1}{2}$) is tossed at each step and the particle moves to the left or to the right depending on the outcome (H or T). We can use the same sample space as above $\Omega = \{H, T\}^n$, with probability P defined by (2.14).

4. We want to compute probability
$$P(\text{particle ends up at } x = kh \text{ after } n \text{ steps, starting from zero}) =: P(kh, n\tau) =: P[k, n]. \tag{2.19}$$
   - Define the random variable $S_n = \xi_1 + \ldots + \xi_n$, s.t. $S_n h$ is the position of the particle after n steps.
   - We have $S_n = (+1)R_n + (-1)(n - R_n)$, hence $R_n = (n + S_n)/2$. So the event $\{S_n = k\} := \{\omega \in \Omega : S_n(\omega) = k\}$ is the same as $\{R_n(\omega) = (n + k)/2\} := \{\omega \in \Omega : R_n = (n + k)/2\}$.
   - Then $P[k, n] = P(S_n = k) = P(R_n = (n + k)/2)$. Since $R_n \sim \text{Binomial}(n, 1/2)$, we have
$$P[k, n] := P(R_n = (n + k)/2) = \binom{n}{\frac{1}{2}(n + k)} \frac{1}{2^n}, \tag{2.20}$$
   provided $\frac{1}{2}(n + k)$ is integer. Otherwise $P[k, n] = 0$.

5. Let us now try to take the continuum limit $h, \tau \to 0$ s.t. $t := n\tau, \cancel{x := kh}$, $D := \frac{h^2}{2\tau}$ stay constant. Position of the particle after n steps is $x = hS_n = \sqrt{2\tau D} S_n = \sqrt{\frac{2Dt}{n}} S_n$. We restrict attention to n of the form $n = t\tilde{n}$ for $t \in \mathbb{N}$ and $\tilde{n} \in \mathbb{N}$. Then $x = \sqrt{\frac{2D}{\tilde{n}}} S_{\tilde{n}t}$ is still position of the particle after n steps at time t expressed in terms of $\tilde{n}$ which we will take to ∞. This motivates the definition (setting $D = 1/2$ and renaming $\tilde{n} \to n$):
$$B_t^{(n)} := \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor}, \tag{2.21}$$
with the convention $S_0 = 0$.

6. Fact: By the CLT the limit
$$B_t := \lim_{n \to \infty} B_t^{(n)} \tag{2.22}$$
exists in distribution and defines a family of Gaussian random variables $B_t \sim N(0, t)$.

   Def: The family $\{B_t\}_{t \in \mathbb{R}_+}$ is called the Brownian motion.

   Proof of the fact: In our case $X_j = \xi_j$.
   - Let us compute the law of ξ_i. Let $A \in \mathcal{B}(\mathbb{R})$. We note that $\xi_i(\omega) = \tilde{\xi}_i(\omega_i)$.
$$\mu_{\xi_i}(A) := P(\omega : \xi_i(\omega) \in A) = P(\omega : \tilde{\xi}_i(\omega_i) \in A). \tag{2.23}$$
   - Independence: For $n = 2$
$$\begin{aligned} \mu_{\xi_1, \xi_2}(A_1 \times A_2) &= P(\omega : (\xi_1(\omega), \xi_2(\omega)) \in A_1 \times A_2) \\ &= P(\omega : (\tilde{\xi}_1(\omega_1), \tilde{\xi}_2(\omega_2)) \in A_1 \times A_2) \\ &= P(\omega : \tilde{\xi}_1(\omega_1) \in A_1, \tilde{\xi}_2(\omega_2) \in A_2) \\ &= P(\omega : \tilde{\xi}_1(\omega_1) \in A_1) P(\omega : \tilde{\xi}_2(\omega_2) \in A_2) = \mu_{\xi_1}(A_1)\mu_{\xi_2}(A_2), \end{aligned} \tag{2.24}$$
   where in the next to the last step we made use of the independence of coin tosses, i.e. (2.14). For arbitrary n the argument is analogous.

- We compute

$$m_j = E(\xi_j) = \int \xi_j(\omega)dP(\omega) = \int \xi_j d\mu_{\xi_j} = (+1)\frac{1}{2} + (-1)\frac{1}{2} = 0, \quad (2.25)$$

$$\sigma_j^2 = E(\xi_j^2) = \int \xi_j(\omega_j)^2 dP(\omega_j) = \int \xi_j^2 d\mu_{\xi_j} = (+1)\frac{1}{2} + (+1)\frac{1}{2} = 1. \quad (2.26)$$

- Thus, by the CLT, $Y_n = \frac{S_n}{\sqrt{n}}$ converges in distribution to a Gaussian random variable $Y = B_{t=1}$ with law $N(0,1)$.
- Similarly, for $t \in \mathbb{N}$, $\frac{S_{nt}}{\sqrt{n}} = \sqrt{t}\frac{S_{nt}}{\sqrt{nt}} = \sqrt{t}Y_{nt} \to \sqrt{t}Y =: B_t \sim N(0,t)$. Its law is

$$\mu_{\sqrt{t}Y}(A) = P(\sqrt{t}Y \in A) = P(Y \in A/\sqrt{t}) = \int_{A/\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \int_A \frac{1}{\sqrt{2\pi t}} e^{-y^2/(2t)} dy, \quad (2.27)$$

that is $B_t \sim N(0,t)$. (I leave for the reader the extension to $t \in \mathbb{R}_+$). $\square$

7. Remark: We know from the proof that $E(B_t) = 0$, $E(B_t^2) = t$. We would also like to show $E((B_0 - B_t)(B_t - B_s)) = 0$, $s > t$, but so far we do not have a joint probability space for B_t and B_s, $t \neq s$. However, already now, such an outcome is suggested by

$$E((B_0^{(n)} - B_t^{(n)})(B_t^{(n)} - B_s^{(n)})) = 0. \quad (2.28)$$

The two factors have the form

$$B_t^{(n)} - B_s^{(n)} = \frac{1}{\sqrt{n}}(\xi_{\lfloor nt \rfloor + 1} + \cdots + \xi_{\lfloor ns \rfloor}), \quad (2.29)$$

$$B_0^{(n)} - B_t^{(n)} = \frac{1}{\sqrt{n}}(\xi_1 + \cdots + \xi_{\lfloor nt \rfloor}). \quad (2.30)$$

They involve independent coin tosses. For independent random variables we have $E(XY) = E(X)E(Y)$ by (2.9), so (2.28) follows.

8. We can also compute $E(B_t^{(n)} B_s^{(n)})$. By (2.28), for $s > t$,

$$E(B_t^{(n)} B_s^{(n)}) = E(B_t^{(n)2}) = \frac{1}{n} E(\xi_1^2 + \cdots + \xi_{\lfloor nt \rfloor}^2) = \frac{\lfloor nt \rfloor}{n}. \quad (2.31)$$

We used $E(\xi_i \xi_j) = E(\xi)E(\xi_j) = 0$ for $i \neq j$. Since the roles of t, s can be exchanged

$$E(B_t^{(n)} B_s^{(n)}) = \min(\frac{\lfloor nt \rfloor}{n}, \frac{\lfloor ns \rfloor}{n}). \quad (2.32)$$

As we will see, in the limit $E(B_t B_s) = \min(t,s)$.

9. Let us consider again the continuum limit $h, \tau \to 0$ s.t. $t := n\tau, x := kh$, $D := \frac{h^2}{2\tau}$ stay constant. There is another way to see (heuristically) that the probability distribution of the position of the particle must be Gaussian:

- Recall: $\binom{n+1}{m} = \binom{n}{m} + \binom{n}{m-1}$.
- Hence: $P[k, n+1] = \frac{1}{2}P[k-1, n] + \frac{1}{2}P[k+1, n]$. In fact,

$$P[k, n+1] = \binom{n+1}{\frac{1}{2}(n+1+k)} \frac{1}{2^{n+1}} = \frac{1}{2}\binom{n}{\frac{1}{2}(n+1+k)} \frac{1}{2^n} + \frac{1}{2}\binom{n}{\frac{1}{2}(n+1+k)-1} \frac{1}{2^n}. \quad (2.33)$$

Alternatively, we can say that the event { particle at point k after $n+1$ steps } is a union of two disjoint events { particle at point $k-1$ after and n steps, then jump to the right } and { particle at point $k+1$ after n steps, then jump to the left }

- Thus, we have

$$\frac{P(kh,(n+1)\tau)-P(kh,n\tau)}{\tau}=\frac{h^2}{2\tau}\frac{P((k+1)h,n\tau)-2P(kh,n\tau)+P((k-1)h,n\tau)}{h^2} \tag{2.34}$$

If the continuum limit $\tilde{p}(x,t) := \lim_{h,\tau\to 0} P(kh,n\tau)$ exists, then $\int_I \tilde{p}(x,t)dx$ is the probability of finding the particle in the interval $I\subset\mathbb{R}$. The probability distribution $\tilde{p}$ satisfies the heat equation

$$\frac{\partial\tilde{p}(x,t)}{\partial t}=D\frac{\partial^2\tilde{p}(x,t)}{\partial x^2}. \tag{2.35}$$

This is an example of a Fokker-Planck equation, describing the probability distribution of a stochastic process (in our case $\{B_t\}_{t\in\mathbb{R}_+}$).

- In the continuum limit 'starting' the process at $x=0$ corresponds to the initial condition $\tilde{p}(x,0)=\delta_0(x)$ (Dirac delta) as this means zero probability of finding the particle in any I not containing $\{0\}$. Then the solution of the heat equation is:

$$\tilde{p}(x,t)=\frac{1}{\sqrt{4\pi Dt}}\exp\bigg(-\frac{x^2}{4Dt}\bigg). \tag{2.36}$$

Setting $D=1/2$ (as above) this is the probability distribution of B_t.

Lecture 2

The goal is now to construct a joint probability space for all B_t, $t\in\mathbb{R}_+$.

2.4 Stochastic processes

1. Def. A stochastic process is a parametrized collection of random variables

$$\{X_t\}_{t\in\mathcal{T}} \tag{2.37}$$

defined on a probability space $(\Omega,\mathcal{F},P)$ and assuming values in $\mathbb{R}^d$. ($\mathcal{T}$ is a set. Typical choices: $\mathbb{R}_+ := [0,\infty)$, $[a,b]$, $\mathbb{N}$...).

2. Remark: For fixed $t\in\mathcal{T}$ we have a random variable $\Omega\ni\omega\mapsto X_t(\omega)\in\mathbb{R}^d$.

3. Def: For fixed ω the function $\mathcal{T}\ni t\mapsto X_t(\omega)$ is called a path of X_t. (Interpretation: result of an experiment ω at time t. Eg. position of a random walker at time t).

4. Remark: For fixed stochastic process $\{X_t\}_{t\in\mathcal{T}}$ we may associate with each ω a function $t\mapsto X_t(\omega)=:\omega(t)$ from $\mathcal{T}$ into $\mathbb{R}^d$. (Recall that for random walk $\omega=HHT\dots TH$ was uniquely determining the path).

   - Then, we regard Ω as a subset of $\tilde{\Omega}=(\mathbb{R}^d)^{\mathcal{T}}$. Namely, with any ω we associate a path $\{t\to X_t(\omega)\}$. We denote this embedding by $\iota:\Omega\to\tilde{\Omega}$.
   - The σ-algebra $\tilde{\mathcal{F}}$ on $\tilde{\Omega}$ is generated by 'cylinder sets' of the form

$$C_{t_1,\dots,t_k;F_1,\dots,F_k} := \{\tilde{\omega}\in(\mathbb{R}^d)^{\mathcal{T}}\ :\ \tilde{\omega}(t_1)\in F_1,\dots,\tilde{\omega}(t_k)\in F_k\},\quad F_i\subset\mathbb{R}^d \text{ Borel}. \tag{2.38}$$

   Fact: For $\mathcal{T}$ countable, $\tilde{\mathcal{F}}$ is the Borel algebra on $\tilde{\Omega}$ provided that $\tilde{\Omega}$ is equipped with the product topology. (The product topology is the smallest topology containing sets (2.38) with F_i open). [There is a statement in [Ok, p.10] that the same is true for $\mathcal{T}=[0,\infty)$, but I have doubts if it really holds. Remains to be checked.]

- We define the measure $\tilde{P}$ on $(\tilde{\Omega}, \mathcal{B})$ as follows

$$\tilde{P}(C) = P(\iota^{-1}(C)) \tag{2.39}$$

for cylinder sets C.

In view of the above, one can see the stochastic process as a measure $\tilde{P}$ on the space of paths $(\tilde{\Omega}, \tilde{\mathcal{F}})$. We include all possible paths in $\tilde{\Omega} = (\mathbb{R}^d)^{\mathcal{T}}$, but some of them may have probability zero. For example, for the random walk, the path for which after time τ the walker is at position $2h$ has probability zero.

5. Def: Suppose that $\{X_t\}_{t\in\mathcal{T}}$ and $\{Y_t\}_{t\in\mathcal{T}}$ are stochastic processes on $(\Omega, \mathcal{F}, P)$. Then, we say that $\{X_t\}_{t\in\mathcal{T}}$ is a version of (or modification of) $\{Y_t\}_{t\in\mathcal{T}}$ if

$$P(\{\omega \in \Omega \,:\, X_t(\omega) = Y_t(\omega)\}) = 1 \text{ for all } t. \tag{2.40}$$

6. Def: The finite-dimensional distributions of the process $X = \{X_t\}_{t\in\mathcal{T}}$ are the measures $\nu_{t_1,\ldots,t_k}$ defined on $\mathbb{R}^{dk}$, $k = 1, 2, \ldots,$ by

$$\begin{aligned} \nu_{t_1,\ldots,t_k}(F_1 \times F_2 \times \cdots \times F_k) &= P[X_{t_1} \in F_1, \ldots, X_{t_k} \in F_k] \\ &= \int \chi(X_{t_1}(\omega) \in F_1) \ldots \chi(X_{t_k}(\omega) \in F_k) dP(\omega), \end{aligned} \tag{2.41}$$

where $t_i \in \mathcal{T}$ and $F_1, \ldots F_k$ denote Borel sets in $\mathbb{R}^d$.

7. Under certain conditions, it is possible to reconstruct the stochastic process $\{X_t\}_{t\in\mathcal{T}}$ from the distributions:

Theorem 2.1. *(Kolmogorov's extension theorem). For all $t_1, \ldots, t_k \in \mathcal{T}$ let $\nu_{t_1,\ldots,t_k}$ be probability measures on $\mathbb{R}^{dk}$ s.t.*

$$\nu_{t_{\sigma(1)},\cdots,t_{\sigma(k)}}(F_1 \times \cdots \times F_k) = \nu_{t_1,\cdots,t_k}(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(k)}) \tag{K1}$$

for all permutations σ on $\{1, 2, \ldots, k\}$ and

$$\nu_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = \nu_{t_1,\ldots,t_k,t_{k+1},\ldots,t_{k+m}}(F_1 \times \cdots \times F_k \times \mathbb{R}^d \times \cdots \times \mathbb{R}^d). \tag{K2}$$

Then, there exists a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process $\{X_t\}_{t\in\mathcal{T}}$ on Ω s.t.

$$\nu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k) = P(X_{t_1} \in F_1, \cdots, X_{t_k} \in F_k), \tag{2.42}$$

for all $t_i \in \mathcal{T}$, $k \in \mathbb{N}$ and all Borel sets F_i.

2.5 Brownian motion from Kolmogorov theorem

1. To construct $\{B_t\}_{t\geq 0}$ it suffices to specify a family of probability measures $\{\nu_{t_1,\ldots t_k}\}_{t_1,\ldots t_k \geq 0}$ satisfying the conditions from the Kolmogorov theorem.

   - For fixed $x \in \mathbb{R}^d$ define

$$p(t, x, y) = (2\pi t)^{-d/2} \exp\big(-\frac{|x-y|^2}{2t}\big) \quad \text{for } y \in \mathbb{R}^d, t > 0. \tag{2.43}$$

   - If $0 \leq t_1 \leq t_2 \leq \cdots \leq t_k$ define a measure $\nu_{t_1,\ldots t_k}$ on $\mathbb{R}^{dk}$ by

$$\begin{aligned} &\nu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k) \\ &= \int_{F_1\times\cdots\times F_k} p(t_1, x, x_1) p(t_2 - t_1, x_1, x_2) \cdots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \ldots dx_k, \end{aligned} \tag{2.44}$$

   where we use the convention $p(0, x, y)dy = \delta_x(y)dy$ (unit point mass at x in physicist's notation).

- Extend to all finite sequences using (K1). Since $\int p(t,x,y)dy = 1$ for all $t \geq 0$, also (K2) holds.

2. Remark: Note that for $x = 0$, $d = 1$, we have

$$\nu_{t_1}(F_1) = \int_{F_1} p(t_1, 0, x_1)dx_1 = \int_{F_1} \tilde{p}(t_1, x_1)dx_1 \tag{2.45}$$

where $\tilde{p}$ appeared in (2.36) as a limit of random walk probability distributions, satisfying $\frac{\partial \tilde{p}(x,t)}{\partial t} = D\frac{\partial^2 \tilde{p}(x,t)}{\partial x^2}$.

3. Consequently, there exists a probability space $(\Omega, \mathcal{F}, P^x)$ and a stochastic process $\{B_t\}_{t\geq 0}$ on Ω s.t. the finite-dimensional distributions B_t are given by:

$$P^x(B_{t_1} \in F_1, \cdots, B_{t_k} \in F_k) = \int_{F_1\times\cdots\times F_k} p(t_1, x, x_1)\cdots p(t_k - t_{k-1}, x_{k-1}, x_k)dx_1 \cdots dx_k. \tag{2.46}$$

This formula determines the law of a family of random variables in the sense of (2.6). In the notation of (2.6) it reads

$$P^x(B_{t_1} \in F_1, \cdots, B_{t_k} \in F_k) = \mu_{B_{t_1},\dots,B_{t_k}}(F_1 \times \ldots \times F_k). \tag{2.47}$$

Hence, the measure on the l.h.s. of (2.46) is absolutely continuous w.r.t. the Lebesgue measure on the r.h.s. with density given by $p(t_1, x, x_1)\cdots p(t_k - t_{k-1}, x_{k-1}, x_k)$. Thus, for any Borel measurable function f on $\mathbb{R}^d$ we have

$$\begin{aligned}
&\int f(B_{t_1}, \ldots, B_{t_k})P^x(B_{t_1} \in dx_1, \cdots, B_{t_k} \in dx_k) \\
&:= \int f(x_1, \ldots, x_k)\mu_{B_{t_1},\dots,B_{t_k}}(x_1, \ldots, x_k) \\
&= \int f(x_1, \ldots, x_k)\{p(t_1, x, x_1)\cdots p(t_k - t_{k-1}, x_{k-1}, x_k)\}dx_1 \cdots dx_k.
\end{aligned} \tag{2.48}$$

The notation $P^x(B_{t_1} \in dx_1, \cdots, B_{t_k} \in dx_k)$, popular in probability theory, is defined by the above relations.

4. Def: The process constructed above is called (a version of) a Brownian motion starting at x.

5. Properties of the Brownian motion:

   (i) $\{B_t\}_{t\in\mathbb{R}_+}$ is a Gaussian process, i.e., for all $0 \leq t_1 \leq \cdots \leq t_k$ the random variable $Z = (B_{t_1}, \ldots, B_{t_k}) \in \mathbb{R}^{dk}$ has a multi-normal distribution. (To be defined below.)

   (ii) B_t has independent increments, i.e.,

   $$B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}}, \tag{2.49}$$

   are independent for all $0 \leq t_1 < t_2 \cdots < t_k$.

   (iii) Brownian motion has a version with continuous paths. By considerations above (2.38), Brownian motion can be seen as the space $C([0,\infty), \mathbb{R}^d)$ with probability measure P^x.

These properties will be precisely stated and proven in Section 2.7 after preparatory Section 2.6.

2.6 Gaussian random variables

This appendix is based on [Ok, Appendix A].

Definition 2.2. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A random variable $X : \Omega \to \mathbb{R}$ is (non-degenerate) normal if the distribution of X has a density of the form*

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(- \frac{(x-m)^2}{2\sigma^2}\right) \tag{2.50}$$

where $\sigma > 0$ and $m \in \mathbb{R}$. That is

$$P(X \in G) = \int_G p_X(x)dx \quad \text{for} \quad G \subset \mathbb{R} \text{ Borel.} \tag{2.51}$$

By Gaussian integration:

$$E(X) = \int_\Omega X dP = \int_\mathbb{R} x p_X(x)dx = \int_\mathbb{R} (x+m)\frac{1}{\sigma\sqrt{2\pi}} \exp\left(- \frac{x^2}{2\sigma^2}\right)dx = m, \tag{2.52}$$

$$\operatorname{var}(X) = E[(X-m)^2] = \int_\mathbb{R} (x-m)^2 p_X(x)dx = \sigma^2. \tag{2.53}$$

Lecture 3

Definition 2.3. *A random variable $X : \Omega \to \mathbb{R}^d$ is called (non-degenerate) multi-normal $N(m, C)$ if the distribution of X has a density of the form*

$$p_X(\underbrace{x_1, \ldots, x_d}_{x}) = \frac{\sqrt{\det(A)}}{(2\pi)^{n/2}} \exp\left(- \frac{1}{2}\sum_{j,k} (x_j - m_j)a_{j,k}(x_k - m_k)\right) \tag{2.54}$$

$$= \frac{\sqrt{\det(A)}}{(2\pi)^{n/2}} \exp\left(- \frac{1}{2}(x-m)^T A(x-m)\right) \tag{2.55}$$

$$= \frac{\sqrt{\det(A)}}{(2\pi)^{n/2}} \exp\left(- \frac{1}{2}\langle (x-m), A(x-m)\rangle\right) \tag{2.56}$$

where $m = (m_1, \ldots, m_d) \in \mathbb{R}^d$ and $C = A^{-1} = [c_{j,k}]_{1\le j,k\le d} \in \mathbb{R}^{d\times d}$ is a symmetric positive definite matrix (i.e. all eigenvalues strictly positive).

Similarly as for $d = 1$, by Gaussian integration,

$$E(X_j) = m_j, \tag{2.57}$$

$$E[(X_j - m_j)(X_k - m_k)] = c_{j,k}. \tag{2.58}$$

Definition 2.4. *The characteristic function of a random variable $X : \Omega \to \mathbb{R}^d$ is the function $\phi_X : \mathbb{R}^d \to \mathbb{C}$ defined by*

$$\phi_X(u_1, \ldots, u_d) = E[\exp(i(u_1X_1 + \cdots + u_dX_d))] = \int_{\mathbb{R}^d} \mathrm{e}^{i\langle u,x\rangle}\mu_X(dx), \tag{2.59}$$

where $\mu_X(G) = P(X \in G) = P((X_1, \ldots X_d) \in G)$ and $\langle u, x\rangle = u_1x_1 + \cdots + u_dx_d$.

Remark 2.5. *ϕ_X determines the distribution μ_X uniquely. Otherwise, we would have*

$$\int_{\mathbb{R}^d} \mathrm{e}^{i\langle u,x\rangle}\mu_{X_1}(dx) = \int_{\mathbb{R}^d} \mathrm{e}^{i\langle u,x\rangle}\mu_{X_2}(dx), \tag{2.60}$$

for all $u \in \mathbb{R}^d$ and two different measures μ_{X_1}, μ_{X_2}. By integrating both sides w.r.t. u with some $f \in S(\mathbb{R}^d)$ [1] *we would get*

$$\int_{\mathbb{R}^d} \hat{f}(x)\mu_{X_1}(dx) = \int_{\mathbb{R}^d} \hat{f}(x)\mu_{X_2}(dx), \qquad \textit{where} \quad \hat{f}(x) := \int_{\mathbb{R}^d} \mathrm{e}^{i\langle u,x\rangle} f(u)du \tag{2.61}$$

is the Fourier transform. Since Fourier transform maps $S(\mathbb{R}^d)$ onto $S(\mathbb{R}^d)$ and $S(\mathbb{R}^d)$ is dense in $C_0(\mathbb{R}^d)$ [2] *in the supremum norm, the equality extends to functions from $C_0(\mathbb{R}^d)$. Now the assumption that the measures are different is in conflict with the Riesz-Markov theorem. (A positive, linear functional on $C_0(\mathbb{R}^d)$ is given by integration against a uniquely given measure, cf. [Ru]).*

Theorem 2.6. *If $X : \Omega \to \mathbb{R}^d$ is multi-normal, i.e., $N(m, C)$, then*

$$\phi_X(u_1, \ldots, u_d) = \exp\Big(-\frac{1}{2}\sum_{j,k} u_j c_{j,k} u_k + i\sum_j u_j m_j\Big) \tag{2.62}$$

$$= \exp\Big(-\frac{1}{2}\langle u, Cu\rangle + i\langle u, m\rangle\Big). \tag{2.63}$$

- Proof by Gaussian integration. (HS2).
- By Remark 2.5, we can equivalently define a Gaussian random variable $X : \Omega \to \mathbb{R}^d$ as having the characteristic functional (2.62).
- This latter definition makes sense also for non-negative definite C (i.e., having eigenvalues larger or equal to zero). We will use the concept of multi-normal distribution $N(m, C)$ in this extended sense, unless stated otherwise. If some eigenvalues of C are zero, the multi-normal distribution is called degenerate.

