

# CONTENTS

1	Real analysis	3
1.1	Convergence of sequences and series	4
1.2	Differentiation	5
1.2.1	Differentiation of functions with real powers	5
1.2.2	Example of $f \in C^\infty \setminus C^\omega$	5
1.2.3	Taylor series for multivariate functionals	5
2	Probability theory	7
2.1	Elementary ideas	8
2.1.1	$\sigma$ -algebras	8
2.1.2	Examples	8
2.1.3	Measurability of $\inf, \sup, \liminf, \limsup$ TODO	8
2.1.4	Method of substitution	8
2.2	Borel–Cantelli lemmas	9
2.3	Modes of convergence	10
2.4	Conditioning	11
2.4.1	In $L^2$ , conditional expectation is a projection	11
2.4.2	Uncorrelated does not imply independence	11
2.4.3	$\phi_{aX+bY} = \phi_{aX}\phi_{bY} \forall (a, b) \in \mathbb{R}^2$ implies independence	11
2.5	Limiting behaviour of $\bar{X}_n$	12
3	Stochastic analysis	13
3.1	Classification of stochastic processes	14
3.2	Martingales	15
3.2.1	New martingales from old	15
3.3	Markov processes	16
3.4	Itô calculus	17
3.4.1	Step 1: $X \in \mathcal{A} \cap \mathcal{S}$ a.s.	17
3.4.2	Step 2: $X \in \mathcal{A} \cap \mathcal{B} \cap \mathcal{C}$ a.s.	17
3.4.3	Step 3: $X \in \mathcal{A} \cap \mathcal{B} \cap L^0(T \times \Omega)$	17
3.4.4	Step 4: $X \in \mathcal{A} \cap L^2(T \times \Omega)$	17
3.4.5	Step 5: $X \in \mathcal{A} \cap \left\{ X \in \mathbb{C}^{T \times \Omega} : \forall t \geq 0, \int_0^t X(s, \cdot) ds < \infty \right\}$ a.s.	17
3.4.6	Properties of the Itô integral	17
3.4.7	Itô formula for multidimensional processes	18

	2	
3.5	Examples	20
3.5.1	Find $\mathbb{E} \left( \int_0^1 B_t^2 \, dt \right)^2$ .	20
3.5.2	Find $\mathbb{V} \left( \int_0^1 t^2 B_t \, dt \right)$ .	20
4	Infinite-dimensional analysis	21
4.1	Abstract Wiener spaces	22
4.2	White noise distribution theory	23
4.2.1	Characterization theorem	23
	Bibliography	24



## 1.1 CONVERGENCE OF SEQUENCES AND SERIES

- We can only talk of *convergence of sequences* in Hausdorff topological spaces.
- We can only talk of *series* in commutative groups, because we need  $+$  to be defined.
- We can only talk of *convergence of series* in commutative Hausdorff topological groups.
- We can only talk of *absolute convergence of series* in normed commutative Hausdorff topological groups.
- This is from [Wikipedia](#). Let  $S$  be the vector space of sequences. Then the partial summation  $\Sigma : S \rightarrow S, (a_n) \mapsto \left(\sum_{j=1}^n a_j\right)$  is a linear operator on  $S$ , whose inverse is the finite difference operator,  $\Delta$ . These behave as discrete analogs of integration and differentiation, only for series (functions of a natural number) instead of functions of a real variable. For example, the sequence  $(1, 1, 1, \dots)$  has series  $(1, 2, 3, \dots)$  as its partial summation, which is analogous to the fact that  $\int_0^x 1 \, dt = x$ .
- Classification of convergence of series
  1. Pointwise or uniform convergence
  2. Absolute, unconditional and conditional convergence
    - *Absolute convergence* means  $\sum \|a_n\| < \infty$ .
    - *Unconditional convergence* means all rearrangements of the series are convergent to the same value. That is, if  $\sigma : \mathbb{N} \rightarrow \mathbb{N}$  is a permutation, then  $\sum_n a_n = \sum_n a_{\sigma(n)}$ .  
In complete spaces, absolute convergence  $\implies$  unconditional convergence, but the converse is not true in general. In finite dimensional spaces, the converse is true by Riemann rearrangement theorem. But the Dvoretzky–Rogers theorem asserts that every infinite-dimensional Banach space admits an unconditionally convergent series that is not absolutely convergent. (see this [Wikipedia article](#))
    - *Conditional convergence* means convergent but not absolutely convergent.
  3. Depending on the space of values, for example, real number, arithmetic progression, trigonometric function, etc.

## 1.2 DIFFERENTIATION

### 1.2.1 Differentiation of functions with real powers

This idea is by Prof Sundar.

