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 12-13 (Except Tuesday 7.10) or by appointment.
- Textbook: In general Øksendal [Ok]. But first lecture: Roepstorff [Ro, Ch 1].

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

$$m\frac{d^2x(t)}{dt^2} = -kx(t), \quad k > 0. \tag{1}$$

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

$$x(t) = x_0\cos(\omega t) + \frac{v_0}{\omega}\sin(\omega t), \quad \omega = \sqrt{\frac{k}{m}}. \tag{2}$$

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

$$m\frac{d^2x(t)}{dt^2} = -kx(t) - \gamma\frac{dx(t)}{dt}, \quad \gamma > 0. \tag{3}$$

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):

$$m\frac{d^2x(t)}{dt^2} = -kx(t) - \gamma\frac{dx(t)}{dt} + \sigma\frac{dB_t}{dt}, \quad \sigma > 0. \tag{4}$$

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,

$$\int_0^t ds\Big\{\text{independent random kicks at time } s\Big\} \tag{5}$$

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,

$$1 = \frac{d}{dt}t = \frac{d}{dt}E(B_t^2) = 2E(B_t \frac{dB_t}{dt}) = \lim_{\Delta t \to 0} 2E\big(B_t \frac{B_{t+\Delta t} - B_t}{\Delta t}\big) = 0, \tag{6}$$

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:
   $$\frac{\Delta X}{X} = r\Delta t. \tag{7}$$
   - If we take the idealization of a continuous interest rate, this gives
   $$\frac{dX_t}{dt} = rX_t, \tag{8}$$
   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:
   $$\frac{dX_t}{dt} = \left(r + \alpha \frac{dB_t}{dt}\right)X_t, \quad r, \alpha > 0. \tag{9}$$
   Naive solution has the form:
   $$X_t = X_0 e^{rt + \alpha B_t} \tag{10}$$
   One can make a mathematical sense out of it. But one can also make mathematical sense of the solution:
   $$X_t = X_0 e^{(r - \frac{1}{2}\alpha^2)t + \alpha B_t}, \tag{11}$$
   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 $\{X_s\}_{0 \le s \le t}$ 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 (9) (a stock) and a safe investment (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 CLT. (Picture from the notes of Bertsekas and Tsitsiklis [BT, p.6]).]

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
   $$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \tag{12}$$
   [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 $I \subset \{1, \ldots n\}$
   $$P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i). \tag{13}$$
6. Def. Conditional probability:
   $$P(A|B) := \frac{P(A \cap B)}{P(B)} \text{ for } P(B) \neq 0, \tag{14}$$
   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
$$\mu_X(B) = P(X^{-1}(B)) = P(X \in B). \tag{15}$$
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
$$\mu_X(A) = \int_A \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}} dx, \tag{16}$$
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 $X_1, \dots, X_n$ the joint law is the measure on $\mathbb{R}^{dn}$ given by
$$\mu_{X_1,\dots,X_n}(A) = P((X_1, \dots, X_n) \in A), \quad A \in \mathcal{B}(\mathbb{R}^{dn}). \tag{17}$$

10. Def. If $\int |X(\omega)| dP(\omega) < \infty$ then the number
$$E(X) := \int_\Omega X(\omega) dP(\omega) = \int_{\mathbb{R}} x \, d\mu_X(x) \tag{18}$$
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
$$\mathrm{var}(X) = E(X^2) - E(X)^2 = E[(X - E(X))^2] \tag{19}$$
is called the variance.

12. Def. We say that real-valued random variables $X_1, \dots, X_n$ are independent if
$$\mu_{X_1,\dots,X_n} = \mu_{X_1} \otimes \dots \otimes \mu_{X_n}. \tag{20}$$

13. Fact: If two random variables X, Y are independent, then they are uncorrelated
$$E[XY] = E[X]E[Y] \tag{21}$$
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 $X = \chi_{F_1}$, $Y = \chi_{F_2}$.
$$E[\chi_{F_1}\chi_{F_2}] = \int \chi_{F_1}(\omega)\chi_{F_2}(\omega) dP(\omega) = P(F_1 \cap F_2) = P(F_1)P(F_2). \tag{22}$$
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 $\mathbb{R}$
$$\int f(x) d\mu_{Y_n}(x) \to \int f(x) d\mu_Y(x). \tag{23}$$
(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

$$\mathcal{F}_0 = \{\emptyset, \{H\}, \{T\}, \{H, T\}\} \tag{24}$$

    - 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,

$$P(\{\omega\}) = \prod_{i=1}^{n} P_0(\{\omega_i\}) = p^{R_n(\omega)}(1-p)^{n-R_n(\omega)}, \tag{25}$$

      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.

$$\xi_i(\omega) = \begin{cases} +1 & \text{if } \omega_i = H, \\ -1 & \text{if } \omega_i = T. \end{cases} \tag{26}$$

4. Then $R_n = \#\{\, i : \xi_i = +1\}$ is the number of heads in n trials.

5. Fact: We have

$$P(R_n = \ell) = P(\ell \text{ heads come up in an } n\text{-toss sequence}) = \binom{n}{\ell} p^{\ell}(1-p)^{n-\ell}. \tag{27}$$

   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,

$$P(\text{HTTHTT}) = P(\text{H})P(\text{H})P(\text{T})P(\text{H})P(\text{T})P(\text{T}) = p^{\ell}(1-p)^{n-\ell}. \tag{28}$$

    - 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

$$P(\text{HHTHTT or HTTHTT}) = P(\text{HHTHTT}) + P(\text{HTTHTT}). \tag{29}$$

    - 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 (25).

4. We want to compute probability

$$P(\text{particle ends up at } x = kh \text{ after } n \text{ steps, starting from zero}) =: P(kh, n\tau) =: P[k, n]. \tag{30}$$

   - 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

$$P[k, n] := P(R_n = (n + k)/2) = \binom{n}{\frac{1}{2}(n + k)} \frac{1}{2^n}, \tag{31}$$

   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 $\tilde{n} \to n$):

$$B_t^{(n)} := \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor}, \tag{32}$$

   with the convention $S_0 = 0$.