Theorem 2.7. *Let $X_j : \Omega \to \mathbb{R}$ be random variables, $1 \le j \le d$. Then*

$$X = (X_1, \ldots, X_d) \tag{2.64}$$

is multi-normal iff

$$Y = \lambda_1 X_1 + \cdots + \lambda_d X_d =: \langle \lambda, X\rangle \ \textit{is normal for all}\ \lambda_1, \ldots, \lambda_d \in \mathbb{R}. \tag{2.65}$$

The equivalence also holds for (multi-)normal non-degenerate random variables if we restrict to $\lambda \neq 0$ in (2.65).

Proof. If X is multi-normal, by Theorem 2.6

$$E[\exp(i\tilde{u}Y)] = E[\exp(i\tilde{u}\langle\lambda, X\rangle)] = \exp\Big(-\frac{1}{2}\tilde{u}^2\langle\lambda, C\lambda\rangle + i\tilde{u}\langle\lambda, m\rangle\Big), \tag{2.66}$$

so Y is normal with $E(Y) = \langle\lambda, m\rangle$ and $\mathrm{var}(Y) = \langle\lambda, C\lambda\rangle$.

Conversely, suppose that $Y = \langle\lambda, X\rangle$ is normal with $E(Y) = m$ and $\mathrm{var}(Y) = \sigma^2$, i.e.,

$$E[\exp(i\tilde{u}Y)] = \exp\big(-\frac{1}{2}\tilde{u}^2\sigma^2 + i\tilde{u}m\big). \tag{2.67}$$

Then, by (2.52), (2.53)

$$m = E(Y) = \sum_j \lambda_j E(X_j), \tag{2.68}$$

$$\sigma^2 = E[(Y - E(Y))^2] = E\big[\big(\sum_j \lambda_j (X_j - m_j)\big)^2\big] = \sum_{i,j} \lambda_i\lambda_j \underbrace{E((X_i - m_i)(X_j - m_j))}_{c_{i,j}}, \tag{2.69}$$

[1] Schwartz class functions: smooth and bounded together with their derivatives even after multiplication by polynomials.
[2] Continuous functions tending to zero at infinity.

where $m_j = E(X_j)$. As λ arbitrary, by comparing with (2.62), (2.57), (2.58) we obtain that X is multi-normal.

To prove the non-degenerate case, we use that a symmetric matrix C has all eigenvalues strictly positive iff $\langle \lambda, C\lambda \rangle > 0$ for all $\lambda \neq 0$. $\square$

Theorem 2.8. *Let $Y_0, Y_1 \dots Y_d$ be real random variables on Ω. Assume that $X = (Y_0, Y_1, \dots Y_d)$ is multi-normal and Y_0 and Y_j are uncorrelated for $j \geq 1$, i.e.,*

$$E[(Y_0 - E[Y_0])(Y_j - E[Y_j])] = 0, \quad 1 \leq j \leq n. \tag{2.70}$$

Then Y_0 is independent of $\{Y_1, \dots Y_m\}$.

Remark 2.9. *Recall that independent random variables are always uncorrelated. But the converse is not true in general: Such random variables X, Y that $A := \{X \neq 0\}$ and $B := \{Y \neq 0\}$ are disjoint, are not independent[3] ($P(A \cap B) \neq P(A)P(B)$). But, if we choose them so that $E(X) = E(Y) = 0$, we obtain $E(XY) = 0$, hence uncorrelated.*

Proof. We have to prove that

$$P[Y_0 \in G_0, Y_1 \in G_1, \dots, Y_d \in G_d] = P[Y_0 \in G_0] \cdot P[Y_1 \in G_1, \dots, Y_d \in G_d], \tag{2.71}$$

for all Borel sets $G_0, \dots G_d \subset \mathbb{R}$. By (2.70), the covariance matrix

$$c_{k,j} = E[(Y_k - E[Y_k])(Y_j - E[Y_j])] \tag{2.72}$$

has only the first entry $c_{0,0}$ non-zero in the first row and first column. Thus, writing $u = (u_0, \vec{u})$ we have

$$\langle u, Cu \rangle = c_{0,0} u_0^2 + \langle \vec{u}, \tilde{C} \vec{u} \rangle. \tag{2.73}$$

Therefore, by (2.62),

$$\begin{aligned} \phi_X(u) &= \exp\Big(-\frac{1}{2}\langle u, Cu \rangle + i\langle u, m \rangle \Big) = \exp\Big(-\frac{1}{2} c_{0,0} u_0^2 - \frac{1}{2}\langle \vec{u}, \tilde{C}\vec{u} \rangle + i\langle u_0, m_0 \rangle + i\langle \vec{u}, \vec{m} \rangle \Big) \\ &= \phi_{Y_0}(u_0)\phi_{(Y_1,\dots,Y_d)}(u_1, \dots, u_d). \end{aligned} \tag{2.74}$$

By Remark 2.5, ϕ_X determines the law uniquely, so (2.71) follows. $\square$

2.7 Proofs of basic properties of the Brownian motion

Let us come back to the Brownian motion, starting at x, which is a family of random variables $\{B_t\}_{t \in [0,\infty)}$ on a probability space $(\Omega, \mathcal{F}, P^x)$.

Theorem 2.10. *$\{B_t\}_{t \in [0,\infty)}$ is a Gaussian process, i.e., for all $0 \leq t_1 \leq \dots \leq t_k$ the random variable $Z = (B_{t_1}, \dots B_{t_k}) \in \mathbb{R}^{dk}$ has a multi-normal distribution with mean and covariance*

$$m = (x, x, \dots, x), \quad c_{(j,\ell),(j',\ell')} = \min(t_j, t_{j'})\delta_{\ell,\ell'}, \tag{2.75}$$

where $j, j' = 1, \dots, k$, $\ell, \ell' = 1, \dots, d$. (In particular, $E^x[(B_{t_i} - x) \cdot (B_{t_j} - x)] = d \min(t_i, t_j)$ in agreement with random walk considerations.)

[3] If two random variables X, Y are independent, then $X^{-1}(F_1)$, $Y^{-1}(F_2)$, $F_1, F_2 \in \mathcal{B}(\mathbb{R}^d)$, are independent events. (We will come to this reformulation of independence later).

Proof. Recall that

$$\begin{aligned}&P^x(B_{t_1}\in F_1,\cdots,B_{t_k}\in F_k)\\&\quad=\int_{F_1\times\cdots\times F_k}p(t_1,x,x_1)\cdots p(t_{k-1}-t_{k-2},x_{k-2},x_{k-1})p(t_k-t_{k-1},x_{k-1},x_k)dx_1\cdots dx_k,\end{aligned}\tag{2.76}$$

where

$$p(t,x,y)=(2\pi t)^{-d/2}\exp\big(-\frac{|x-y|^2}{2t}\big)\qquad\text{for } y\in\mathbb{R}^d, t>0. \tag{2.77}$$

Consequently

$$\begin{aligned}E[\exp(i\langle u,Z\rangle)]&=\int\exp(iu_1B_{t_1}+\cdots+iu_kB_{t_k})P^x(B_{t_1}\in dx_1,\dots B_{t_k}\in dx_k)\\&=\int\exp(iu_1x_1+\cdots+iu_kx_k)p(t_1,x,x_1)\cdots p(t_k-t_{k-1},x_{k-1},x_k)dx_1\cdots dx_k\\&=\int\exp(iu_1x_1+\cdots+iu_kx_k)\tilde p(\Delta t_1,\underbrace{x-x_1}_{z_1})\cdots\tilde p(\Delta t_k,\underbrace{x_{k-1}-x_k}_{z_k})dx_1\cdots dx_k\\&=\mathrm{e}^{iU_1x}\int\exp(iU_1z_1+\cdots+iU_kz_k)\tilde p(\Delta t_1,z_1)\cdots\tilde p(\Delta t_{k-1},z_{k-1})\tilde p(\Delta t_k,z_k)dz_1\cdots dz_k\\&=\mathrm{e}^{iU_1x}\prod_{\tilde m=1}^k\widehat{\tilde p}(\Delta t_{\tilde m},U_{\tilde m})=e^{i(\sum_{j=1}^k u_j)\cdot x-\frac12\sum_{\tilde m=1}^k(\Delta t_{\tilde m})U_{\tilde m}^2},\end{aligned}\tag{2.78}$$

where $\widehat{\tilde p}$ is the Fourier transform of $\tilde p$ (cf. (2.36), $U_{\tilde m}:=\sum_{j=\tilde m}^k u_j$ and $\Delta t_{\tilde m}:=t_{\tilde m}-t_{\tilde m-1}$ $t_0:=0$. From this we read off the mean $m=(x,x,\dots,x)$. Now we read off the covariance:

$$\begin{aligned}\sum_{\tilde m=1}^k\Delta t_{\tilde m}U_{\tilde m}^2&=\sum_{\tilde m=1}^k\Delta t_{\tilde m}(\sum_{j=\tilde m}^k u_j)^2\\&=\sum_{\tilde m=1}^k\Delta t_{\tilde m}\sum_{j,j'=\tilde m}^k u_j\cdot u_{j'}=\sum_{\tilde m=1}^k\Delta t_{\tilde m}\sum_{j,j'=1}^k\chi(j\ge\tilde m)\chi(j'\ge\tilde m)u_j\cdot u_{j'}\\&=\sum_{j,j'=1}^k u_j\cdot u_{j'}\sum_{\tilde m=1}^{\min(j,j')}\Delta t_{\tilde m}=\sum_{j,j'=1}^k u_j\cdot u_{j'}(t_1-t_0+t_2-t_1+\cdots+t_{\min(j,j')})\\&=\sum_{j,j'=1}^k u_j\cdot u_{j'}t_{\min(j,j')}=\sum_{j,j'=1}^k u_j\cdot u_{j'}\min(t_j,t_{j'}),\end{aligned}\tag{2.79}$$

where in the last step we used that the times are ordered $t_1\le t_2\le\cdots\le t_j\le\cdots\le t_k$. Thus, $c_{(j,\ell),(j',\ell')}:=\min(t_j,t_{j'})\delta_{\ell,\ell'}$. □

Lecture 4

Theorem 2.11. $\{B_t\}_{t\in\mathbb{R}_+}$ *has independent increments, i.e.,*

$$B_{t_1},B_{t_2}-B_{t_1},\dots,B_{t_k}-B_{t_{k-1}}, \tag{2.80}$$

are independent for all $0\le t_1<t_2\cdots<t_k$*. Moreover,* $(B_t-B_s)^i\sim N(0,|t-s|)$*. (Upper index* i *in* B_t^i *denotes component, not power.)*

Proof. By Theorem 2.8, it suffices to check that the variables are uncorrelated. By Theorem 2.10 $E^x(\tilde{B}_{t_1}\tilde{B}_{t_2}) = d\min(t_1,t_2)$, where $\tilde{B}_t = B_t - x$. Then,

$$\begin{aligned} E[(B_{t_2} - B_{t_1})^i(B_{t_3} - B_{t_2})^i] &= E[(\tilde{B}_{t_2} - \tilde{B}_{t_1})^i(\tilde{B}_{t_3} - \tilde{B}_{t_2})^i] \\ &= E[\tilde{B}^i_{t_2}\cdot\tilde{B}^i_{t_3}] - E[\tilde{B}^i_{t_2}\cdot\tilde{B}^i_{t_2}] - E[\tilde{B}^i_{t_1}\cdot\tilde{B}^i_{t_3}] + E[\tilde{B}^i_{t_1}\cdot\tilde{B}^i_{t_2}] \\ &= (t_2 - t_2 - t_1 + t_1) = 0 \end{aligned} \tag{2.81}$$

and analogously for other pairs.

As for the last statement: By Theorem 2.7, $(B_t - B_s)^i$, $i = 1,\dots,d$, are Gaussian random variables. We have $E(B_t - B_s)^i = 0$ and

$$E(\{(B_t - B_s)^i\}^2) = E((B^i_t)^2 - 2B^i_tB^i_s + (B^i_s)^2) = t + s - 2\min(t,s) = |t-s|, \tag{2.82}$$

which concludes the proof. □

Now we justify the continuity property (iii) of the Brownian motion:

Theorem 2.12. *(Kolmogorov's continuity theorem). Suppose that the process $\{X_t\}_{t\in\mathbb{R}_+}$ satisfies the following condition: For all $T > 0$ there exist constants $\alpha, \beta, D > 0$ s.t.*

$$E[|X_t - X_s|^\alpha] \le D|t-s|^{1+\beta}, \quad 0 \le s,t \le T. \tag{2.83}$$

Then there exists a continuous version of X.

To verify the assumptions of Theorem 2.12 for $X_t = B_t$, we prove the following lemma:

Lemma 2.13. *There holds, that*

$$E^x[|B_t - B_s|^4] = E^x[\{(B_t - B_s)\cdot(B_t - B_s)\}^2] \le D|t-s|^2. \tag{2.84}$$

for some $D > 0$.

Remark 2.14. *We could replace above E^x with $E^{x=0}$, since $B_t - B_s = \tilde{B}_t - \tilde{B}_s$, where $\tilde{B}_t = B_t - x$.*

Proof. Let $X^i := B^i_t - B^i_s \sim N(0,|t-s|)$, cf. Theorem 2.11. Then $Z^i = \frac{1}{|t-s|^{1/2}}X^i \sim N(0,1)$, cf. (2.27). We have

$$E^x[|B_t - B_s|^4] = E^x[|X|^4] = |t-s|^2E^x[\{\sum_{i=1}^d (Z^i)^2\}^2]. \tag{2.85}$$

The last factor is finite, since

$$E^x[\{\sum_{i=1}^d (Z^i)^2\}^2] = \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d} dz\{\sum_{i=1}^d (z^i)^2\}^2 \mathrm{e}^{-\frac{1}{2}|z|^2}. \tag{2.86}$$

This concludes the proof. (Actually, (2.86) $= d(d+2)$ see HS3). □

We summarize our discussion of the Brownian motion with the following axiomatic definition of a particular version of the Brownian motion:

Theorem 2.15. *Let $\Omega = C(\mathbb{R}_+;\mathbb{R}^d)$ be equipped with the topology of uniform convergence on compact sets and let $\mathcal{B}$ be the resulting Borel σ-algebra. Let $B_t(\omega) = \omega(t)$ be the coordinate processes. For any $x \in \mathbb{R}^d$ there exists a unique probability measure P^x on $(\Omega,\mathcal{B})$ such that*

1. *$P(B_0 = x) = 1$.*
2. *$(B_t)_{t\in\mathbb{R}_+}$ has independent increments.*
3. *For all $0 \le s < t$, $B_t - B_s \sim N(0,(t-s)I_d)$, where I_d is the $d\times d$ identity matrix.*

It is called the canonical Brownian motion.

For proof cf. [Pa, Chapter VII].

2.8 Brownian motion as a path integral

Recall the formula

$$
\begin{aligned}
&E^x(f(B_{t_1},\dots,B_{t_k}))\\
&=\int f(B_{t_1}(\omega),\dots,B_{t_k}(\omega))dP^x(\omega)=\int f(x_1,\dots,x_k)p(t_1,x,x_1)\cdots p(t_k-t_{k-1},x_{k-1},x_k)dx_1\cdots dx_k,
\end{aligned}
$$

where

$$
\prod_{\ell=1}^{k} p(t_\ell-t_{\ell-1},x_{\ell-1},x_\ell)=\Big[\prod_{\ell'=1}^{k}\frac{1}{(2\pi(t_{\ell'}-t_{\ell'-1}))^{d/2}}\Big]\exp\Big(-\sum_{\ell=1}^{k}\frac{|x_{\ell-1}-x_\ell|^2}{2(t_\ell-t_{\ell-1})^2}(t_\ell-t_{\ell-1})\Big). \tag{2.87}
$$

Suppose $t_{\ell'}-t_{\ell'-1}=\varepsilon$ and we consider the heuristic continuum limit $\varepsilon\to 0$ and $k\to\infty$ for $0\le t\le T$. In this limit:

- f depends on B_t for all $t\in[0,T]$, thus becomes a function of a path starting at x

$$
f(B_{t_1}(\omega),\dots,B_{t_k}(\omega))=f(\omega(t_1),\dots,\omega(t_k))\to f(\{\omega(t)\}_{t\in[0,T]})=f(\omega). \tag{2.88}
$$

- $\exp\big(-\sum_{\ell=1}^{k}\frac{|x_{\ell-1}-x_\ell|^2}{2(t_\ell-t_{\ell-1})^2}(t_\ell-t_{\ell-1})\big)\to\exp\big(-\int_0^T\frac12\big(\frac{d\omega(t)}{dt}\big)^2dt\big)$.
- $\big[\prod_{\ell'=1}^{k}\frac{dx_{\ell'}}{(2\pi(t_{\ell'}-t_{\ell'-1}))^{d/2}}\big]\to\frac{1}{\mathcal N}D\omega$, where $D\omega$ is (non-existent) infinite product of Lebesgue measures and $\mathcal N$ is a normalization constant.
- Altogether,

$$
\int f(B_{t_1}(\omega),\dots,B_{t_k}(\omega))dP^x(\omega)\to\int_{\omega(0)=x} f(\omega)\frac{1}{\mathcal N}\exp\left(-\int_0^T\frac12\left(\frac{d\omega(t)}{dt}\right)^2dt\right)D\omega, \tag{2.89}
$$

  which is an example of a path integral.

- $D\omega$ does not exist and the paths ω are a.s. not differentiable. But the product $\frac{1}{\mathcal N}\exp\left(-\int_0^T\frac12\big(\frac{d\omega(t)}{dt}\big)^2dt\right)D\omega$ makes sense as a Gaussian measure on the Banach space $C([0,T];\mathbb{R}^d)$.
- Since $\int dP^x(\omega)=1$,

$$
\mathcal N=\int\exp\left(-\int_0^T\frac12\left(\frac{d\omega(t)}{dt}\right)^2dt\right)D\omega. \tag{2.90}
$$

  Thus (2.96) can be written as a quotient of two path integrals.

Connection to quantum mechanics:

- Let $f(\omega)=\delta_y(\omega(T))$ so that it fixes the final value of the path to y. Then, the quantity

$$
K_E(T;y,x):=\int_{\omega(0)=x}^{\omega(T)=y}\frac{1}{\mathcal N}\exp\left(-\int_0^T\frac12\left(\frac{d\omega(t)}{dt}\right)^2dt\right)D\omega \tag{2.91}
$$

  is called the Euclidean (imaginary time) propagator (or Green function). It is a propagator of the heat equation, i.e.:

$$
(\partial_T-\frac12\Delta_x)K_E(T;y,x)=\delta(T)\delta(x-y). \tag{2.92}
$$

- Let us change variables $t = i\tau$, $T = i\mathcal{T}$ and interpret $q(\tau) := \omega(i\tau)$ as the position of the quantum particle. Then the (real time) propagator is given by

$$K(\mathcal{T}; y, x) := \int_{q(0)=x}^{q(\mathcal{T})=y} \frac{1}{\mathcal{N}'} \exp\left(i \int_0^{\mathcal{T}} \frac{1}{2}\left(\frac{dq(\tau)}{d\tau}\right)^2 d\tau\right) Dq. \tag{2.93}$$

  This is the propagator of the Schrödinger equation, i.e.:

$$(i\partial_{\mathcal{T}} + \frac{1}{2}\Delta_x) K(\mathcal{T}; y, x) = \delta(\mathcal{T})\delta(x - y). \tag{2.94}$$

- Consider a free quantum mechanical particle described at $\tau = 0$ by a wave function $\mathbb{R}^d \ni x \mapsto \psi_0(x)$. Then, the probability of finding the particle in the region $A \subset \mathbb{R}^d$ is $\int_A |\psi_0(x)|^2 dx$.

- Making use of (2.94), we obtain that at time $\tau = \mathcal{T} > 0$ the quantum mechanical particle is described by the wave function

$$\psi_{\mathcal{T}}(y) := \int K(\mathcal{T}; y, x)\psi_0(x) dx. \tag{2.95}$$

  Then, the probability of finding the particle in the region $A \subset \mathbb{R}^d$ is $\int_A |\psi_{\mathcal{T}}(y)|^2 dy$. (These quantum probabilities do not come from the probability space $(\Omega, \mathcal{B}, P^x)$. There is no generally accepted probability space for quantum mechanics, but there are some proposals, called hidden variables theories).

- More generally, for a quantum mechanical particle moving in an external potential V

$$K_V(T; y, x) := \int_{q(0)=x}^{q(\mathcal{T})=y} \frac{1}{\mathcal{N}'} \exp\big(iS[q]\big)\big) Dq, \tag{2.96}$$

  where $S[q] := \int_0^T [\frac{1}{2}\big(\frac{dq(\tau)}{d\tau}\big)^2 - V(q(\tau))] d\tau$ is called action. The minimum of the action is given by the trajectory of the classical particle in potential V. It satisfies the Newton equations

$$\frac{d^2 q(\tau)}{d\tau^2} = -(\nabla V)(q(\tau)). \tag{2.97}$$

  This is a special case of a certain paradigm in physics: Suppose we are given a theory which in the regime of small fluctuations (i.e. variances) is described by a variational principle (like minimizing $q \mapsto S[q]$). Then, using path integrals of schematic form (2.96), one can cover the case of larger fluctuations. This applies to the step from classical mechanics to quantum mechanics. Similar relation holds between thermodynamics and statistical physics.

3 Stochastic integral

Unless stated otherwise, in this section we consider Brownian motion with values in $\mathbb{R}$ starting at $x = 0$. We will write $P = P^{x=0}$, $E = E^{x=0}$.

3.1 Motivation

Recall the risky investment equation from (1.9)

$$\frac{dX_s}{ds} = \left(r + \alpha \frac{dB_s}{ds}\right) X_s, \quad r, \alpha > 0. \tag{3.1}$$

As we suspect from (1.6) that $s \mapsto B_s(\omega)$ is not differentiable in the usual sense, let us integrate both sides w.r.t $s \in [0,t]$:

$$X_t = X_0 + \int_0^t rX_s ds + \text{"} \int_0^t \alpha X_s dB_s\text{"}. \tag{3.2}$$

This is a special case of a general class of equations

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \text{"} \int_0^t \sigma(s, X_s) dB_s\text{"}. \tag{3.3}$$

How to give meaning to the last integral?

Lecture 5

3.2 Riemann-Stieltjes (RS) integral

RS integral is the first possibility which comes to mind. It gives meaning to integrals of the form

$$\int_a^b f(s) dg(s) \tag{3.4}$$

e.g. when f is continuous on $[a,b]$ [4] and g has bounded total variation.

1. Def. For a partition $\Pi := \{a = s_0 < s_1 < \cdots < s_n = b\}$ write $|\Pi| = \max_{1 \le i \le n} |s_i - s_{i-1}|$ and $\Delta g_i := g(s_i) - g(s_{i-1})$.
2. Def. The variation of g on $[a,b]$ over Π is

$$V_a^b(g, \Pi) := \sum_{i=1}^n |\Delta g_i| \in [0, \infty). \tag{3.5}$$

3. Def. The total variation of g on $[a,b]$ is

$$V_a^b(g) := \sup_{\Pi} V_a^b(g, \Pi) \in [0, \infty]. \tag{3.6}$$

   We say that g has bounded total variation if $V_a^b(g) < \infty$.

4. Examples:
   - Any monotone function (continuous or not) is of bounded total variation. E.g. suppose g is increasing. Then, for any Π

$$\sum_{i=1}^n |\Delta g_i| = \sum_{i=1}^n (\Delta g_i) = g(b) - g(a). \tag{3.7}$$

   - Fact. (Jordan decomposition) A function g of bounded total variation can be decomposed $g = g_+ - g_-$, where $g_\pm$ are increasing. Recall that any monotonous function is differentiable almost everywhere w.r.t. the Lebesgue measure. This is the first indication that RS-integral may not be a good stochastic integral, since we don't expect $s \mapsto B_s(\omega)$ to be differentiable in the usual sense, cf. (1.6).

[4] That is, f is continuous in (a,b), the limits $\lim_{s \downarrow a} f(s)$ resp. $\lim_{s \uparrow b} f(s)$ exist and equal $f(a)$ resp. $f(b)$.

- A continuous function may have an unbounded total variation: E.g.

$$g(s) = \begin{cases} s\sin\left(\frac{1}{s}\right), & s \neq 0, \\ 0, & s = 0. \end{cases} \tag{3.8}$$

on the interval $[-1,1]$. (Homework).

- If $g \in C^1(\mathbb{R})$ then $V_a^b(g) = \int_a^b |g'(s)|ds < \infty$. (Homework). In this case

$$\int_a^b f(s)dg(s) = \int_a^b f(s)g'(s)ds, \tag{3.9}$$

so RS integral is reduced to a Riemann integral.