For each  $n \in \mathbb{N}$ , we can differentiate  $f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto x^n$  using the limit definition of the derivative, by using the factorization  $(x+h)^n - x^n = (x+h-x) \sum_{j=0}^{n-1} h^j x^{n-1-j}$ , which gives

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \sum_{j=0}^{n-1} (x+h)^j x^{n-1-j} = \sum_{j=0}^{n-1} x^j x^{n-1-j} = (n-1)x^{n-1}.$$

But this does not work for exponents  $r \in \mathbb{R}$  in general. How can we do it?

One needs to think outside the box for this. We cannot go by definition here. We note that  $x^r = e^{r \log x}$ . Now use chain rule.

### 1.2.2 Example of $f \in C^\infty \setminus C^\omega$

How does one construct an example of a function which is smooth but not analytic? The idea is to find  $f \neq 0$  such that  $f^{(n)}(0) = 0 \forall n$ .

Note that the graph of  $x \mapsto x^n$  around  $x = 0$  become flatter and flatter as  $n \rightarrow \infty$ . Consider  $f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto e^{-\frac{1}{x}} \mathbb{1}_{x>0}(x)$ . Then  $f'(x) = \frac{1}{x^2} e^{-\frac{1}{x}} \mathbb{1}_{x>0}(x)$  (do the computations separately for  $x > 0$  and  $x = 0$ ). In this way,  $f'(0) = 0$ . Continuing, we see that  $f^{(n)}(0) = 0 \forall n$ . Therefore, the sum  $\sum_n \frac{f^{(n)}(0)}{n!} x^n$  converges to 0, which is not the same as  $f$ .

See also [this Wikipedia article](#).

### 1.2.3 Taylor series for multivariate functionals

**Theorem (Taylor)** Let  $\Omega \subseteq \mathbb{R}^d$  be an open set and  $f : \Omega \rightarrow \mathbb{R}$  be an infinitely differentiable function at the point  $p \in \Omega$ . Then the Taylor series of  $f$  around  $p$  for  $v \in T_p \Omega$  is given by

$$T(p+v) = \sum_{n=0}^{\infty} \frac{1}{n!} (\langle v, D_v \rangle^n f)(p), \quad \text{where } \langle v, D_v \rangle = \sum_{j=1}^d v_j \frac{\partial}{\partial v_j}.$$

*Proof* Is this a  $\mathbb{R}$  life, or is this just imaginary, caught in a landslide, no escape from  $\mathbb{C}$ . Open your eyes, look up to the sky and see. I'm trivial come, trivial go, little high, little low.

ToDo, really! □

In the differential form, this becomes

$$dT(p) = \sum_{n=1}^{\infty} \frac{1}{n!} (\langle dv, D_v \rangle^n f)(p)$$

There is another form with multi-indexes. But the above form seems way more natural to me. See more in [this](#) and [this](#) Wikipedia articles.



## 2.1 ELEMENTARY IDEAS

### 2.1.1 $\sigma$ -algebras

- $\mathcal{I} \subset \mathcal{B} \subset \overline{\mathcal{B}} = \mathcal{L} \subset 2^{\mathbb{R}}$
- $|\mathcal{I}| = |\mathcal{B}| = |\mathbb{R}| = \aleph_1, |\overline{\mathcal{B}}| = |2^{\mathbb{R}}| = \aleph_2$

### 2.1.2 Examples

- $\left(x \mapsto \frac{1}{\sqrt{x}} \mathbb{1}_{[0,1]}(x)\right) \in L^1 \setminus L^2$
- $\left(x \mapsto \frac{1}{x} \mathbb{1}_{[1,\infty)}(x)\right) \in L^2 \setminus L^1$

### 2.1.3 Measurability of $\inf, \sup, \liminf, \limsup$ TODO

Let  $X_n$  be a discrete-time stochastic process. Then  $\{\inf X_t \geq c\} = \bigcap_t \{X_t \geq c\}$

### 2.1.4 Method of substitution

See Folland - Real Analysis Theorem (2.43).



## 2.2 BOREL–CANTELLI LEMMAS

**Theorem (Borel–Cantelli 1)** *Let  $(E_n) \subset \mathcal{F}$  such that  $\sum \mathbb{P}(E_n) < \infty$ . Then  $\mathbb{P}(E_n \text{ i.o.}) = 0$ .*

*Proof* Since  $\sum \mathbb{P}(E_n) < \infty$ , for any fixed  $n \in \mathbb{N}$ , we have

$$\mathbb{P}(E_n \text{ i.o.}) = \mathbb{P}\left(\bigcap_n \bigcup_{m \geq n} E_m\right) \leq \mathbb{P}\left(\bigcup_{m \geq n} E_m\right) \leq \sum_{m \geq n} \mathbb{P}(E_m) \rightarrow 0.$$

□

Counterexample of the converse of BC1: Take  $((0, 1], \lambda)$  as the probability space, and  $E_n = (0, \frac{1}{n^2})$ . Then  $\mathbb{P}(E_n \text{ i.o.}) = 1$ , but  $\sum \mathbb{P}(E_n) < \infty$ .