6. Fact: By the CLT the limit

$$B_t := \lim_{n \to \infty} B_t^{(n)} \tag{33}$$

   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 $\xi_i(\omega) = \tilde{\xi}_i(\omega_i)$.

$$\mu_{\xi_i}(A) := P(\omega : \xi_i(\omega) \in A) = P(\omega : \tilde{\xi}_i(\omega_i) \in A). \tag{34}$$

   - Independence: For $n = 2$

$$\begin{aligned} \mu_{\xi_1, \xi_2}(A_1 \times A_2) &= P(\omega : (\xi_1(\omega), \xi_2(\omega)) \in A_1 \times A_2) \\ &= P(\omega : (\tilde{\xi}_1(\omega_1), \tilde{\xi}_2(\omega_2)) \in A_1 \times A_2) \\ &= P(\omega : \tilde{\xi}_1(\omega_1) \in A_1, \tilde{\xi}_2(\omega_2) \in A_2) \\ &= P(\omega : \tilde{\xi}_1(\omega_1) \in A_1) P(\omega : \tilde{\xi}_2(\omega_2) \in A_2) = \mu_{\xi_1}(A_1) \mu_{\xi_2}(A_2), \end{aligned} \tag{35}$$

   where in the next to the last step we made use of the independence of coin tosses, i.e. (25). For arbitrary n the argument is analogous.

- We compute

$$m_j = E(\xi_j) = \int \xi_j(\omega)dP(\omega) = \int \xi_j d\mu_{\xi_j} = (+1)\frac{1}{2} + (-1)\frac{1}{2} = 0, \quad (36)$$

$$\sigma_j^2 = E(\xi_j^2) = \int \xi_j(\omega_j)^2 dP(\omega_j) = \int \xi_j^2 d\mu_{\xi_j} = (+1)\frac{1}{2} + (+1)\frac{1}{2} = 1. \quad (37)$$

- 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

$$\mu_{\sqrt{t}Y}(A) = P(\sqrt{t}Y \in A) = P(Y \in A/\sqrt{t}) = \int_{A/\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \int_A \frac{1}{\sqrt{2\pi t}} e^{-y^2/(2t)} dy, \quad (38)$$

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

$$E((B_0^{(n)} - B_t^{(n)})(B_t^{(n)} - B_s^{(n)})) = 0. \quad (39)$$

The two factors have the form

$$B_t^{(n)} - B_s^{(n)} = \frac{1}{\sqrt{n}}(\xi_{\lfloor nt \rfloor + 1} + \cdots + \xi_{\lfloor ns \rfloor}), \quad (40)$$

$$B_0^{(n)} - B_t^{(n)} = \frac{1}{\sqrt{n}}(\xi_1 + \cdots + \xi_{\lfloor nt \rfloor}). \quad (41)$$

They involve independent coin tosses. For independent random variables we have $E(XY) = E(X)E(Y)$ by (20), so (39) follows.

8. We can also compute $E(B_t^{(n)} B_s^{(n)})$. By (39), for $s > t$,

$$E(B_t^{(n)} B_s^{(n)}) = E(B_t^{(n)2}) = \frac{1}{n} E(\xi_1^2 + \cdots + \xi_{\lfloor nt \rfloor}^2) = \frac{\lfloor nt \rfloor}{n}. \quad (42)$$

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

$$E(B_t^{(n)} B_s^{(n)}) = \min(\frac{\lfloor nt \rfloor}{n}, \frac{\lfloor ns \rfloor}{n}). \quad (43)$$

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,

$$P[k, n+1] = \binom{n+1}{\frac{1}{2}(n+1+k)} \frac{1}{2^{n+1}} = \frac{1}{2}\binom{n}{\frac{1}{2}(n+1+k)} \frac{1}{2^n} + \frac{1}{2}\binom{n}{\frac{1}{2}(n+1+k) - 1} \frac{1}{2^n}. \quad (44)$$

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

$$\frac{P(kh,(n+1)\tau)-P(kh,n\tau)}{\tau}=\frac{h^2}{2\tau}\frac{P((k+1)h,n\tau)-2P(kh,n\tau)+P((k-1)h,n\tau)}{h^2} \tag{45}$$

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

$$\frac{\partial \tilde{p}(x,t)}{\partial t} = D\frac{\partial^2 \tilde{p}(x,t)}{\partial x^2}. \tag{46}$$

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:

$$\tilde{p}(x,t) = \frac{1}{\sqrt{4\pi Dt}}\exp\Big(-\frac{x^2}{4Dt}\Big). \tag{47}$$

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

$$\{X_t\}_{t\in\mathcal{T}} \tag{48}$$

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

$$C_{t_1,\dots,t_k;F_1,\dots,F_k} := \{\tilde{\omega} \in (\mathbb{R}^d)^{\mathcal{T}} : \tilde{\omega}(t_1) \in F_1, \dots, \tilde{\omega}(t_k) \in F_k\}, \quad F_i \subset \mathbb{R}^d \text{ Borel.} \tag{49}$$

   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 (49) 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

$$\tilde{P}(C) = P(\iota^{-1}(C)) \tag{50}$$

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

$$P(\{\omega \in \Omega \,:\, X_t(\omega) = Y_t(\omega)\}) = 1 \text{ for all } t. \tag{51}$$

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

$$\begin{aligned} \nu_{t_1,\dots,t_k}(F_1 \times F_2 \times \cdots \times F_k) &= P[X_{t_1} \in F_1, \dots, X_{t_k} \in F_k] \\ &= \int \chi(X_{t_1}(\omega) \in F_1) \dots \chi(X_{t_k}(\omega) \in F_k) dP(\omega), \end{aligned} \tag{52}$$

where $t_i \in \mathcal{T}$ and $F_1, \dots F_k$ 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 $t_1, \dots, t_k \in \mathcal{T}$ let $\nu_{t_1,\dots,t_k}$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 $\{1, 2, \dots, k\}$ and

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

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

$$\nu_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k) = P(X_{t_1} \in F_1, \cdots, X_{t_k} \in F_k), \tag{53}$$

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 $\{\nu_{t_1,\dots t_k}\}_{t_1,\dots t_k \geq 0}$ satisfying the conditions from the Kolmogorov theorem.