5. A tagged partition of $[a,b]$ is a pair (Π, ξ) consisting of a partition Π together with a choice of tags

$$\xi_i \in [s_{i-1}, s_i], \qquad i = 1, \dots, n. \tag{3.10}$$

6. Def. Given two tagged partitions Π, Π' we define their common refinement $\hat{\Pi}$ that is the union of all the points in increasing order $\{a = \hat{s}_0 < \hat{s}_1 < \dots < \hat{s}_n = b\}$ (Picture).

   Each interval $\hat{I} = [\hat{s}_{\ell-1}, \hat{s}_\ell]$ of $\hat{\Pi}$ lies inside exactly one interval of $I \in \Pi$ resp $I' \in \Pi'$. We define $\xi(\hat{I})$ resp. $\xi'(\hat{I})$ to be the corresponding tags $\xi_i \in I$ resp. $\xi_{i'} \in I'$. (Note that $\xi(\hat{I})$, $\xi'(\hat{I})$ may not belong to $\hat{I}$ and we don't consider $\hat{\Pi}$ a tagged partition).

   Suppose that $s'_{i'} \in [s_{i-1}, s_i]$. Then

$$\begin{aligned} f(\xi_i)(g(s_i) - g(s_{i-1})) &= f(\xi_i)(g(s_i) - g(s'_{i'}) + g(s'_{i'}) - g(s_{i-1}) \\ &= f(\xi(\hat{I}_1))\Delta g(\hat{I}_1) + f(\xi(\hat{I}_2))\Delta g(\hat{I}_2), \end{aligned} \tag{3.11}$$

   where $\Delta g(\hat{I}) = g(s_i) - g(s_{i-1})$. Consequently

$$\sum_{i=1}^{n} f(\xi_i)(g(s_i) - g(s_{i-1})) = \sum_{\hat{I}\in\hat{\Pi}} f(\xi(\hat{I}))\Delta g(\hat{I}), \tag{3.12}$$

$$\sum_{i'=1}^{n'} f(\xi'_{i'})(g(s_{i'}) - g(s_{i'-1})) = \sum_{\hat{I}\in\hat{\Pi}} f(\xi'(\hat{I}))\Delta g(\hat{I}). \tag{3.13}$$

Theorem 3.1. *Suppose that f is continuous on $[a,b]$ and g has bounded total variation. Then the following limit exists and defines the RS integral:*

$$\int_a^b f(s)dg(s) = \lim_{|\Pi|\to 0} \sum_{i=1}^{n} f(\xi_i)(g(s_i) - g(s_{i-1})). \tag{3.14}$$

The limit is over a sequence of (finer and finer) tagged partitions (Π, ξ) and is independent of the choice of the tags within the partitions. (For stochastic integrals tags will matter).

Proof.

1. For a tagged partition (Π, ξ) with tags $\xi_i \in [s_{i-1}, s_i]$, define the RS sum

$$S(f, g; \Pi, \xi) := \sum_{i=1}^{n} f(\xi_i)\,\Delta g_i, \qquad |\Pi| := \max_{1\le i\le n}(s_i - s_{i-1}). \tag{3.15}$$

2. Let $w_f(\delta) := \sup\{|f(s_1) - f(s_2)| : |s_1 - s_2| \le \delta,\ s_1, s_2 \in [a,b]\}$ be the modulus of continuity of f. Since f is continuous on $[a,b]$, $w_f(\delta) \to 0$ as $\delta \downarrow 0$.

3. If (Π, ξ) and (Π', ξ') are two tagged partitions with $\max\{|\Pi|, |\Pi'|\} \le \delta$, let $\widehat{\Pi}$ be a common refinement. Then, by (3.12), (3.13),

$$S(f, g; \Pi, \xi) - S(f, g; \Pi', \xi') = \sum_{\hat{I} \in \widehat{\Pi}} \left(f(\xi(\hat{I})) - f(\xi'(\hat{I})) \right) \Delta g(\hat{I}), \tag{3.16}$$

hence

$$\left| S(f, g; \Pi, \xi) - S(f, g; \Pi', \xi') \right| \le w_f(\delta) \sum_{\hat{I} \in \widehat{\Pi}} |\Delta g(\hat{I})| \le w_f(\delta)\, V_a^b(g). \tag{3.17}$$

4. Given $\varepsilon > 0$, choose $\delta > 0$ so that $w_f(\delta)\, V_a^b(g) < \varepsilon$. Then any two sums with mesh at most δ differ by less than ε. The net $S(f, g; \Pi, \xi)$ is Cauchy as $|\Pi| \to 0$ and therefore convergent. Define

$$\int_a^b f\, dg := \lim_{|\Pi| \to 0} S(f, g; \Pi, \xi), \tag{3.18}$$

which is independent of the choice of tags. (To see this last point choose in (3.17) $\Pi = \Pi'$ but ξ different from ξ' and observe that the difference goes to zero in the limit $|\Pi| \to 0$.)

3.3 Why Riemann-Stjeltjes integral is not a good stochastic integral?

Short answer: Because the Brownian motion has unbounded total variation.

Theorem 3.2. *Let $(B_s)_{s \in [0,T]}$ be the variant of Brownian motion with continuous paths. Then*

$$V_0^T(B) := \sup_{\Pi} \sum_i |B_{s_i} - B_{s_{i-1}}| = \infty \quad P\text{-almost surely.} \tag{3.19}$$

Remark 3.3. *If g is of unbounded total variation, there exist continuous f for which the RS integral $\int_a^b f(s) dg(s)$ does not exist. (We skip the proof of this fact). Thus, $\Omega \ni \omega \mapsto \int f(s) dB_s(\omega)$, understood as a family of RS integrals, is not a well defined random variable. In fact, even if f can be integrated against $dB_s(\omega_1)$, it may not be possible to integrate it against $dB_s(\omega_2)$ for some $\omega_2 \neq \omega_1$ forming a set of non-zero P-measure.*

To prove Theorem 3.2 we need some preparatory results:

Lemma 3.4. *Let $g : [0,T] \to \mathbb{R}$ be continuous and s.t. $V_0^T(g) < \infty$. Then,*

$$\lim_{|\Pi| \to 0} \sum_{i=1}^{n} \left(g(s_i) - g(s_{i-1}) \right)^2 = 0. \tag{3.20}$$

Proof. For any partition $\Pi = \{0 = s_0 < s_1 < \cdots < s_n = T\}$,

$$\sum_{i=1}^{n} \left(g(s_i) - g(s_{i-1}) \right)^2 \le \Big(\max_{1 \le i \le n} |g(s_i) - g(s_{i-1})| \Big) \sum_{i=1}^{n} |g(s_i) - g(s_{i-1})|. \tag{3.21}$$

Since g is continuous, $\max_i |g(s_i) - g(s_{i-1})| \to 0$ as $|\Pi| \to 0$, while $\sum_i |g(s_i) - g(s_{i-1})| \le V_0^T(g)$. Hence, (3.20) holds. $\square$

Theorem 3.5. *(Kolmogorov's Strong Law of Large Numbers) Let $(X_n)_{n\geq 1}$ be independent and identically distributed (i.i.d) real-valued random variables such that*

$$E[X_1^2] < \infty. \tag{3.22}$$

Define the average

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k. \tag{3.23}$$

Then

$$\bar{X}_n \longrightarrow E[X_1] \quad P\text{-almost surely.} \tag{3.24}$$

That is, $\bar{X}_n(\omega) \to E[X_1]$ for $\omega \in \Omega \backslash N$, where $P(N) = 0$. In other words, $P(\omega \in \Omega : \bar{X}_n(\omega) \to E[X_1]) = 1$.

We skip the proof of SLLN. It can be proven, e.g., using Borel-Cantelli Lemma from HS1, Problem 7.

Proof of Theorem 3.2.

1. Let $\Pi_n = \{kT/2^n : k = 0, \dots, 2^n\}$ and define

$$S_n := \sum_{k=1}^{2^n} \left(B_{kT/2^n} - B_{(k-1)T/2^n}\right)^2. \tag{3.25}$$

The increments $\left(B_{kT/2^n} - B_{(k-1)T/2^n}\right)$ are independent and distributed as $N(0, T/2^n)$ by Theorem 2.11. Write $Z_k := \sqrt{\frac{2^n}{T}}\left(B_{kT/2^n} - B_{(k-1)T/2^n}\right) \sim N(0,1)$, cf. (2.27). Then

$$\frac{2^n}{T} S_n = \sum_{k=1}^{2^n} Z_k^2 \quad \Rightarrow \quad S_n = T\left[\frac{1}{2^n}\sum_{k=1}^{2^n} Z_k^2\right]. \tag{3.26}$$

By the SLLN (Theorem 3.5),

$$\frac{1}{2^n}\sum_{k=1}^{2^n} Z_k^2 \to E[Z_1^2] = 1 \quad P\text{-almost surely,} \tag{3.27}$$

and therefore

$$S_n \to T \neq 0. \quad P\text{-almost surely.} \tag{3.28}$$

By comparing this with (3.20), we will now try to obtain a contradiction.

2. Suppose there exists a set $A \subset \Omega$ of positive probability s.t. for all $\omega \in A$ we have $V_0^T(B(\omega)) < \infty$. Lemma 3.4 applied to the continuous path $s \mapsto B_s(\omega)$ gives

$$\lim_{n\to\infty} \sum_{k=1}^{2^n} \left(B_{kT/2^n}(\omega) - B_{(k-1)T/2^n}(\omega)\right)^2 = 0. \tag{3.29}$$

But by (3.28) this limit equals $T > 0$ almost surely, a contradiction. Hence $P(A) = 0$. It follows that

$$P\left(V_0^T(B) = \infty\right) = 1, \tag{3.30}$$

which proves (3.19). □

For future reference, we note the following lemma:

Lemma 3.6. *$S_n \to T$ in $L^2(\Omega, P)$.*

See HS6 for a proof.

3.4 Itô integral. Introduction

Suppose $0 \le S < T$ and $\mathbb{R}_+ \times \Omega \ni (t,\omega) \mapsto f(t,\omega)$ is given.

1. We want to define $\int_S^T f(t,\omega)dB_t(\omega)$ as a random variable.
2. First assume that f has the form:

$$\varphi(t,\omega) = \sum_{j\ge 0} e_j(\omega)\chi_{[j2^{-n},(j+1)2^{-n}]}(t), \tag{3.31}$$

   that is, it is a step function in t with some unspecified dependence on ω.

3. For such functions we define

$$\int_S^T \varphi(t,\omega)dB_t(\omega) = \sum_{j\ge 0} e_j(\omega)[B_{t_{j+1}} - B_{t_j}](\omega), \tag{3.32}$$

   where

$$t_j := t_j^{(n)} := \begin{cases} j2^{-n} & \text{if} \quad S \le j2^{-n} \le T, \\ S & \text{if} \quad j2^{-n} < S, \\ T & \text{if} \quad j2^{-n} > T. \end{cases} \tag{3.33}$$

4. Example: Let us approximate $f(t,\omega) = B_t(\omega)$ by step functions in two different ways:

$$\varphi_1(t,\omega) = \sum_{j\ge 0} B_{t_j}(\omega)\chi_{[t_j,t_{j+1}]}(t), \tag{3.34}$$

$$\varphi_2(t,\omega) = \sum_{j\ge 0} B_{t_{j+1}}(\omega)\chi_{[t_j,t_{j+1}]}(t). \tag{3.35}$$

   Then

$$E\left[\int_0^T \varphi_1(t,\omega)dB_t(\omega)\right] = \sum_{j>0} E[B_{t_j}(B_{t_{j+1}} - B_{t_j})] = 0 \tag{3.36}$$

   by independence of increments (Theorem 2.11) and $E(B_t) = 0$. But

$$\begin{aligned} E\left[\int_0^T \varphi_2(t,\omega)dB_t(\omega)\right] &= \sum_{j>0} E[B_{t_{j+1}}(B_{t_{j+1}} - B_{t_j})] \\ &= \sum_{j>0} E[B_{t_j}(B_{t_{j+1}} - B_{t_j})] \\ &+ \sum_{j>0} E[(B_{t_{j+1}} - B_{t_j})^2] = T, \end{aligned} \tag{3.37}$$

   also by Theorem 2.11. So two reasonable approximations to f remain far apart, also for large n. We see that variations of the paths are too large to define the integral in the RS sense. This example corroborates Theorem 3.2.

5. In order to define $\int_S^T f(s,\omega)dB_s(\omega)$ we approximate f by step functions:

$$f(t,\omega) \simeq \sum_j f(\xi_j,\omega)\chi_{[t_j,t_{j+1})}(t), \tag{3.38}$$

   where $\xi_j \in [t_j, t_{j+1}]$. The following choices proved useful:

- $\xi_j = t_j$ (the left end point). If the limit $n \to \infty$ exists (in the sense to be specified below), it is called the *Ito integral*, from now on denoted by

$$\int_S^T f(t,\omega) dB_t(\omega). \tag{3.39}$$

- $\xi_j = (t_j + t_{j+1})/2$ (the mid point), which leads to the *Stratonovich integral* denoted by

$$\int_S^T f(t,\omega) \circ dB_t(\omega). \tag{3.40}$$

Remark 3.7. *As mentioned in the first lecture, the heuristic equation*

$$\frac{dX_t}{dt} = \left(r + \alpha \frac{dB_t}{dt}\right) X_t, \quad r, \alpha > 0 \tag{3.41}$$

can be given meaning in two different ways with solutions

$$X_t = X_0 \mathrm{e}^{(r - \frac{1}{2}\alpha^2)t + \alpha B_t}, \quad X_t = X_0 \mathrm{e}^{rt + \alpha B_t}. \tag{3.42}$$

They correspond to the choice of the Itô and Stratonovich integral, respectively.

Lecture 6

3.5 Itô integral. Preparations

Our goal now is to specify the class of functions, which we want to integrate. The first step is to prevent certain pathologies, cf. Remark 3.9 below.

3.5.1 Completeness of probability spaces

1. Def. A probability space $(\Omega, \mathcal{F}, P)$ is called complete[5], if $\mathcal{F}$ contains all subsets of P-null sets[6].

2. Fact: Any probability space can be made complete by adding to $\mathcal{F}$ all subsets of sets of measure 0 and by demanding that P is zero on these subsets (see HS6).

3. Remark: We assume that the probability space $(\Omega, \mathcal{F}, P)$ of the Brownian motion has been completed. This does not change the properties (i),(ii),(iii) of Section 2.5, as all of them are reduced to computing expectations (cf. Theorem 2.12 for (iii)). Integrals w.r.t. dP do not change by adding sets of measure zero.

4. Warning: $([0,1], \mathcal{B}([0,1]), dx)$ is not complete. Its completion is called the Lebesgue probability space. The Cantor set C is to blame for this, see HS6. It has Lebesgue measure zero but cardinality $\mathfrak{c}$ (continuum). Thus the set of all subsets of C has cardinality $2^{\mathfrak{c}}$. But $\mathcal{B}([0,1])$ has only cardinality $\mathfrak{c}$, equal to the cardinality of the set $\mathcal{O}$ of open sets in $[0,1]$. Some ideas behind the latter statement:

    - The set $\mathcal{I}$ of open intervals with rational endpoints has cardinality of natural numbers $\aleph_0$. Any open set is a union of such intervals. So the cardinality of $\mathcal{O}$ is equal to the cardinality of the set of subsets of $\mathcal{I}$ which is $2^{\aleph_0} = \mathfrak{c}$.
    - There is a countable algorithm of generating Borel sets from open sets (iterate countable sums, countable intersections and complements). Thus the cardinality stays $\mathfrak{c}$.

[5] This has nothing to do with the functional analytic concept of complete spaces, i.e. s.t. all Cauchy sequences converge.

[6] $N \subset \Omega$ is a P-null set if $N \in \mathcal{F}$ and $P(N) = 0$.

For similar reasons the canonical Brownian motion of Theorem 2.15 does not have a complete probability space.

5. Remark: In the context of Itô integration it is important to complete the probability space $(\Omega, \mathcal{F}, P)$ containing the *domain* Ω of the relevant random variables (e.g. $\Omega \ni \omega \mapsto B_t \in \mathbb{R}$) cf. Remark 3.9 below. It is not necessary to complete their *range*, which is usually $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$. That's why we can live with Borel sigma algebras.

3.5.2 σ-algebras generated by families of sets and random variables

Let us start with some motivating remarks:

- Heuristically, the idea of the Itô integral ("the left end point") is to keep the random variable $\omega \mapsto f(t, \omega)$ independent of the increment ΔB_t by which it is multiplied.
- But, assuming just this, would be too weak. Abstractly, the problem is the following: Even if random variables X and Y are, separately, independent of Z, their product XY may not be independent of Z. This would undermine the proof of a crucial Itô isometry property (Lemma 3.11 below) which is an L^2 property.
- Therefore, we will formulate a condition on f which is slightly stronger than independence of ΔB_t, but stable under taking products. Heuristically, this condition will say that f is a function of $B_s(\omega)$, $0 \leq s \leq t$.

Now we move on to formulating this condition precisely:

1. Def. Let $\mathcal{A} \subset \Omega$ be some collection of sets. Then, $\sigma[\mathcal{A}]$ denotes the smallest σ-algebra containing $\mathcal{A}$. It is called the σ-algebra generated by $\mathcal{A}$.

2. Def. Let $X : \Omega \to \mathbb{R}$ be a random variable on $(\Omega, \mathcal{F}, P)$. Then $\sigma(X)$ denotes the smallest σ-algebra w.r.t. which X is measurable. It is called the σ-algebra generated by X.

   In other words, it is the smallest σ-algebra on Ω containing

   $$X^{-1}(A) := \{\, \omega \in \Omega \,:\, X(\omega) \in A \,\}, \quad A \in \mathcal{B}(\mathbb{R}). \tag{3.43}$$

   Equivalently, $\sigma(X) = \sigma[X^{-1}(A), A \in \mathcal{B}(\mathbb{R})]$.

3. Fact: (Doob-Dynkin lemma). If $X, Y : \Omega \to \mathbb{R}$ are two given functions then Y is $\sigma(X)$-measurable iff there exists a Borel measurable function $g : \mathbb{R} \to \mathbb{R}$ s.t. $Y = g(X)$. (Thus measurability w.r.t. $\sigma(X)$ is a rather restrictive condition).

4. Fact. A collection of random variables $\{X_i \,:\, i \in \mathcal{T}\}$ is independent if the collection of generated σ-algebras $\{\, \sigma(X_i) \,:\, i \in \mathcal{T}\}$, is independent. That is, for any $A_{i_1} \in \sigma(X_{i_1}), \ldots A_{i_k} \in \sigma(X_{i_k})$, where $i_1, \ldots, i_k$ is a finite collection of distinct indices,

   $$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}). \tag{3.44}$$

5. Def. Let $\{X_t\}_{t \in [0,T]}$, be a family of random variables (stochastic process)[7]. The σ-algebra generated by $\{X_t\}_{t \in [0,T]}$, denoted $\sigma(\{X_t\}_{t \in [0,T]})$, is the smallest σ-algebra w.r.t. which all these random variables are measurable.

   In other words, it is the smallest σ-algebra on Ω containing

   $$\bigcup_{t \in [0,T]} \sigma(X_t). \tag{3.45}$$

   Equivalently, $\sigma(\{X_t\}_{t \in [0,T]}) = \sigma\big[\bigcup_{t \in [0,T]} \sigma(X_t)\big]$.

[7] We admit $T = \infty$ in which case $[0, T]$ stands for $[0, \infty)$.

6. Def. $\mathcal{F}_t^X := \sigma(\{X_s\}_{s\in[0,t]})$, $t \in [0,T]$, is called the natural filtration of the stochastic process $\{X_t\}_{t\in[0,T]}$.

7. Def. More abstractly, a filtration is a family of σ-algebras $\{\mathcal{G}_t\}_{t\in[0,T]}$, $\mathcal{G}_t \subset \mathcal{F}$, s.t. $\mathcal{G}_s \subset \mathcal{G}_t$ for $s \leq t$.

8. Def. A process $\{Y_t\}_{t\in[0,T]}$ is called $\{\mathcal{G}_t\}_{t\in[0,T]}$-adapted if for each $s \in [0,T]$ the function Y_s is $\mathcal{G}_s$-measurable.

   For example, $Y_s := X_{s/2}$ is $\mathcal{F}_t^X$-adapted, but $\tilde{Y}_s := X_{2s}$ not necessarily.

9. Def. Suppose $(\Omega, \mathcal{F}, P)$ is complete and $\mathcal{N} := \{N \in \mathcal{F} : P(N) = 0\}$. The completion of the filtration $\{\mathcal{G}_t\}_{t\in[0,T]}$ is a new filtration defined by

$$\mathcal{G}_t^{\text{comp}} := \sigma(\mathcal{G}_t \cup \mathcal{N}). \tag{3.46}$$

10. Def. The completion of the natural filtration of the Brownian motion will be denoted $\{\mathcal{F}_t\}_{t\in[0,T]}$. (Completing the filtration helps avoiding pathologies in the construction of the Itô integral, cf Remark 3.9).

11. Fact: A function $f : \Omega \to \mathbb{R}$ is $\mathcal{F}_t$-measurable iff it can be written as the pointwise limit (for a.a. ω) of sums of functions of the form

$$g_1(B_{t_1}) \cdots g_k(B_{t_k}), \tag{3.47}$$

    where $g_1, \ldots g_k$ are bounded continuous and $t_j \leq t$ for $j \leq k$. Can be proven using the Doob-Dynkin lemma. Intuitively, values of $f(\omega)$ can be computed from values of $B_s(\omega)$ for $s \leq t$. Therefore, we think of $\mathcal{F}_t$ as "the history of $\{B_s\}_{s\in\mathbb{R}_+}$ up to time t".

12. Fact: If Y_t is $\mathcal{F}_t$-measurable then it is independent of $B_s - B_t$, $s \geq t$. (Follows from the previous item and independence of increments.)

3.5.3 Admissible functions f

General strategy to identify f for which $\int_S^T f(t,\omega)dB_t(\omega)$ makes sense:

- $\{B_t\}_{t\in\mathbb{R}_+}$ determines the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$.
- We admit only f which are adapted to this filtration, i.e., $f(t,\cdot)$ is $\mathcal{F}_t$-measurable.
- We also require that the resulting random variable has finite variance. (Recall that the quadratic variation of the Brownian motion (3.25) has a finite limit in expectation, cf. Lemma 3.6. This will stabilize the variance.)

Now formally:

Definition 3.8. *Let $\mathcal{V}(S,T)$ be the class of functions*

$$(t,\omega) \mapsto f(t,\omega) : \mathbb{R}_+ \times \Omega \to \mathbb{R} \tag{3.48}$$

such that

(i) f *is* $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$*-measurable.*

(ii) f *is* $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$*-adapted.*

(iii) $E\big[\int_S^T f(t,\cdot)^2 dt\big] < \infty$.

Remark 3.9. *If $(\Omega, \mathcal{F}, P)$ and $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$ were not complete, by changing the process $\{f(t,\cdot)\}_{t\in\mathbb{R}_+}$ on a set of measure zero we could escape from the class $\mathcal{V}(S,T)$:*

- *Let us start with a process $f_0 \equiv 0$, which is certainly in $\mathcal{V}(S,T)$.*
- *Choose a P-null set $N \in \mathcal{F}$ with $A \subset N$ such that $A \notin \mathcal{F}$, hence $A \notin \mathcal{F}_t$ for any t.*
- *Set $f_1(\omega, t) = \chi_A(\omega)\chi_{[0,1]}(t)$.*
- *For $t = 0$ the inverse image of 1 is A which is not in $\mathcal{F}$. Thus f_1 is not measurable, although $f_1 = f_0$ except on the set N of P-measure zero. Thus equality of two functions 'almost everywhere' does not guarantee that they are both measurable. This is one pathology of probability spaces $(\Omega, \mathcal{F}, P)$ which are not complete. (We note, however, that f_1 is not a version[8] of f_0, since a version is by definition measurable).*
- *If we completed $(\Omega, \mathcal{F}, P)$, the process f_1 would become measurable, but still it could happen that $A \notin \mathcal{F}_0$, in which case f_1 is not adapted. To ensure that f_1 is adapted, we have to complete also the filtration. In other words, one pathology of incomplete filtrations is that we can destroy adaptedness by changing a version of the process.*

Lecture 7.

