Stochastic processes

Péter Kevei

February 14, 2022

Contents

1 Discrete time martingales

First part of discrete time martingales are mainly from Durrett [3].

1.1 Regular conditional probabilities

This is from Durrett [3].

Let $\mathcal{G} \subset \mathcal{A}$ be a sub-σ-algebra, and consider for $A \in \mathcal{A}$ the conditional probabilities $\mathbf{P}(A|\mathcal{G}) = \mathbf{E}[\mathbf{I}_A|\mathcal{G}]$. Since $0 \leq \mathbf{I}_A \leq 1$,

$$\mathbf{P}(A|\mathcal{G}) \in [0,1] \quad \text{a.s.} \tag{1}$$

Furthermore, for disjoint $A_i \in \mathcal{A}$, $i = 1, 2, \ldots$, we have

$$\mathbf{P}(\cup_{i=1}^{\infty} A_i|\mathcal{G}) = \sum_{i=1}^{\infty} \mathbf{P}(A_i|\mathcal{G}) \quad \text{a.s.} \tag{2}$$

Therefore, $\mathbf{P}(\cdot|\mathcal{G})$ behaves as a probability measure. However, both (1) and (2) hold almost surely. That is, there is an exceptional set N_A, which is a $\mathbf{P}$-null set, $\mathbf{P}(N_A) = 0$, such that (1) holds for $\omega \notin N_A$. The $\mathbf{P}$-null set N_A may depend on A. In general, a σ-algebra has more than countable infinitely many sets (indeed, it is either finite, or at least continuum). Thus these null set may pile up to a large set. The same problem appears in (2). Under some general conditions we can guarantee that the bad points cannot pile up.

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space, $X : \Omega \to S$ a random element in $(S, \mathcal{S})$, and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{A}$. The *regular conditional distribution for X given* $\mathcal{G}$ is a function $\mu : \Omega \times \mathcal{S} \to [0,1]$ such that

(i) for each $A \in \mathcal{A}$ fix, $\mathbf{P}(X \in A|\mathcal{G})(\omega) = \mu(\omega, A)$ a.s.;

(ii) Almost surely $A \mapsto \mu(\omega, A)$ is a probability measure on $(S, \mathcal{S})$.

If $S = \Omega$ and X is the identity map, μ is a *regular conditional probability.*

A measurable space $(S, \mathcal{S})$ is *nice*, if there is a 1-1 map $\phi : S \to \mathbb{R}$ such that ϕ, ϕ^{-1} are measurable. If S is a Borel subset of a complete separable metric space, and $\mathcal{S}$ are the Borel sets, then $(S, \mathcal{S})$ is nice.

Theorem 1. *Regular conditional distribution exist if $(S, \mathcal{S})$ is nice.*

Proof. We prove only in the special case $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The general case is almost identical, with some technical difficulties.

Consider the conditional probabilities $\mathbf{P}(X \leq q|\mathcal{G})$, $q \in \mathbf{Q}$. For each $q \in \mathbf{Q}$ there is $\mathbf{P}$-null set N_q, such that $\mathbf{P}(X \leq q|\mathcal{G})(\omega) \in [0,1]$ for $\omega \notin N_q$. Similarly, for each $q < r$ there exists a $\mathbf{P}$-null set $N_{q,r}$ such that for each $\omega \notin N_{q,r}$

$$\mathbf{P}(X \leq q|\mathcal{G})(\omega) \leq \mathbf{P}(X \leq r|\mathcal{G})(\omega).$$

Set

$$N = \cup_{q \in \mathbf{Q}} N_q \cup \cup_{q<r \in \mathbf{Q}} N_{q,r}.$$

Then $\mathbf{P}(N) = 0$ and $\mathbf{P}(X \leq q|\mathcal{G})(\omega) \in [0,1]$, and it is nondecreasing in $q \in \mathbf{Q}$ for $\omega \notin N$. Let

$$G(x,\omega) = \inf\{\mathbf{P}(X \leq q|\mathcal{G})(\omega) : q > x\}.$$

If $\omega \notin N$ then $G(x,\omega)$ is a distribution function in x.

Furthermore, since $\mathbf{P}(X \leq q_n|\mathcal{G}) \downarrow \mathbf{P}(X \leq x|\mathcal{G})$ as $q_n \downarrow x$ we see that $G(x,\omega) = \mathbf{P}(X \leq x|\mathcal{G})(\omega)$ a.s. □

Taking $\mathcal{G} = \sigma(Y)$, we see that $\mathbf{P}(X \in A|\mathcal{G})$ is a measurable function of Y, for each A. This can be done also simultaneously, as above.

Theorem 2. *Let X, Y be random elements in the nice space $(S, \mathcal{S})$, and let $\mathcal{G} = \sigma(Y)$. Then there exists $\mu : S \times \mathcal{S} \to [0,1]$ such that*

(i) for each $A \in \mathcal{S}$, $\mu(Y(\omega), A) = \mathbf{P}(X \in A|Y)(\omega)$ a.s.

(ii) almost surely $A \mapsto \mu(Y(\omega), A)$ is a probability measure on $(S, \mathcal{S})$.

Proof. The proof is similar to the previous one. Again assume $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

As above, we can find random variables $G(q,\omega)$ nondecreasing in q outside of $\mathbf{P}$-null set N, such that $G(q,\omega) = \mathbf{P}(X \leq q|Y)(\omega)$, a.s., $q \in \mathbf{Q}$. Since the latter is $\sigma(Y)$-measurable, $G(q,\omega) = H(q, Y(\omega))$. Choosing

$$F(x,y) = \inf\{H(q,y) : q > x\},$$

we can show that $F(x, Y(\omega)) = \mathbf{P}(X \leq x|Y)(\omega)$. This defines a measure, since $F(\cdot, y)$ is nondecreasing. □

Existence of regular conditional distribution allows us to compute conditional expectations simultaneously, and also shows the connection to usual expectation.

Theorem 3. *Let μ be a regular conditional distribution for X given $\mathcal{G}$. Let $f : S \to \mathbb{R}$ measurable with $\mathbf{E}|f(X)| < \infty$. Then*

$$\mathbf{E}[f(X)|\mathcal{G}] = \int f(x)\mu(\omega, \mathrm{d}x) \quad a.s.$$

Proof. The result holds for indicators, by definition. Linearity and monotone convergence implies the statement, as usual. □

1.2 Martingales: definition, properties

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space. A *filtration* is an increasing sequence of sub-σ-algebras $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}_n \subset \ldots$. A sequence of random variables $(X_n)_n$ is adapted to the filtration $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_n$-measurable. The sequence $(X_n, \mathcal{F}_n)$ is a *martingale*, or (X_n) is martingale with respect to $(\mathcal{F}_n)$, if

(i) $\mathbf{E}|X_n| < \infty$;

(ii) (X_n) is adapted;

(iii) $\mathbf{E}[X_{n+1}|\mathcal{F}_n] = X_n$.