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

$$p(t, x, y) = (2\pi t)^{-d/2} \exp\big(-\frac{|x-y|^2}{2t}\big) \quad \text{for } y \in \mathbb{R}^d, t > 0. \tag{54}$$

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

$$\begin{aligned} &\nu_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k) \\ &= \int_{F_1\times\cdots\times F_k} p(t_1, x, x_1) p(t_2 - t_1, x_1, x_2) \cdots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \dots dx_k, \end{aligned} \tag{55}$$

   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

$$\nu_{t_1}(F_1) = \int_{F_1} p(t_1, 0, x_1)dx_1 = \int_{F_1} \tilde{p}(t_1, x_1)dx_1 \tag{56}$$

where $\tilde{p}$ appeared in (47) 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:

$$P^x(B_{t_1} \in F_1, \cdots, B_{t_k} \in F_k) = \int_{F_1\times\cdots\times F_k} p(t_1, x, x_1)\cdots p(t_k - t_{k-1}, x_{k-1}, x_k)dx_1\cdots dx_k. \tag{57}$$

This formula determines the law of a family of random variables in the sense of (17). In the notation of (17) it reads

$$P^x(B_{t_1} \in F_1, \cdots, B_{t_k} \in F_k) = \mu_{B_{t_1},\ldots,B_{t_k}}(F_1 \times \ldots \times F_k). \tag{58}$$

Hence, the measure on the l.h.s. of (57) 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

$$\begin{aligned}
&\int f(B_{t_1},\ldots,B_{t_k})P^x(B_{t_1} \in dx_1, \cdots, B_{t_k} \in dx_k) \\
&:= \int f(x_1,\ldots,x_k)\mu_{B_{t_1},\ldots,B_{t_k}}(x_1,\ldots,x_k) \\
&= \int f(x_1,\ldots,x_k)\{p(t_1,x,x_1)\cdots p(t_k - t_{k-1}, x_{k-1}, x_k)\}dx_1\cdots dx_k.
\end{aligned} \tag{59}$$

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.,

   $$B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}}, \tag{60}$$

   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 (49), 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*

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2} \right) \tag{61}$$

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

$$P(X \in G) = \int_G p_X(x)dx \quad \text{for} \quad G \subset \mathbb{R} \text{ Borel.} \tag{62}$$

By Gaussian integration:

$$E(X) = \int_\Omega X dP = \int_\mathbb{R} x p_X(x)dx = \int_\mathbb{R} (x+m)\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2} \right) dx = m, \tag{63}$$

$$\text{var}(X) = E[(X-m)^2] = \int_\mathbb{R} (x-m)^2 p_X(x)dx = \sigma^2. \tag{64}$$

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*

$$p_X(\underbrace{x_1, \ldots, x_d}_{x}) = \frac{\sqrt{\det(A)}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{j,k}(x_j - m_j)a_{j,k}(x_k - m_k) \right) \tag{65}$$

$$= \frac{\sqrt{\det(A)}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(x-m)^T A(x-m) \right) \tag{66}$$

$$= \frac{\sqrt{\det(A)}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\langle (x-m), A(x-m)\rangle \right) \tag{67}$$

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,

$$E(X_j) = m_j, \tag{68}$$

$$E[(X_j - m_j)(X_k - m_k)] = c_{j,k}. \tag{69}$$

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*

$$\phi_X(u_1, \ldots, u_d) = E[\exp(i(u_1X_1 + \cdots + u_dX_d))] = \int_{\mathbb{R}^d} \mathrm{e}^{i\langle u,x\rangle}\mu_X(dx), \tag{70}$$

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*

$$\int_{\mathbb{R}^d} \mathrm{e}^{i\langle u,x\rangle}\mu_{X_1}(dx) = \int_{\mathbb{R}^d} \mathrm{e}^{i\langle u,x\rangle}\mu_{X_2}(dx), \tag{71}$$

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*

$$\int_{\mathbb{R}^d} \hat{f}(x)\mu_{X_1}(dx) = \int_{\mathbb{R}^d} \hat{f}(x)\mu_{X_2}(dx), \quad \text{where} \quad \hat{f}(x) := \int_{\mathbb{R}^d} \mathrm{e}^{i\langle u,x\rangle} f(u)du \tag{72}$$

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*

$$\phi_X(u_1, \ldots, u_d) = \exp\Big(-\frac{1}{2}\sum_{j,k} u_j c_{j,k} u_k + i \sum_j u_j m_j \Big) \tag{73}$$

$$= \exp\Big(-\frac{1}{2}\langle u, Cu\rangle + i\langle u, m\rangle \Big). \tag{74}$$

- 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 (73).
- 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*

$$X = (X_1, \ldots, X_d) \tag{75}$$

is multi-normal iff

$$Y = \lambda_1 X_1 + \cdots + \lambda_d X_d =: \langle \lambda, X\rangle \text{ is normal for all } \lambda_1, \ldots, \lambda_d \in \mathbb{R}. \tag{76}$$

The equivalence also holds for (multi-)normal non-degenerate random variables if we restrict to $\lambda \neq 0$ *in (76).*

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

$$E[\exp(i\tilde{u}Y)] = E[\exp(i\tilde{u}\langle\lambda, X\rangle)] = \exp\Big(-\frac{1}{2}\tilde{u}^2\langle\lambda, C\lambda\rangle + i\tilde{u}\langle\lambda, m\rangle \Big), \tag{77}$$

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.,

$$E[\exp(i\tilde{u}Y)] = \exp\big(-\frac{1}{2}\tilde{u}^2\sigma^2 + i\tilde{u}m\big). \tag{78}$$

Then, by (63), (64)

$$m = E(Y) = \sum_j \lambda_j E(X_j), \tag{79}$$

$$\sigma^2 = E[(Y - E(Y))^2] = E\big[\big(\sum_j \lambda_j (X_j - m_j)\big)^2\big] = \sum_{i,j} \lambda_i \lambda_j \underbrace{E((X_i - m_i)(X_j - m_j))}_{c_{i,j}}, \tag{80}$$