3.6 Itô integral. Definition

1. Def. A function $\varphi \in \mathcal{V}(S,T)$ is called elementary if it has the form

$$\varphi(t,\omega) = \sum_{j\geq 0} e_j(\omega)\chi_{[t_j,t_{j+1})}(t). \tag{3.49}$$

   In particular, since φ is adapted, e_j must be $\mathcal{F}_{t_j}$-measurable. For elementary functions, we set

$$\int_S^T \varphi(t,\omega)dB_t(\omega) = \sum_{j\geq 0} e_j(\omega)[B_{t_{j+1}} - B_{t_j}](\omega). \tag{3.50}$$

2. Def. Let $f \in \mathcal{V}(S,T)$. Then the Itô integral of f is defined by

$$\int_S^T f(t,\omega)dB_t(\omega) = \lim_{n\to\infty} \int_S^T \varphi_n(t,\omega)dB_t(\omega), \quad \text{(limit in } L^2(\Omega,P)), \tag{3.51}$$

   where $\{\varphi_n\}_{n\in\mathbb{N}}$ is a sequence of elementary functions s.t.

$$E\left[\int_S^T |f(t,\cdot) - \varphi_n(t,\cdot)|^2 dt\right] \to 0 \quad \text{as } n \to \infty. \tag{3.52}$$

   Theorem 3.10. *(Approximation theorem). For any $f \in \mathcal{V}(S,T)$, a sequence $\{\varphi_n\}_{n\in\mathbb{N}}$ satisfying (3.51), (3.52) exists.*

   We will prove this theorem in Subsection 3.8 below.

3. Recall example above, where we approximated $f(t,\omega) = B_t(\omega)$ by step functions in two different ways:

$$\varphi_1(t,\omega) = \sum_{j\geq 0} B_{t_j}(\omega)\chi_{[t_j,t_{j+1})}(t), \tag{3.53}$$

$$\varphi_2(t,\omega) = \sum_{j\geq 0} B_{t_{j+1}}(\omega)\chi_{[t_j,t_{j+1})}(t). \tag{3.54}$$

   We note that φ_1 is elementary whereas φ_2 may not be elementary because $B_{t_{j+1}}$ may not be (actually is not) $\mathcal{F}_{t_j}$-measurable.

[8] Versions were defined in Section 2.4.

3.7 The Itô isometry

On the one hand, this subsection can be seen as Step 0 of the proof of Theorem 3.10. On the other hand we will see here how different ingredients of the intricate definition of the space $\mathcal{V}(S,T)$ operate in practice.

Lemma 3.11. *(The Itô isometry) If $\varphi \in \mathcal{V}(S,T)$ is elementary, then*

$$E\bigg[\bigg(\int_S^T \varphi(t,\,\cdot\,)dB_t(\,\cdot\,)\bigg)^2\bigg] = E\bigg[\int_S^T \varphi(t,\,\cdot\,)^2 dt\bigg]. \tag{3.55}$$

Remark 3.12. *Suppose the limit (3.52) exists. Then the sequence $\{\varphi_n\}_{n\in\mathbb{N}}$ is Cauchy in the norm $(E[\int_S^T |\,\cdot\,|^2 dt])^{1/2}$. Hence, by (3.55), the sequence of random variables $\int_S^T \varphi_n(t,\,\cdot\,)dB_t(\,\cdot\,)$ is Cauchy in $L^2(\Omega,P)$. Thus the limit (3.51) exists.*

Remark 3.13. *After we prove Theorem 3.10, the Itô isometry will extend to $\mathcal{V}(S,T)$.*

We start with some preparatory facts.

Lemma 3.14. *$\Delta B_j := B_{t_{j+1}} - B_{t_j}$ is independent of any function f, which is $\mathcal{F}_{t_j}$-measurable. The same is true if ΔB_j above is replaced with $g(\Delta B_j)$ for some Borel function $g:\mathbb{R}\to\mathbb{R}$.*

Proof. By Theorem 2.11, we have that $B_{t_{j+1}} - B_{t_j}$ is independent of B_s, $0\le s\le t_j$. By definition of independence in terms of σ-algebras, we have that $\sigma(\Delta B_j)$ is independent of $\sigma(\{B_s\}_{s\in[0,t_j]}) =: \mathcal{F}_{t_j}$. Since f is $\mathcal{F}_{t_j}$-measurable, inverse images of Borel sets w.r.t. f are in $\mathcal{F}_{t_j}$, which gives the first claim.

Regarding the second statement, by the Doob-Dynkin lemma, $\sigma(g(\Delta B_j))\subset\sigma(\Delta B_j)$ and we know already that $\sigma(\Delta B_j)$ is independent of $\mathcal{F}_{t_j}$. Actually, one can also give an elementary argument here: For any two random variables X,Y, if $P(X\in F_1, Y\in F_2) = P(X\in F_1)P(Y\in F_2)$, then

$$\begin{aligned} P(g(X)\in F_1, Y\in F_2) &= P(X\in g^{-1}(F_1), Y\in F_2) \\ &= P(X\in g^{-1}(F_1))P(Y\in F_2) = P(g(X)\in F_1)P(Y\in F_2). \end{aligned} \tag{3.56}$$

This concludes the proof. □

Lemma 3.15. *$e_ie_j\Delta B_i$ and ΔB_j are independent for $i<j$.*

Remark 3.16. *If we only knew that e_i, e_j and ΔB_i are (separately) independent of ΔB_j we could not conclude that the product $e_ie_j\Delta B_i$ is independent of ΔB_j. Here we need the stronger assumption of adaptation to a filtration.*

Proof. Since e_i is $\mathcal{F}_{t_i}$-measurable and $\mathcal{F}_{t_i}\subset\mathcal{F}_{t_j}$, we observe that e_i is $\mathcal{F}_{t_j}$-measurable. Considering that $\Delta B_{t_i} = B_{t_{i+1}} - B_{t_i}$ and $i+1\le j$, we see that ΔB_{t_i} is $\mathcal{F}_{t_j}$ measurable. (By definition of $\mathcal{F}_{t_j}$, all B_s, $0\le s\le t_j$, are measurable w.r.t. this σ-algebra, hence so is $B_{t_{i+1}} - B_{t_i}$.) Since a product of functions measurable w.r.t. $\mathcal{F}_{t_j}$ is also measurable w.r.t. $\mathcal{F}_{t_j}$, we obtain that $e_ie_j\Delta B_{t_i}$ is $\mathcal{F}_{t_j}$-measurable. Now the claim follows from Lemma 3.14. □

Proof of Lemma 3.11. Put $\Delta B_j = B_{t_{j+1}} - B_{t_j}$. Then, by Lemma 3.15,

$$E[e_ie_j\Delta B_i\Delta B_j] = \begin{cases} 0 & \text{if} \quad i\neq j, \\ E[e_j^2](t_{j+1}-t_j) & \text{if} \quad i=j. \end{cases} \tag{3.57}$$

Here in the first line we used Lemma 3.15, the fact that independent random variables are uncorrelated: $E[e_ie_j\Delta B_i\Delta B_j] = E(e_ie_j\Delta B_i)E(\Delta B_j)$ and $E(\Delta B_j)=0$. In the second line we use Lemma 3.14 to obtain independence of e_j^2 and $(\Delta B_j)^2$. Hence, $E[e_j^2(\Delta B_j)^2] = E[e_j^2]E[(\Delta B_j)^2] = E[e_j^2](t_{j+1}-t_j)$, where in the last step we used Theorem 2.11. By (3.57),

$$E\bigg[\bigg(\int_S^T \varphi dB\bigg)^2\bigg] = \sum_{i,j} E[e_ie_j\Delta B_i\Delta B_j] = \sum_j E[e_j^2](t_{j+1}-t_j). \tag{3.58}$$

On the other hand

$$\begin{aligned} E\left[\int_S^T \varphi(t,\cdot)^2 dt\right] &= E\left[\int_S^T |e_j|^2 \chi_{[t_j,t_{j+1})}(t)dt\right] \\ &= E\left[\sum_{j\geq 0} |e_j|^2 (t_{j+1}-t_j)\right] = \sum_{j\geq 0} E(|e_j|^2)(t_{j+1}-t_j), \end{aligned} \tag{3.59}$$

which completes the proof. □

3.8 Itô integral. Proof of Theorem 3.10

We will approximate as follows:

$$\{f\in\mathcal{V}(S,T)\}\approx\{h\in\mathcal{V}(S,T),\text{bd.}\}\approx\{g\in\mathcal{V}(S,T),\text{bd., cont. in } t \text{ for any fixed } \omega\}\approx\{\varphi \text{ elementary.}\} \tag{3.60}$$

Lemma 3.17. *Let $f\in\mathcal{V}(S,T)$. Then there exists a sequence $\{h_n\}_{n\in\mathbb{N}}\subset\mathcal{V}(S,T)$ s.t. h_n is bounded for each n and*

$$E\left[\int_S^T (f-h_n)^2 dt\right] \underset{n\to\infty}{\to} 0. \tag{3.61}$$

Proof. Put

$$h_n(t,\omega)=\begin{cases} -n & \text{if} \quad f(t,\omega)<-n \\ f(t,\omega) & \text{if} \quad -n\leq f(t,\omega)\leq n \\ n & \text{if} \quad f(t,\omega)>n. \end{cases} \tag{3.62}$$

Since $f\in\mathcal{V}(S,T)$, by Definition 3.8 we have

$$E\left[\int_S^T f^2 dt\right]<\infty. \tag{3.63}$$

By definition (3.62), $|h_n(t,\omega)|\leq|f(t,\omega)|$. Thus the claim follows by dominated convergence. □

Lemma 3.18. *Let $h\in\mathcal{V}(S,T)$ be bounded. Then there exist bounded functions $g_n\in\mathcal{V}(S,T)$ such that $g_n(\cdot,\omega)$ is continuous for all ω and n, and*

$$E\left[\int_S^T (h-g_n)^2 dt\right] \underset{n\to\infty}{\to} 0. \tag{3.64}$$

Proof. Suppose $|h(t,\omega)|\leq M$ for all (t,ω). For each n let ψ_n be a non-negative, continuous function on $\mathbb{R}$ s.t.

(i) $\psi_n(x)=0$ for $x\leq-\frac{1}{n}$ and $x\geq 0$,

(ii) $\int_{-\infty}^{\infty}\psi_n(x)dx=1$,

i.e. a certain Dirac delta approximating sequence. The functions

$$g_n(t,\omega):=\int_0^t \psi_n(s-t)h(s,\omega)ds \overset{(i)}{=} \int_0^\infty \psi_n(s-t)h(s,\omega)ds \tag{3.65}$$

have the following properties:

(a) g_n are bounded uniformly in n:

$$|g_n(t,\omega)| \leq \int_0^\infty \psi_n(s-t)|h(s,\omega)|ds \leq \sup_{s\in\mathbb{R}_+} |h(s,\omega)| \leq M. \tag{3.66}$$

(b) g_n are continuous for each fixed ω and n. In fact, suppose $t_\ell \underset{\ell\to\infty}{\to} t$. Then, by boundedness of h, continuity of ψ_n and dominated convergence

$$\lim_{\ell\to\infty} g_n(t_\ell,\omega) := \int_0^\infty \lim_{\ell\to\infty} \psi_n(s-t_\ell)h(s,\omega)ds = g_n(t,\omega). \tag{3.67}$$

We also made use of the fact that, since t_ℓ is a bounded sequence, we can replace the upper boundary of integration in (3.65) by a sufficiently large constant.

(c) Since $h \in \mathcal{V}(S,T)$, one can show that $g_n(t,\cdot)$ is $\mathcal{F}_t$-measurable for all t. In fact, we note that $\psi_n(s-t)h(s,\omega)$ is non-zero only for $s \leq t$ and $h(s,\omega)$ is, by assumption, $\mathcal{F}_s$-measurable. Hence, it is $\mathcal{F}_t$-measurable as $\mathcal{F}_s \subset \mathcal{F}_t$. Since sums of measurable functions are measurable, also Riemann sums $g_{n,\Pi}(t,\omega) := \sum_{i=1}^{\tilde{n}} \psi_n(s_i-t)h(s_i,\omega)\Delta s_i$ are $\mathcal{F}_t$-measurable. The pointwise limit $g_n(t,\omega) = \lim_{|\Pi|\to 0} g_{n,\Pi}(t,\omega)$ of $\mathcal{F}_t$-measurable functions is $\mathcal{F}_t$-measurable.

(d) There holds (HS7, Problem 2)

$$\int_S^T (g_n(t,\omega)-h(t,\omega))^2 dt \underset{n\to\infty}{\to} 0 \text{ for each } \omega. \tag{3.68}$$

Finally, by the uniform boundedness of g_n (see (a)), boundedness of h and dominated convergence, we have

$$E\left[\int_S^T dt\,(h-g_n)^2 dt\right] \underset{n\to\infty}{\to} 0. \tag{3.69}$$

This concludes the proof. □

Lemma 3.19. *Let $g \in \mathcal{V}(S,T)$ be bounded and $g(\,\cdot\,,\omega)$ be continuous for each ω. Then, there exist elementary functions $\varphi_n \in \mathcal{V}(S,T)$ s.t.*

$$E\left[\int_S^T (g-\varphi_n)^2 dt\right] \underset{n\to\infty}{\to} 0. \tag{3.70}$$

Proof. Set $\varphi_n(t,\omega) = \sum_j g(t_j,\omega)\chi_{[t_j,t_{j+1})}(t)$. Then φ_n is elementary since $g \in \mathcal{V}(S,T)$. (In fact, for $t_j \leq t < t_{j+1}$ we have $\varphi_n(t,\omega) = g(t_j,\omega)$ which is $\mathcal{F}_{t_j}$-measurable. Since $\mathcal{F}_{t_j} \subset \mathcal{F}_t$, it is also $\mathcal{F}_t$-measurable). Moreover,

$$\int_S^T (g-\varphi_n)^2 dt \underset{n\to\infty}{\to} 0 \quad \text{for each } \omega \tag{3.71}$$

since g is continuous. In fact, $g(t,\omega)-\varphi_n(t,\omega) = \sum_j [g(t,\omega)-g(t_j,\omega)]\chi_{[t_j,t_{j+1})}(t)$, for $S \leq t \leq T$, hence

$$\begin{aligned} \int |g(t,\omega)-\varphi_n(t,\omega)|^2 dt &= \int_S^T \sum_j [g(t,\omega)-g(t_j,\omega)]^2 \chi_{[t_j,t_{j+1})}(t)dt \\ &\leq w_{g(\,\cdot\,,\omega)}(2^{-n})^2 \int_S^T \sum_j \chi_{[t_j,t_{j+1})}(t)dt = \omega_{g(\,\cdot\,,\omega)}(2^{-n})^2(T-S), \end{aligned} \tag{3.72}$$

where the modulus of continuity

$$w_{g(\cdot,\omega)}(\delta) := \sup\{|g(t_1,\omega)-g(t_2,\omega)| : |t_1-t_2| \leq \delta,\ t_1,t_2 \in [S,T]\} \tag{3.73}$$

satisfies $\lim_{n\to\infty} w_{g(\,\cdot\,,\omega)}(2^{-n}) = 0$ for any fixed ω. This gives (3.71). Finally, by the boundedness of g and dominated convergence we obtain the claim. □

3.9 Itô integral. Example

Lemma 3.20. *Assume $B_0 = 0$. Then*

$$\int_0^T B_t dB_t = \frac{1}{2}B_T^2 - \frac{1}{2}T. \tag{3.74}$$

Proof. Put $\varphi_n(t,\omega) = \sum_j B_j(\omega) \cdot \chi_{[t_j,t_{j+1})}(t)$, where $B_j := B_{t_j}$. Then

$$\begin{aligned} E\Big[\int_0^T (\varphi_n - B_t)^2 dt\Big] &= E\Big[\sum_j \int_{t_j}^{t_{j+1}} (B_j - B_t)^2 dt\Big] \\ &= \sum_j \int_{t_j}^{t_{j+1}} (t - t_j) dt = \frac{1}{2}\sum_j (t_{j+1} - t_j)^2. \end{aligned} \tag{3.75}$$

Recall from (3.33) that $t_j = j2^{-n} \le T$. Then

$$(3.75) \le \frac{1}{2} \sum_{0 \le j2^{-n} \le T} (t_{j+1} - t_j)^2 = \frac{1}{2} \sum_{0 \le j \le T2^n} 2^{-2n} = \frac{1}{2} 2^{-2n}(T2^n + 1) \underset{n\to\infty}{\to} 0. \tag{3.76}$$

Thus we checked (3.52). Consequently, by (3.51),

$$\int_0^T B_t dB_t = \lim_{n\to\infty} \int_0^T \varphi_n dB_t = \lim_{n\to\infty} \sum_j B_j \Delta B_j \tag{3.77}$$

in $L^2(\Omega, P)$. Note that

$$\begin{aligned} \Delta(B_j^2) &:= B_{j+1}^2 - B_j^2 = (B_{j+1} - B_j)^2 + 2B_j(B_{j+1} - B_j) \\ &= (\Delta B_j)^2 + 2B_j \Delta B_j. \end{aligned} \tag{3.78}$$

Since $B_0 = 0$,

$$B_T^2 = \sum_j \Delta(B_j^2) = \sum_j (\Delta B_j)^2 + 2\sum_j B_j \Delta B_j. \tag{3.79}$$

By Lemma 3.6, we have that $\sum_j (\Delta B_j)^2 \to T$ in $L^2(\Omega, P)$ as $n \to \infty$, thus the result follows. □

Lecture 8

3.10 Itô integral. Properties

Theorem 3.21. *Let $f, g \in \mathcal{V}(S,T)$ and let $0 \le S < U < T$. Then:*

(i) $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t$ *for a.a.* ω.

(ii) $\int_S^T (c_1 f + c_2 g) dB_t = c_1 \int_S^T f dB_t + c_2 \int_S^T g dB_t$, *($c_1, c_2$ constants), for a.a.* ω.

(iii) $E[\int_S^T f dB_t] = 0$.

(iv) $\int_S^T f dB_t$ *is* $\mathcal{F}_T$ *- measurable.*

Proof. This holds true for all elementary functions so, by taking limits, we obtain the theorem for $f, g \in \mathcal{V}(S,T)$. □

Theorem 3.22. *Let $f \in \mathcal{V}(0,T)$. Then there exists a t-continuous version of*

$$\int_0^t f(s,\omega)dB_s(\omega), \quad 0 \le t \le T, \tag{3.80}$$

i.e., there exists a t-continuous stochastic process J_t on $(\Omega, \mathcal{F}, P)$ s.t.

$$P\big[\, J_t(\,\cdot\,) = \int_0^t f(s,\,\cdot\,)dB_s(\,\cdot\,)\big] = 1 \tag{3.81}$$

for all $t \in [0,T]$. In other words,

$$J_t(\omega) = \int_0^t f(s,\omega)dB_s(\omega), \tag{3.82}$$

for all ω outside of some set of measure zero N. (A priori N depends on t, but in the proof we will construct a t-independent set N).

To prove this lemma it will be important to know that the Itô integral is a martingale. Martingales are the main topic of the next chapter.

4 Conditional expectations and martingales

4.1 Conditional expectations

References for this section are [Ok, Appendix B] and [JP, Section 23].

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X \in L^1(\Omega, \mathcal{F}, P; \mathbb{R}^d)$ be a random variable.

1. Def: Conditional probability $P(C|A) = \frac{P(C\cap A)}{P(A)}$, $P(A) \neq 0$, $A, C \in \mathcal{F}$.

2. Def: Conditional expectation on one event A is naturally defined by

$$E[X|A] = \int X(\omega)dP(\omega|A). \tag{4.1}$$

   Obviously, it is a number (or constant random variable). If $P(A) = 0$ we set $E[X|A] = 0$ by convention.

3. Let us approximate X by a step function given $E[X|A_i]$, where $A_i \in \mathcal{F}$, $i = 1, \dots n$, form a disjoint partition of Ω:

$$E[X|A_1, \dots, A_n](\omega) := \sum_{i=1}^n E[X|A_i]\chi_{A_i}(\omega). \tag{4.2}$$

4. Question: Let $\mathcal{H} \subset \mathcal{F}$ be a σ-algebra. Can we, similarly, give meaning to $E[X|\mathcal{H}]$?

5. Answer for $\mathcal{H}$ finite:

   - Def: A non-empty set $A \in \mathcal{H}$ is an atom if for any non-empty $H \in \mathcal{H}$ we have $H \subset A \Rightarrow H = A$. We denote the set of all atoms in $\mathcal{H}$ by $\mathrm{Atoms}(\mathcal{H})$.
   - Fact: Any $H \in \mathcal{H}$ is a union of some collection of atoms.
   - Fact: Atoms are disjoint and form a disjoint partition of Ω.
   - Fact: X is $\mathcal{H}$-measurable iff it is constant on atoms.

Then we define the random variable:

$$E[X|\mathcal{H}](\omega) = \sum_{A\in\text{Atoms}(\mathcal{H})} E[X|A]\chi_A(\omega). \tag{4.3}$$

Proposition 4.1. *Let $X \in L^1(\Omega, \mathcal{F}, P; \mathbb{R}^d)$. It holds that:*

(a) $E[X|\mathcal{H}]$ is $\mathcal{H}$-measurable.

(b) $\int_H E[X|\mathcal{H}]dP = \int_H XdP$ for all $H \in \mathcal{H}$.

(c) $E[X|\mathcal{H}] = X$ if X is $\mathcal{H}$-measurable.

(d) $E[X|\mathcal{H}] = E[X]$ if $\sigma(X)$ is independent of $\mathcal{H}$.

(e) $E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{G}]$ for a σ-algebra $\mathcal{G} \subset \mathcal{H}$. ("Tower property").

(f) $E[E[X|\mathcal{H}]] = E[X]$. (Clearly, (b) $\Rightarrow$ (f) by setting $H = \Omega$, or (e) $\Rightarrow$ (f), by setting $\mathcal{G} = \{\emptyset, \Omega\}$).

Remark 4.2. *Some comments:*

- *It turns out that (a), (b) define the conditional expectation uniquely. (See below). Then linearity of $X \mapsto E[X|\mathcal{H}]$ follows from (b) and uniqueness.*
- *As for (c), clearly, if a function is constant on atoms, we recover it exactly by (4.3).*
- *As for (d), an example of X which is independent of $\mathcal{H}$ is a function which has exactly the same behaviour in each atom. (Think of X as periodic of period T and atoms as intervals of length T.)*
- *Item (e) says that if we first approximate X with step functions on a finer partition and then approximate the result $E[X|\mathcal{H}]$ with step functions on a coarser partition then this amounts to approximating X directly on the coarser partition.*

Proof. Part (a) is obvious, as χ_A are $\mathcal{H}$-measurable. Regarding part (b), we write

$$\begin{aligned}\int_H E[X|\mathcal{H}]dP &= \sum_{A\in\text{Atoms}(\mathcal{H})} E[X|A]P(A\cap H) \\ &= \sum_{\substack{A\in\text{Atoms}(\mathcal{H})\\ A\subset H}} E[X|A]P(A),\end{aligned} \tag{4.4}$$

where we used that since A is an atom, $A \cap H = A$ or $A \cap H = \emptyset$. Now we note that for any $C \in \mathcal{F}$

$$\sum_{\substack{A\in\text{Atoms}(\mathcal{H})\\ A\subset H}} P(C|A)P(A) = \sum_{\substack{A\in\text{Atoms}(\mathcal{H})\\ A\subset H}} P(C\cap A) = P(C\cap H), \tag{4.5}$$

using the fact that atoms contained in H form a partition of H. Thereby we complete the proof if X is a step function. The general case follows by pointwise approximating X with step functions.

We will show (c), (d), (e) below in the context of arbitrary σ-algebras. (f) is justified in the statement of the proposition. $\square$

6. Answer for arbitrary $\mathcal{H}$:

Definition 4.3. *$E[X|\mathcal{H}]$ is the (P-a.s. unique) function from Ω to $\mathbb{R}^d$ satisfying properties (a), (b) from Proposition 4.1, i.e.,*

(a) $E[X|\mathcal{H}]$ is $\mathcal{H}$-measurable.

(b) $\int_H E[X|\mathcal{H}]dP = \int_H XdP$ for all $H \in \mathcal{H}$.

Theorem 4.4. *$E[X|\mathcal{H}]$ exists and is P-a.s. unique (i.e., unique up to modifications on sets of P-measure zero).*

To prove it, we need some preparations:

Definition 4.5. *Let ν be a signed measure*[9] *and μ a (positive) measure on the measurable space $(\Omega, \mathcal{F})$. We say that ν is absolutely continuous w.r.t. μ ($\nu \ll \mu$) if*

$$\{\, \mu(F) = 0 \,\} \quad \Rightarrow \quad \{\, \nu(F) = 0 \,\} \tag{4.6}$$

for all $F \in \mathcal{F}$.

Theorem 4.6. *(Radon-Nikodym). Let μ be a σ-finite*[10] *positive measure and ν be a σ-finite signed measure on $(\Omega, \mathcal{F})$. If $\nu \ll \mu$ then there exists a measurable function f s.t.*

$$\nu(F) = \int_F f d\mu, \quad \text{for all} \quad F \in \mathcal{F}. \tag{4.7}$$

We write $f = \frac{d\nu}{d\mu}$. This function is μ-a.s. unique.

