Contents

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 $\sum : S \to S, (a_n) \mapsto \left(\sum_{j=1}^{n} a_j\right)$ is a *linear operator* on S, whose inverse is the finite difference operator, Δ. 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 $(1, 1, 1, \ldots)$ has series $(1, 2, 3, \ldots)$ as its partial summation, which is analogous to the fact that $\int_0^x 1 \, \mathrm{d}t = x$.
- 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 $\sigma : \mathbb{N} \to \mathbb{N}$ is a permutation, then $\sum_n a_n = \sum_n a_{\sigma(n)}$.
       In complete spaces, absolute convergence $\Longrightarrow$ 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 in 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 $f : \mathbb{R} \to \mathbb{R} : x \mapsto x^n$ 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

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

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 $n \to \infty$. Consider $f : \mathbb{R} \to \mathbb{R} : x \mapsto e^{-\frac{1}{x}} \mathbb{1}_{x>0}(x)$. 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 $f : \Omega \to \mathbb{R}$ 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*

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

Proof Is this a $\mathbb{R}$ life, or is this just *i*maginary, 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

$$\mathrm{d}T(p) = \sum_{n=1}^{\infty} \frac{1}{n!} \left(\langle \mathrm{d}v, \mathrm{D}_v \rangle^n f \right)(p)$$

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 σ-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, lim inf, lim sup 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

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

□

Counterexample of the converse of BC1: Take $((0,1], \lambda)$ as the probability space, and $E_n = \left(0, \frac{1}{n^2}\right)$. 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 $\mathbb{P}\left((E_n \text{ i.o.})^{\mathsf{C}}\right) = \mathbb{P}\left(E_n^{\mathsf{C}} \text{ ev}\right)$, it is equivalent to prove that $\mathbb{P}\left(E_n^{\mathsf{C}} \text{ ev}\right) = 0$. Using independence and the fact that $1 - x < e^{-x}$, for each fixed $n \in \mathbb{N}$, we have

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

Taking $N \to \infty$, we get $\mathbb{P}\left(\bigcap_{m \geq n} E_n^{\mathsf{C}}\right) \to 0$. Therefore

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

□

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.

$$
\begin{tikzcd}
unif \arrow[r] \arrow[d] & L^\infty \arrow[r] \arrow[d] & L^2 \arrow[r] \arrow[dl, "(X_n \perp) \wedge (\mathbb{E} X_n = 0) \wedge (\sum \mathbb{V} X_n < \infty)"'] & L^1 \arrow[r] & \mathbb{P} \arrow[r] \arrow[l, bend right, "\text{u.i.} \vee \text{unif bound}"'] \arrow[dlll, bend left, "\forall \text{ subseq } \exists \text{ subsubseq}"] & w^* \arrow[l, bend right, "const"'] \\
pt \arrow[r] & a.s. \arrow[urrr] \arrow[urr, dotted, "LDCT?"] & & & &
\end{tikzcd}
$$

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-σ-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 $\forall E \in \mathcal{G}$,

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

Therefore, $\mathbb{E}(X \mid \mathcal{G}) = \pi_{L^2(\mathcal{G})} X$ 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 μ, 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 ($\bar{X}_n \overset{a.s.}{\to} m$),
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 ($\sqrt{n}(\bar{X}_n - m) \overset{w^*}{\to} \mathcal{N}(0, \Sigma)$).

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 $n \to \infty$, 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

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

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.