[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 (73), (68), (69) 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.,*

$$E[(Y_0 - E[Y_0])(Y_j - E[Y_j])] = 0, \quad 1 \leq j \leq n. \tag{81}$$

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

$$P[Y_0 \in G_0, Y_1 \in G_1, \dots, Y_d \in G_d] = P[Y_0 \in G_0] \cdot P[Y_1 \in G_1, \dots, Y_d \in G_d], \tag{82}$$

for all Borel sets $G_0, \dots G_d \subset \mathbb{R}$. By (81), the covariance matrix

$$c_{k,j} = E[(Y_k - E[Y_k])(Y_j - E[Y_j])] \tag{83}$$

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

$$\langle u, Cu \rangle = c_{0,0} u_0^2 + \langle \vec{u}, \tilde{C}\vec{u} \rangle. \tag{84}$$

Therefore, by (73),

$$\begin{aligned} \phi_X(u) &= \exp\Big(-\frac{1}{2}\langle u, Cu \rangle + i\langle u, m \rangle \Big) = \exp\Big(-\frac{1}{2} c_{0,0} u_0^2 - \frac{1}{2}\langle \vec{u}, \tilde{C}\vec{u} \rangle + i\langle u_0, m_0 \rangle + i\langle \vec{u}, \vec{m} \rangle \Big) \\ &= \phi_{Y_0}(u_0)\phi_{(Y_1,\dots,Y_d)}(u_1, \dots, u_d). \end{aligned} \tag{85}$$

By Remark 2.5, ϕ_X determines the law uniquely, so (82) 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 $0 \leq t_1 \leq \cdots \leq t_k$ the random variable $Z = (B_{t_1}, \dots B_{t_k}) \in \mathbb{R}^{dk}$ has a multi-normal distribution with mean and covariance*

$$m = (x, x, \dots, x), \quad c_{(j,\ell),(j',\ell')} = \min(t_j, t_{j'})\delta_{\ell,\ell'}, \tag{86}$$

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

$$P^x(B_{t_1}\in F_1,\cdots,B_{t_k}\in F_k)$$
$$=\int_{F_1\times\cdots\times F_k} p(t_1,x,x_1)\cdots p(t_{k-1}-t_{k-2},x_{k-2},x_{k-1})p(t_k-t_{k-1},x_{k-1},x_k)dx_1\cdots dx_k, \quad (87)$$

where

$$p(t,x,y)=(2\pi t)^{-d/2}\exp\big(-\frac{|x-y|^2}{2t}\big) \qquad \text{for } y\in\mathbb{R}^d, t>0. \quad (88)$$

Consequently

$$\begin{aligned}
E[\exp(i\langle u,Z\rangle)] &= \int \exp(iu_1B_{t_1}+\cdots+iu_kB_{t_k})P^x(B_{t_1}\in dx_1,\dots B_{t_k}\in dx_k)\\
&= \int \exp(iu_1x_1+\cdots+iu_kx_k)p(t_1,x,x_1)\cdots p(t_k-t_{k-1},x_{k-1},x_k)dx_1\cdots dx_k\\
&= \int \exp(iu_1x_1+\cdots+iu_kx_k)\tilde{p}(\Delta t_1,\underbrace{x-x_1}_{z_1})\cdots\tilde{p}(\Delta t_k,\underbrace{x_{k-1}-x_k}_{z_k})dx_1\cdots dx_k\\
&= \mathrm{e}^{iU_1x}\int \exp(iU_1z_1+\cdots+iU_kz_k)\tilde{p}(\Delta t_1,z_1)\cdots\tilde{p}(\Delta t_{k-1},z_{k-1})\tilde{p}(\Delta t_k,z_k)dz_1\cdots dz_k\\
&= \mathrm{e}^{iU_1x}\prod_{\tilde{m}=1}^{k}\widehat{\tilde{p}}(\Delta t_{\tilde{m}},U_{\tilde{m}}) = e^{i(\sum_{j=1}^k u_j)\cdot x-\frac{1}{2}\sum_{\tilde{m}=1}^k(\Delta t_{\tilde{m}})U_{\tilde{m}}^2},
\end{aligned} \quad (89)$$

where $\widehat{\tilde{p}}$ is the Fourier transform of $\tilde{p}$, $U_{\tilde{m}}:=\sum_{j=\tilde{m}}^k u_j$ 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:

$$\begin{aligned}
\sum_{\tilde{m}=1}^k \Delta t_{\tilde{m}}U_{\tilde{m}}^2 &= \sum_{\tilde{m}=1}^k \Delta t_{\tilde{m}}(\sum_{j=\tilde{m}}^k u_j)^2\\
&= \sum_{\tilde{m}=1}^k \Delta t_{\tilde{m}}\sum_{j,j'=\tilde{m}}^k u_j\cdot u_{j'} = \sum_{\tilde{m}=1}^k \Delta t_{\tilde{m}}\sum_{j,j'=1}^k \chi(j\ge\tilde{m})\chi(j'\ge\tilde{m})u_j\cdot u_{j'}\\
&= \sum_{j,j'=1}^k u_j\cdot u_{j'}\sum_{\tilde{m}=1}^{\min(j,j')}\Delta t_{\tilde{m}} = \sum_{j,j'=1}^k u_j\cdot u_{j'}(t_1-t_0+t_2-t_1+\cdots+t_{\min(j,j')})\\
&= \sum_{j,j'=1}^k u_j\cdot u_{j'}t_{\min(j,j')} = \sum_{j,j'=1}^k u_j\cdot u_{j'}\min(t_j,t_{j'}),
\end{aligned} \quad (90)$$

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.,*

$$B_{t_1}, B_{t_2}-B_{t_1},\dots,B_{t_k}-B_{t_{k-1}}, \quad (91)$$

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,

$$\begin{aligned} E[(B_{t_2} - B_{t_1})^i (B_{t_3} - B_{t_2})^i] &= E[(\tilde{B}_{t_2} - \tilde{B}_{t_1})^i (\tilde{B}_{t_3} - \tilde{B}_{t_2})^i] \\ &= E[\tilde{B}^i_{t_2} \cdot \tilde{B}^i_{t_3}] - E[\tilde{B}^i_{t_2} \cdot \tilde{B}^i_{t_2}] - E[\tilde{B}^i_{t_1} \cdot \tilde{B}^i_{t_3}] + E[\tilde{B}^i_{t_1} \cdot \tilde{B}^i_{t_2}] \\ &= (t_2 - t_2 - t_1 + t_1) = 0 \end{aligned} \tag{92}$$