Proof of Theorem 4.4. The existence and uniqueness of $E[X|\mathcal{H}]$ comes from the Radon-Nikodym theorem: Let ν be the (signed) measure on $\mathcal{H}$ defined by

$$\nu(H) = \int_H X dP, \quad H \in \mathcal{H}. \tag{4.8}$$

Then ν is absolutely continuous w.r.t. $P|_{\mathcal{H}}$ (restriction of P to the σ-algebra $\mathcal{H}$). So there exists a $P|_{\mathcal{H}}$-a.s. unique $\mathcal{H}$-measurable function f s.t.

$$\nu(H) = \int_H f dP|_{\mathcal{H}}, \quad H \in \mathcal{H}. \tag{4.9}$$

By setting $E[X|\mathcal{H}] := f$ we fulfill the conditions of Definition 4.3. □

Remark 4.7. *It suffices to consider $d = 1$ in the proofs in the remaining part of this subsection. The general case is obtained using that $L^2(\Omega, \mathcal{F}, P; \mathbb{R}^d) = L^2(\Omega, \mathcal{F}, P) \otimes \mathbb{R}^d$ and, denoting by $\{e_i\}_{i=1,\ldots,d}$ the canonical basis in $\mathbb{R}^d$,*

$$\begin{aligned} \int_H E[X|\mathcal{H}] dP = \int_H X dP &= \sum_{i=1}^d \Big(\int_H X_i dP\Big) \otimes e_i \\ &= \sum_{i=1}^d \Big(\int_H E[X_i|\mathcal{H}] dP\Big) \otimes e_i = \int_H \sum_{i=1}^d (E[X_i|\mathcal{H}] \otimes e_i) dP. \end{aligned} \tag{4.10}$$

Thus, uniqueness gives $E[X|\mathcal{H}] = \sum_{i=1}^d (E[X_i|\mathcal{H}] \otimes e_i)$. Using this formula we obtain the claims for $d > 1$.

Proof of Proposition 4.1 (c), (d), (e) for arbitrary σ-fields:

(c) $E[X|\mathcal{H}] = X$ if X is $\mathcal{H}$-measurable. Follows from (4.9), (4.8) and the uniqueness statement from the Radon-Nikodym theorem.

[9] That is, $\nu = \nu^+ - \nu^-$, where $\nu^\pm$ are measures and at least one of the measures $\nu^\pm$ is finite.

[10] That is, Ω is a union of a countable number of sets of finite measure.

(d) $E[X|\mathcal{H}] = E[X]$ if X is independent of $\mathcal{H}$. In fact, then X and χ_H, $H \in \mathcal{H}$, are uncorrelated: $E[X\chi_H] = E[X]E[\chi_H]$. In other words:

$$\int_H X dP = \int X\chi_H dP = \int X dP \int \chi_H dP = \int_H E[X] dP|_{\mathcal{H}}. \tag{4.11}$$

Now we refer to (4.9), (4.8).

(e) $E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{G}]$ for $\mathcal{G} \subset \mathcal{H}$. In fact, by (4.9), (4.8), for any $G \in \mathcal{G}$

$$\int_G X dP = \int_G E[X|\mathcal{H}] dP|_{\mathcal{H}} = \int_G E[E[X|\mathcal{H}]|\mathcal{G}] dP|_{\mathcal{G}}. \tag{4.12}$$

On the other hand

$$\int_G X dP = \int_G E[X|\mathcal{G}] dP|_{\mathcal{G}}. \tag{4.13}$$

The uniqueness statement of the Radon-Nikodyn theorem gives the claim. □

Lecture 9.

The above L^1 theory of conditional expectations, based on the Radon-Nikodym theorem, is efficient for proving lemmas but conceptually not very transparent. Fortunately, there is a nice L^2-theory of conditional expectations:

Proposition 4.8. *For $p \geq 1$, it holds that:*

(a) If $X \in L^p$ then $E[X|\mathcal{H}] \in L^p$.

(b) If $X_n \to X$ in L^p then $E[X_n|\mathcal{H}] \to E[X|\mathcal{H}]$ in L^p.

where we set $L^p := L^p(\Omega, \mathcal{F}, P; \mathbb{R}^d)$.

Proof. To prove (a), we use the conditional Jensen inequality, see HS10, which says that if $g : \mathbb{R}^d \to \mathbb{R}$ is convex and $E[|g(X)|] < \infty$ then

$$g(E[X|\mathcal{H}]) \leq E[g(X)|\mathcal{H}]. \tag{4.14}$$

By choosing $g(X) = |X|^p$, we obtain

$$|E[X|\mathcal{H}]|^p \leq E[|X|^p|\mathcal{H}]. \tag{4.15}$$

Hence, for any $H \in \mathcal{H}$

$$\int_H |E[X|\mathcal{H}]|^p dP \leq \int_H E[|X|^p|\mathcal{H}] dP = \int_H |X|^p dP, \tag{4.16}$$

by Definition 4.3. Choosing $H = \Omega$ we obtain (a). Regarding (b), by replacing X with $X_n - X$ in (4.16), we obtain the claim. □

Proposition 4.9. *Let $X, Y \in L^2(\Omega, \mathcal{F}, P; \mathbb{R}^d)$ and $\mathcal{H} \subset \mathcal{F}$ be a σ-field. Then:*

(a) Let $Q_{\mathcal{H}}$ be the orthogonal projection on the closed subspace $L^2(\Omega, \mathcal{H}, P; \mathbb{R}^d) \subset L^2(\Omega, \mathcal{F}, P; \mathbb{R}^d)$. Then $E[X|\mathcal{H}] = Q_{\mathcal{H}}X$.

(b) $E[X \cdot Y|\mathcal{H}] = X \cdot E[Y|\mathcal{H}]$ if $X \in L^2(\Omega, \mathcal{H}, P; \mathbb{R}^d)$.

Remark 4.10. *Let us write for brevity $L^p(\mathcal{G}) := L^p(\Omega, \mathcal{G}, P)$. To have an example how the subspace $L^2(\mathcal{H})$ looks like, suppose for a moment that $\mathcal{H} = \sigma(Y)$, for some $Y \in L^2(\mathcal{F})$. Then, by the Doob-Dynkin lemma, $L^2(\mathcal{H})$ is spanned by $g(Y(\cdot))$, $g : \mathbb{R}^d \to \mathbb{R}^d$ measurable and $g(Y(\cdot)) \in L^2(\mathcal{F})$.*

Remark 4.11. *Note that for $X \in L^2(\mathcal{F})$ the tower property*

$$E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{G}], \quad \mathcal{G} \subset \mathcal{H}, \tag{4.17}$$

simply follows from $Q_{\mathcal{H}} Q_{\mathcal{G}} = Q_{\mathcal{G}}$, where $Q_{\mathcal{H}}, Q_{\mathcal{G}}$ are projections on $L^2(\mathcal{H}), L^2(\mathcal{G})$, respectively. This always holds for orthogonal projections on two closed subspaces, one included in the other. In our case $L^2(\mathcal{G}) \subset L^2(\mathcal{H})$.

Proof. (a) For any $X \in L^2(\mathcal{F})$ and any closed subspace $\mathfrak{h} \subset L^2(\mathcal{F})$ we have a unique decomposition

$$X = X^{\parallel} + X^{\perp}, \tag{4.18}$$

where $X^{\parallel} \in \mathfrak{h}$, $X^{\perp} \in \mathfrak{h}^{\perp}$ and $\mathfrak{h}^{\perp}$ is the orthogonal complement of $\mathfrak{h}$. The self-adjoint operator Q on $L^2(\mathcal{F})$, s.t. $QX = X^{\parallel}$, is called the orthogonal projection on $\mathfrak{h}$.

We set $\mathfrak{h} = L^2(\mathcal{H})$ and want to show that decomposition (4.18) has the form

$$X = E[X|\mathcal{H}] + (X - E[X|\mathcal{H}]). \tag{4.19}$$

We know from Proposition 4.8, that $E[X|\mathcal{H}] \in L^2(\mathcal{H})$. It suffices to show that $(X - E[X|\mathcal{H}]) \in L^2(\mathcal{H})^{\perp}$. By Definition 4.3, for $H \in \mathcal{H}$,

$$\int_H E[X|\mathcal{H}] dP = \int_H X dP \quad \Leftrightarrow \quad \int \chi_H E[X|\mathcal{H}] dP = \int \chi_H X dP \quad \Leftrightarrow \quad \langle \chi_H, E[X|\mathcal{H}] \rangle = \langle \chi_H, X \rangle. \tag{4.20}$$

In other words,

$$\langle \chi_H, (X - E[X|\mathcal{H}]) \rangle = 0 \tag{4.21}$$

so $(X - E[X|\mathcal{H}])$ is orthogonal to all χ_H, $H \in \mathcal{H}$. To conclude that $(X - E[X|\mathcal{H}])$ is orthogonal to all $Y \in L^2(\mathcal{H})$, we use that χ_H, for $H \in \mathcal{H}$, span a dense set in $L^2(\mathcal{H})$. In fact, Y can be approximated in L^2 by step functions $\sum_i c_i \chi_{H_i}$ like in the construction of the Lebesgue integral. (This is similar, but much simpler, than what we have seen in the case of the Itô integral).

(b) First, let us assume that $X \in L^{\infty}(\mathcal{H})$. We observe that X is a bounded, self-adjoint operator on the Hilbert space $L^2(\mathcal{F})$ which leaves the subspace $L^2(\mathcal{H}) \subset L^2(\mathcal{F})$ invariant, i.e. $X : L^2(\mathcal{H}) \to L^2(\mathcal{H})$. Hence, $X : L^2(\mathcal{H})^{\perp} \to L^2(\mathcal{H})^{\perp}$, where $\perp$ is the orthogonal complement. Consequently, X commutes with the orthogonal projection $Q_{\mathcal{H}}$ on $L^2(\mathcal{H})$. In fact, for any $Y \in L^2(\mathcal{H})$,

$$(Q_{\mathcal{H}} X) Y = Q_{\mathcal{H}} X (Y^{\parallel} + Y^{\perp}) = Q_{\mathcal{H}} X Y^{\parallel} = X Y^{\parallel}, \tag{4.22}$$

$$(X Q_{\mathcal{H}}) Y = X Q_{\mathcal{H}} (Y^{\parallel} + Y^{\perp}) = X Y^{\parallel}, \tag{4.23}$$

where $Y^{\parallel}$ is the component of Y in $L^2(\mathcal{H})$ defined by (4.18).

Now suppose $X \in L^2(\mathcal{H})$ but not necessarily in $L^{\infty}(\mathcal{H})$. We introduce the approximating family like in the proof of Lemma 3.17

$$X_n(\omega) = \begin{cases} -n & \text{if} \quad X(\omega) < -n \\ X(\omega) & \text{if} \quad -n \le X(\omega) \le n \\ n & \text{if} \quad X(\omega) > n, \end{cases} \tag{4.24}$$

s.t. $X_n \in L^{\infty}(\mathcal{H})$. By definition, $|X_n| \le |X| \in L^2(\mathcal{H})$. Hence, by dominated convergence,

$$\lim_{n \to \infty} \int |X_n \tilde{Y} - X \tilde{Y}| dP = 0, \quad \tilde{Y} \in L^2(\mathcal{F}). \tag{4.25}$$

In particular, $X_n Y \to XY$ and $X_n E[Y|\mathcal{H}] \to XE[Y|\mathcal{H}]$, where the convergence is in $L^1(\mathcal{F})$, since $E[Y|\mathcal{H}] \in L^2(\mathcal{H})$ by Proposition 4.8 (a). Due to Proposition 4.8 (b), boundedness of X_n and the first part of the proof:

$$E[XY|\mathcal{H}] = \lim_{n\to\infty} E[X_n Y|\mathcal{H}] = \lim_{n\to\infty} X_n E[Y|\mathcal{H}] = XE[Y|\mathcal{H}], \tag{4.26}$$

which concludes the proof. □

4.2 Martingales

The word 'martingale' has interesting origins, from the French town of Martigues, through fastening of a horse (strap system to control the height of the horse's head) up to a class of betting strategies in 18th century France [Ma].

- As usually, $(\Omega, \mathcal{F}, P)$ is a probability space.
- Def. Let $\mathcal{T}$ be an ordered set (e.g. $\mathbb{N}_0$ or $\mathbb{R}_+$). Then a family if σ-algebras $\mathcal{H}_t \subset \mathcal{F}$, $t \in \mathcal{T}$, is called a filtration if $t \le s \Rightarrow \mathcal{H}_t \subset \mathcal{H}_s$.

Definition 4.12. *A d-dimensional stochastic process $\{M_t\}_{t\in\mathcal{T}}$ on $(\Omega, \mathcal{F}, P)$ is called a martingale w.r.t. a filtration $\{\mathcal{H}_t\}_{t\in\mathcal{T}}$ if:*

(i) M_t is $\mathcal{H}_t$-measurable for all $t \in \mathcal{T}$.

(ii) $E[|M_t|] < \infty$ for all $t \in \mathcal{T}$.

(iii) $E[M_s|\mathcal{H}_t] = M_t$ for all $s \ge t$.

The usual gambling interpretation of the martingale property (iii) of Definition 4.15 is that in a fair game the best estimate of the fortune of tomorrow (M_s), given by the knowledge of today ($\mathcal{H}_t$), is the fortune of today M_t. In fact, by taking expectations of both sides of (4.15)

$$E[M_s] = E[M_t] \quad s \ge t. \tag{4.27}$$

So, on average, the capital does not change.

1. Example: The original martingale betting strategy was the following: You toss a fair coin. To start with, H means winning 1 EUR, T means loosing 1 EUR. But, whenever you lose, you double the stake and you reset the game upon the first H:

$$\begin{pmatrix} T & T & T & \dots & T & H \\ -1 & -2 & -2^2 & \dots & -2^N & 2^{N+1} \end{pmatrix} \tag{4.28}$$

   Loss: $-(1 + 2 + 2^2 + \dots + 2^N) = -(2^{N+1} - 1)$. Gain: 2^{N+1}. So, eventually, you always win 1 EUR! How to reconcile it with (4.27)? Suppose first H happened at $s \in \mathcal{T}$. This only means that $M_s(\omega) = 1$ EUR for a specific $\omega \in \Omega$. It is compensated by $M_s(\omega') < 0$ for many other $\omega' \in \Omega$.

   In real live this strategy is impractical because your debt 'between' the heads grows exponentially which may lead to bankruptcy. In a game described by a martingale 'almost sure' gain requires very high risk.

   Lecture 10

2. Example: Symmetric random walk, driven by tosses of a fair coin, can be reinterpreted as a fair game if 'Step to the right' means 'Winning 1 EUR' and 'Step to the left' means 'Losing 1 EUR'. It is easy to check that a random walk is a martingale iff it is symmetric, i.e., the coin is fair.

3. Example: The martingale property of the symmetric random walk is inherited by its continuum limit - the Brownian motion $\{B_t\}_{t\in\mathbb{R}_+}$ with its natural filtration $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$: Property (i) follows by definition of $\mathcal{F}_t := \sigma\{\, B_s \,:\, 0 \le s \le t\}$. As for (i), assuming $B_0 = 0$,

$$E[|B_t|]^2 \le E[B_t^2] = t. \tag{4.29}$$

As for (iii), for $s > t$,

$$E[B_s|\mathcal{F}_t] = E[B_s - B_t|\mathcal{F}_t] + E[B_t|\mathcal{F}_t] = E[B_s - B_t] + B_t = B_t, \tag{4.30}$$

where we used Proposition 4.1 (c), (d) and the fact that $\sigma(B_s - B_t)$ is independent of $\mathcal{F}_t$.

4.3 Itô integral as a martingale

By the last example, $B_t = \int_0^t 1 dB_{t'}$ is a martingale. More generally:

Theorem 4.13. *Let $f \in \mathcal{V}(0,T)$ for all $T \ge 0$. Consider, like in the construction of the Itô integral,*

$$I(t,\omega) := \int_0^t f(t',\omega) dB_{t'}(\omega) = \lim_{n\to\infty} \int_0^t \varphi_n(t',\omega) dB_{t'}(\omega) =: \lim_{n\to\infty} I_n(t,\omega). \tag{4.31}$$

Then $\{I(t,\,\cdot\,)\}_{t\in\mathbb{R}_+}$ and $\{I_n(t,\,\cdot\,)\}_{t\in\mathbb{R}_+}$ are martingales w.r.t. $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$.

Proof. Let us first show that $I_n(t,\,\cdot\,)$ is a martingale w.r.t. $\mathcal{F}_t$ for all n. We check the three properties from Definition 4.15:

(i) We know that $I_n(t,\,\cdot\,)$ is $\mathcal{F}_t$-measurable (Theorem 3.21).

(ii) $(E[|I_n(t,\,\cdot\,)|])^2 \le E[I_n(t,\,\cdot\,)^2] < \infty$ by the Cauchy-Schwarz inequality and the Itô lemma (Lemma 3.11).

(iii) Let us show that $E[I_n(s,\,\cdot\,)|\mathcal{F}_t] = I_n(t,\,\cdot\,)$ for $t \le s \le T$.

$$\begin{aligned} E[I_n(s,\,\cdot\,)|\mathcal{F}_t] &= E\Big[\int_0^t \varphi_n dB + \int_t^s \varphi_n dB \Big| \mathcal{F}_t\Big] \\ &= E\Big[\int_0^t \varphi_n dB \Big| \mathcal{F}_t\Big] + E\Big[\int_t^s \varphi_n dB \Big| \mathcal{F}_t\Big] \\ &= \int_0^t \varphi_n dB + E\Big[\int_t^s \varphi_n dB \Big| \mathcal{F}_t\Big], \end{aligned} \tag{4.32}$$

where we used linearity and $E[X|\mathcal{H}] = X$ for $\mathcal{H}$-measurable X (Proposition 4.1 (c)). It suffices to show that the last term vanishes:

$$\begin{aligned} E\Big[\int_t^s \varphi_n dB \Big| \mathcal{F}_t\Big] &= \sum_{t \le s_j \le s} E[e_j \Delta B_j|\mathcal{F}_t] \\ &= \sum_{t \le s_j \le s} E\big[E[e_j \Delta B_j|\mathcal{F}_{s_j}]|\mathcal{F}_t\big], \end{aligned} \tag{4.33}$$

where we used the tower property $E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{G}]$ for $\mathcal{G} \subset \mathcal{H}$. Next, we exploit $E[e_j \Delta B_j|\mathcal{F}_{s_j}] = e_j E[\Delta B_j|\mathcal{F}_{s_j}]$, since e_j is $\mathcal{F}_{s_j}$-measurable (Proposition 4.9 (b)) and $E[\Delta B_j|\mathcal{F}_{s_j}] = E[\Delta B_j] = 0$, where we used that ΔB_j is independent of $\mathcal{F}_{s_j}$ and applied Proposition 4.1 (d).

Now we check the martingale property for $I(t,\cdot)$. First two properties from Definition 4.15 follow like for $I_n(t,\cdot)$. We check the third property using the L^2-continuity of conditional expectations (Proposition 4.8 (b)) and the martingale property of I_n:

$$E[I(s,\cdot)|\mathcal{F}_t] = \lim_{n\to\infty} E[I_n(s,\cdot)|\mathcal{F}_t] = \lim_{n\to\infty} I_n(t,\cdot) = I(t,\cdot). \tag{4.34}$$

This concludes the proof. □

Actually, there is also a partial inverse, which we state here without proof (cf. [Ok, Theorem 4.3.4]), because the proof requires Itô calculus.

Theorem 4.14. *Let $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$ be the natural filtration of the Brownian motion. Suppose $(M_t,\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is a martingale s.t. $E[|M_t|^2]<\infty$ for all $t\in\mathbb{R}_+$. Then there exists a unique stochastic process $\{g(s,\cdot)\}_{s\in\mathbb{R}_+}$ s.t. $g\in\mathcal{V}(0,T)$ for all $t\geq 0$ and*

$$M_t(\omega) = E[M_0] + \int_0^t g(s,\omega)dB_s(\omega) \tag{4.35}$$

a.s. for all $t\geq 0$.

Having established that the Itô integral is a martingale, we will use this information to prove the continuity of a version of $t\mapsto \int_0^t f(t,\omega)dB_t(\omega)$, cf. Theorem 3.22. This will be achieved in Subsection 4.6 after some preliminaries on sub- and supermartingales and stopping times.

4.4 Submartingales and supermartingales

Definition 4.15. *An $\mathbb{R}$-valued stochastic process $\{M_t\}_{t\in\mathcal{T}}$ on $(\Omega,\mathcal{F},P)$ is called a submartingale (resp. supermartingale) w.r.t. a filtration $\{\mathcal{H}_t\}_{t\in\mathcal{T}}$ if:*

(i) M_t is $\mathcal{H}_t$-measurable for all $t\in\mathcal{T}$.

(ii) $E[|M_t|]<\infty$ for all $t\in\mathcal{T}$.

(iii) $E[M_s|\mathcal{H}_t]\geq M_t$ (resp. $E[M_s|\mathcal{H}_t]\leq M_t$) for all $s\geq t$.

Lemma 4.16. *If $(M_t,\mathcal{H}_t)_{t\in\mathbb{R}_+}$ is a martingale, g is convex and $E\big[|g(M_t)|\big]<\infty$ for each $t\in\mathbb{R}_+$, then $\{g(M_t)\}_{t\in\mathbb{R}_+}$ is a submartingale.*

Proof. Properties (i), (ii) of Definition 4.16 are covered by assumptions. Let us prove (iii): Let $t\leq s$. Then $E[M_s|\mathcal{H}_t]=M_t$ a.s. so $g(E[M_s|\mathcal{H}_t])=g(M_t)$ a.s. By the conditional Jensen inequality (cf. HS10)

$$E[g(M_s)|\mathcal{H}_t]\geq g(E[M_s|\mathcal{H}_t]) = g(M_t). \quad \square \tag{4.36}$$

4.5 Stopping times

In the context of gambling, a stopping time is the moment at which a gambler decides to quit the game based solely on the history of wins and losses observed so far. Such a stopping time is a random variable. Now formally: let $(\Omega,\mathcal{F},\{\mathcal{H}_t\}_{t\in\mathcal{T}},P)$ be a filtered probability space. Let $\mathcal{T}$ be an ordered subset of $\mathbb{R}_+$, e.g. $\mathbb{R}_+$, $\mathbb{N}_0$. We also set $\overline{\mathcal{T}}:=\mathcal{T}\cup\{\infty\}$.

Definition 4.17. *A random variable $\tau:\Omega\to\overline{\mathcal{T}}$ is called a stopping time if $\{\tau\leq t\}\in\mathcal{H}_t$ for all $t\in\mathcal{T}$.*

- *We say that a stopping time is bounded by $c\geq 0$ if $P(\tau\leq c)=1$.*
- *We write $X_\tau(\omega):=X_{\tau(\omega)}(\omega)$.*

Lemma 4.18. *Let $\mathcal{T}$ be a discrete set. Let τ be a stopping time bounded by c and let $(X_t, \mathcal{H}_t)_{t\in\mathcal{T}}$ be a submartingale. Then $E[X_\tau] \le E[X_c]$. (For martingales there holds equality).*

Remark 4.19. *One might think that for martingales $E[X_\tau] = E[X_c]$ immediately follows from (4.27), which says $E[X_t] = E[X_0]$ for any $t \ge 0$. Such an argument would not be correct, because τ may depend on ω, whereas t is an ω-independent parameter.*

Proof. We follow [JP, Proof of Th. 24.2]. We first give a proof for $\mathcal{T} = \mathbb{N}_0$. We write $X_\tau(\omega) = \sum_{t=0}^{\infty} X_t(\omega)\chi_{\{\tau(\omega)=t\}}$. Therefore, assuming that $c \in \mathbb{N}$,

$$E[X_\tau] \; = \; E\Big[\sum_{t=0}^{\infty} X_t\, \chi_{\{\tau=t\}}\Big] = E\Big[\sum_{t=0}^{c} X_t \chi_{\{\tau=t\}}\Big] = \sum_{t=0}^{c} E\Big[X_t \chi_{\{\tau=t\}}\Big] \le \sum_{t=0}^{c} E[E[X_c|\mathcal{H}_t]\chi_{\{\tau=t\}}], \quad (4.37)$$

where in the last step we used the submartingale property. Since $\{\tau = t\} = \{\tau \le t\}\backslash\{\tau \le t-1\}$ we see $\{\tau = t\} \in \mathcal{H}_t$. Therefore, $E[X_c|\mathcal{H}_t]\chi_{\{\tau=t\}} = E[\chi_{\{\tau=t\}}X_c|\mathcal{H}_t]$ by a variant of Proposition 4.9 (b), in which $X \in L^\infty(\Omega, \mathcal{F}, P)$ and $Y \in L^1(\Omega, \mathcal{F}, P)$, (Actually, since X is an indicator function in our case, this follows directly from the defining relation $\int_H E[XY|\mathcal{H}]dP = \int_H XY dP$, $H \in \mathcal{H}$). Making also use of Proposition 4.1 (f), (i.e., $E[E[X|\mathcal{H}_t]] = E[X]$), we obtain from (4.37)

$$E[X_\tau] \le \sum_{t=0}^{c} E[E[\chi_{\{\tau=t\}}X_c|\mathcal{H}_t]] = \sum_{t=0}^{c} E[\chi_{\{\tau=t\}}X_c] = E[\sum_{t=0}^{c} \chi_{\{\tau=t\}}X_c] = E[X_c]. \quad (4.38)$$

Clearly, the same proof works for any discrete $\mathcal{T}$. □

The key technical input for the next subsection is the following theorem:

Theorem 4.20. *(Doob's martingale inequality). If $(M_t, \mathcal{H}_t)_{t\in\mathbb{R}_+}$ is a martingale s.t. $t \mapsto M_t(\omega)$ is continuous a.s., then for all $p \ge 1$, $T \ge 0$ and $\lambda > 0$*

$$P[\sup_{0\le t\le T} |M_t| \ge \lambda] \le \frac{1}{\lambda^p} E[|M_T|^p]. \quad (4.39)$$

Remark 4.21. *If we replaced the event $\{\sup_{0\le t\le T} |M_t| \ge \lambda\}$ on the l.h.s. with the smaller event $\{|M_T| \ge \lambda\}$ the inequality would trivially follow from the Markov inequality (HS2, Problem 3). However, the supremum is important for us for proving continuity of $t \mapsto \int_0^t f(t,\omega)dB_t(\omega)$ because continuity is preserved under the supremum norm.*

Lecture 11

Proof. We follow [JP, Proof of Th. 26.1], [Ba, Proof of Th. 5.9]. We first prove an analogous statement for $\mathcal{T} = \mathbb{N}_0$ (in which case the continuity assumption is empty).