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

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

- If $\mathbb{T} = \{1\}, \Xi = \mathbb{R}$, then X is a *random variable*.
- If $\mathbb{T} = \{1, \ldots, n\}, \Xi = \mathbb{R}$, then X is a *random vector*.
- If $\mathbb{T} = \{1\}, \Xi = \mathbb{R}^d$, then X is a *random vector*.
- If $\mathbb{T} = \mathbb{N}, \Xi = \mathbb{R}$, then X is a *one-sided random sequence* or *one-sided discrete-time stochastic process*.
- If $\mathbb{T} = \mathbb{Z}, \Xi = \mathbb{R}$, then X is a *two-sided random variable* or *two-sided discrete-time stochastic process*.
- If $\mathbb{T} = \mathbb{Z}^d, \Xi = \mathbb{R}$, then X is a *spatial random variable*.
- If $\mathbb{T} = \mathbb{R}, \Xi = \mathbb{R}$, then X is a *continuous-time random variable*.
- If $\mathbb{T} = \mathcal{B}, \Xi = [0, \infty]$, then X is a *random set function on the reals*.
- If $\mathbb{T} = \mathcal{B} \times \mathbb{N}, \Xi = [0, \infty]$, then X is a *one-sided random sequence of set function on the reals*.
- *Emperical measures.* Let (Z_n) be an i.i.d. random sequence and define $\hat{\mathbb{P}}_n : \mathcal{B} \times \Omega \to \mathcal{P} : (B, \omega) \mapsto \hat{\mathbb{P}}_n(B, \omega) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_B(Z_j(\omega))$. 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 $\mathbb{T} = \mathcal{B}^d, \Xi = [0, \infty]$, 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 $\mathbb{T} = \mathcal{B}^d, |\Xi| < \infty$, then X is a *point process*.
- If $\mathbb{T} = [0, \infty), \Xi = \mathbb{R}^d < \infty$. A Ξ-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.

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

Now, let X_n be a stochastic process and τ 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* $A_n = \mathbb{1}_{\{\tau > n\}} = 1 - \mathbb{1}_{\{\tau \leqslant n\}}$ is adapted to $(\mathcal{F}_n)$ and is bounded by n. Now,

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

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:

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

Martingale vs Markov

See djalil.chafai.net and MathSx: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, $W : T \times \Omega \to \mathbb{C}$ be a $\mathbb{F}$-adapted Wiener martingale, and $X : T \times \Omega \to \mathbb{C}$ 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 $X \in \mathcal{C}$, $|X_n(t, \omega) - X(t, \omega)| \to 0$ (pointwise convergence) (t, ω)-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, ω)-a.s, so $|X_n(t, \omega) - X(t, \omega)|^2 < \varepsilon^2 < \infty$ (t, ω)-a.s. Therefore, by the bounded convergence theorem, $\|X_n - X\| \to 0$.
Therefore, (X_n) is Cauchy in $L^2(T \times \Omega)$, that is, $\|X_n - X_m\| \to 0$. Now, by linearity and Itô isometry for the Itô integral for simple processes, $\|\mathcal{I}(X_n) - \mathcal{I}(X_m)\| = \|\mathcal{I}(X_n - X_m)\| = \|X_n - X_m\| \to 0$. 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, $\|\mathcal{I}(X_n) - \mathcal{I}(X)\| \to 0$.

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: $X \in \mathcal{A} \cap \left\{X \in \mathbb{C}^{T \times \Omega} : \forall t \geq 0, \int_0^t X(s, \cdot)\, \mathrm{d}s < \infty\right\}$ *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 $\|X_n - X\| \to 0$ and $\|Y_n - Y\| \to 0$. 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 $\|(zX_n+Y_n)-(zX+Y)\| \leq |z|\,\|X_n-X\|+\|Y_n-Y\| \to 0$. Now, by the linearity of the integral $\mathcal{I}:\mathcal{A}\cap\mathcal{S}\to L^2(\Omega)$, we have

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

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

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

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

Note that the 'Itô isometry' is actually a unitary transformation.