**Theorem (Borel–Cantelli 2)** *Let  $(E_n) \subset \mathcal{F}$  be (mutually) independent such that  $\sum \mathbb{P}(E_n) = \infty$ . Then  $\mathbb{P}(E_n \text{ i.o.}) = 1$ .*

*Proof* Since  $\mathbb{P}((E_n \text{ i.o.})^c) = \mathbb{P}(E_n^c \text{ ev})$ , it is equivalent to prove that  $\mathbb{P}(E_n^c \text{ ev}) = 0$ . Using independence and the fact that  $1 - x < e^{-x}$ , for each fixed  $n \in \mathbb{N}$ , we have

$$\mathbb{P}\left(\bigcap_{m=n}^N E_m^c\right) = \prod_{m=n}^N \mathbb{P}(E_m^c) = \prod_{m=n}^N (1 - \mathbb{P}(E_m)) \leq \prod_{m=n}^N e^{-\mathbb{P}(E_m)} = e^{-\sum_{m=n}^N \mathbb{P}(E_m)}.$$

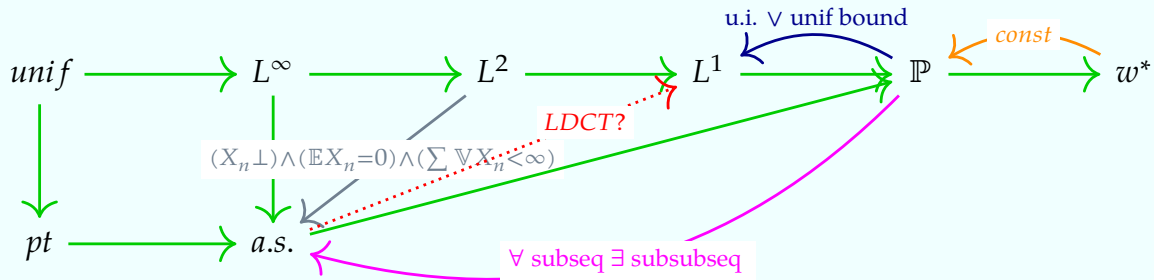
Taking  $N \rightarrow \infty$ , we get  $\mathbb{P}(\bigcap_{m \geq n} E_m^c) \rightarrow 0$ . Therefore

$$\mathbb{P}(E_n^c \text{ ev}) = \mathbb{P}\left(\bigcup_n \bigcap_{m \geq n} E_m^c\right) \leq \sum_n \mathbb{P}\left(\bigcap_{m=n}^N E_m^c\right) = 0.$$

□

## 2.3 MODES OF CONVERGENCE

Study this part from [Robert L Wolpert - Convergence in  \$\mathbb{R}^d\$  and in metric spaces](#). In this diagram, the top row represents 'point independent' modes of convergence and the bottom row represents the 'point dependent' modes of convergence.



Legend

- green: automatic implication
- any other color: depends on the mentioned condition

## 2.4 CONDITIONING

### 2.4.1 In $L^2$ , conditional expectation is a projection

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space,  $\mathcal{G} \subseteq \mathcal{F}$  be a sub- $\sigma$ -algebra, and  $X \in L^2(\mathcal{F})$  be a random variable. Let  $\pi_{L^2(\mathcal{G})}$  denote the projection operator onto  $L^2(\mathcal{G})$ . We know that projection operators are self-adjoint. So  $\forall E \in \mathcal{G}$ ,

$$\begin{aligned} \int_E \mathbb{E}(X | \mathcal{G}) d\mathbb{P} &= \int_E X d\mathbb{P} = \int \mathbb{1}_E X d\mathbb{P} = \langle \mathbb{1}_E, X \rangle && \text{[definitions]} \\ &= \langle \pi_{L^2(\mathcal{G})} \mathbb{1}_E, X \rangle && [E \in \mathcal{G} \Rightarrow \mathbb{1}_E \in L^2(\mathcal{G})] \\ &= \langle \mathbb{1}_E, \pi_{L^2(\mathcal{G})}^* X \rangle = \langle \mathbb{1}_E, \pi_{L^2(\mathcal{G})} X \rangle && \text{[self-adjointness]} \\ &= \int \mathbb{1}_E \pi_{L^2(\mathcal{G})} X d\mathbb{P} = \int_E \pi_{L^2(\mathcal{G})} X d\mathbb{P}. && \text{[definitions]} \end{aligned}$$

Therefore,  $\mathbb{E}(X | \mathcal{G}) = \pi_{L^2(\mathcal{G})} X$  a.s.