- Def. Let $M_s^* := \sup_{0\le t\le s} |M_t|$.
- Def. Let $\tau = \inf\{\, t : |M_t| \ge \lambda\}$ with the convention $\tau(\omega) = \infty$ if the condition $|M_t(\omega)| \ge \lambda$ fails for all $t \in \mathcal{T}$.
- Fact: τ is a stopping time, i.e., $\{\tau \le s\} \in \mathcal{H}_s$.

  Proof: We write $\{\tau \le s\} = \bigcup_{0\le s'\le s}\{\tau = s'\}$ and show that $\{\tau = s'\} \in \mathcal{H}_{s'}$. Clearly, $\{\tau = s'\} = \{|M_{s'}| \ge \lambda$ and $|M_{s''}| < \lambda$ for $0 \le s'' \le s'\,\}$.
- Fact: $\{\tau \le s\} = \{M_s^* \ge \lambda\}$.

  Proof: The latter event means "the process has crossed level λ by time s". The former event means "the first crossing time has occurred by time s". This is exactly the same event.

- We use the last fact and exploit $\tau(\omega) \leq s \Rightarrow |M_{\tau(\omega)}(\omega)| \geq \lambda$. (The condition $\tau \leq s$ excludes $\tau(\omega) = \infty$ which corresponds to $|M_t(\omega)| < \lambda$ for all t).

$$\begin{aligned} P(M_s^* \geq \lambda) &= P(\tau \leq s) = P(\{\tau \leq s\} \cap \{|M_\tau| \geq \lambda\}) \\ &= E[\chi_{\{\tau \leq s\}}\chi_{\{|M_\tau| \geq \lambda\}}] \leq E\left[\chi_{\{\tau \leq s\}}\chi_{\{|M_\tau| \geq \lambda\}}\frac{|M_\tau|^p}{\lambda^p}\right] \leq E\left[\chi_{\{\tau \leq s\}}\frac{|M_\tau|^p}{\lambda^p}\right]. \end{aligned} \tag{4.40}$$

- Using that $M_\tau = M_{\min(\tau,s)}$ on $\{\tau \leq s\}$

$$P(M_s^* \geq \lambda) \leq E\left[\frac{|M_{\min(\tau,s)}|^p}{\lambda^p}\right] \leq E\left[\frac{|M_s|^p}{\lambda^p}\right], \tag{4.41}$$

where in the last step we noted that $\min(\tau, s)$ is a stopping time bounded by s (Homework), applied Lemma 4.18, and the fact that $(|M_s|^p, \mathcal{H}_s)_{s\in\mathcal{T}}$ is a submartingale for $p \geq 1$ (Lemma 4.16). Setting $s = T$ we conclude the proof for $\mathcal{T} = \mathbb{N}_0$.

- Clearly, analogous arguments apply for any discrete $\mathcal{T} = \{0 = t_0 < t_1 \ldots < t_n = T\}$.

- We let $\{t_0, \ldots, t_n\}$ be numbered elements of $\mathbb{Q} \cap [0, T]$. Such partition increases as $n \to \infty$ to $\mathbb{Q} \cap [0, T]$. Then we have that $\sup_{0\leq k\leq n} |M_{t_k}|$ increases to $\sup_{t\in\mathbb{Q}\cap[0,T]} |M_t|$. Using monotone convergence to exchange this limit with P, we get

$$P[\sup_{t\in\mathbb{Q}\cap[0,T]} |M_t| \geq \lambda] \leq \frac{1}{\lambda^p} E[|M_T|^p]. \tag{4.42}$$

- Using a.s. continuity of $t \mapsto M_t$, we get $\sup_{t\in\mathbb{Q}\cap[0,T]} |M_t| = \sup_{t\in[0,T]} |M_t|$ almost surely. □

4.6 Continuity of the Itô integral: Proof of Theorem 3.22

We will use Theorem 4.20 to prove Theorem 3.22. We restate here the latter theorem in a shortened form:

Theorem 4.22. *Let $f \in \mathcal{V}(0, T)$. Then there exists a t-continuous version of*

$$\int_0^t f(s, \omega) dB_s(\omega), \quad 0 \leq t \leq T, \tag{4.43}$$

i.e., there exists a t-continuous stochastic process $\{J_t\}_{t\in\mathbb{R}_+}$ on $(\Omega, \mathcal{F}, P)$ s.t.

$$P\big[\, J_t = \int_0^t f(s, \cdot\,) dB_s(\,\cdot\,)\big] = 1 \tag{4.44}$$

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

Proof. Let $\varphi_n(s, \omega) = \sum_j e_j(\omega)\chi_{[s_j,s_{j+1})}(s)$ be elementary functions s.t.

$$E\left[\int_0^T (f - \varphi_n)^2 ds\right] \underset{n\to\infty}{\to} 0. \tag{4.45}$$

Define

$$I_n(t, \omega) = \int_0^t \varphi_n(s, \omega) dB_s(\omega) \tag{4.46}$$

and

$$I(t, \omega) = \int_0^t f(s, \omega) dB_s(\omega) = \lim_{n\to\infty} I_n(t, \omega), \quad 0 \leq t \leq T. \tag{4.47}$$

Let us now proceed in steps:

1. Fact: $t \mapsto I_n(t,\omega)$ is continuous for all n. To prove it, recall the partition:

$$s_j := \begin{cases} j2^{-n} & \text{if} \quad 0 \le j2^{-n} \le t, \\ t & \text{if} \quad j2^{-n} > t, \end{cases} \tag{4.48}$$

i.e. $s_j = \min(t, j2^{-n})$. Then, we have,

$$I_n(t,\,\cdot\,) = \sum_j e_j(B_{s_{j+1}} - B_{s_j})(\,\cdot\,) = \sum_j e_j(B_{\min(t,(j+1)2^{-n})} - B_{\min(t,j2^{-n})})(\,\cdot\,). \tag{4.49}$$

This is manifestly continuous, because $t \mapsto \min(t, j2^{-n})$ is continuous and $s \mapsto B_s(\omega)$ is continuous.

2. Fact: $(I_n(t,\,\cdot\,), \mathcal{F}_t)_{t\in\mathbb{R}_+}$ is a martingale for all n by Theorem 4.13.

3. Fact: There is a subsequence $\{n_k\}_{k\in\mathbb{N}}$ s.t. for almost all ω, $\{I_{n_k}(t,\omega)\}_{k\in\mathbb{N}}$ converges uniformly, i.e., in the norm $\|\,\cdot\,\|_\infty := \sup_{t\in[0,T]}|\,\cdot\,|$. Let us show this fact in steps:

   - From the previous fact, $((I_n - I_m)(t,\,\cdot\,), \mathcal{F}_t)_{t\in\mathbb{R}_+}$ is also a martingale, so by the Doob's martingale inequality (Theorem 4.20)

   $$\begin{aligned} P[\|I_n - I_m\|_\infty \ > \ \varepsilon] &\le \frac{1}{\varepsilon^2}E\big[|I_n(T,\,\cdot\,) - I_m(T,\,\cdot\,)|^2\big] \\ &= \frac{1}{\varepsilon^2}E\left[\Big|\int_0^T (\varphi_n(s,\,\cdot\,) - \varphi_m(s,\,\cdot\,))dB_s\Big|^2\right] \\ &= \frac{1}{\varepsilon^2}E\left[\int_0^T (\varphi_n - \varphi_m)^2 ds\right] \underset{m,n\to\infty}{\to} 0, \end{aligned} \tag{4.50}$$

   where we applied the Itô lemma and (4.45).

   - Thus, we can choose a subsequence $n_k \uparrow \infty$ (increasing monotonously to infinity) s.t.

   $$P[\|I_{n_{k+1}} - I_{n_k}\|_\infty \ > \ 2^{-k}] < 2^{-k}. \tag{4.51}$$

   - Recall the Borel-Cantelli lemma (HS1): Let $\{A_k\}_{k\in\mathbb{N}}$ be a sequence of events. If $\sum_{k=1}^\infty P(A_k) < \infty$, then

   $$P(\omega \in \Omega \,:\, \omega \in A_k \text{ for infinitely many } k\ ) = 0. \tag{4.52}$$

   - We set $A_k = \{\,\|I_{n_{k+1}} - I_{n_k}\|_\infty > 2^{-k}\}$ and obtain $\sum_{k=1}^\infty P(A_k) < \infty$ by (4.51). Thus, by (4.52),

   $$P(\|I_{n_{k+1}} - I_{n_k}\|_\infty \ > \ 2^{-k} \text{ for infinitely many } k\ ) = 0. \tag{4.53}$$

   Let us call N the null set under $P(\ldots)$ above. For all $\omega \notin N$ (i.e., for almost all ω) there holds the opposite condition: $\|I_{n_{k+1}}(\,\cdot\,,\omega) - I_{n_k}(\,\cdot\,\omega)\|_\infty > 2^{-k}$ for finitely many k. So there exists $k_1(\omega)$ s.t.

   $$\|I_{n_{k+1}}(\,\cdot\,,\omega) - I_{n_k}(\,\cdot\,,\omega)\|_\infty \le 2^{-k} \quad \text{for} \quad k \ge k_1(\omega). \tag{4.54}$$

   By the Cauchy criterion, and the telescopic argument we complete the proof of the fact.

4. By the preceding facts, $J_t(\omega) := \lim_{k\to\infty} I_{n_k}(t,\omega)$, $\omega \notin N$, is continuous as the uniform limit of continuous functions. In particular, for any fixed t, the functions $I_{n_k}(t,\,\cdot\,)$ converge almost surely to J_t. Hence, by HS9, they converge also in probability. On the other hand, by definition of the Itô integral, $I_{n_k}(t,\,\cdot\,) \to I(t,\,\cdot\,)$ in $L^2(P)$, hence also in probability. By the uniqueness of the limit in probability, we obtain $I_t = J_t$ almost surely, for all $t \in [0,T]$. This concludes the proof of the theorem. □

4.7 Extension of the Itô integral to more general filtrations

We generalize Definition 3.8 of admissible functions as follows:

Definition 4.23. *Let* $\cancel{\mathcal{V}(S,T)} \to \mathcal{V}_{\mathcal{H}}(S,T)$ *be the class of functions*

$$(t,\omega) \mapsto f(t,\omega) : \mathbb{R}_+ \times \Omega \to \mathbb{R} \tag{4.55}$$

such that

(i) f *is* $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$*-measurable.*

(ii') ~~f *is* $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$*-adapted.*~~ $\to$ *There exists a filtration* $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$ *s.t.* $(B_t, \mathcal{H}_t)_{t\in\mathbb{R}_+}$ *is a martingale and* f *is* $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$*-adapted.*

(iii) $E\big[\int_S^T f(t,\cdot)^2 dt\big] < \infty$.

Comments

1. Since $\{B_t\}_{t\in\mathbb{R}_+}$ is a martingale w.r.t. $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$, we have that B_t is $\mathcal{H}_t$-measurable. But $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$ is an increasing family, so also B_s is $\mathcal{H}_t$-measurable for $0 \le s \le t$. Therefore $\mathcal{F}_t = \sigma(B_s : 0 \le s \le t) \subset \mathcal{H}_t$.

   So we can allow $f(t,\cdot)$ to depend on $\mathcal{H}_t$, which is larger than $\mathcal{F}_t$, as long as the Brownian motion remains a martingale w.r.t. $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$.

2. Using that, for $s \ge t$, $E[B_s|\mathcal{H}_t] = B_t = E[B_t|\mathcal{H}_t]$ we get $E[B_s - B_t|\mathcal{H}_t] = 0$ which suffices for the construction of the Itô integral (Homework).

3. Definition 4.23 applies in the following situation (in which Definition 3.8 does not apply):

   - Denote by $B_t^{(k)}(\omega)$ by the k-th coordinate of the d-dimensional Brownian motion: $(B^{(1)},\dots,B^{(d)})$.
   - Def. $\mathcal{F}_t^{(d)}$ is the σ-algebra generated by $B_{s_1}^{(1)},\dots,B_{s_d}^{(d)}$, $s_k \le t$.
   - Fact. $(B_t^{(k)}, \mathcal{F}_t^{(d)})_{t\in\mathbb{R}_+}$ is a martingale. Indeed, $B_s^{(k)} - B_t^{(k)}$ is independent of $\mathcal{F}_t^{(d)}$ for $s > t$ by Theorem 2.11. Thus, with $\mathcal{H}_t = \mathcal{F}_t^{(d)}$ in Definition 4.23, we cover, e.g.,

     $$\int_S^T B_t^{(2)} dB_t^{(1)}, \quad \int_S^T \sin((B_t^{(1)})^2 + (B_t^{(2)})^2) dB_t^{(2)}, \tag{4.56}$$

     where in $(B_t^{(k)})^2$ the superscript k is an index and the superscript 2 is a power.

4. More generally, we define the multi-dimensional Itô integral as follows:

   Definition 4.24. *Let* $B = (B^{(1)},\dots,B^{(d)})$. *Then* $\mathcal{V}_{\mathcal{H}}^{m\times d}(S,T)$ *denotes the set of* $m \times d$ *matrices* $v(t,\omega) = [v_{i,j}(t,\omega)]_{\substack{1\le i\le m,\\ 1\le j\le d}}$*, where each* $v_{i,j}$ *satisfies the requirements of Definition 4.23 w.r.t. some filtration* $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$*. We define:*

   $$\int_S^T v dB = \int_S^T \begin{pmatrix} v_{1,1} & \dots & v_{1,d} \\ \vdots & & \vdots \\ v_{m,1} & \dots & v_{m,d} \end{pmatrix} \begin{pmatrix} dB^{(1)} \\ \vdots \\ dB^{(d)} \end{pmatrix} = \begin{pmatrix} \vdots \\ \sum_{j=1}^n \int_S^T v_{i,j} dB^{(j)} \\ \vdots \end{pmatrix}. \tag{4.57}$$

5. Fact. The integral of Definition 4.24 is a martingale w.r.t. $\{\mathcal{F}_t^{(d)}\}_{t\in\mathbb{R}_+}$. (Slight generalization of Theorem 4.13).

4.8 Extension of the Itô integral to a.s. square-integrable processes

Now we move on to the next generalization of Definition 3.8. As [Ok, Section 3.3] covers it only briefly, we follow [Ba, Section 7.5]

Definition 4.25. *Let* $\cancel{\mathcal{V}(S,T)} \rightarrow \mathcal{W}_{\mathcal{H}}(S,T)$ *be the class of functions*

$$(t,\omega) \mapsto f(t,\omega) : \mathbb{R}_+ \times \Omega \to \mathbb{R} \tag{4.58}$$

such that

(i) f *is* $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$*-measurable.*

(ii') $\cancel{f \text{ is } \{\mathcal{F}_t\}_{t\in\mathbb{R}_+}\text{-adapted.}} \rightarrow$ *There exists a filtration* $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$ *s.t.* $(B_t, \mathcal{H}_t)_{t\in\mathbb{R}_+}$ *is a martingale and* f *is* $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$*-adapted.*

(iii') $\cancel{E\big[\int_S^T f(t,\cdot)^2 dt\big] < \infty.} \rightarrow P\bigg[\int_S^T f(t,\cdot)^2 dt < \infty\bigg] = 1.$

1. Condition (iii') is also better suited for the Itô formula (see next section). In a simple case:

$$g(B_t) = g(0) + \int_0^t \frac{dg}{dx}(B_s)dB_s + \int_0^t \frac{1}{2}\frac{d^2g}{dx^2}(B_s)ds. \tag{4.59}$$

   We would like it to make sense for all $g \in C^2(\mathbb{R})$. But condition (iii) imposed on $\frac{dg}{dx}$ would exclude rapidly growing functions, like $g(x) = \exp(\exp(x))$, for which

$$E(\frac{dg}{dx}(B_s)^2) = \int [\exp(\exp(x))\exp(x)]^2 \frac{1}{\sqrt{2\pi s}}\exp(-\frac{x^2}{2s})dx = \infty, \tag{4.60}$$

   since $B_s \sim N(0,s)$.

2. Clearly (iii') is weaker than (iii). Let us argue that a variant of Itô integral can still be defined by an approximation argument.

   Naive idea: Define $A_n := \bigg\{\omega \in \Omega \ : \ \int_S^T f^2(t,\omega)dt \le n\bigg\}$ and set $f_n(t,\omega) := f(t,\omega)\chi_{A_n}(\omega)$. Then $f_n(\cdot,\omega) \to f(\cdot,\omega)$ in $L^2([S,T],dt)$ for almost all ω. Also

$$E\bigg[\int_S^T f_n(t,\omega)^2 dt\bigg] \le n < \infty, \tag{4.61}$$

   so, it seems, by the Itô lemma we could control $\int_S^T f_n(t,\omega)dB_t(\omega)$ and define the new Itô integral as its limit as $n \to \infty$. However, this idea fails as it stands, because $f_n = f\chi_{A_n}$ may not be adapted to the filtration $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$.

   Let us refine this idea as follows:

   - We can set $S = 0$ for simplicity, as multiplying the integrand $f(t,\omega)$ by $\chi_{[S,T]}(t)$ does not affect the requirements of Definition 4.25.
   - Define the sequence of stopping times

$$\tau_n(\omega) := \inf\{t' \in [0,T] \ : \ \int_0^{t'} f(s,\omega)^2 ds \ge n\}. \tag{4.62}$$

     with convention $\inf(\emptyset) = T$.

- Note that

$$\{\tau_n \le t\} = \{\int_0^t f(s,\,\cdot\,)^2 ds \ge n\} \in \mathcal{H}_t. \tag{4.63}$$

This follows from the fact that $f(s,\,\cdot\,)$ is $\mathcal{H}_s$-measurable, $s \le t$ implies $\mathcal{H}_s \subset \mathcal{H}_t$ and measurability is preserved by all the operations involved. Hence τ_n are stopping times in the sense of Definition 4.17.

- By (4.63), the processes $f_n(t,\omega) = f(t,\omega)\chi_{[0,\tau_n(\omega)]}(t) = f(t,\omega)\chi_{[t,\infty)}(\tau_n(\omega))$ are adapted to the filtration $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$.
- For all ω

$$\int_0^T f_n(t,\omega)^2 dt = \int_0^{\tau_n(\omega)} f(t,\omega)^2 dt = n \quad \Rightarrow \quad E\Big[\int_0^T f_n(t,\omega)^2 dt\Big] = n, \tag{4.64}$$

where $\int_0^{\tau_n(\omega)} f(t,\omega)^2 dt = n$ comes from the fact that $\tau_n(\omega)$ is the smallest time for which $\int_0^{\tau_n(\omega)} f(t,\omega)^2 dt \ge n$ and the Riemann integral is continuous in the upper boundary of integration. Thus, $I_n(t) := I_n(t,\,\cdot\,) := \int_0^t f_n(s,\,\cdot\,)dB_s$ is well defined as the Itô integral for processes satisfying Definition 4.23.

Theorem 4.26. *For $f \in \mathcal{W}_{\mathcal{H}}(S,T)$ the following limit exists in probability*

$$\int_S^T f(s,\,\cdot\,)dB_s = \lim_{n\to\infty} \int_S^T f_n(s,\,\cdot\,)dB_s. \tag{4.65}$$

It is called the Itô integral for $f \in \mathcal{W}_{\mathcal{H}}(S,T)$.

Proof. We look at the case $S = 0$. We proceed in steps:

- Fix $t \in [0,T]$. Choose integers $m, n \ge k$.
- On the event $\{\tau_k \ge t\}$ we have the following: Clearly, $\tau_n \ge \tau_k$ and $\tau_m \ge \tau_k$. Thus, for all $t \ge s$ we have $\tau_n \ge \tau_k \ge s$ and $\tau_m \ge \tau_k \ge s$. Hence, by $f_n(s,\omega) = f(s,\omega)\chi_{[0,\tau_n(\omega)]}(s)$,

$$f_n(s,\,\cdot\,) = f(s,\,\cdot\,) = f_m(s,\,\cdot\,) \qquad \text{for } 0 \le s \le t. \tag{4.66}$$

Therefore, $I_n(t) = I_m(t)$ on $\{\tau_k \ge t\}$.

- Thus, for any $\varepsilon > 0$,

$$\{|I_n(t) - I_m(t)| > \varepsilon\} \subset \{\tau_k < t\} \quad \Rightarrow \quad P(|I_n(t) - I_m(t)| > \varepsilon) \le P(\tau_k < t). \tag{4.67}$$

- Recall that $\{\tau_k \le t\} = \{\int_0^t f(s,\,\cdot\,)^2 ds \ge k\} \subset \{\int_0^T f(s,\,\cdot\,)^2 ds \ge k\}$. By dominated convergence

$$\lim_{k\to\infty} P\big(\int_0^T f(t,\,\cdot\,)^2 dt \ge k\big) = \lim_{k\to\infty} \int dP\, \chi_{[k,\infty)}\big(\int_0^T f(t,\,\cdot\,)^2 dt\big) = 0. \tag{4.68}$$

Thus $\{I_n(t)\}_{n\in\mathbb{N}}$ is Cauchy in probability, which concludes the proof. □

Theorem 4.27. *Let $f \in \mathcal{W}_{\mathcal{H}}(0,T)$. Then the following integral is t-continuous*

$$\int_0^t f(s,\omega)dB_s(\omega), \quad 0 \le t \le T \tag{4.69}$$

for ω in some set $\tilde{\Omega}$ of measure one.

Proof. Note that $\tau_n(\omega) \uparrow T$ for any fixed ω. Thus, there exists n s.t. for all t in a small neighbourhood N_{t_0} of some t_0

$$\int_0^t f(s,\omega)dB_s(\omega) = \int_0^{\min(t,\tau_n(\omega))} f(s,\omega)dB_s(\omega) = \int_0^t \underbrace{f(s,\omega)\chi_{[0,\tau_n(\omega)]}(s)}_{f_n(s,\omega)} dB_s(\omega). \tag{4.70}$$

By Theorem 4.22, this last expression is continuous in this neighbourhood, provided that $\omega \in \tilde{\Omega}_n \subset \Omega$, where $\tilde{\Omega}_n$ is the probability one subset on which the continuous version on $\int_0^t f_n(s,\omega)dB_s(\omega)$ 'lives'. Fortunately, $\tilde{\Omega} = \bigcap_n \tilde{\Omega}_n$ is also probability one, so we can restrict attention to this set from the beginning. □

Remark 4.28. *The second equality in (4.70) may look obvious, but for a stochastic integral it actually requires an argument. We omit it here and refer to [SP, Theorem 15.7].*

Remark 4.29. *We conclude with some comments:*

1. *From now on, $\int_0^t f(s,\omega)dB_s(\omega)$ denotes a t-continuous version.*
2. *For $f \in \bigcap_{T\geq 0} \mathcal{W}_{\mathcal{H}}(0,T)$ the process $M_t := \int_0^t f(s,\cdot)dB_s$ may not be a martingale w.r.t. $\mathcal{H}_t$. But it is clearly a local martingale in the following sense:*
   - *There exists a sequence of stopping times $\{\tau_n\}_{n\in\mathbb{N}}$ s.t. $\tau_n \uparrow \infty$ almost surely.*
   - *For each n, the stopped process $M_t^{\tau_n} := M_{\min(t,\tau_n)}$ is a martingale w.r.t. $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$.*

   In the case of the Itô integral, $M_{\min(t,\tau_n)}$ appears in the second step of (4.70).

5 Itô formula

5.1 Itô formula: simple case

In this subsection we follow [SP, Chapter 16].