For a submartingale (supermartingale) conditions (i) and (ii) hold, and (iii) hold with $\geq$ ($\leq$).

If the filtration is not specified then (X_n) is martingale meant as it is martingale with respect to the natural filtration $(\mathcal{F}_n)$, where $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$.

Clearly, if (X_n) is a submartingale then $(-X_n)$ is a supermartingale, therefore it is enough to prove statements for submartingales.

Example 1. Let $X, X_1, \ldots$ iid random variables with $\mathbf{E}X = 0$, and put $S_n = X_1 + \ldots + X_n$. Then (S_n) is a martingale. If $\mathbf{E}X^2 < \infty$ then (S_n^2) is a submartingale.

If $X, X_1, \ldots$ are iid nonnegative random variables, $\mathbf{E}X = 1$, then $(R_n)_n$ is a martingale, where $R_n = \prod_{i=1}^n X_i$.

Proposition 1. *If (X_n) is a submartingale then $\mathbf{E}[X_n|\mathcal{F}_m] \geq X_m$ for any $n > m$. Equality holds for martingales.*

Proposition 2. *(i) Let $(X_n, \mathcal{F}_n)$ be a martingale and φ be a convex function such that $\mathbf{E}|\varphi(X_n)| < \infty$. Then $\varphi(X_n)$ is a martingale.*

(ii) Let $(X_n, \mathcal{F}_n)$ be a submartingale and φ be a nondecreasing convex function such that $\mathbf{E}|\varphi(X_n)| < \infty$. Then $\varphi(X_n)$ is a submartingale.

Proof. It follows from Jensen's inequality. □

Corollary 1. *If (X_n) is a submartingale then $((X_n - a)_+)$ is a submartingale. If (X_n) is a supermartingale then $(X_n \wedge a)_n$ is a supermartingale.*

1.3 Martingale convergence theorem

A sequence (H_n) is *predictable* if H_n is $\mathcal{F}_{n-1}$-measurable. Let (H_n) be predictable and (X_n) be adapted. Then

$$(H \cdot X)_n = \sum_{k=1}^{n} H_m (X_m - X_{m-1}).$$

Note that this is a discrete stochastic integral.

Theorem 4. *Let (X_n) be a submartingale and (H_n) be predictable, nonnegative, and bounded. Then $(H \cdot X)_n$ is a submartingale.*

Proof. Follows from the submartingale property. □

An integer valued random variable τ is a *stopping time*, if $\{\tau = n\} \in \mathcal{F}_n$.

Corollary 2. *Let τ be a stopping time, and (X_n) a submartingale. Then $(X_{\tau \wedge n})$ is a submartingale.*

Proof. Let $H_n = \mathbf{I}(\tau \geq n)$. Then (H_n) is predictable, thus $((H \cdot X)_n = X_{\tau \wedge n} - X_0)_n$ is a submartingale. □

Let (X_n) be a submartingale, $a < b$. Let $\tau_0 = -1$, and

$$\begin{aligned} \tau_{2k-1} &= \min\{m > \tau_{2k-2} : X_m \leq a\}, \\ \tau_{2k} &= \min\{m > \tau_{2k-1} : X_m \geq b\}. \end{aligned}$$

Then τ_k's are stopping times. So

$$H_m = \begin{cases} 1, & \text{if } \tau_{2k-1} < m \leq \tau_{2k} \text{ for some } k, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

is predictable. By definition $X(\tau_{2k-1}) \leq a$ and $X(\tau_{2k}) \geq b$, thus between τ_{2k-1} and τ_{2k} the process (X_n) crosses the strip $[a, b]$. This is an *upcrossing.* Let $U_n = \max\{k : \tau_{2k} \leq n\}$ is the number of upcrossings up to time n.

Lemma 1 (Upcrossing lemma)**.** *Let* (X_n) *be a submartingale,* $a < b$*. Then*

$$(b-a)\mathbf{E}U_n \leq \mathbf{E}(X_n - a)_+ - \mathbf{E}(X_0 - a)_+.$$

Proof. Define $Y_n = a + (X_n - a)_+$. This is a submartingale, which upcrosses $[a,b]$ the same number of times as (X_n) does. Recalling the definition of H from (3) we have $(b-a)U_n \leq (H \cdot Y)_n$. Indeed, each upcrossing has at least $b-a$ contribution, and the last incomplete one has a nonnegative contribution (because of changing X to Y).

Let $K_n = 1 - H_n$. Since $Y_n - Y_0 = (H \cdot Y)_n + (K \cdot Y)_n$, and

$$\mathbf{E}(K \cdot Y)_n \geq \mathbf{E}(K \cdot Y)_0 = 0,$$

we have

$$\mathbf{E}(H \cdot Y)_n \leq \mathbf{E}(Y_n - Y_0),$$

and the result follows. □

A consequence of the previous lemma we obtain the following.

Theorem 5 (Martingale convergence theorem)**.** *Let* (X_n) *be a submartingale with* $\sup \mathbf{E}X_n^+ < \infty$*. Then* $\lim_{n\to\infty} X_n$ *converges almost surely to a* X *with* $\mathbf{E}|X| < \infty$*.*

Proof. Fix $a < b$. Since $(X - a)_+ \leq X_+ + |a|$ we have

$$\mathbf{E}U_n \leq \frac{|a| + \mathbf{E}X_n^+}{b-a}.$$

Let $U = \lim_{n\to\infty} U_n$. Clearly, U_n is nondecreasing, so the limit exists. By the assumptions $\mathbf{E}U < \infty$, in particular U is finite almost surely. This holds for any $a < b$, the set

$$A = \cup_{a,b\in\mathbf{Q}}\{\liminf X_n < a < b < \limsup X_n\}$$

has probability 0. If $\omega \notin A$ then $\lim X_n(\omega)$ exists. By Fatou, $\mathbf{E}X^* \leq \liminf \mathbf{E}X_n^+ < \infty$, so X is finite a.s. Furthermore,

$$\mathbf{E}X_n^- = \mathbf{E}X_n^+ - \mathbf{E}X_n \leq \mathbf{E}X_n^+ - \mathbf{E}X_0,$$

which implies

$$\mathbf{E}X^- \leq \liminf_{n\to\infty} \mathbf{E}X_n^- \leq \sup \mathbf{E}X_n^+ - \mathbf{E}X_0 < \infty.$$

□

1.4 Doob's decomposition

A submartingale is informally a stochastically increasing sequence. It can be decomposed to a martingale part, which corresponds to a fair game, and a predictable almost surely nondecreasing part.