### 2.4.2 Uncorrelated does not imply independence

See the wikipedia entries on [uncorrelated random variables](#) and [normally distributed and uncorrelated does not imply independent](#).

### 2.4.3 $\phi_{aX+bY} = \phi_{aX}\phi_{bY} \forall (a,b) \in \mathbb{R}^2$ implies independence

See [MathSx:1802289](#).

## 2.5 LIMITING BEHAVIOUR OF $\bar{X}_n$

A great resource for this chapter is the Wikipedia article on [central limit theorem](#).

Let  $(X_n)$  be a sequence of independent and identically distributed random variables with mean  $m$  and distribution  $\mu$ , and let  $\bar{X}_n = \frac{1}{n}S_n = \frac{1}{n} \sum_{j=1}^n X_j$ .

There are three scales for which we have three theorems, namely

1. law of large numbers ( $\bar{X}_n \xrightarrow{a.s.} m$ ),
2. law of the iterated logarithm ( $\overline{\lim} \frac{\bar{X}_n}{\sqrt{\frac{2 \log \log n}{n}}} = 1$  a.s.), and
3. central limit theorem ( $\sqrt{n}(\bar{X}_n - m) \xrightarrow{w^*} \mathcal{N}(0, \Sigma)$ ).

The idea is that we have an asymptotic expansion of  $\bar{X}_n$  given (in law) by

$$\bar{X}_n \sim m + \frac{1}{\sqrt{n}} \mathcal{N}(0, \Sigma), \quad \text{where } \Sigma \text{ is the covariance operator.}$$

The convergence to  $m$  is given by the law of large numbers, and the convergence to  $\frac{1}{\sqrt{n}} \mathcal{N}(0, \Sigma)$  is given by the central limit theorem. As  $n \rightarrow \infty$ , the dependence on  $\mathcal{N}(0, \Sigma)$  goes to zero, so this is consistent with the law of large numbers.

**Central limit theorem: how to remember in 1-dim**

$$\frac{\bar{X}_n - \mathbb{E}\bar{X}_n}{\sqrt{\mathbb{V}\bar{X}_n}} = \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{w^*} \mathcal{N}(0, 1)$$



### 3.1 CLASSIFICATION OF STOCHASTIC PROCESSES

This is well written in Cosma Rohilla Shalizi - Almost None of the Theory of Stochastic (2010), Chapter 1. Let  $X$  be a stochastic process given by

$$\begin{aligned} X : \mathbb{T} \times \Omega &\rightarrow \mathbb{E} \\ \mathcal{F} &\rightarrow \mathcal{X} \\ (t, \omega) &\mapsto X(t, \omega). \end{aligned}$$

The spaces are as follows.

$\mathbb{T}$  The *index set*. Can be finite, discrete (countable) or continuous (uncountable). Can be one-sided, two-sided, spatially distributed, or sets.

$(\mathbb{E}, \mathcal{X})$  The *state space*. Requirements: measurable. Can be finite, discrete or continuous.

$(\Omega, \mathcal{F}, \mathbb{P})$  The *probability space*.

- If  $\mathbb{T} = \{1\}$ ,  $\mathbb{E} = \mathbb{R}$ , then  $X$  is a *random variable*.
- If  $\mathbb{T} = \{1, \dots, n\}$ ,  $\mathbb{E} = \mathbb{R}$ , then  $X$  is a *random vector*.
- If  $\mathbb{T} = \{1\}$ ,  $\mathbb{E} = \mathbb{R}^d$ , then  $X$  is a *random vector*.
- If  $\mathbb{T} = \mathbb{N}$ ,  $\mathbb{E} = \mathbb{R}$ , then  $X$  is a *one-sided random sequence* or *one-sided discrete-time stochastic process*.
- If  $\mathbb{T} = \mathbb{Z}$ ,  $\mathbb{E} = \mathbb{R}$ , then  $X$  is a *two-sided random variable* or *two-sided discrete-time stochastic process*.
- If  $\mathbb{T} = \mathbb{Z}^d$ ,  $\mathbb{E} = \mathbb{R}$ , then  $X$  is a *spatial random variable*.
- If  $\mathbb{T} = \mathbb{R}$ ,  $\mathbb{E} = \mathbb{R}$ , then  $X$  is a *continuous-time random variable*.
- If  $\mathbb{T} = \mathcal{B}$ ,  $\mathbb{E} = [0, \infty]$ , then  $X$  is a *random set function on the reals*.
- If  $\mathbb{T} = \mathcal{B} \times \mathbb{N}$ ,  $\mathbb{E} = [0, \infty]$ , then  $X$  is a *one-sided random sequence of set function on the reals*.
- *Empirical measures*. Let  $(Z_n)$  be an i.i.d. random sequence and define  $\hat{\mathbb{P}}_n : \mathcal{B} \times \Omega \rightarrow \mathcal{P} : (B, \omega) \mapsto \hat{\mathbb{P}}_n(B, \omega) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_B(Z_j(\omega))$ . Then  $\hat{\mathbb{P}}_n$  is a *one-sided random sequence of set function on the reals*, which are in fact *probability measures*. [ $\mathcal{P}$  is the space of probability measures on  $\mathbb{R}$ .]
- If  $\mathbb{T} = \mathcal{B}^d$ ,  $\mathbb{E} = [0, \infty]$ , then  $X$  is the class of set functions on  $\mathbb{R}^d$ . Let  $\mathcal{M}$  be the subclass of measures. Then a random set function with realizations in  $\mathcal{M}$  is called a *random measure*.
- If  $\mathbb{T} = \mathcal{B}^d$ ,  $|\mathbb{E}| < \infty$ , then  $X$  is a *point process*.
- If  $\mathbb{T} = [0, \infty)$ ,  $\mathbb{E} = \mathbb{R}^d$ ,  $|\mathbb{E}| < \infty$ . A  $\mathbb{E}$ -valued random process on  $\mathbb{T}$  with paths in  $C(\mathbb{T})$  is a *continuous random process*. E.g. Wiener process.