Theorem 5.1. *Let $\{B_t\}_{t\in\mathbb{R}_+}$ be one-dimensional Brownian motion and let $g \in C^2(\mathbb{R})$. Then, we have for all $t \geq 0$ almost surely*

$$g(B_t) - g(B_0) = \int_0^t g'(B_s)dB_s + \frac{1}{2}\int_0^t g''(B_s)ds. \tag{5.1}$$

(As usually, we set $B_0 = 0$. The stochastic integral in the sense of Subsection 4.8). In a short-hand notation

$$dg(B_t) = g'(B_t)dB_t + \frac{1}{2}g''(B_t)dt. \tag{5.2}$$

We will prove this Theorem in Subsection 5.3. For now, we apply it to compute $I_t := \int_0^t B_s^\ell dB_s$, $\ell \in \mathbb{N}$.

- Guess a relevant g s.t. $dg(B_t) = dI_t + \cdots$, where dots denote some correction terms, hopefully simpler than the relevant integral.
- Inspired guessing may go as follows: Let $t \mapsto \tilde{B}_t$ be a differentiable function. Then $\tilde{I}_t = \int_0^t \tilde{B}_s^\ell d\tilde{B}_s = \frac{1}{\ell+1}(\tilde{B}_t)^{\ell+1}$. Accordingly, take $g(x) = \frac{1}{\ell+1}x^{\ell+1}$.
- Compute $dg(B_t)$ by the Itô formula (5.2):

$$dg(B_t) = B_t^\ell dB_t + \frac{1}{2}\ell B_t^{\ell-1}dt \tag{5.3}$$

- Using $B_0 = 0$ and integrating both sides of (5.3), we get

$$\frac{1}{\ell+1}B_t^{\ell+1} = \int_0^t B_s^\ell dB_s + \frac{1}{2}\ell\int_0^t B_s^{\ell-1}ds. \tag{5.4}$$

  Thus we expressed the stochastic integral $\int_0^t B_s^\ell dB_s$ by a Riemann integral.

5.2 Itô formula: more general case

Definition 5.2. *Let B_t be one-dimensional (d=1) Brownian motion on $(\Omega, \mathcal{F}, P)$. A one-dimensional Itô process is a stochastic process $\{X_t\}_{t\in\mathbb{R}_+}$ on $(\Omega, \mathcal{F}, P)$ of the form*

$$X_t = X_0 + \int_0^t u_s(\omega)ds + \int_0^t v_s(\omega)dB_s, \tag{5.5}$$

where

(a) $v \in \mathcal{W}_{\mathcal{H}}$, cf. Definition 4.25.

(b) u is $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$-adapted, where $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$ is as in Definition 4.25 and

$$P\left[\int_0^t |u_s(\omega)|ds < \infty \text{ for all } t \geq 0\right] = 1. \tag{5.6}$$

Comments:

- As a short-hand notation, equation (5.5) can be restated as

$$dX_t = u_t dt + v_t dB_t. \tag{5.7}$$

- Typically, $\mathcal{H}_t = \mathcal{F}_t$, i.e., the natural filtration of the Brownian motion. The generalization to the d-dimensional Itô formula (see Homework) requires $\mathcal{H}_t = \mathcal{F}_t^{(d)}$, cf. Subsection 4.7.

Theorem 5.3. *(The one-dimensional Itô formula). Let $\{X_t\}_{t\in\mathbb{R}_+}$ be an Itô process given by*

$$dX_t = u_t dt + v_t dB_t. \tag{5.8}$$

Let $\{(t,x) \mapsto g(t,x)\} \in C^2([0,\infty)\times\mathbb{R})$, i.e. g is twice continuously differentiable on $[0,\infty)\times\mathbb{R}$. Then

$$Y_t = g(t, X_t) \tag{5.9}$$

is again an Itô process, and

$$dY_t = \frac{\partial g}{\partial t}(t,X_t)dt + \frac{\partial g}{\partial x}(t,X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,X_t)(dX_t)^2, \tag{5.10}$$

where

$$dt\cdot dt = dt\cdot dB_t = dB_t\cdot dt = 0, \quad dB_t\cdot dB_t = dt. \tag{5.11}$$

As an example, let us compute $I_t = \int_0^t f(s)dB_s$ for $f \in C^1(\mathbb{R}_+)$.

- Again, let $s \mapsto \tilde{B}_s$ be in $C^1(\mathbb{R}_+)$. Here $\tilde{I}_t = \int_0^t f(s)d\tilde{B}_s$ is not computable, but integration by parts gives:

$$\tilde{I}_t = \int_0^t f(s)d\tilde{B}_s = f(t)\tilde{B}_t - \int_0^t \tilde{B}_s df(s), \tag{5.12}$$

  where the last integral is just a Riemann-Stieltjes integral. We set $Y_t = f(t)B_t$, $X_t = B_t$ and aim at finding a similar formula for I_t.

- $d(f(t)B_t) = B_t\frac{df}{dt}dt + f(t)dB_t$, where $\frac{1}{2}\frac{\partial^2 g}{\partial x^2}$-term from the Itô formula vanishes.

- As a consequence we exactly reproduce (5.12):

$$\int_0^t f(s)dB_s = f(t)B_t - \int_0^t B_s df(s). \tag{5.13}$$

5.3 Proof of Theorem 5.1

We proceed in steps:

1. We can assume that $\mathrm{supp}(g)$ is compact (see Lemma 5.4 below). Then
$$C_g := \|g\|_\infty + \|g'\|_\infty + \|g''\|_\infty < \infty. \tag{5.14}$$

2. By a telescopic argument, applied to our usual partition:
$$g(B_t) - g(B_0) = \sum_j (g(B_{j+1}) - g(B_j)), \tag{5.15}$$
where $B_j := B_{s_j}$, $s_j = j2^{-n} \wedge t$, (where $a \wedge b := \min(a, b)$).

3. Now we apply Taylor's theorem to each difference
$$g(B_{j+1}) - g(B_j) = g'(B_j)\Delta B_j + \frac{1}{2} g''(\xi_j)(\Delta B_j)^2 =: J_{1,j} + \frac{1}{2} J_{2,j}, \tag{5.16}$$
where $\Delta B_j = B_{j+1} - B_j$ and $\xi_j \in [B_j, B_{j+1}]$. We note that $g''(\xi_j)$ is measurable as a function of ω, since we can solve (5.16) for $g''(\xi_j)$ and thus express it by measurable functions.

4. Fact. The following limit exists in $L^2(P)$ and gives the first term on the r.h.s. of (5.1)
$$J_1 := \sum_j J_{1,j} = \sum_j g'(B_j)\Delta B_j \underset{n\to\infty}{\to} \int_0^t g'(B_s) dB_s. \tag{5.17}$$
See HS11, Problem 4.

5. Now we treat the second order terms in (5.16):
$$J_2 \;=\; \sum_j J_{2,j} = J_2^{(1)} + J_2^{(2)} := \sum_j g''(B_j)(\Delta B_j)^2 + \sum_j (g''(\xi_j) - g''(B_j))(\Delta B_j)^2. \tag{5.18}$$

   - Fact. $J_2^{(1)} := \sum_j g''(B_j)(\Delta B_j)^2 \underset{n\to\infty}{\to} \int_0^t g''(B_s)ds$ in probability and gives the second term on the r.h.s. of (5.1).
   Proof. We write
   $$J_2^{(1)} - \int_0^t g''(B_s)ds = \big[J_2^{(1)} - \sum_j g''(B_j)\Delta s_j\big] + \big[\sum_j g''(B_j)\Delta s_j - \int_0^t g''(B_s)ds\big]. \tag{5.19}$$
   Let us show that the first bracket on the r.h.s. of (5.19) converges in $L^2(P)$ to zero. We write:
   $$\begin{aligned} &E\Big[\Big(\sum_j g''(B_j)\{(\Delta B_j)^2 - \Delta s_j\}\Big)^2\Big] \\ &\qquad = \sum_{j,k} E\big[g''(B_j)g''(B_k)\{(\Delta B_j)^2 - \Delta s_j\}\{(\Delta B_k)^2 - \Delta s_k\}\big] \\ &\qquad = \sum_j E\big[g''(B_j)^2\{(\Delta B_j)^2 - \Delta s_j\}^2\big]. \end{aligned} \tag{5.20}$$
   In the last step we used that the sum over $j \neq k$ vanishes. This can be seen as follows: suppose $j < k$. Then $\{(\Delta B_k)^2 - \Delta s_k\}$ is independent of $g''(B_j)g''(B_k)\{(\Delta B_j)^2 - \Delta s_j\}$, as the latter random variable involves only B_s with $s \leq s_k$. Therefore,
   $$\begin{aligned} &E\big[\big(g''(B_j)g''(B_k)\{(\Delta B_j)^2 - \Delta s_j\}\{(\Delta B_k)^2 - \Delta s_k\}\big) \\ &\qquad = E\big[g''(B_j)g''(B_k)\{(\Delta B_j)^2 - \Delta s_j\}\big]E\big[\{(\Delta B_k)^2 - \Delta s_k\}\big] = 0, \end{aligned} \tag{5.21}$$

where in the last step we used $E[(\Delta B_k)^2] = \Delta s_k$. Coming back to (5.20),

$$\begin{aligned}\sum_j E\big[g''(B_j)^2\{(\Delta B_j)^2 - \Delta s_j\}^2\big] &\le \|g''\|_\infty \sum_j E\big[\{(\Delta B_j)^2 - \Delta s_j\}^2\big] \\ &\le \|g''\|_\infty \sum_j 4(\Delta s_j)^2 \underset{n\to\infty}{\to} 0,\end{aligned} \tag{5.22}$$

where we used $E[(\Delta B_j)^4] = 3(\Delta s_j)^2$.

Regarding the second bracket in (5.19), by the theory of the Riemann integral

$$\lim_{n\to\infty}\Big[\sum_j g''(B_j)\Delta s_j - \int_0^t g''(B_s)ds\Big] = 0 \tag{5.23}$$

almost surely, hence in probability. As the L^2-convergence implies the convergence in probability, the statement follows.

- Fact: $J_2^{(2)} := \sum_j (g''(\xi_j) - g''(B_j))(\Delta B_j)^2 \underset{n\to\infty}{\to} 0$ in $L^1(P)$, hence in probability.

  Proof. We estimate

$$|J_2^{(2)}| \le \sum_j |g''(\xi_j) - g''(B_j)|(\Delta B_j)^2 \le \sup_{j'} |g''(\xi_{j'}) - g''(B_{j'})| \sum_j (\Delta B_j)^2. \tag{5.24}$$

  Recall that $S_n := \sum_j (\Delta B_j)^2$ converges to t in $L^2(P)$ as $n \to \infty$ (HS6). Consequently, $\{E[S_n^2]\}_{n\in\mathbb{N}}$ is a bounded sequence. Hence, by the Cauchy-Schwarz inequality applied to the $\int \ldots dP$ integration

$$E\big[|J_2^{(2)}|\big] \le \Big[E\big(\sup_j |g''(\xi_j) - g''(B_j)|^2\big)\Big]^{1/2}\big[E[S_n^2]\big]^{1/2}. \tag{5.25}$$

  By the uniform continuity of $[0,t] \ni s \mapsto g''(B_s)$, we have $\lim_{n\to\infty} \sup_j |g''(\xi_j) - g''(B_j)|^2 = 0$. By the dominated convergence and $\|g''\|_\infty \le C_f$ assumed in (5.14), we have

$$\lim_{n\to\infty} E[|J_2^{(2)}|] = 0. \tag{5.26}$$

  So this part converges in $L^1(P)$ to zero. By the Markov inequality, the $L^1(P)$ convergence implies the convergence in probability.

Using (5.15), (5.16), (5.18) and the three facts above we conclude the proof. □

Now we prove a lemma which was used in the above proof:

Lemma 5.4. *Suppose formula (5.1) holds for compactly supported $g \in C^2(\mathbb{R})$. Then it holds for all $g \in C^2(\mathbb{R})$.*

Proof. We fix $g \in C^2(\mathbb{R})$, $\ell \ge 0$ and pick a smooth cut-off function $\tilde{\chi}_\ell$ satisfying $\chi_{B(0,\ell)} \le \tilde{\chi}_\ell \le \chi_{B(0,\ell+1)}$, where $\ell \in \mathbb{N}$ and $B(0,r)$ is the closed ball of radius r centered at zero. We set $g_\ell(x) := g(x)\tilde{\chi}_\ell(x)$. As g_ℓ is compactly supported, we have

$$g_\ell(B_t) - g_\ell(B_0) = \int_0^t g_\ell'(B_s)dB_s + \frac{1}{2}\int_0^t g_\ell''(B_s)ds. \tag{5.27}$$

Consider the stopping times

$$\tau_\ell := \inf\{s > 0 : |B_s| \ge \ell\}, \tag{5.28}$$

and note that

$$\frac{d^j}{dx^j} g_\ell(B_{t\wedge\tau_\ell}) = g^{(j)}(B_{t\wedge\tau_\ell}), \tag{5.29}$$

as for $t \leq \tau_\ell$ we have $|B_t| \leq \ell$ thus $\tilde{\chi}_\ell(B_t) = 1$. Recalling that we chose continuous versions of the Itô integrals, it is meaningful to replace in (5.27) t with the random variable $t \wedge \tau_\ell$:

$$\begin{aligned} g(B_{t\wedge\tau_\ell}) - g(B_0) &= \int_0^{t\wedge\tau_\ell} g_\ell'(B_s)dB_s + \frac{1}{2}\int_0^{t\wedge\tau_\ell} g_\ell''(B_s)ds \\ &= \int_0^t g_\ell'(B_s)\chi_{[0,\tau_\ell)}(s)dB_s + \frac{1}{2}\int_0^t g_\ell''(B_s)\chi_{[0,\tau_\ell)}(s)ds \\ &= \int_0^t g'(B_s)\chi_{[0,\tau_\ell)}(s)dB_s + \frac{1}{2}\int_0^t g''(B_s)\chi_{[0,\tau_\ell)}(s)ds \\ &= \int_0^{t\wedge\tau_\ell} g'(B_s)dB_s + \frac{1}{2}\int_0^{t\wedge\tau_\ell} g''(B_s)ds. \end{aligned} \tag{5.30}$$

The third equality follows from (5.29).

Since the Brownian motion does not explode in finite time, $\lim_{\ell\to\infty}\tau_\ell = \infty$ almost surely. Using that g is continuous and the stochastic integrals are continuous as functions of their upper boundary, we conclude the proof by taking the limit $\ell \to \infty$ in (5.30).

6 Stochastic differential equations

6.1 Example

Let us come back to the heuristic equation (3.1)

$$\frac{dZ_s}{ds} = \left(r + \alpha\frac{dB_s}{ds}\right)Z_s, \quad r, \alpha > 0. \tag{6.1}$$

We interpret it as follows

$$Z_t = Z_0 + r\int_0^t Z_s ds + \alpha\int_0^t Z_s dB_s \tag{6.2}$$

in terms of the Itô integral. In the short-hand notation:

$$dZ_t = rZ_t dt + \alpha Z_t dB_t. \tag{6.3}$$

Let us now solve this equation by inspired guessing:

- Consider a counterpart of (6.1) with $s \mapsto \tilde{B}_s \in C^1(\mathbb{R})$ and possibly different coefficients $\tilde{r}$, $\tilde{\alpha}$:

$$\frac{d\tilde{Z}_s}{ds} = \left(\tilde{r} + \tilde{\alpha}\frac{d\tilde{B}_s}{ds}\right)\tilde{Z}_s, \quad \tilde{B}_0 = 0. \tag{6.4}$$

- Solve it by separation of variables:

$$\frac{d\tilde{Z}_s}{\tilde{Z}_s} = \tilde{r}ds + \tilde{\alpha}d\tilde{B}_s, \tag{6.5}$$

$$\int_0^t \frac{d\tilde{Z}_s}{\tilde{Z}_s} = \tilde{r}t + \tilde{\alpha}\tilde{B}_t, \tag{6.6}$$

$$\log(\tilde{Z}_t) - \log(\tilde{Z}_0) = \tilde{r}t + \tilde{\alpha}\tilde{B}_t, \tag{6.7}$$

$$\tilde{Z}_t = \tilde{Z}_0\exp(\tilde{r}t + \tilde{\alpha}\tilde{B}_t). \tag{6.8}$$

- Accordingly, define $Z_t := Z_0 \exp(\tilde{r}t+\tilde{\alpha}B_t)$, with Z_0 a constant, and check if it satisfies (6.3) for suitable $\tilde{r}, \tilde{\alpha}$. We use the Itô formula (6.9)

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)(dX_t)^2, \tag{6.9}$$

with $X_t = B_t$ and $Y_t = Z_t$:

$$\begin{aligned} dZ_t &= Z_t\bigg(\tilde{r}dt + \tilde{\alpha}dB_t + \frac{1}{2}\tilde{\alpha}^2(dB_t)^2\bigg) \\ &= Z_t\bigg((\tilde{r} + \frac{1}{2}\tilde{\alpha}^2)dt + \tilde{\alpha}dB_t\bigg). \end{aligned} \tag{6.10}$$

Thus for $\tilde{r} := r - \frac{1}{2}\tilde{\alpha}^2$ and $\tilde{\alpha} := \alpha$ we obtain agreement with (6.3). Thus the solution of (6.2) has the form:

$$Z_t := Z_0 \exp\big((r - \frac{1}{2}\alpha^2)t + \alpha B_t\big). \tag{6.11}$$

- Remark. We refrained here from choosing Z_0 as a t-independent random variable as this could affect adaptedness assumptions.

6.2 Existence and uniqueness of solutions

Theorem 6.1. *Let $(\Omega, \mathcal{F}, P)$ be the probability space of the d-dimensional Brownian motion from Theorem 2.1. Let $T > 0$ and $b(\,\cdot\,,\cdot\,) : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma(\,\cdot\,,\cdot\,) : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times d}$ be measurable functions satisfying*

$$|b(t,x)| + |\sigma(t,x)| \leq C(1 + |x|), \quad x \in \mathbb{R}^n, t \in [0,T] \tag{6.12}$$

for some constant C and s.t. the Lipschitz property holds

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq D|x-y|, \quad x,y \in \mathbb{R}^d, t \in [0,T] \tag{6.13}$$

for some constant D. Let Z be a random variable which is independent of the σ-algebra $\mathcal{F}_\infty^{(d)}$ generated by $B_s(\,\cdot\,)$, $s \geq 0$, and s.t.

$$E[|Z|^2] < \infty. \tag{6.14}$$

Then the stochastic differential equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad t \in [0,T], \quad X_0 = Z, \tag{6.15}$$

which is a shorthand for

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s, \quad t \in [0,T], \quad X_0 = Z, \tag{6.16}$$

has a unique t-continuous solution $\{X_t\}_{t\in[0,T]}$ which is adapted to the filtration $\mathcal{F}_t^Z$ generated by Z and $B_s(\,\cdot\,)$, $s \leq t$, and satisfies

$$E\bigg[\int_0^T |X_t|^2 dt\bigg] < \infty. \tag{6.17}$$

6.2.1 Proof of Theorem 6.1: Main line of the argument.

Proof. We proceed in steps:

1. *(Uniqueness).* Let X_t, $\hat{X}_t$ be two solutions with initial value Z.

   - Put $a(s,\omega) := b(s,X_s) - b(s,\hat{X}_s)$ and $\gamma(s,\omega) := \sigma(s,X_s) - \sigma(s,\hat{X}_s)$. Then, using (6.17),

$$\begin{aligned} E\big[|X_t - \hat{X}_t|^2\big] &= E\bigg[\bigg(\int_0^t a ds + \int_0^t \gamma dB_s\bigg)^2\bigg] \\ &\overset{(c_1+c_2)^2 \le 2(c_1^2+c_2^2)}{\le} 2E\bigg[\bigg(\int_0^t a ds\bigg)^2\bigg] + 2E\bigg[\bigg(\int_0^t \gamma dB_s\bigg)^2\bigg] \\ &\overset{\text{Cauchy-Schwarz+Itô isometry}}{\le} 2tE\bigg[\bigg(\int_0^t a^2 ds\bigg)\bigg] + 2E\bigg[\bigg(\int_0^t \gamma^2 ds\bigg)\bigg] \\ &\overset{\text{Lipschitz (6.13)}}{\le} 2(1+t)D^2 \int_0^t E\big[|X_s - \hat{X}_s|^2\big] ds. \end{aligned} \tag{6.18}$$

   - Thus, the function $v(t) := E\big[|X_t - \hat{X}_t|^2\big]$, $t \in [0,T]$, satisfies

$$v(t) \le A \int_0^t v(s) ds, \quad A := 2(1+T)D^2. \tag{6.19}$$

   Recall the Grönwall inequality (cf. [Ok], Problem 5.17): Let $\tilde{v}$ be a non-negative function on $[0,T]$ Then,

$$\tilde{v}(t) \le \tilde{C} + \tilde{A} \int_0^t \tilde{v}(s) ds \quad \Rightarrow \quad \tilde{v}(t) \le \tilde{C} \exp(\tilde{A} t). \tag{6.20}$$

   In (6.19) we have $\tilde{C} = 0$ thus $\tilde{v} \equiv 0$.

   - From $v(t) := E\big[|X_t - \hat{X}_t|^2\big] = 0$ we get $X_t = \hat{X}_t$ almost surely, i.e., on some set $\tilde{\Omega}_t \subset \Omega$, $P(\tilde{\Omega}_t) = 1$. That is, $\{\hat{X}_t\}_{t\in[0,T]}$ is a version of $\{X_t\}_{t\in[0,T]}$. We postpone further discussion of the uniqueness to Subsection 6.2.2 below.

2. *(Construction of a candidate $\{Y_t\}_{t\in[0,T]}$ for the solution).* Let us restate (6.16):

$$X_t = X_0 + \int_0^t b(s,X_s) ds + \int_0^t \sigma(s,X_s) dB_s, \quad t \in [0,T], \quad X_0 = Z. \tag{6.21}$$

   We proceed by the Picard iteration:

   - Define $Y_t^{(0)} = X_0$ and

$$Y_t^{(k+1)} = X_0 + \int_0^t b(s, Y_s^{(k)}) ds + \int_0^t \sigma(s, Y_s^{(k)}) dB_s. \tag{6.22}$$

   (We will show that $\sigma(s, Y_s^{(k)})$ is Itô integrable in the next subsection).

   - For $k = 0$

$$Y_t^{(1)} - Y_t^{(0)} = \int_0^t b(s, X_0) ds + \int_0^t \sigma(s, X_0) dB_s. \tag{6.23}$$

Proceeding as in (6.18),

$$\begin{aligned} E\big[|Y_t^{(1)} - Y_t^{(0)}|^2\big] &\le 2tE\Big[\Big(\int_0^t b(s,X_0)^2 ds\Big)\Big] + 2E\Big[\Big(\int_0^t \sigma(s,X_0)^2 ds\Big)\Big] \\ &\le 2tE\Big[\Big(\int_0^t 2C^2(1+X_0^2)ds\Big)\Big] + 2E\Big[\Big(\int_0^t 2C^2(1+X_0^2)ds\Big)\Big] \\ &= 4t^2C^2(1+E[X_0^2]) + 4tC^2(1+E[X_0^2]) \le A_1 t, \end{aligned} \tag{6.24}$$

where A_1 depends on $C, T, E[X_0^2]$.

- For $k \ge 1$, we write using (6.22)

$$Y_t^{(k+1)} - Y_t^{(k)} = \int_0^t \big(b(s,Y_s^{(k)}) - b(s,Y_s^{(k-1)})\big)ds + \int_0^t \big(\sigma(s,Y_s^{(k)}) - \sigma(s,Y_s^{(k-1)})\big)dB_s. \tag{6.25}$$

Proceeding as in (6.18) we get

$$E\big[|Y_t^{(k+1)} - Y_t^{(k)}|^2\big] \le 2(1+T)D^2 \int_0^t E\big[|Y_s^{(k)} - Y_s^{(k-1)}|^2\big]ds. \tag{6.26}$$

- Fact. For any $A_2 \ge \max\{2(1+T)D^2, A_1\}$

$$E\big[|Y_t^{(k+1)} - Y_t^{(k)}|^2\big] \le \frac{A_2^{k+1}t^{k+1}}{(k+1)!}. \tag{6.27}$$

Proof of the fact. We proceed by induction. For $k=0$ this follows from (6.24) for any $A_2 \ge A_1$. Suppose it is true for $k-1$. Then, by (6.26),

$$\begin{aligned} E\big[|Y_t^{(k+1)} - Y_t^{(k)}|^2\big] &\le 2(1+T)D^2 \int_0^t E\big[|Y_s^{(k)} - Y_s^{(k-1)}|^2\big]ds \\ &\le 2(1+T)D^2 \int_0^t \frac{A_2^k s^k}{k!} ds \\ &\le 2(1+T)D^2 \frac{A_2^k t^{k+1}}{(k+1)!} \le \frac{A_2^{k+1}t^{k+1}}{(k+1)!}, \end{aligned} \tag{6.28}$$

since $A_2 \ge 2(1+T)D^2$. This verifies (6.27).