Theorem 6 (Doob's decomposition)**.** *Let (X_n) be a submartingale. There exists a unique martingale (M_n), and a predictable nondecreasing sequence (A_n), with $A_0 = 0$ such that $X_n = M_n + A_n$.*

Proof. Existence. Under the stated properties

$$\mathbf{E}[X_n|\mathcal{F}_{n-1}] = M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n.$$

Therefore, we must have

$$A_n = \sum_{k=1}^{n} \mathbf{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}],$$

and $M_n = X_n - A_n$. It is easy to see that this indeed works.

Uniqueness follows easily. □

1.5 Doob's maximal inequality

This part is mainly from Csörgő [2].

Our first optional stopping theorem is the following.

Theorem 7. *Let $(X_n)_n$ be a submartingale and let N be a bounded stopping time, i.e. $N \leq k$ a.s. for some $k \in \mathbb{N}$. Then*

$$\mathbf{E}X_0 \leq \mathbf{E}X_N \leq \mathbf{E}X_k.$$

Proof. We proved that the stopped process $(X_{n\wedge N})_n$ is submartingale, thus

$$\mathbf{E}X_0 = \mathbf{E}X_{N\wedge 0} \leq \mathbf{E}X_{N\wedge k} = \mathbf{E}X_N.$$

For the other direction, put $K_n = \mathbf{I}(N < n) = \mathbf{I}(N \leq n-1)$. Then K_n is $\mathcal{F}_{n-1}$-measurable, so $(K_n)_n$ is predictable. Therefore $(K \cdot X)_n$ is submartingale, where

$$(K \cdot X)_n = \sum_{i=1}^{n} \mathbf{I}(N \leq i-1)(X_i - X_{i-1}) = X_n - X_{N\wedge n}.$$

That is

$$\mathbf{E}X_k - \mathbf{E}X_N = \mathbf{E}(K \cdot X)_k \geq \mathbf{E}(K \cdot X)_0 = 0.$$

□

An easy consequence is Doob's maximal inequality.

Theorem 8 (Doob's maximal inequality). *Let $(X_k, \mathcal{F}_k)_k$ be a submartingale, and put*

$$M_n = \max_{1 \leq k \leq n} X_k.$$

Then for any $x > 0$

$$x\mathbf{P}(M_n \geq x) \leq \int_{\{M_n \geq x\}} X_n \mathrm{d}\mathbf{P} \leq \mathbf{E}X_n^+,$$

where $a^+ = \max\{a, 0\}$.

Proof. The second inequality is obvious.

Let $N = \min\{\min\{k : X_k \geq x, k = 1, 2, \ldots, n\}, n\}$. Then N is a bounded stopping time. Since $X_N \geq x$ on $\{M_n \geq x\}$

$$x\mathbf{P}\{M_n \geq x\} \leq \int_{\{M_n \geq x\}} X_N \mathrm{d}\mathbf{P}.$$

By Theorem 7 $\mathbf{E}X_N \leq \mathbf{E}X_n$, and $X_N = X_n$ on the event $\{M_n < x\}$, thus

$$\int_{\{M_n \geq x\}} X_N \mathrm{d}\mathbf{P} \leq \int_{\{M_n \geq x\}} X_n \mathrm{d}\mathbf{P},$$

proving the statement. □

We obtain a new proof for Kolmogorov's maximal inequality.

Example 2 (Kolmogorov's maximal inequality). Let $\xi, \xi_1, \ldots$ be independent random variables with $\mathbf{E}\xi_i = 0$, and $\mathbf{E}\xi_i^2 = \sigma_i^2 < \infty$. Then $X_n = \sum_{i=1}^n \xi_i$ is a martingale with respect to the natural filtration. Therefore $(X_n^2)_n$ is a submartingale and

$$\begin{aligned} \mathbf{P}\left(\max_{1 \leq k \leq n} |X_k| \geq x\right) &= \mathbf{P}\left(\max_{1 \leq k \leq n} X_k^2 \geq x^2\right) \\ &\leq x^{-2}\mathbf{E}X_n^2 = x^{-2} \sum_{i=1}^n \sigma_i^2. \end{aligned}$$

For an infinite sequence we obtain the following.

Corollary 3. *If* $(X_k, \mathcal{F}_k)$ *is a submartingale and* $x > 0$*, then*

$$\mathbf{P}(\sup_n X_n \geq x) \leq \frac{1}{x} \sup_n \mathbf{E} X_n^+.$$

Proof. Follows from the previous result combined with the monotone convergence theorem. □

Exercise 1. Prove the corollary.

For the L^p version we need a lemma.

Lemma 2. *Let* X, Y *be nonnegative random variables such that*

$$\mathbf{P}(X \geq x) \leq \frac{1}{x} \int_{\{X \geq x\}} Y \mathrm{d}\mathbf{P}, \quad x > 0.$$

Then for any $p > 1$

$$\mathbf{E} X^p \leq \left(\frac{p}{p-1} \right)^p \mathbf{E} Y^p.$$

Proof. Note the for a nonnegative random variable X

$$\mathbf{E} X^p = \int_0^\infty p x^{p-1} [1 - F(x)] \mathrm{d}x,$$

where $F(x) = \mathbf{P}(X \leq x)$ is the distribution function of X. Indeed,

$$\begin{aligned}\mathbf{E} X^p = \int_\Omega X^p \mathrm{d}\mathbf{P} &= \int_\Omega \int_0^\infty \mathbf{I}(x < X(\omega)) p x^{p-1} \mathrm{d}x \mathrm{d}\mathbf{P}(\omega) \\ &= \int_0^\infty p x^{p-1} [1 - F(x)] \mathrm{d}x.\end{aligned}$$