## 3.2 MARTINGALES

### 3.2.1 New martingales from old

A stochastic process  $A = (A_n)$  is called adapted if  $\forall n \in \mathbb{N}, A_n \in L^0(\mathcal{F}_n)$ . Let  $M = (M_n)$  be a martingale. Then process  $\tilde{M} = (\tilde{M}_n)$  defined by  $(A \cdot M)_n = \tilde{M}_n = \sum_{j=0}^{n-1} A_j \Delta M_j$ , where  $\Delta M_j = M_{j+1} - M_j$ , is called the *martingale transform* of  $M$  by  $A$ .

Theorem (martingale transform theorem):  $\tilde{M}$  is a martingale.

Proof.

$$\mathbb{E}(\Delta \tilde{M}_n | \mathcal{F}_n) = \mathbb{E}(A_n \Delta M_n | \mathcal{F}_n) = A_n \mathbb{E}(\Delta M_n | \mathcal{F}_n) = 0.$$

Now, let  $X_n$  be a stochastic process and  $\tau$  be a stopping time. Define the stopped process  $X_\tau = \sum_{j=0}^{\infty} \mathbb{1}_{\{\tau=j\}} X_j$  when  $\mathbb{P}(\tau < \infty) = 1$ .

Theorem (stopping time theorem): Let  $(M_n)$  be a martingale with respect to  $(\mathcal{F}_n)$ . Then  $(M_{n \wedge \tau})$  is also a martingale with respect to  $(\mathcal{F}_n)$ .

Proof. Without loss of generality, assume  $M_0 = 0$ , otherwise we can translate by  $M_0$  as  $\tilde{M}_n = M_n - M_0$ . Now, the *stake process*  $A_n = \mathbb{1}_{\{\tau > n\}} = 1 - \mathbb{1}_{\{\tau \leq n\}}$  is adapted to  $(\mathcal{F}_n)$  and is bounded by  $n$ . Now,

$$\begin{aligned} (A \cdot M)_n &= \sum_{j=0}^{n-1} A_j \Delta M_j \\ &= \sum_{j=0}^{n-1} \Delta M_j - \sum_{j=0}^{n-1} \mathbb{1}_{\{\tau \leq j\}} (M_{j+1} - M_j) \\ &= M_n - M_0 - M_n \mathbb{1}_{\{\tau \leq n\}} + \sum_{j=0}^{n-1} (\mathbb{1}_{\{\tau \leq j\}} M_j - \mathbb{1}_{\{\tau \leq j-1\}} M_j) \\ &= M_n \mathbb{1}_{\{\tau > n\}} + \sum_{j=0}^{n-1} M_j \mathbb{1}_{\{\tau=j\}} \\ &= M_n \mathbb{1}_{\{\tau > n\}} + \sum_{j=0}^{n-1} M_\tau \mathbb{1}_{\{\tau=j\}} \\ &= M_n \mathbb{1}_{\{\tau > n\}} + M_\tau \sum_{j=0}^{n-1} \mathbb{1}_{\{\tau=j\}} \\ &= M_n \mathbb{1}_{\{\tau > n\}} + M_\tau \mathbb{1}_{\{\tau \leq n\}} \\ &= M_{n \wedge \tau}. \end{aligned}$$

Therefore,  $(M_{n \wedge \tau})$  is a martingale transform of  $(M_n)$ . Since  $(A_n)$  is bounded and adapted, by the martingale transform theorem,  $(M_{n \wedge \tau})$  is a martingale.