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

$$E(\{(B_t - B_s)^i\}^2) = E((B^i_t)^2 - 2B^i_t B^i_s + (B^i_s)^2) = t + s - 2\min(t, s) = |t - s|, \tag{93}$$

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.*

$$E[|X_t - X_s|^\alpha] \leq D|t - s|^{1+\beta}, \quad 0 \leq s, t \leq T. \tag{94}$$

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*

$$E^x[|B_t - B_s|^4] = E^x[\{(B_t - B_s) \cdot (B_t - B_s)\}^2] \leq D|t - s|^2. \tag{95}$$

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. (38). We have

$$E^x[|B_t - B_s|^4] = E^x[|X|^4] = |t - s|^2 E^x[\{\sum_{i=1}^d (Z^i)^2\}^2]. \tag{96}$$

The last factor is finite, since

$$E^x[\{\sum_{i=1}^d (Z^i)^2\}^2] = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} dz \{\sum_{i=1}^d (z^i)^2\}^2 \mathrm{e}^{-\frac{1}{2}|z|^2}. \tag{97}$$

This concludes the proof. (Actually, (97) $= d(d + 2)$ 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 $0 \leq s < t$, $B_t - B_s \sim N(0, (t - s)I_d)$, 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

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

where

$$\prod_{\ell=1}^{k} p(t_\ell - t_{\ell-1}, x_{\ell-1}, x_\ell) = \Big[\prod_{\ell'=1}^{k} \frac{1}{(2\pi(t_{\ell'}-t_{\ell'-1}))^{d/2}}\Big] \exp\Big(-\sum_{\ell=1}^{k} \frac{|x_{\ell-1}-x_\ell|^2}{2(t_\ell-t_{\ell-1})^2}(t_\ell-t_{\ell-1})\Big). \tag{98}$$

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

$$f(B_{t_1}(\omega),\dots,B_{t_k}(\omega)) = f(\omega(t_1),\dots,\omega(t_k)) \to f(\{\omega(t)\}_{t\in[0,T]}) = f(\omega). \tag{99}$$

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

$$\int f(B_{t_1}(\omega),\dots,B_{t_k}(\omega))dP^x(\omega) \to \int_{\omega(0)=x} f(\omega)\frac{1}{\mathcal{N}}\exp\Big(-\int_0^T \frac{1}{2}\Big(\frac{d\omega(t)}{dt}\Big)^2 dt\Big)D\omega, \tag{100}$$

  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 $\frac{1}{\mathcal{N}}\exp\Big(-\int_0^T\frac{1}{2}\big(\frac{d\omega(t)}{dt}\big)^2 dt\Big)D\omega$ makes sense as a Gaussian measure on the Banach space $C([0,T];\mathbb{R}^d)$.
- Since $\int dP^x(\omega) = 1$,

$$\mathcal{N} = \int \exp\Big(-\int_0^T \frac{1}{2}\Big(\frac{d\omega(t)}{dt}\Big)^2 dt\Big)D\omega. \tag{101}$$

  Thus (107) 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 x. Then, the quantity

$$K_E(T;y,x) := \int_{\omega(0)=x}^{\omega(T)=y} \frac{1}{\mathcal{N}}\exp\Big(-\int_0^T \frac{1}{2}\Big(\frac{d\omega(t)}{dt}\Big)^2 dt\Big)D\omega \tag{102}$$

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

$$(\partial_T - \frac{1}{2}\Delta_x)K_E(T;y,x) = \delta(T)\delta(x-y). \tag{103}$$

- 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

$$K(\mathcal{T}; y, x) := \int_{q(0)=x}^{q(\mathcal{T})=y} \frac{1}{\mathcal{N}'} \exp\left(i \int_0^{\mathcal{T}} \frac{1}{2}\left(\frac{dq(\tau)}{d\tau}\right)^2 d\tau\right) Dq. \tag{104}$$

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

$$(i\partial_{\mathcal{T}} + \frac{1}{2}\Delta_x)K(\mathcal{T}; y, x) = \delta(\mathcal{T})\delta(x - y). \tag{105}$$

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

$$\psi_{\mathcal{T}}(y) := \int K(\mathcal{T}; y, x)\psi_0(x)dx. \tag{106}$$

  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

$$K_V(T; y, x) := \int_{q(0)=x}^{q(\mathcal{T})=y} \frac{1}{\mathcal{N}'} \exp\big(iS[q]\big)\big) Dq, \tag{107}$$

  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

$$\frac{d^2q(\tau)}{d\tau^2} = -(\nabla V)(q(\tau)). \tag{108}$$

  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 (107), 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 (9)

$$\frac{dX_s}{ds} = \left(r + \alpha\frac{dB_s}{ds}\right)X_s, \quad r, \alpha > 0. \tag{109}$$

As we suspect from (6) that $s \mapsto B_s(\omega)$ is not differentiable in the usual sense, let us integrate both sides w.r.t $s \in [0,t]$:

$$X_t = X_0 + \int_0^t rX_s ds + \text{"} \int_0^t \alpha X_s dB_s \text{"}. \tag{110}$$

This is a special case of a general class of equations

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \text{"} \int_0^t \sigma(s, X_s) dB_s \text{"}. \tag{111}$$

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

$$\int_a^b f(s)dg(s) \tag{112}$$

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

$$V_a^b(g, \Pi) := \sum_{i=1}^n |\Delta g_i| \in [0, \infty). \tag{113}$$

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

$$V_a^b(g) := \sup_\Pi V_a^b(g, \Pi) \in [0, \infty]. \tag{114}$$

   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

$$\sum_{i=1}^n |\Delta g_i| = \sum_{i=1}^n (\Delta g_i) = g(b) - g(a). \tag{115}$$

   - 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. (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.

$$g(s) = \begin{cases} s \sin\left(\frac{1}{s}\right), & s \neq 0, \\ 0, & s = 0. \end{cases} \tag{116}$$

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

$$\int_a^b f(s) dg(s) = \int_a^b f(s) g'(s) ds, \tag{117}$$

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

$$\xi_i \in [s_{i-1}, s_i], \qquad i = 1, \dots, n. \tag{118}$$

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 $\{a = \hat{s}_0 < \hat{s}_1 < \cdots < \hat{s}_n = b\}$ (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

$$\begin{aligned} f(\xi_i)(g(s_i) - g(s_{i-1})) &= f(\xi_i)(g(s_i) - g(s'_{i'}) + g(s'_{i'}) - g(s_{i-1}) \\ &= f(\xi(\hat{I}_1))\Delta g(\hat{I}_1) + f(\xi(\hat{I}_2))\Delta g(\hat{I}_2), \end{aligned} \tag{119}$$