The result follows using Hölder's inequality as

$$\begin{aligned}
\mathbf{E}X^p &= \int_0^\infty p x^{p-1}[1-F(x)]\mathrm{d}x \\
&\le \int_0^\infty p x^{p-1} \frac{1}{x} \int_{\{X\ge x\}} Y(\omega)\mathrm{d}\mathbf{P}(\omega)\,\mathrm{d}x \\
&= \int_0^\infty \int_\Omega p x^{p-2} \mathbf{I}(X(\omega)\ge x) Y(\omega)\mathrm{d}\mathbf{P}(\omega)\mathrm{d}x \\
&= \int_\Omega Y(\omega) \left(\int_0^{X(\omega)} p x^{p-2}\mathrm{d}x \right) \mathrm{d}\mathbf{P}(\omega) \\
&= \int_\Omega Y X^{p-1} \frac{p}{p-1} \mathrm{d}\mathbf{P} \\
&\le \frac{p}{p-1} \left(\mathbf{E}Y^p\right)^{1/p} \left(\mathbf{E}X^{(p-1)q}\right)^{1/q} \\
&= \frac{p}{p-1} \left(\mathbf{E}Y^p\right)^{1/p} \left(\mathbf{E}X^p\right)^{1/q},
\end{aligned}$$

where p and q are conjugate exponents, i.e. $1/p + 1/q = 1$. $\square$

Theorem 9 (L^p maximal inequality)**.** *(i) Let $(X_k)_{k=1}^n$ be a nonnegative submartingale and $p \in (1,\infty)$. Then*

$$\mathbf{E}\max\{X_1^p,\dots,X_n^p\} \le \left(\frac{p}{p-1}\right)^p \mathbf{E}X_n^p.$$

(ii) Let $(X_k)_{k=1}^\infty$ be a nonnegative submartingale and $p \in (1,\infty)$. Then

$$\mathbf{E}\sup_{n\in\mathbb{N}} X_n^p \le \left(\frac{p}{p-1}\right)^p \sup_{n\in\mathbb{N}} \mathbf{E}X_n^p.$$

Proof. Statement (i) follows from Doob's maximal inequality and Lemma 2.

(ii) follows from (i) and the monotone convergence theorem as

$$\begin{aligned}
\mathbf{E}\sup_n X_n^p &= \lim_{n\to\infty} \mathbf{E} \max_{1\le k\le n} X_k^p \\
&\le \liminf_{n\to\infty} \left(\frac{p}{p-1}\right)^p \mathbf{E}X_n^p \\
&\le \left(\frac{p}{p-1}\right)^p \sup_n \mathbf{E}X_n^p.
\end{aligned}$$

$\square$

1.6 Optional stopping theorem

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability measure and $(\mathcal{F}_n)_n$ a filtration on it. Recall that a random variable $\tau : \Omega \to \mathbb{N}$ is *stopping time*, if $\{\tau \leq n\} \in \mathcal{F}_n$ for each $n \in \mathbb{N}$.

We already used the following simple observation.

Proposition 3. *The following are equivalent.*

(i) τ *is stopping time;*

(ii) $\{\tau > n\} \in \mathcal{F}_n$ *for each* $n \in \mathbb{N}$*;*

(iii) $\{\tau = n\} \in \mathcal{F}_n$ *for each* $n \in \mathbb{N}$*.*

Exercise 2. Prove this result.

Let τ be a stopping time. The *σ-algebra of the events prior to* τ, or short *pre-τ-sigma algebra* is defined as

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : \, A \cap \{\tau \leq n\} \in \mathcal{F}_n, \; n = 1, 2, \ldots\}. \tag{4}$$

It is easy to see that $\mathcal{F}_\tau$ is indeed a σ-algebra. Clearly, $\Omega \in \mathcal{F}_\tau$, and if $A \in \mathcal{F}_\tau$, then

$$A^c \cap \{\tau \leq n\} = (\Omega - A) \cap \{\tau \leq n\} = \{\tau \leq n\} - (A \cap \{\tau \leq n\}) \in \mathcal{F}_n, \; n \in \mathbb{N}.$$

Finally, if $A_1, A_2, \ldots \in \mathcal{F}_\tau$, then

$$(\cup_{k=1}^\infty A_k) \cap \{\tau \leq n\} = \cup_{k=1}^\infty (A_k \cap \{\tau \leq n\}) \in \mathcal{F}_n$$

for any $n = 1, 2, \ldots$.

Exercise 3. Show that if $\tau \equiv k$ for some $k \in \mathbb{N}$ then $\mathcal{F}_\tau = \mathcal{F}_k$, so the notation is consistent.

Some simple properties are summarized in the next statement.

Lemma 3. *Let* σ, τ *be stopping times.*

(i) τ *is* $\mathcal{F}_\tau$*-measurable.*

(ii) $\sigma \wedge \tau = \min(\sigma, \tau)$ *and* $\sigma \vee \tau = \max(\sigma, \tau)$ *are stopping times.*

(iii) If $\sigma \leq \tau$*, then* $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$*.*

(iv) If $(X_n)_n$ *is an adapted sequence then* X_τ *is* $\mathcal{F}_\tau$*-measurable.*

Theorem 10 (Optional stopping theorem, Doob). *Let $(X_n)_n$ be a supermartingale, and $\sigma \le \tau$ stopping times such that*

$$\mathbf{E}(|X_\sigma|) < \infty\,, \qquad \mathbf{E}(|X_\tau|) < \infty \tag{5}$$

and

$$\liminf_{n\to\infty} \int_{\{\tau>n\}} |X_n| \,\mathrm{d}\mathbf{P} = 0. \tag{6}$$

Then $\mathbf{E}(X_\tau|\mathcal{F}_\sigma) \le X_\sigma$ almost surely.

Furthermore, if $(X_n)_n$ is martingale then $\mathbf{E}(X_\tau|\mathcal{F}_\sigma) = X_\sigma$.

Clearly, conditions (5) and (6) hold if the stopping times are bounded.