## 3.3 MARKOV PROCESSES

### Equivalent definitions

Let  $s \in [0, t]$ . Then  $X$  is a Markov process if any of the following are true:

- $\forall E \in \mathcal{F}, \mathbb{P}(X_t \in E | \mathcal{F}_s) = \mathbb{P}(X_t \in E | X_s)$ , or
- $\forall E \in \mathcal{F}, \forall f \in L^0 \cap \mathcal{B}, \mathbb{E}(f(X_t) | \mathcal{F}_s) = \mathbb{E}(f(X_t) | X_s)$ .

### Martingale vs Markov

See [djalil.chafai.net](http://djalil.chafai.net) and [MathSx:763645](https://math.stackexchange.com/questions/763645).



## 3.4 ITÔ CALCULUS

Notation:

In what follows,  $T = [0, \infty)$ ,  $\mathcal{A}$  means adapted,  $\mathcal{B}$  means bounded,  $\mathcal{C}$  means continuous, and  $\|\cdot\|$  denotes the  $L^2$ -norm.

Let  $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t), \mathbb{P})$  be a filtered probability space,  $W : T \times \Omega \rightarrow \mathbb{C}$  be a  $\mathbb{F}$ -adapted Wiener martingale, and  $X : T \times \Omega \rightarrow \mathbb{C}$  be a stochastic process.

### 3.4.1 Step 1: $X \in \mathcal{A} \cap \mathcal{S}$ a.s.

Let  $X(t, \omega) = \sum_{j \geq 0} \xi_j(\omega) \mathbb{1}_{[t_j, t_{j+1})}(t)$ , where  $\xi_j \in L^0(\mathcal{F}_{t_j})$ .

### 3.4.2 Step 2: $X \in \mathcal{A} \cap \mathcal{B} \cap \mathcal{C}$ a.s.

Define  $X_n(t, \omega) = X\left(\frac{\lfloor nt \rfloor}{n}, \omega\right)$ ,  $n \in \mathbb{N}$ . Note that  $\forall n, X_n \in \mathcal{A} \cap \mathcal{S}$ , and since  $X \in \mathcal{C}$ ,  $|X_n(t, \omega) - X(t, \omega)| \rightarrow 0$  (pointwise convergence)  $(t, \omega)$ -a.s. Then  $\forall \varepsilon > 0$ , there exists a sufficiently large  $n \in \mathbb{N}$  such that  $|X_n(t, \omega) - X(t, \omega)| < \varepsilon < \infty$  (bounded)  $(t, \omega)$ -a.s., so  $|X_n(t, \omega) - X(t, \omega)|^2 < \varepsilon^2 < \infty$   $(t, \omega)$ -a.s. Therefore, by the bounded convergence theorem,  $\|X_n - X\| \rightarrow 0$ .

Therefore,  $(X_n)$  is Cauchy in  $L^2(T \times \Omega)$ , that is,  $\|X_n - X_m\| \rightarrow 0$ . Now, by linearity and Itô isometry for the Itô integral for simple processes,  $\|\mathcal{I}(X_n) - \mathcal{I}(X_m)\| = \|\mathcal{I}(X_n - X_m)\| = \|X_n - X_m\| \rightarrow 0$ . Therefore, for  $t \in T$  fixed,  $(\mathcal{I}(X_n))$  is Cauchy in  $L^2(\Omega)$ . Since  $L^2(\Omega)$  is complete, the sequence converges. Denote the limit by  $\mathcal{I}(X)$ , that is,  $\|\mathcal{I}(X_n) - \mathcal{I}(X)\| \rightarrow 0$ .

### 3.4.3 Step 3: $X \in \mathcal{A} \cap \mathcal{B} \cap L^0(T \times \Omega)$

### 3.4.4 Step 4: $X \in \mathcal{A} \cap L^2(T \times \Omega)$

### 3.4.5 Step 5: $X \in \mathcal{A} \cap \left\{ X \in \mathbb{C}^{T \times \Omega} : \forall t \geq 0, \int_0^t X(s, \cdot) ds < \infty \right\}$ a.s.

### 3.4.6 Properties of the Itô integral

In what follows, assume the following. Let  $X, Y \in \mathcal{A} \cap L^2(T \times \Omega)$ ;  $(X_n), (Y_n) \subset \mathcal{A} \cap \mathcal{S}$  such that  $\|X_n - X\| \rightarrow 0$  and  $\|Y_n - Y\| \rightarrow 0$ . Let  $z \in \mathbb{C}$ .