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

$$\sum_{i=1}^{n} f(\xi_i)(g(s_i) - g(s_{i-1})) = \sum_{\hat{I} \in \hat{\Pi}} f(\xi(\hat{I}))\Delta g(\hat{I}), \tag{120}$$

$$\sum_{i'=1}^{n'} f(\xi'_{i'})(g(s_{i'}) - g(s_{i'-1})) = \sum_{\hat{I} \in \hat{\Pi}} f(\xi'(\hat{I}))\Delta g(\hat{I}). \tag{121}$$

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:*

$$\int_a^b f(s) dg(s) = \lim_{|\Pi| \to 0} \sum_{i=1}^{n} f(\xi_i)(g(s_i) - g(s_{i-1})). \tag{122}$$

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

$$S(f, g; \Pi, \xi) := \sum_{i=1}^{n} f(\xi_i)\, \Delta g_i, \qquad |\Pi| := \max_{1 \leq i \leq n} (s_i - s_{i-1}). \tag{123}$$

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 (120), (121),

$$S(f, g; \Pi, \xi) - S(f, g; \Pi', \xi') = \sum_{\hat{I} \in \widehat{\Pi}} \big(f(\xi(\hat{I})) - f(\xi'(\hat{I}))\big)\, \Delta g(\hat{I}), \tag{124}$$

hence

$$\big|S(f, g; \Pi, \xi) - S(f, g; \Pi', \xi')\big| \le w_f(\delta) \sum_{\hat{I} \in \widehat{\Pi}} |\Delta g(\hat{I})| \le w_f(\delta)\, V_a^b(g). \tag{125}$$

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

$$\int_a^b f\, dg := \lim_{|\Pi| \to 0} S(f, g; \Pi, \xi), \tag{126}$$

which is independent of the choice of tags. (To see this last point choose in (125) $\Pi = \Pi'$ 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*

$$V_0^T(B) := \sup_{\Pi} \sum_i |B_{s_i} - B_{s_{i-1}}| = \infty \quad P\text{-almost surely.} \tag{127}$$

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,*

$$\lim_{|\Pi| \to 0} \sum_{i=1}^{n} \big(g(s_i) - g(s_{i-1})\big)^2 = 0. \tag{128}$$

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

$$\sum_{i=1}^{n} \big(g(s_i) - g(s_{i-1})\big)^2 \le \Big(\max_{1 \le i \le n} |g(s_i) - g(s_{i-1})|\Big) \sum_{i=1}^{n} |g(s_i) - g(s_{i-1})|. \tag{129}$$

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, (128) holds. □

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*

$$E[X_1^2] < \infty. \tag{130}$$

Define the average

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k. \tag{131}$$

Then

$$\bar{X}_n \longrightarrow E[X_1] \quad P\text{-almost surely.} \tag{132}$$

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 $\Pi_n = \{kT/2^n : k = 0, \ldots, 2^n\}$ and define

$$S_n := \sum_{k=1}^{2^n} \big(B_{kT/2^n} - B_{(k-1)T/2^n}\big)^2. \tag{133}$$

The increments $\big(B_{kT/2^n} - B_{(k-1)T/2^n}\big)$ are independent and distributed as $N(0, T/2^n)$ by Theorem 2.11. Write $Z_k := \sqrt{\frac{2^n}{T}}\big(B_{kT/2^n} - B_{(k-1)T/2^n}\big) \sim N(0,1)$, cf. (38). Then

$$\frac{2^n}{T} S_n = \sum_{k=1}^{2^n} Z_k^2 \quad \Rightarrow \quad S_n = T\left[\frac{1}{2^n}\sum_{k=1}^{2^n} Z_k^2\right]. \tag{134}$$

By the SLLN (Theorem 3.5),

$$\frac{1}{2^n}\sum_{k=1}^{2^n} Z_k^2 \to E[Z_1^2] = 1 \quad P\text{-almost surely,} \tag{135}$$

and therefore

$$S_n \to T \neq 0. \quad P\text{-almost surely.} \tag{136}$$

By comparing this with (128), 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

$$\lim_{n\to\infty}\sum_{k=1}^{2^n} \big(B_{kT/2^n}(\omega) - B_{(k-1)T/2^n}(\omega)\big)^2 = 0. \tag{137}$$

But by (136) this limit equals $T > 0$ almost surely, a contradiction. Hence $P(A) = 0$. It follows that

$$P\big(V_0^T(B) = \infty\big) = 1, \tag{138}$$

which proves (127). □

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:

$$\varphi(t,\omega) = \sum_{j\ge 0} e_j(\omega)\chi_{[j2^{-n},(j+1)2^{-n}]}(t), \tag{139}$$

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

3. For such functions we define

$$\int_S^T \varphi(t,\omega)dB_t(\omega) = \sum_{j\ge 0} e_j(\omega)[B_{t_{j+1}} - B_{t_j}](\omega), \tag{140}$$

   where

$$t_j := t_j^{(n)} := \begin{cases} j2^{-n} & \text{if} \quad S \le j2^{-n} \le T, \\ S & \text{if} \quad j2^{-n} < S, \\ T & \text{if} \quad j2^{-n} > T. \end{cases} \tag{141}$$

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

$$\varphi_1(t,\omega) = \sum_{j\ge 0} B_{t_j}(\omega)\chi_{[t_j,t_{j+1}]}(t), \tag{142}$$

$$\varphi_2(t,\omega) = \sum_{j\ge 0} B_{t_{j+1}}(\omega)\chi_{[t_j,t_{j+1}]}(t). \tag{143}$$

   Then

$$E\Big[\int_0^T \varphi_1(t,\omega)dB_t(\omega)\Big] = \sum_{j>0} E[B_{t_j}(B_{t_{j+1}} - B_{t_j})] = 0 \tag{144}$$