Proof. Since X_σ is $\mathcal{F}_\sigma$-measurable, $X_\sigma = \mathbf{E}(X_\sigma|\mathcal{F}_\sigma)$, therefore it is enough to show that

$$\mathbf{E}(X_\tau - X_\sigma|\mathcal{F}_\sigma) \le 0.$$

This is the same as

$$\int_A (X_\tau - X_\sigma)\,\mathrm{d}\mathbf{P} \le 0 \ \text{ for all } A \in \mathcal{F}_\sigma. \tag{7}$$

First assume that τ is bounded, that is $\tau \le m$ for some m. For any $A \in \mathcal{F}_\sigma$

$$A \cap \{\sigma < k \le \tau\} = A \cap \{\sigma \le k-1\} \cap \{\tau > k-1\} \in \mathcal{F}_{k-1}\,, \quad k \ge 2,$$

thus

$$\begin{aligned}
&\int_A (X_\tau - X_\sigma)\,\mathrm{d}\mathbf{P}\\
&= \int_A \left(\sum_{k=\sigma+1}^{\tau} (X_k - X_{k-1}) \right) \mathrm{d}\mathbf{P}\\
&= \int_A \left(\sum_{k=2}^{m} \mathbf{I}(\sigma < k \le \tau)(X_k - X_{k-1}) \right) \mathrm{d}\mathbf{P}\\
&= \sum_{k=2}^{m} \int_{A\cap\{\sigma<k\le\tau\}} (X_k - X_{k-1})\,\mathrm{d}\mathbf{P}\\
&= \sum_{k=2}^{m} \int_{A\cap\{\sigma<k\le\tau\}} \mathbf{E}(X_k - X_{k-1}|\mathcal{F}_{k-1})\,\mathrm{d}\mathbf{P} \le 0,
\end{aligned}$$

proving (7).

Consider the general case. For any n we can write

$$\begin{aligned}
&\int_A (X_\tau - X_\sigma)\mathrm{d}\mathbf{P} \\
&= \int_A (X_{\tau\wedge n} - X_{\sigma\wedge n})\mathrm{d}\mathbf{P} + \int_A (X_\tau - X_{\tau\wedge n})\mathrm{d}\mathbf{P} - \int_A (X_\sigma - X_{\sigma\wedge n})\mathrm{d}\mathbf{P}.
\end{aligned}$$

On the event $\{\sigma \geq n\}$ we have $X_{\tau\wedge n} = X_n = X_{\sigma\wedge n}$, therefore

$$\int_A (X_{\tau\wedge n} - X_{\sigma\wedge n})\mathrm{d}\mathbf{P} = \int_{A\cap\{\sigma<n\}} (X_{\tau\wedge n} - X_{\sigma\wedge n})\mathrm{d}\mathbf{P} \leq 0\,, \quad n \in \mathbb{N}, \tag{8}$$

where the inequality follows from the previous case.

By condition (6) there exists a sequence $n_k \to \infty$ such that

$$\lim_{k\to\infty} \int_{\{\tau>n_k\}} |X_{n_k}|\,\mathrm{d}\mathbf{P} = 0.$$

It is enough to show that on this subsequence the second and third terms in decomposition (8) tends to 0. For the second term

$$\begin{aligned}
\left| \int_A (X_\tau - X_{\tau\wedge n_k})\mathrm{d}\mathbf{P} \right| &= \left| \int_{A\cap\{\tau>n_k\}} (X_\tau - X_{\tau\wedge n_k})\mathrm{d}\mathbf{P} \right| \\
&\leq \int_{A\cap\{\tau>n_k\}} (|X_\tau| + |X_{n_k}|)\mathrm{d}\mathbf{P} \\
&\leq \int_{\{\tau>n_k\}} |X_\tau|\,\mathrm{d}\mathbf{P} + \int_{\{\tau>n_k\}} |X_{n_k}|\,\mathrm{d}\mathbf{P}.
\end{aligned}$$

Similarly, for the third term

$$\begin{aligned}
\left| \int_A (X_\sigma - X_{\sigma\wedge n_k})\mathrm{d}\mathbf{P} \right| &= \left| \int_{A\cap\{\sigma>n_k\}} (X_\sigma - X_{n_k})\mathrm{d}\mathbf{P} \right| \\
&\leq \int_{\{\sigma>n_k\}} |X_\sigma|\,\mathrm{d}\mathbf{P} + \int_{\{\tau>n_k\}} |X_{n_k}|\,\mathrm{d}\mathbf{P}.
\end{aligned}$$

Using (5) both upper bounds tend to 0. □

Corollary 4. *Assume that (X_n) is (super-, sub-) martingale, τ is a stopping time, $\mathbf{E}(|X_\tau|) < \infty$ and (6) holds. Then* *(i)* $\mathbf{E}(X_\tau|\mathcal{F}_1) \le X_1$ *and* $\mathbf{E}(X_\tau) \le \mathbf{E}(X_1)$ *for supermartingales;*

(ii) $\mathbf{E}(X_\tau|\mathcal{F}_1) \ge X_1$ *and* $\mathbf{E}(X_\tau) \ge \mathbf{E}(X_1)$ *for submartingales;*

(iii) $\mathbf{E}(X_\tau|\mathcal{F}_1) = X_1$ *and* $\mathbf{E}(X_\tau) = \mathbf{E}(X_1)$ *for martingales.*

Some conditions are needed for the optional stopping to hold.

Example 3 (Simple symmetric random walk)**.** Let $\xi, \xi_1, \xi_2, \dots$ are iid random variables with $\mathbf{P}(\xi = \pm 1) = 1/2$. Let $S_0 = 1$ and $S_n = S_{n-1} + \xi_n$. Then (S_n) is martingale. Let $\tau = \min\{n : S_n = 0\}$. Then τ is a stopping time and the martingale $(S_{\tau\wedge n})_n$ tends to 0 a.s. The optional stopping does not hold as $S_\tau \equiv 0$ a.s., while $S_0 = 1$. Clearly, condition (6) does not hold.

Theorem 11 (Wald identity)**.** *Let $X, X_1, X_2, \dots$ be iid random variables with $\mathbf{E}X = \mu \in \mathbb{R}$, and let τ be a stopping time with $\mathbf{E}(\tau) < \infty$. Put $S_n = X_1 + \cdots + X_n$, $n \in \mathbb{N}$. Then $\mathbf{E}(S_\tau) = \mu\mathbf{E}(\tau)$.*

Proof. First assume $X \ge 0$. We have

$$\begin{aligned}
\mathbf{E}(S_\tau) &= \mathbf{E}\left(\sum_{k=1}^{\infty} \mathbf{I}(\tau \ge k) X_k\right) \\
&= \sum_{k=1}^{\infty} \mathbf{E}(\mathbf{I}(\tau \ge k) X_k) \\
&= \sum_{k=1}^{\infty} \mathbf{E}\mathbf{I}(\tau \ge k)\mathbf{E}(X_k) \\
&= \mu \sum_{k=1}^{\infty} \mathbf{P}(\tau \ge k) \\
&= \mu\mathbf{E}(\tau).
\end{aligned}$$

To see the general case consider the decomposition $S_\tau = S_\tau^{(+)} - S_\tau^{(-)}$ where

$$S_\tau^{(+)} = \sum_{k=1}^{\infty} X_k^+ \mathbf{I}(\tau \ge k)$$

and

$$S_\tau^{(-)} = \sum_{k=1}^{\infty} X_k^- \mathbf{I}(\tau \ge k).$$

$\square$

As a simple application of the optional stopping problem we consider the gambler's ruin problem. There is an elementary but longer way to derive these formulas.