### 3.4.6.1 Linearity: $\|z\mathcal{I}(X) + \mathcal{I}(Y) - \mathcal{I}(zX + Y)\| = 0$

First, note that  $\|(zX_n + Y_n) - (zX + Y)\| \leq |z|\|X_n - X\| + \|Y_n - Y\| \rightarrow 0$ . Now, by the linearity of the integral  $\mathcal{I} : \mathcal{A} \cap \mathcal{S} \rightarrow L^2(\Omega)$ , we have

$$\begin{aligned} & \|z\mathcal{I}(X) + \mathcal{I}(Y) - \mathcal{I}(zX + Y)\| \\ &= \|z\mathcal{I}(X) + \mathcal{I}(Y) - z\mathcal{I}(X_n) - \mathcal{I}(Y_n) + \mathcal{I}(zX_n + Y_n) - \mathcal{I}(zX + Y)\| \\ &\leq |z|\|\mathcal{I}(X) - \mathcal{I}(X_n)\| + \|\mathcal{I}(Y) - \mathcal{I}(Y_n)\| + \|\mathcal{I}(zX_n + Y_n) - \mathcal{I}(zX + Y)\| \rightarrow 0. \end{aligned}$$

### 3.4.6.2 Itô isometry: $\|\mathcal{I}(X)\| = \|X\|$

Using the isometry of the integral  $\mathcal{I} : \mathcal{A} \cap \mathcal{S} \rightarrow L^2(\Omega)$ , we have

$$\begin{aligned} \|\mathcal{I}(X)\| &\leq \|\mathcal{I}(X) - \mathcal{I}(X_n)\| + \|\mathcal{I}(X_n)\| \\ &= \|\mathcal{I}(X) - \mathcal{I}(X_n)\| + \|X_n\| \\ &= \|\mathcal{I}(X) - \mathcal{I}(X_n)\| + \|X_n - X\| + \|X\| \rightarrow \|X\|. \end{aligned}$$

Note that the ‘Itô isometry’ is actually a unitary transformation.

### 3.4.6.3 Martingale property: $\mathbb{E}(\mathcal{I}_t(X) | \mathcal{F}_s) = \mathcal{I}_s(X)$ a.s.

The martingale property of the integral  $\mathcal{I} : \mathcal{A} \cap \mathcal{S} \rightarrow L^2(\Omega)$  gives  $\mathbb{E}(\mathcal{I}_t(X_n) - \mathcal{I}_s(X_n) | \mathcal{F}_s) = 0$ . Using this and the unitariness of the Itô isometry, we get

$$\begin{aligned} & \|\mathbb{E}(\mathcal{I}_t(X) - \mathcal{I}_s(X) | \mathcal{F}_s)\|^2 \\ &= \mathbb{E} |\mathbb{E}(\mathcal{I}_t(X) - \mathcal{I}_t(X_n) + \mathcal{I}_s(X_n) - \mathcal{I}_s(X) | \mathcal{F}_s) + \mathbb{E}(\mathcal{I}_t(X_n) - \mathcal{I}_s(X_n) | \mathcal{F}_s)|^2 \\ &= \mathbb{E} |\mathbb{E}(\mathcal{I}_t(X) - \mathcal{I}_t(X_n) + \mathcal{I}_s(X_n) - \mathcal{I}_s(X) | \mathcal{F}_s) + 0|^2 \\ &\leq \mathbb{E} \mathbb{E}(|\mathcal{I}_t(X) - \mathcal{I}_t(X_n) + \mathcal{I}_s(X_n) - \mathcal{I}_s(X)|^2 | \mathcal{F}_s) \\ &= \mathbb{E} |\mathcal{I}_t(X) - \mathcal{I}_t(X_n) + \mathcal{I}_s(X_n) - \mathcal{I}_s(X)|^2 \\ &= \|\mathcal{I}_t(X - X_n) + \mathcal{I}_s(X_n - X)\|^2 \\ &\leq 2(\|\mathcal{I}_t(X - X_n)\|^2 + \|\mathcal{I}_s(X_n - X)\|^2) \quad [\|a + b\|^2 \leq 2(\|a\|^2 + \|b\|^2)] \\ &= 2(\|X - X_n\|^2 + \|X_n - X\|^2) = 4\|X_n - X\|^2 \rightarrow 0. \end{aligned}$$

## 3.4.7 Itô formula for multidimensional processes

Recall the discussion of Taylor series for multivariate functionals in section 1.2.3. This part is from (Sundar & Kallianpur, 2014), § 5.3, 5.4 and 5.6.

A (local, continuous) semimartingale is a process  $X_t$  that can be written as  $X_t = X_0 + M_t + A_t$ , where

1.  $M_t$  is a mean-zero (local, continuous) martingale, and
2.  $A_t$  is an right-continuous adapted process of locally bounded variation.

This is equivalently represented in the differential form as  $dX_t = dM_t + dA_t$ .