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

$$\begin{aligned} E\Big[\int_0^T \varphi_2(t,\omega)dB_t(\omega)\Big] &= \sum_{j>0} E[B_{t_{j+1}}(B_{t_{j+1}} - B_{t_j})] \\ &= \sum_{j>0} E[B_{t_j}(B_{t_{j+1}} - B_{t_j})] \\ &+ \sum_{j>0} E[(B_{t_{j+1}} - B_{t_j})^2] = T, \end{aligned} \tag{145}$$

   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:

$$f(t,\omega) \simeq \sum_j f(\xi_j,\omega)\chi_{[t_j,t_{j+1})}(t), \tag{146}$$

   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

$$\int_S^T f(t,\omega) dB_t(\omega). \tag{147}$$

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

$$\int_S^T f(t,\omega) \circ dB_t(\omega). \tag{148}$$

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

$$\frac{dX_t}{dt} = \left(r + \alpha \frac{dB_t}{dt} \right) X_t, \quad r, \alpha > 0 \tag{149}$$

can be given meaning in two different ways with solutions

$$X_t = X_0 \mathrm{e}^{(r - \frac{1}{2}\alpha^2)t + \alpha B_t}, \quad X_t = X_0 \mathrm{e}^{rt + \alpha B_t}. \tag{150}$$

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 $B_s(\omega)$, $0 \le s \le t$.

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
   $$X^{-1}(A) := \{\, \omega \in \Omega \,:\, X(\omega) \in A \,\}, \quad A \in \mathcal{B}(\mathbb{R}). \tag{151}$$
   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,
   $$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}). \tag{152}$$
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
   $$\bigcup_{t \in [0,T]} \sigma(X_t). \tag{153}$$
   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 $s \le t$.

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

$$\mathcal{G}_t^{\text{comp}} := \sigma(\mathcal{G}_t \cup \mathcal{N}). \tag{154}$$

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

$$g_1(B_{t_1}) \cdots g_k(B_{t_k}), \tag{155}$$

   where $g_1, \dots g_k$ are bounded continuous and $t_j \le t$ for $j \le k$. Can be proven using the Doob-Dynkin lemma. Intuitively, values of $f(\omega)$ can be computed from values of $B_s(\omega)$ for $s \le t$. 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 $B_s - B_t$, $s \ge t$. (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 (133) 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*

$$(t,\omega) \mapsto f(t,\omega) : \mathbb{R}_+ \times \Omega \to \mathbb{R} \tag{156}$$

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
$$\varphi(t,\omega) = \sum_{j\geq 0} e_j(\omega)\chi_{[t_j,t_{j+1})}(t). \tag{157}$$
In particular, since φ is adapted, e_j must be $\mathcal{F}_{t_j}$-measurable. For elementary functions, we set
$$\int_S^T \varphi(t,\omega)dB_t(\omega) = \sum_{j\geq 0} e_j(\omega)[B_{t_{j+1}} - B_{t_j}](\omega). \tag{158}$$

2. Def. Let $f \in \mathcal{V}(S,T)$. Then the Itô integral of f is defined by
$$\int_S^T f(t,\omega)dB_t(\omega) = \lim_{n\to\infty} \int_S^T \varphi_n(t,\omega)dB_t(\omega), \quad \text{(limit in } L^2(\Omega,P)\text{)}, \tag{159}$$
where $\{\varphi_n\}_{n\in\mathbb{N}}$ is a sequence of elementary functions s.t.
$$E\bigg[\int_S^T |f(t,\cdot\,) - \varphi_n(t,\cdot\,)|^2 dt\bigg] \to 0 \quad \text{as } n \to \infty. \tag{160}$$
Theorem 3.10. *(Approximation theorem). For any $f \in \mathcal{V}(S,T)$, a sequence $\{\varphi_n\}_{n\in\mathbb{N}}$ satisfying (159), (160) 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:
$$\varphi_1(t,\omega) = \sum_{j\geq 0} B_{t_j}(\omega)\chi_{[t_j,t_{j+1})}(t), \tag{161}$$
$$\varphi_2(t,\omega) = \sum_{j\geq 0} B_{t_{j+1}}(\omega)\chi_{[t_j,t_{j+1})}(t). \tag{162}$$
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*

$$E\Big[\Big(\int_S^T \varphi(t,\,\cdot\,)dB_t(\,\cdot\,)\Big)^2\Big] = E\Big[\int_S^T \varphi(t,\,\cdot\,)^2 dt\Big]. \tag{163}$$

Remark 3.12. *Suppose the limit (160) 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 (163), 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 (159) 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

$$\begin{aligned} P(g(X) \in F_1, Y \in F_2) &= P(X \in g^{-1}(F_1), Y \in F_2) \\ &= P(X \in g^{-1}(F_1))P(Y \in F_2) = P(g(X) \in F_1)P(Y \in F_2). \end{aligned} \tag{164}$$

This concludes the proof. □

Lemma 3.15. *$e_i e_j \Delta B_i$ 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 $e_i e_j \Delta B_i$ 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 $e_i e_j \Delta B_{t_i}$ 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,

$$E[e_i e_j \Delta B_i \Delta B_j] = \begin{cases} 0 & \text{if} \quad i \ne j, \\ E[e_j^2](t_{j+1} - t_j) & \text{if} \quad i = j. \end{cases} \tag{165}$$

Here in the first line we used Lemma 3.15, the fact that independent random variables are uncorrelated: $E[e_i e_j \Delta B_i \Delta B_j] = E(e_i e_j \Delta B_i)E(\Delta B_j)$ 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 (165),

$$E\Big[\Big(\int_S^T \varphi dB\Big)^2\Big] = \sum_{i,j} E[e_i e_j \Delta B_i \Delta B_j] = \sum_j E[e_j^2](t_{j+1} - t_j). \tag{166}$$

On the other hand

$$
\begin{aligned}
E\bigg[\int_S^T \varphi(t,\,\cdot\,)^2 dt\bigg] &= E\bigg[\int_S^T |e_j|^2 \chi_{[t_j,t_{j+1})}(t)dt\bigg] \\
&= E\bigg[\sum_{j\geq 0} |e_j|^2 (t_{j+1}-t_j)\bigg] = \sum_{j\geq 0} E(|e_j|^2)(t_{j+1}-t_j), \qquad (167)
\end{aligned}
$$

which completes the proof. □

3.8 Itô integral. Proof of Theorem 3.10

We will approximate as follows:

$$\{f \in \mathcal{V}(S,T)\} \approx \{h \in \mathcal{V}(S,T), \text{bd.}\} \approx \{g \in \mathcal{V}(S,T), \text{bd., cont. in } t \text{ for any fixed } \omega\} \approx \{\varphi \text{ elementary.}\} \qquad (168)$$