Example 4 (Gambler's ruin)**.** Let $X, X_1, X_2, \ldots$ be iid random variables such that $\mathbf{P}(X = 1) = p = 1 - \mathbf{P}(X = -1)$, $0 < p < 1$, and put $S_n = X_1 + \cdots + X_n$, $n \in \mathbb{N}$. Fix $a, b \in \mathbb{N}$ and let

$$\tau = \tau_{a,b}(p) = \inf\{n : S_n \geq b \text{ or } S_n \leq -a\},$$

with the convention $\inf \emptyset = \infty$. Let $(\mathcal{F}_n)$ be the natural filtration, i.e. $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, $n \in \mathbb{N}$.

It is easy to show that $\mathbf{P}(\tau < \infty) = 1$, and τ is a stopping time. Furthermore, $|S_\tau| \leq \max(a, b)$, in particular $\mathbf{E}|S_\tau| < \infty$ and

$$\liminf_{n \to \infty} \int_{\{\tau > n\}} |S_n| \, \mathrm{d}\mathbf{P} \leq \liminf_{n \to \infty} \max(a, b) \mathbf{P}(\tau > n) = 0.$$

First assume that $p = 1/2$. Then $\mathbf{E}X = 0$ and (S_n) is a martingale. Therefore, by the optional stopping theorem

$$\begin{aligned} 0 = \mathbf{E}S_0 = \mathbf{E}S_\tau &= -a\mathbf{P}(S_\tau = -a) + b\mathbf{P}(S_\tau = b) \\ &= -a(1 - \mathbf{P}(S_\tau = b)) + b\mathbf{P}(S_\tau = b). \end{aligned}$$

Thus

$$\mathbf{P}(S_\tau = b) = \frac{a}{a+b} \quad \text{and} \quad \mathbf{P}(S_\tau = -a) = \frac{b}{a+b}.$$

Using that $(S_n^2 - n)$ is a martingale, we can determine $\mathbf{E}\tau$. Since

$$0 = \mathbf{E}(S_0^2 - 0) = \mathbf{E}(S_\tau^2 - \tau)$$

we obtain

$$\mathbf{E}\tau = \mathbf{E}S_\tau^2 = a^2\mathbf{P}(S_\tau = -a) + b^2\mathbf{P}(S_\tau = b) = a^2\frac{b}{a+b} + b^2\frac{a}{a+b} = ab.$$

The case $p \neq 1/2$ is different. Introduce

$$Z_n = s^{S_n} = \prod_{k=1}^n s^{X_k}$$

with $s = (1-p)/p = 1/r$. Then (Z_n) is a martingale and

$$Z_\tau = s^b \mathbf{I}(S_n = b) + s^{-a}\mathbf{I}(S_n = -a) \le s^b + s^{-a},$$

thus $\mathbf{E}Z_\tau < \infty$ and

$$\liminf_{n\to\infty} \int_{\{\tau>n\}} |Z_n| \,\mathrm{d}\mathbf{P} \le (s^b + s^{-a}) \liminf_{n\to\infty} \mathbf{P}\{\tau > n\} = 0.$$

Again, by the optional sampling theorem

$$\begin{aligned}
&s^{-a}\mathbf{P}(S_\tau = -a) + s^b\left(1 - \mathbf{P}(S_\tau = -a)\right) \\
&= s^{-a}\mathbf{P}(S_\tau = -a) + s^b\mathbf{P}(S_\tau = b) \\
&= \mathbf{E}(s^{S_\tau}) = \mathbf{E}(Z_\tau) \\
&= \mathbf{E}(Z_1) = \mathbf{E}(s^X) = 1.
\end{aligned}$$

Rearranging we obtain

$$\mathbf{P}(S_\tau = -a) = \frac{1-s^b}{s^{-a}-s^b}\,\frac{r^b}{r^b} = \frac{r^b - 1}{r^{a+b}-1} = \frac{1-r^b}{1-r^{a+b}}.$$

To obtain $\mathbf{E}\tau$, using the Wald identity

$$\mathbf{E}S_\tau = (2p-1)\mathbf{E}\tau,$$

from which

$$\mathbf{E}\tau = \frac{1}{2p-1}\mathbf{E}S_\tau = \frac{1}{2p-1}\left[-a\mathbf{P}(S_\tau = -a) + b\mathbf{P}(S_\tau = b)\right].$$

Exercise 4. Show that $\tau < \infty$ a.s.

2 Continuous time martingales

2.1 Definitions and simple properties

This part is from Karatzas and Shreve [5].

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and $(\mathcal{F}_t)_{t\ge 0}$ a *filtration*, i.e. an increasing sequence of σ-algebras. The time horizon is either finite or infinite, $t \in [0,T]$ or $t \in [0,\infty)$.

In what follows we *always* assume that the filtration satisfies the *usual conditions*:

(i) $\mathcal{F}_0$ contains the $\mathbf{P}$-null sets;

(ii) $(\mathcal{F}_t)_t$ is right-continuous, i.e. $\cap_{s>t}\mathcal{F}_s =: \mathcal{F}_{t+} = \mathcal{F}_t$.