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

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

$$\begin{aligned}
&\left\|\mathbb{E}\big(\mathcal{I}_t(X)-\mathcal{I}_s(X)\mid\mathcal{F}_s\big)\right\|^2 \\
&=\mathbb{E}\left|\mathbb{E}\big(\mathcal{I}_t(X)-\mathcal{I}_t(X_n)+\mathcal{I}_s(X_n)-\mathcal{I}_s(X)\mid\mathcal{F}_s\big)+\mathbb{E}\big(\mathcal{I}_t(X_n)-\mathcal{I}_s(X_n)\mid\mathcal{F}_s\big)\right|^2 \\
&=\mathbb{E}\left|\mathbb{E}\big(\mathcal{I}_t(X)-\mathcal{I}_t(X_n)+\mathcal{I}_s(X_n)-\mathcal{I}_s(X)\mid\mathcal{F}_s\big)+0\right|^2 \\
&\leq \mathbb{E}\mathbb{E}\left(\left|\mathcal{I}_t(X)-\mathcal{I}_t(X_n)+\mathcal{I}_s(X_n)-\mathcal{I}_s(X)\right|^2\mid\mathcal{F}_s\right) \\
&=\mathbb{E}\left|\mathcal{I}_t(X)-\mathcal{I}_t(X_n)+\mathcal{I}_s(X_n)-\mathcal{I}_s(X)\right|^2 \\
&=\left\|\mathcal{I}_t(X-X_n)+\mathcal{I}_s(X_n-X)\right\|^2 \\
&\leq 2\left(\left\|\mathcal{I}_t(X-X_n)\right\|^2+\left\|\mathcal{I}_s(X_n-X)\right\|^2\right) \qquad \left[\|a+b\|^2\leq 2\left(\|a\|^2+\|b\|^2\right)\right] \\
&=2\left(\|X-X_n\|^2+\|X_n-X\|^2\right)=4\,\|X_n-X\|^2 \to 0.
\end{aligned}$$

3.4.7 Itô formula for multidimensional processes

Recall the discussion of Taylor series for multivariate functionals in section 1.2.3.
Let $X_t=(X_t^{(1)},\ldots,X_t^{(d)})$ be a d-dimensional process and $f(t,x)$ be a functional of (t,X_t).
Then the Itô formula becomes

$$\mathrm{d}f(t, X_t) = (\langle\, \mathrm{d}t,\, \mathrm{D}_t \rangle f)\,(t, X_t) + (\langle\, \mathrm{d}x,\, \mathrm{D}_x \rangle f)\,(t, X_t) + \frac{1}{2}\left(\langle\, \mathrm{d}x,\, \mathrm{D}_x \rangle^2 f\right)(t, X_t),$$

or in short, $\mathrm{d}f = \left(\langle\, \mathrm{d}t,\, \mathrm{D}_t \rangle + \langle\, \mathrm{d}x,\, \mathrm{D}_x \rangle + \frac{1}{2}\langle\, \mathrm{d}x,\, \mathrm{D}_x \rangle^2\right) f.$

3.5 Examples

3.5.1 Find $\mathbb{E}\left(\int_0^1 B_t^2\,\mathrm{d}t\right)^2$.

By Fubini's theorem,

$$\mathbb{E}\left(\int_0^1 B_t^2\,\mathrm{d}t\right)^2 = \mathbb{E}\left(\int_0^1 B_t^2\,\mathrm{d}t\int_0^1 B_s^2\,\mathrm{d}s\right) = \mathbb{E}\left(\int_0^1\int_0^1 B_t^2 B_s^2\,\mathrm{d}s\,\mathrm{d}t\right) = \int_0^1\int_0^1 \mathbb{E}(B_t^2 B_s^2)\,\mathrm{d}s\,\mathrm{d}t.$$

Now, $\forall s \in [0,t]$,

$$\begin{aligned}\mathbb{E}(B_t^2 B_s^2) &= \mathbb{E}(\mathbb{E}(B_t^2 B_s^2 \mid \mathcal{F}_s)) = \mathbb{E}(B_s^2\mathbb{E}((B_t^2 - t) + t \mid \mathcal{F}_s)) \\ &= \mathbb{E}(B_s^2((B_s^2 - s) + t)) = \mathbb{E}(B_s^2((B_s^2 - s) + t)) \\ &= \mathbb{E}(B_s^4 - sB_s^2 + tB_s^2) = 3s^2 - s^2 + ts = 2s^2 + ts.\end{aligned}$$

So

$$\mathbb{E}\left(\int_0^1 B_t^2\,\mathrm{d}t\right)^2 = 2\int_0^1\int_0^t (2s^2 + ts)\,\mathrm{d}s\,\mathrm{d}t = \frac{7}{9}.$$

3.5.2 Find $\mathbb{V}\left(\int_0^1 t^2 B_t\,\mathrm{d}t\right)$.

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

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

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