- Fact. For any $t \in [0,T]$, the sequence $\{Y_t^{(n)}\}_{n\in\mathbb{N}}$ converges in $L^2(P)$ as $n \to \infty$ to a limit Y_t. Proof of the fact. We estimate for $m > n \ge 0$ and use (6.27)

$$\begin{aligned} \|Y_t^{(m)} - Y_t^{(n)}\|_{L^2(P)} &= \Big\|\sum_{k=n}^{m-1}(Y_t^{(k+1)} - Y_t^{(k)})\Big\|_{L^2(P)} \\ &\le \sum_{k=n}^{m-1}\|Y_t^{(k+1)} - Y_t^{(k)}\|_{L^2(P)} \le \sum_{k=n}^{\infty}\left[\frac{(A_2t)^{k+1}}{(k+1)!}\right]^{1/2} \underset{n\to\infty}{\to} 0. \end{aligned} \tag{6.29}$$

3. (*Proof that the candidate $\{Y_t\}_{t\in[0,T]}$ satisfies the equation*).
   For all n we have

$$Y_t^{(n+1)} = X_0 + \int_0^t b(s,Y_s^{(n)})ds + \int_0^t \sigma(s,Y_s^{(n)})dB_s. \tag{6.30}$$

Now we take the $L^2(P)$-limit of both sides. We know that $Y_t^{(n+1)} \underset{n\to\infty}{\to} Y_t$. Let us consider the first integral on the r.h.s. of (6.30). Using the Lipschitz property (6.13), we can write

$$\begin{aligned} E\Big[\Big(\int_0^t \big(b(s,Y_s^{(n)}) - b(s,Y_s)\big)ds\Big)^2\Big] &\le D^2 E\Big[\Big(\int_0^t |Y_s^{(n)} - Y_s|\,ds\Big)^2\Big] \\ &\le D^2 t\, E\Big[\int_0^t |Y_s^{(n)} - Y_s|^2\,ds\Big] \\ &= D^2 t\Big[\int_0^t E\big(|Y_s^{(n)} - Y_s|^2\big)\,ds\Big] \underset{n\to\infty}{\to} 0, \end{aligned} \tag{6.31}$$

where in the last step we used the dominated convergence. Let us now analyze the second integral on the r.h.s. of (6.30). By the Itô isometry and analogous steps as in (6.31)

$$\begin{aligned} E\Big[\Big(\int_0^t \big(\sigma(s,Y_s^{(n)}) - \sigma(s,Y_s)\big)dB_s\Big)^2\Big] &= E\Big[\int_0^t \big(\sigma(s,Y_s^{(n)}) - \sigma(s,Y_s)\big)^2 ds\Big] \\ &\le D^2\, E\Big[\int_0^t \big|Y_s^{(n)} - Y_s\big|^2 ds\Big] \underset{n\to\infty}{\to} 0. \end{aligned} \tag{6.32}$$

Thus, we obtained

$$Y_t = X_0 + \int_0^t b(s,Y_s)ds + \int_0^t \sigma(s,Y_s)dB_s. \tag{6.33}$$

(We used in (6.32) that $\{\sigma(s,Y_s)\}_{s\in[0,T]}$ is Itô integrable. This will be justified in the next subsection.)

4. (*Square integrability of the solution*). Fact. There holds (6.17) from the theorem:

$$E\Big[\int_0^T |Y_t|^2 dt\Big] < \infty. \tag{6.34}$$

Remark: We need $\mathcal{B}(0,T)\times\mathcal{F}$-measurability of $\{Y_t\}_{t\in[0,T]}$ for the integral (6.34) to make sense. This will be justified in the next subsection.

Proof of the fact. To check (6.34), we write

$$Y_t^{(n)}(\omega) = Y_t^{(0)}(\omega) + \sum_{k=0}^{n-1}(Y_t^{(k+1)}(\omega) - Y_t^{(k)}(\omega)). \tag{6.35}$$

Hence, by (6.27),

$$\begin{aligned} \|Y_t^{(n)}\|_{L^2(P)} &\le \|Z\|_{L^2(P)} + \sum_{k=0}^{n-1}\|Y_t^{(k+1)} - Y_t^{(k)}\|_{L^2(P)} \\ &\le \|Z\|_{L^2(P)} + \sum_{k=0}^{n-1}\frac{A_2^{k+1}t^{k+1}}{(k+1)!}. \end{aligned} \tag{6.36}$$

Taking the limit $n\to\infty$ of both sides, we get

$$\|Y_t\|_{L^2(P)} \le \|Z\|_{L^2(P)} + \sum_{k=0}^{\infty}\frac{A_2^{k+1}t^{k+1}}{(k+1)!} \le \|Z\|_{L^2(P)} + \exp(A_2 t). \tag{6.37}$$

Using this bound, we obtain

$$\int_0^T \|Y_t\|_{L^2(P)}^2 dt < \infty. \tag{6.38}$$

5. *(Continuity of the solution in t).* We will proceed similarly as in the proof of t-continuity of the Itô integral by combining the Doob martingale inequality with the Borel-Cantelli lemma. This allows to exploit the fact that continuity is preserved under the limits in the supremum norm.

   - From (6.25), we obtain

$$\sup_{t\in[0,T]} |Y_t^{(k+1)} - Y_t^{(k)}| \le \int_0^T \big|b(s,Y_s^{(k)}) - b(s,Y_s^{(k-1)})\big| ds + \sup_{t\in[0,T]} \Big| \int_0^t \big(\sigma(s,Y_s^{(k)}) - \sigma(s,Y_s^{(k-1)})\big) dB_s \Big|. \tag{6.39}$$

   - Note that for any positive random variables Z, Z_1, Z_2 s.t. $Z \le Z_1 + Z_2$ and $\varepsilon > 0$ we have

$$\{Z_1 \le \varepsilon/2\} \cap \{Z_2 \le \varepsilon/2\} \Rightarrow \{Z_1 + Z_2 \le \varepsilon\} \Rightarrow \{Z \le \varepsilon\}. \tag{6.40}$$

   By negating both sides

$$\begin{aligned} \{Z > \varepsilon\} &\Rightarrow \{Z_1 > \varepsilon/2\} \cup \{Z_2 > \varepsilon/2\}, \quad \text{hence,} \\ P(Z > \varepsilon) &\le P(Z_1 > \varepsilon/2) + P(Z_2 > \varepsilon/2). \end{aligned} \tag{6.41}$$

   Applying this consideration to inequality (6.39) with $\varepsilon = 2^{-k}$, we get

$$\begin{aligned} P(\sup_{t\in[0,T]} |Y_t^{(k+1)} - Y_t^{(k)}| > 2^{-k}) &\le P\Big(\Big[\int_0^T \big|b(s,Y_s^{(k)}) - b(s,Y_s^{(k-1)})\big| ds\Big]^2 > \big[2^{-k-1}\big]^2\Big) \\ &+ P\Big(\sup_{t\in[0,T]} \Big| \int_0^t \big(\sigma(s,Y_s^{(k)}) - \sigma(s,Y_s^{(k-1)})\big) dB_s \Big| > 2^{-k-1}\Big). \end{aligned} \tag{6.42}$$

   - We consider the first term on the r.h.s. of (6.42). By Markov's inequality, i.e. $P(Z_1 \ge \varepsilon) \le \frac{E[Z_1]}{\varepsilon}$, the Cauchy-Schwarz inequality, Lipschitz condition (6.13) and the bound (6.27)

$$\begin{aligned} &P\Big(\Big[\int_0^T \big|b(s,Y_s^{(k)}) - b(s,Y_s^{(k-1)})\big| ds\Big]^2 > \big[2^{-k-1}\big]^2\Big) \\ &\le 2^{2(k+1)} E\Big(\Big[\int_0^T \big|b(s,Y_s^{(k)}) - b(s,Y_s^{(k-1)})\big| ds\Big]^2\Big) \\ &\le 2^{2(k+1)} T E\Big(\int_0^T \big|b(s,Y_s^{(k)}) - b(s,Y_s^{(k-1)})\big|^2 ds\Big) \\ &\le 2^{2(k+1)} T D^2 \int_0^T E(\big|Y_s^{(k)} - Y_s^{(k-1)}\big|^2) ds \\ &\le 2^{2(k+1)} T D^2 \int_0^T \frac{A_2^k s^k}{k!} ds \le 2^{2(k+1)} T D^2 \frac{A_2^k T^{k+1}}{(k+1)!} \le \frac{(4A_2T)^{k+1}}{(k+1)!}. \end{aligned} \tag{6.43}$$

   - Now we consider the second term on the r.h.s. of (6.42). Using that the Itô integral is a martingale, we can apply the Doob martingale inequality (Theorem 4.23) $P[\sup_{t\in[0,T]} |M_t| > \varepsilon] \le \frac{1}{\varepsilon^2} E[|M_T|^2]$.

Then we use the Itô isometry, Lipschitz condition (6.13) and the bound (6.27):

$$\begin{aligned}
P\Big(\sup_{t\in[0,T]}\Big|\int_0^t \big(\sigma(s,Y_s^{(k)})-\sigma(s,Y_s^{(k-1)})\big)dB_s\Big| > 2^{-k-1}\Big) \\
&\le 2^{2(k+1)}E\Big(\Big|\int_0^T \big(\sigma(s,Y_s^{(k)})-\sigma(s,Y_s^{(k-1)})\big)dB_s\Big|^2\Big) \\
&= 2^{2(k+1)}E\Big(\int_0^T |\sigma(s,Y_s^{(k)})-\sigma(s,Y_s^{(k-1)})|^2 ds\Big) \\
&\le 2^{2(k+1)}D^2\int_0^T E(|Y_s^{(k)}-Y_s^{(k-1)}|^2)ds \le \frac{(4A_2T)^{k+1}}{(k+1)!},
\end{aligned} \tag{6.44}$$

where in the last step we proceeded analogously as in (6.43).

- Summing up the last three steps:

$$P(\sup_{t\in[0,T]}|Y_t^{(k+1)}-Y_t^{(k)}|>2^{-k}) \;\le\; 2\frac{(4A_2T)^{k+1}}{(k+1)!}. \tag{6.45}$$

- Recall the Borel-Cantelli lemma (HS1): Let $\{E_k\}_{k\in\mathbb{N}}$ be a sequence of events. If $\sum_{k=1}^{\infty}P(E_k)<\infty$, then

$$P(\omega\in\Omega\,:\,\omega\in E_k \text{ for infinitely many } k\,)=0. \tag{6.46}$$

- We set $E_k=\{\sup_{t\in[0,T]}|Y_t^{(k+1)}-Y_t^{(k)}|>2^{-k}\}$ and obtain $\sum_{k=1}^{\infty}P(E_k)<\infty$ by (6.45). Thus, by (6.46),

$$P(\sup_{t\in[0,T]}|Y_t^{(k+1)}-Y_t^{(k)}| \;>\; 2^{-k} \text{ for infinitely many } k\,)=0. \tag{6.47}$$

  Let us call N the null set under $P(\dots)$ above. For all $\omega\notin N$ (i.e., for almost all ω) there holds the opposite condition: $\sup_{t\in[0,T]}|Y_t^{(k+1)}-Y_t^{(k)}|>2^{-k}$ for finitely many k. So there exists $k_1(\omega)$ s.t.

$$\sup_{t\in[0,T]}|Y_t^{(k+1)}-Y_t^{(k)}|\le 2^{-k} \quad \text{for} \quad k\ge k_1(\omega). \tag{6.48}$$

  Therefore, the sequence

$$Y_t^{(n)}(\omega)=Y_t^{(0)}(\omega)+\sum_{k=0}^{n-1}(Y_t^{(k+1)}(\omega)-Y_t^{(k)}(\omega)) \tag{6.49}$$

  is uniformly convergent as $n\to\infty$ in $[0,T]$ for a.a. ω. We denote the limit by X_t.

- Fact. $\{X_t\}_{t\in[0,T]}$ is t-continuous version of $\{Y_t\}_{t\in[0,T]}$.

  Proof of the fact: $Y_t^{(n)}$ are t-continuous for ω in some measure one sets $\tilde\Omega_n$ as Itô integrals (cf. Theorem 4.22). Hence they are also t-continuous on the measure one set $\tilde\Omega:=\bigcap_{n\in\mathbb{N}}\tilde\Omega_n$. Then $\tilde\Omega\backslash N$ is still of measure one. On this set, the sequence (6.49) consists of continuous functions and converges uniformly to $\{X_t\}_{t\in[0,T]}$. Thus $\{X_t\}_{t\in[0,T]}$ is t-continuous.

  For any fixed t we have $Y_t^{(n)}\to Y_t$ in $L^2(P)$ hence in probability. On the other hand, $Y_t^{(n)}\to X_t$ almost surely, hence in probability. By the uniqueness of the limit in probability, we have $X_t=Y_t$ on some measure one sets Ω_t'. Thus $\{X_t\}_{t\in[0,T]}$ is a continuous version of $\{Y_t\}_{t\in[0,T]}$.

6.2.2 Proof of Theorem 6.1: Measure theoretic details.

Remark 6.2. *Equation (6.16) is understood to hold 'up to indistinguishability'. That is, there exists a t-independent $\tilde{\Omega} \subset \Omega$, $P(\tilde{\Omega}) = 1$, s.t. for all $t \in [0,T]$ and $\omega \in \tilde{\Omega}$ (6.16) holds true. Similarly:*

- *We mean the t-continuity of $\{X_t\}_{t\in\mathbb{R}_+}$ up to indistinguishability, i.e. there is $\tilde{\Omega}_1 \subset \Omega$, $P(\tilde{\Omega}_1) = 1$ s.t. $t \mapsto X_t(\omega)$ is continuous for all $t \in [0,T]$ and $\omega \in \tilde{\Omega}$.*
- *We mean here the uniqueness up to indistinguishability, i.e., given two solutions $\{X_t\}_{t\in[0,T]}$ and $\{\hat{X}_t\}_{t\in[0,T]}$, with the same initial value Z, satisfying the conditions of Theorem 6.1, there is $\tilde{\Omega}_2 \subset \Omega$, $P(\tilde{\Omega}_2) = 1$, s.t. $X_t = \hat{X}_t$ for all $t \in [0,T]$ and $\omega \in \tilde{\Omega}$. In other words*

$$P(X_t = \hat{X}_t \text{ for all } t \in [0,T]) = 1. \tag{6.50}$$

The distinction between $\tilde{\Omega}$, $\tilde{\Omega}_1$, $\tilde{\Omega}_2$ is not important, as we can pass to $\tilde{\Omega}\cap\tilde{\Omega}_1\cap\tilde{\Omega}_2$ and use that an intersection of a countable family of sets of measure one is a set of measure one.

Statements 'up to indistinguishability' are stronger than statements 'up to a version'. In fact, for the latter statements, sets $\tilde{\Omega}$ may depend on t. The uncountable intersections $\bigcap_{t\in[0,T]} \tilde{\Omega}_t$ may not be measurable, so it is not automatic to pass from a statement 'up to a version' to a statement 'up to indistinguishability'.

1. (*Uniqueness*).
   - Recall that from $v(t) := E\big[|X_t - \hat{X}_t|^2\big] = 0$ we get $X_t = \hat{X}_t$ almost surely, i.e., on some set $\tilde{\Omega}_t \subset \Omega$, $P(\tilde{\Omega}_t) = 1$. That is, $\{\hat{X}_t\}_{t\in[0,T]}$ is a version of $\{X_t\}_{t\in[0,T]}$.
   - Thus, we have $X_t(\omega) = \hat{X}_t(\omega)$ for $t \in \mathbb{Q} \cap [0,T]$ and $\omega \in \tilde{\Omega} := \bigcap_{t\in\mathbb{Q}\cap[0,T]} \tilde{\Omega}_t$. The latter set is measurable and $P(\tilde{\Omega}) = 1$ (Exercise).
   - We are considering only solutions which are t-continuous up to indistinguishability. That is, there exist $\Omega', \hat{\Omega}'$ s.t. $P(\Omega') = P(\hat{\Omega}') = 1$ and $t \mapsto X_t(\omega)$, $\omega \in \Omega'$, and $t \mapsto \hat{X}_t(\hat{\omega})$, $\hat{\omega} \in \hat{\Omega}'$, are continuous.
   - Since $\mathbb{Q}\cap[0,T]$ is dense in $[0,T]$ and $t \mapsto X_t(\omega)$, $t \mapsto \hat{X}_t(\omega)$, $\omega \in \Omega_1 := \tilde{\Omega}\cap\Omega'\cap\hat{\Omega}'$, are continuous, we obtain $X_t(\omega) = \hat{X}_t(\omega)$ for $t \in [0,T]$ and $\omega \in \Omega_1$, $P(\Omega_1) = 1$. This is the uniqueness up to indistinguishability (6.50).
2. (*Construction of a candidate $\{Y_t\}_{t\in[0,T]}$ for the solution.*) Let us recall the relevant definition of Itô integrable functions:

   Definition 6.3. *Let $\mathcal{V}_{\mathcal{H}}(0,T)$ be the class of functions*

   $$(t,\omega) \mapsto f(t,\omega) : \mathbb{R}_+ \times \Omega \to \mathbb{R} \tag{6.51}$$

   such that

   (i) f is $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$-measurable.

   (ii') There exists a filtration $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$ s.t. $(B_t, \mathcal{H}_t)_{t\in\mathbb{R}_+}$ is a martingale and f is $\{\mathcal{H}_t\}_{t\in\mathbb{R}_+}$-adapted.

   (iii) $E\big[\int_0^T f(t,\cdot)^2 dt\big] < \infty$.

   Lemma 6.4. *For Z independent of $\mathcal{F}_\infty^{(d)}$, the process $(B_t, \mathcal{F}_t^Z)_{t\in\mathbb{R}_+}$ is a martingale.*

   Proof. A generalization of (4.30). (Exercise). □

Lemma 6.5. *Properties of the solution $\{X_t\}_{t\in[0,T]}$ listed in Theorem 6.1, namely adaptedness to $\{\mathcal{F}_t^Z\}_{t\in[0,T]}$, t-continuity, and square integrability, imply that $\{X_t\}_{t\in[0,T]}$ is in $\mathcal{V}_{\mathcal{F}^Z}(0,T)$. Thus, the Itô integral*

$$\int_0^t \sigma(s, X_s) dB_s \tag{6.52}$$

from (6.16), is well defined.

Proof. Properties (ii') and (iii) of Definition 6.3 are obvious, given Lemma 6.4. Property (i) requires an argument: Let $s_j = j2^{-n} \wedge T$ by our usual partition of $[0,T]$. Let $q_n(t)$ be the point of the partition which is the closest to the left from a given t. Let

$$X_t^{(n)}(\omega) := X_{q_n(t)}(\omega) = \sum_j \chi_{[s_j, s_{j+1})}(t) X_{s_j}(\omega). \tag{6.53}$$

It is a finite sum of products of $\chi_{[s_j,s_{j+1})}$ which are $\mathcal{B}(\mathbb{R}_+)$-measurable and X_{s_j} which are $\mathcal{F}$-measurable as they are measurable w.r.t. $\mathcal{F}_{s_j}^Z \subset \mathcal{F}$ by adaptedness (ii'). Hence $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$-measurable. Now, by t-continuity,

$$X_t(\omega) = \lim_{n\to\infty} X_t^{(n)}(\omega) \tag{6.54}$$

so $\{X_t\}_{t\in[0,T]}$ is $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$-measurable as a pointwise limit of measurable functions. □

Lemma 6.6. *Recall the processes $\{Y_t^{(k)}\}_{t\in[0,T]}$ defined by (6.22), i.e.,*

$$Y_t^{(k+1)} = Y_t^{(0)} + \int_0^t b(s, Y_s^{(k)}) ds + \int_0^t \sigma(s, Y_s^{(k)}) dB_s, \tag{6.55}$$

with $Y_t^{(0)} = X_0 = Z$, where $E[Z^2] < \infty$. These processes and $Y_t = L^2(P)\text{-}\lim_{k\to\infty} Y_t^{(k)}$ are in $\mathcal{V}_{\mathcal{F}_t^Z}(0,T)$.

Proof. As for $Y_t^{(0)} = X_0 = Z$, since it is t-independent and $\mathcal{F}_t^Z$-measurable for any t, it clearly belongs to $\mathcal{V}_{\mathcal{F}_t^Z}(0,T)$. As for $Y_t^{(1)}$, we have

$$Y_t^{(1)} = Z + \int_0^t b(s, Z) ds + \int_0^t \sigma(s, Z) dB_s. \tag{6.56}$$

First, we justify that $(s,\omega) \mapsto \sigma(s, Z(\omega))$ is in $\mathcal{V}_{\mathcal{F}_t^Z}(0,T)$.

(i) $\sigma(s, Z(\omega)) = \sigma \circ (\mathrm{id}(s) \times Z(\omega))$ is a composition of measurable functions, hence $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$-measurable.

(ii') For any s,t the function $\omega \mapsto \sigma(s, Z(\omega))$ is $\mathcal{F}_t^Z$ measurable by our measurability assumption on Z.

(iii) Property (6.12) gives

$$E\left[\int_0^T |\sigma(s,Z)|^2 ds\right] \le E\left[\int_0^T |(1+|Z|)|^2 ds\right] \le 2T(1 + E[Z^2]). \tag{6.57}$$

Thus the integral $I_t(\omega) := \int_0^t \sigma(s, Z(t,\omega)) dB_s(\omega)$ in (6.56) is well defined and we can verify that it is in $\mathcal{V}_{\mathcal{F}_t^Z}(0,T)$:

(iii) Follows by the Itô isometry and (6.57).

(ii') For any fixed t the process I_t is an $L^2(P)$-limit of $\mathcal{F}_t^Z$-measurable processes $I_t^{(n)}$, hence $\mathcal{F}_t^Z$-measurable.

(i) Following up on (ii'), one is tempted to use that an $L^2(P)$ convergent sequence has an almost surely convergent subsequence (Exercise). Thus we have a sequence $\{n_k^{(t)}\}_{k\in\mathbb{N}}$ and a measure one set $\tilde{\Omega}_t$ s.t. $I_t^{(n_k^{(t)})}(\omega) \underset{k\to\infty}{\to} I_t(\omega)$, $\omega \in \tilde{\Omega}_t$. However, the t-dependence of $\tilde{\Omega}_t$ and $n_k^{(t)}$ prevents us from concluding the $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$-measurability of $\{I_t\}_{t\in[0,T]}$ from the $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$-measurability of $\{I_t^{(n_k^{(t)})}\}_{t\in[0,T]}$ this way. This is a limitation of the slogan "measurability is preserved under limits".

After this side remark, let us prove (i). As we know from the proof of the t-continuity of the Itô integral (Theorem 4.22) there is a t-independent sequence $\{n_k\}_{k\in\mathbb{N}}$ s.t. $I_t^{(n_k)}(\omega) \to I_t(\omega)$ for all $t \in [0,T]$ and ω in some t-independent measure one set $\tilde{\Omega}'$. From this we can conclude $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$-measurability of $\{I_t\}_{t\in[0,T]}$ from the $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$-measurability of $\{I_t^{(n_k^{(t)})}\}_{t\in[0,T]}$ as this is the usual pointwise convergence, which preserves measurability.

By similar but simpler considerations we obtain that $Z + \int_0^t b(s,Z)ds$ from (6.56) is in $\mathcal{V}_{\mathcal{F}_t^Z}(0,T)$, hence $\{Y_t^{(1)}\}_{t\in[0,T]} \in \mathcal{V}_{\mathcal{F}^Z}(0,T)$. We obtain that $\{Y_t^{(k)}\}_{t\in[0,T]} \in \mathcal{V}_{\mathcal{F}^Z}(0,T)$ using (6.55) and iterating the above arguments. To show that the limiting process $\{Y_t\}_{t\in[0,T]}$ is in $\mathcal{V}_{\mathcal{F}^Z}(0,T)$ we argue similarly as for $\{I_t\}_{t\in[0,T]}$ above. Instead of the proof of Theorem 4.22 we refer to the proof of t-continuity of $\{Y_t\}_{t\in[0,T]}$ from the previous subsection. $\square$

References

[Ba] P. Baldi. *Stochastic Calculus. An Introduction Through Theory and Exercises.* Universitext, Springer, 2017.

[BT] D.P. Bertsekas and J.N. Tsitsiklis. *Introduction to Probability.* Lecture Notes, M.I.T. Fall, 2000.

[JP] J. Jacod and P. Protter. *Probability Essentials.* Universitext, Springer, 2004.

[Ma] R. Mansuy. *The Origins of the Word "Martingale".* Electronic Journ@l for History of Probability and Statistics. Vol 5, no 1; June 2009.

[Ok] B. Øksendal. *Stochastic Differential Equations. An Introduction with Applications.* Springer-Verlag, 2000.

[Pa] K. R. Parthasarathy. *Probability measures on metric spaces.* Academic Press, 1967.

[Ro] G. Roepstorff. *Path Integral Approach to Quantum Physics. An Introduction.* Springer-Verlag, 1991.

[Ru] W. Rudin. *Real and complex analysis.* McGraw-Hill Inc. 1970

[SP] R.L. Schilling and L. Partzsch. *Brownian motion. An introduction to stochastic processes.* De Gruyter, 2012.