Let (X_t) and (Y_t) be stochastic processes. The process Y is a *modification* of X if $X_t = Y_t$ a.s. for any fix t, i.e. $\mathbf{P}(X_t = Y_t) = 1$ for each $t \geq 0$. The processes X and Y are *indistinguishable* if their sample path are the same almost surely, i.e.

$$\mathbf{P}(X_t = Y_t, \ t \geq 0) = 1.$$

They have the *same finite dimensional distributions* if for all $0 \leq t_1 < t_2 < \ldots < t_n < \infty$ and $A \in \mathcal{B}(\mathbb{R}^n)$

$$\mathbf{P}\left((X_{t_1}, \ldots, X_{t_n}) \in A\right) = \mathbf{P}\left((Y_{t_1}, \ldots, Y_{t_n}) \in A\right).$$

Example 5. Let U be uniform$(0,1)$, and $X_t \equiv 0$, $t \in [0,1]$, and $Y_t = \mathbf{I}(U = t)$. Then Y is a modification of X, but they are not indistinguishable, since

$$\mathbf{P}(X_t = Y_t, \ t \geq 0) = 0.$$

The process $(X_t)_t$ is *adapted to the filtration* $(\mathcal{F}_t)_t$, if X_t is $\mathcal{F}_t$-measurable for each $t \geq 0$. The process $(X_t, \mathcal{F}_t)_t$ is a *martingale* if

(i) $(X_t)_t$ is adapted to $(\mathcal{F}_t)_t$;

(ii) $\mathbf{E}|X_t| < \infty$ for all $t \geq 0$;

(iii) $\mathbf{E}[X_t|\mathcal{F}_s] = X_s$ a.s. for all $t \geq s$.

It is sub- or supermartingale if (i) and (ii) holds, and (iii) holds with $\geq$ or $\leq$ instead of $=$.

A random variable $\tau : \Omega \to [0, \infty)$ is a *stopping time* if $\{\tau \leq t\} \in \mathcal{F}_t$. The σ-algebra of the events prior to τ, or pre-τ-σ-algebra is

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\}.$$

Exercise 5. Show that $\mathcal{F}_\tau$ is indeed a σ-algebra.

The next result is obvious, but very useful.

Proposition 4. *Let $(X_t, \mathcal{F}_t)$ be a (sub-, super-) martingale. Then for any sequence $0 \leq t_0 < t_1 < \ldots < t_N < \infty$ the process $(X_{t_n}, \mathcal{F}_{t_n})_{n=0}^N$ is a discrete time martingale.*

Lemma 4. *Let σ, τ be stopping times.*

(i) *τ is $\mathcal{F}_\tau$-measurable.*

(ii) *If $\tau \equiv t$ then $\mathcal{F}_\tau = \mathcal{F}_t$.*

(iii) *$\sigma \wedge \tau = \min(\sigma, \tau)$ and $\sigma \vee \tau = \max(\sigma, \tau)$ are stopping times.*

(iv) *If $\sigma \leq \tau$, then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.*

(v) *If $(X_t)_t$ is right-continuous and adapted then X_τ is $\mathcal{F}_\tau$-measurable.*

Exercise 6. Prove the lemma.

Remark 1. In continuous time the technical details are trickier.

The process $(X_t)_t$ is adapted to $(\mathcal{F}_t)_t$, if X_t is $\mathcal{F}_t$-measurable, and it is *progressively measurable* with respect to $(\mathcal{F}_t)_t$, if for all $t \geq 0$ and $A \in \mathcal{B}(\mathbb{R}^d)$

$$\{(s, \omega) : \ s \leq t, \ X_s(\omega) \in A\} \in \mathcal{B}([0, t]) \otimes \mathcal{F}_t,$$

where $\mathcal{B}$ stands for the Borel sets, and $\otimes$ is the product σ-algebra. In what follows we always need progressive measurability, adaptedness is not enough.

The next statement says that the situation is not too bad.

Proposition 5. *If $(X_t)_t$ is right continuous and adapted, then it is progressively measurable.*

Example 6 (Poisson process)**.** A Poisson process with intensity $\lambda > 0$ is an adapted integer valued RCLL (right continuous with left limits) process $N = (N_t, \mathcal{F}_t)_{t \geq 0}$ such that

(i) N has independent increments, that is $N_t - N_s$ is independent of $\mathcal{F}_s$ for any $s < t$,

(ii) $N_0 = 0$ a.s.,

(iii) $N_t - N_s \sim \text{Poisson}(\lambda(t - s))$.

Exercise 7. Show that $(N_t - \lambda t)$ is martingale.

Proposition 6. *Let (X_t) be a martingale, and φ a convex function such that $\mathbf{E}|\varphi(X_t)| < \infty$ for all $t \geq 0$. Then $(\varphi(X_t))$ is submartingale.*

Furthermore if (X_t) is a submartingale and φ nondecreasing and convex that $\mathbf{E}|\varphi(X_t)| < \infty$ for all $t \geq 0$, then $(\varphi(X_t))$ is a submartingale.

Example 7 (Wiener process)**.** The Wiener process or standard Brownian motion is an adapted process $W = (W_t, \mathcal{F}_t)_{t \geq 0}$ such that

(i) W has independent increments, that is $W_t - W_s$ is independent of $\mathcal{F}_s$ for any $s < t$,

(ii) $W_0 = 0$ a.s.,

(iii) $W_t - W_s \sim \mathrm{N}(0, t-s)$,

(iv) W_t has continuous sample path.

Exercise 8. Show that (W_t) and $(W_t^2 - t)$ are martingales.

2.2 Martingale convergence theorem

Consider an adapted stochastic process $(X_t)_{t\geq 0}$. Fix $a < b$, and a finite set $F \subset [0, \infty)$. Let U_F denote the number of *upcrossings* of the interval $[a, b]$ by the restricted process $(X_t)_{t\in F}$. Formally, let $\tau_0 = 0$, and

$$\begin{aligned} \tau_{2k-1} &= \min\{t \in F : \ t \geq \tau_{2k-2}, X_t < a\}, \\ \tau_{2k} &= \min\{t \in F : \ t \geq \tau_{2k-1}, X_t > b\}. \end{aligned}$$

The number of upcrossings on F is

$$U_F(a, b) = U_F = \max\{k : \tau_{2k} < \infty\}.$$

We can extend the definition of infinite sets $I \subset [0, \infty)$ as

$$U_I = \sup\{U_F : \ F \subset I, \ F \text{ finite}\}.$$

We have the upcrossing inequality.