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*

$$E\bigg[\int_S^T (f-h_n)^2 dt\bigg] \underset{n\to\infty}{\to} 0. \qquad (169)$$

Proof. Put

$$h_n(t,\omega) = \begin{cases} -n & \text{if} \quad f(t,\omega) < -n \\ f(t,\omega) & \text{if} \quad -n \leq f(t,\omega) \leq n \\ n & \text{if} \quad f(t,\omega) > n. \end{cases} \qquad (170)$$

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

$$E\bigg[\int_S^T f^2 dt\bigg] < \infty. \qquad (171)$$

By definition (170), $|h_n(t,\omega)| \leq |f(t,\omega)|$. 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*

$$E\bigg[\int_S^T (h-g_n)^2 dt\bigg] \underset{n\to\infty}{\to} 0. \qquad (172)$$

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

$$g_n(t,\omega) := \int_0^t \psi_n(s-t)h(s,\omega)ds \overset{(i)}{=} \int_0^\infty \psi_n(s-t)h(s,\omega)ds \qquad (173)$$

have the following properties:

(a) g_n are bounded uniformly in n:

$$|g_n(t,\omega)| \le \int_0^\infty \psi_n(s-t)|h(s,\omega)|ds \le \sup_{s\in\mathbb{R}_+} |h(s,\omega)| \le M. \tag{174}$$

(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

$$\lim_{\ell\to\infty} g_n(t_\ell,\omega) := \int_0^\infty \lim_{\ell\to\infty} \psi_n(s-t_\ell)h(s,\omega)ds = g_n(t,\omega). \tag{175}$$

We also made use of the fact that, since t_ℓ is a bounded sequence, we can replace the upper boundary of integration in (173) 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 $s \le t$ 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)

$$\int_S^T (g_n(t,\omega)-h(t,\omega))^2 dt \underset{n\to\infty}{\to} 0 \text{ for each } \omega. \tag{176}$$

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

$$E\bigg[\int_S^T dt\,(h-g_n)^2 dt\bigg] \underset{n\to\infty}{\to} 0. \tag{177}$$

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.*

$$E\bigg[\int_S^T (g-\varphi_n)^2 dt\bigg] \underset{n\to\infty}{\to} 0. \tag{178}$$

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 $t_j \le t < t_{j+1}$ 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,

$$\int_S^T (g-\varphi_n)^2 dt \underset{n\to\infty}{\to} 0 \quad \text{for each } \omega \tag{179}$$

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 $S \le t \le T$, hence

$$\begin{aligned} \int |g(t,\omega)-\varphi_n(t,\omega)|^2 dt &= \int_S^T \sum_j [g(t,\omega)-g(t_j,\omega)]^2 \chi_{[t_j,t_{j+1})}(t)dt \\ &\le w_{g(\cdot,\omega)}(2^{-n})^2 \int_S^T \sum_j \chi_{[t_j,t_{j+1})}(t)dt = \omega_{g(\cdot,\omega)}(2^{-n})^2(T-S), \end{aligned} \tag{180}$$

where the modulus of continuity

$$w_{g(\cdot,\omega)}(\delta) := \sup\{|g(t_1,\omega)-g(t_2,\omega)| : |t_1-t_2| \le \delta,\ t_1,t_2 \in [S,T]\} \tag{181}$$

satisfies $\lim_{n\to\infty} w_{g(\cdot,\omega)}(2^{-n}) = 0$ for any fixed ω. This gives (179). 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*

$$\int_0^T B_t dB_t = \frac{1}{2}B_T^2 - \frac{1}{2}T. \tag{182}$$

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

$$\begin{aligned} E\Big[\int_0^T (\varphi_n - B_t)^2 dt\Big] &= E\Big[\sum_j \int_{t_j}^{t_{j+1}} (B_j - B_t)^2 dt\Big] \\ &= \sum_j \int_{t_j}^{t_{j+1}} (t - t_j)dt = \frac{1}{2}\sum_j (t_{j+1} - t_j)^2. \end{aligned} \tag{183}$$

Recall from (141) that $t_j = j2^{-n} \leq T$. Then

$$(183) \leq \frac{1}{2}\sum_{j2^{-n}\leq T} (t_{j+1} - t_j)^2 = \frac{1}{2}\sum_{j \leq T2^n} 2^{-2n} = \frac{1}{2}2^{-2n}(T2^n) \underset{n\to\infty}{\to} 0. \tag{184}$$

Thus we checked (160). Consequently, by (159),

$$\int_0^T B_t dB_t = \lim_{n\to\infty} \int_0^T \varphi_n dB_t = \lim_{n\to\infty} \sum_j B_j \Delta B_j \tag{185}$$

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

$$\begin{aligned} \Delta(B_j^2) &:= B_{j+1}^2 - B_j^2 = (B_{j+1} - B_j)^2 + 2B_j(B_{j+1} - B_j) \\ &= (\Delta B_j)^2 + 2B_j \Delta B_j. \end{aligned} \tag{186}$$

Since $B_0 = 0$,

$$B_T^2 = \sum_j \Delta(B_j^2) = \sum_j (\Delta B_j)^2 + 2\sum_j B_j \Delta B_j. \tag{187}$$

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. $\square$

References

[Ok] B. Øksendal. *Stochastic Differential Equations. An Introduction with Applications.* Springer-Verlag, 2000.

[Ro] G. Roepstorff. *Path Integral Approach to Quantum Physics. An Introduction.* Springer-Verlag, 1991.

[BT] D.P. Bertsekas and J.N. Tsitsiklis. *Introduction to Probability.* Lecture Notes, M.I.T. Fall, 2000.

[Ru] W. Rudin. *Real and complex analysis.* McGraw-Hill Inc. 1970

[Pa] K. R. Parthasarathy. *Probability measures on metric spaces.* Academic Press, 1967.