Let  $X_t$  be a  $d$ -dimensional semimartingale, and let  $Y_t = f(X_t)$ , where  $f \in C^2(\mathbb{R})$ . Then

$$dY_t = df(X_t) = f'(X_t) dA_t + f'(X_t) dM_t + \frac{1}{2} f''(X_t) d\langle M_t, M_t \rangle,$$

where we use the rule  $d\langle B^{(j)}, B^{(j)} \rangle_t = (dB_t^{(j)})^2 = dt$ , everything else being 0.

Alternatively, as from (Kuo, 2006), we have the following.

Let  $X_t = (X_t^{(1)}, \dots, X_t^{(d)})$  be a  $d$ -dimensional process and  $f(t, x)$  be a functional of  $(t, X_t)$ .

Then the Itô formula becomes

$$df(t, X_t) = (\langle dt, D_t \rangle f)(t, X_t) + (\langle dx, D_x \rangle f)(t, X_t) + \frac{1}{2} (\langle dx, D_x \rangle^2 f)(t, X_t),$$

or in short,  $df = (\langle dt, D_t \rangle + \langle dx, D_x \rangle + \frac{1}{2} \langle dx, D_x \rangle^2) f$ .

## 3.5 EXAMPLES

### 3.5.1 Find $\mathbb{E} \left( \int_0^1 B_t^2 dt \right)^2$ .

By Fubini's theorem,

$$\mathbb{E} \left( \int_0^1 B_t^2 dt \right)^2 = \mathbb{E} \left( \int_0^1 B_t^2 dt \int_0^1 B_s^2 ds \right) = \mathbb{E} \left( \int_0^1 \int_0^1 B_t^2 B_s^2 ds dt \right) = \int_0^1 \int_0^1 \mathbb{E}(B_t^2 B_s^2) ds dt.$$

$$\begin{aligned} \text{Now, } \forall s \in [0, t], \quad \mathbb{E}(B_t^2 B_s^2) &= \mathbb{E}(\mathbb{E}(B_t^2 B_s^2 | \mathcal{F}_s)) = \mathbb{E}(B_s^2 \mathbb{E}((B_t^2 - t) + t | \mathcal{F}_s)) \\ &= \mathbb{E}(B_s^2 ((B_s^2 - s) + t)) = \mathbb{E}(B_s^2 ((B_s^2 - s) + t)) \\ &= \mathbb{E}(B_s^4 - sB_s^2 + tB_s^2) = 3s^2 - s^2 + ts = 2s^2 + ts. \end{aligned}$$

$$\text{So} \quad \mathbb{E} \left( \int_0^1 B_t^2 dt \right)^2 = 2 \int_0^1 \int_0^t (2s^2 + ts) ds dt = \frac{7}{9}.$$

### 3.5.2 Find $\mathbb{V} \left( \int_0^1 t^2 B_t dt \right)$ .

By Fubini's theorem,  $\mathbb{E} \left( \int_0^1 t^2 B_t dt \right) = \int_0^1 t^2 \mathbb{E} B_t dt = 0$ . So by Fubini's theorem (again),

$$\begin{aligned} \mathbb{V} \left( \int_0^1 t^2 B_t dt \right) &= \mathbb{E} \left( \int_0^1 t^2 B_t dt \right)^2 = \mathbb{E} \left( \int_0^1 t^2 B_t dt \int_0^1 s^2 B_s ds \right) \\ &= \mathbb{E} \left( \int_0^1 \int_0^1 t^2 s^2 B_t B_s ds dt \right) = \int_0^1 \int_0^1 t^2 s^2 \mathbb{E}(B_t B_s) ds dt \\ &= \int_0^1 \int_0^1 t^2 s^2 (t \wedge s) ds dt = 2 \int_0^1 \int_0^t t^2 s^2 ds dt = \frac{1}{14}. \end{aligned}$$



## 4.1 ABSTRACT WIENER SPACES

## 4.2 WHITE NOISE DISTRIBUTION THEORY

### 4.2.1 Characterization theorem

Importance and history.

In the following,  $F$  is defined on  $S_{\mathbb{C}}$ , and  $F(z\xi + \eta)$  is entire  $\forall z \in \mathbb{C}$ .

$$(S)_{\beta} \subset (S) \subset (L^2) \subset (S)^* \subset (S)_{\beta}^*$$

$$S \subset L^2 \subset S'$$

## BIBLIOGRAPHY

Kuo, H. H. (2006). *Introduction to Stochastic Integration*. (1 ed.). Springer-Verlag. doi:10.1007/0-387-31057-6

Sundar, P. & Kallianpur, G. (2014). *Stochastic Analysis and Diffusion Processes*. Author. doi:10.1093/acprof:oso/9780199657063.001.0001