Theorem 12 (Upcrossing inequality)**.** *Let (X_t) be a right-continuous submartingale. For any $a < b$ and $0 \leq S \leq T < \infty$*

$$(b-a)\mathbf{E}U_{[S,T]} \leq \mathbf{E}(X_T - a)^+ - \mathbf{E}(X_S - a)^+.$$

Proof. Consider an enumeration of the countable set $\mathbf{Q} \cap [S, T]$ as

$$\mathbf{Q} \cap [S, T] = \{q_1, q_2, \ldots\},$$

and let $F_n = \{q_1, \ldots, q_n\} \cup \{s, t\}$. Then $(X_t, \mathcal{F}_t)_{t\in F_n}$ is a discrete time submartingale, therefore, by the upcrossing inequality

$$(b-a)\mathbf{E}U_{F_n} \leq \mathbf{E}(X_T - a)^+ - \mathbf{E}(X_S - a)^+.$$

Since F_n is increasing, U_{F_n} is increasing, and by the right-continuity of (X_t)

$$\lim_{n\to\infty} U_{F_n} = U_{[S,T]} \quad \text{a.s.}$$

In particular, $U_{[S,T]}$ is measurable, and by the monotone convergence theorem the result follows. $\square$

Theorem 13 (Martingale convergence theorem)**.** *Let (X_t) be a right-continuous submartingale such that*

$$\sup_{t\geq 0} \mathbf{E}(X_t^+) < \infty.$$

Then $\lim_{t\to\infty} X_t = X$ exists a.s. and $\mathbf{E}|X| < \infty$.

Proof. By the upcrossing inequality and the monotone convergence theorem for any $a < b$

$$\mathbf{E}U_{[0,\infty)}(a,b) \leq \frac{\sup_{t\geq 0} \mathbf{E}X_t^+ + |a|}{b-a}.$$

Therefore, for any $a < b$ the upcrossings $U_{[0,\infty)}(a,b)$ are a.s. finite. Thus almost surely the upcrossings are finite for all $a < b$ rationals, implying the existence of the limit.

The integrability of the limit follows from Fatou's lemma. $\square$

Exercise 9. Let (X_t) be a right-continuous nonnegative submartingale. Show that the following are equivalent:

(i) (X_t) is uniformly integrable;

(ii) converges in L^1;

(iii) converges a.s. to an integrable random variable X_∞, such that $(X_t)_{t\in[0,\infty]}$ is a submartingale.

2.3 Inequalities

Theorem 14 (Doob's maximal inequality)**.** *Let (X_t) be a right-continuous submartingale.*

(i) For any $0 < S < T < \infty$, $x > 0$

$$x\mathbf{P}(\sup_{S\leq t\leq T} X_t \geq x) \leq \mathbf{E}X_T^+.$$

(ii) If (X_t) is nonnegative and $p > 1$ then

$$\mathbf{E}\left(\sup_{S\le t\le T} X_t\right)^p \le \left(\frac{p}{p-1}\right)^p \mathbf{E}X_T^p.$$

Proof. (i): Let F_n be as above. Then $(X_t, \mathcal{F}_t)_{t\in F_n}$ is a discrete time martingale. Therefore, by Doob's maximal inequality

$$y\mathbf{P}\left(\sup_{t\in F_n} X_t > y\right) \le \mathbf{E}X_T^+.$$

Right-continuity implies

$$\left\{\sup_{S\le t\le T} X_t > y\right\} = \cup_{n=1}^{\infty}\left\{\sup_{t\in F_n} X_t > y\right\},$$

and the union is increasing. Letting $n \to \infty$

$$y\mathbf{P}\left(\sup_{S\le t\le T} X_t > y\right) \le \mathbf{E}X_T^+.$$

Letting $y \uparrow x$ the result follows.

Part (ii) follows as in the discrete time case. □

Exercise 10. Let N be a Poisson process with intensity $\lambda > 0$. Show that for any $c > 0$

$$\limsup_{t\to\infty} \mathbf{P}\left(\sup_{0\le s\le t}(N_s - \lambda s) \ge c\sqrt{\lambda t}\right) \le \frac{1}{c\sqrt{2\pi}},$$

and

$$\limsup_{t\to\infty} \mathbf{P}\left(\inf_{0\le s\le t}(N_s - \lambda s) \le -c\sqrt{\lambda t}\right) \le \frac{1}{c\sqrt{2\pi}}.$$

Show that for any $0 < S < T < \infty$

$$\mathbf{E}\sup_{S\le t\le T}\left(\frac{N_t}{t} - \lambda\right)^2 \le \frac{4T\lambda}{S^2}.$$

Corollary 5. *Let N be a Poisson process with intensity $\lambda > 0$. Then*

$$\lim_{t\to\infty}\frac{N_t}{t} = \lambda \quad a.s.$$

Proof. By Chebyshev's inequality

$$\mathbf{P}\left(\left|t^{-1}N_t - \lambda\right| > \varepsilon\right) \leq \frac{\mathbf{Var}(N_t)}{t^2\varepsilon^2} = \frac{\lambda}{\varepsilon^2 t}.$$

By the first Borel–Cantelli-lemma almost surely

$$\lim_{n\to\infty} \frac{N_{2^n}}{2^n} = \lambda.$$

So on a subsequence we are done. In between we have

$$\begin{aligned}\mathbf{P}\left(\sup_{2^n\leq t\leq 2^{n+1}} \left|t^{-1}N_t - \lambda\right| > \varepsilon\right) &\leq \frac{\mathbf{E}\left(\sup_{2^n\leq t\leq 2^{n+1}} |t^{-1}N_t - \lambda|\right)^2}{\varepsilon^2} \\ &\leq \frac{4\,2^{n+1}\lambda}{2^{2n}\varepsilon^2} = 2^{-n}\frac{8\lambda}{\varepsilon^2}.\end{aligned}$$

Applying Borel–Cantelli again, we are done. □

2.4 Optional stopping

Let $(X_t, \mathcal{F}_t)_{t\in[0,\infty)}$ be a right-continuous submartingale. It has a *last element* X_∞, if X_∞ is measurable with respect to the σ-algebra $\mathcal{F}_\infty = \sigma\left(\cup_{t\geq 0}\mathcal{F}_t\right)$, $\mathbf{E}|X_\infty| < \infty$ and for all $t \geq 0$ $\mathbf{E}[X_\infty|\mathcal{F}_t] \geq X_t$ a.s.

If we work on the finite time horizon $[0,T]$, $T < \infty$, then the submartingale $(X_t)_{t\in[0,T]}$ has a last element X_T (by definition!).

Theorem 15 (Optional stopping)**.** *Let $(X_t, \mathcal{F}_t)_{t\geq 0}$ be a right-continuous submartingale with last element X_∞. Let $\sigma \leq \tau$ be stopping times. Then*

$$\mathbf{E}[X_\tau|\mathcal{F}_\sigma] \geq X_\sigma \quad a.s.$$

Proof. Assume that τ is bounded, i.e. $\tau \leq K$. Let

$$\sigma_n(\omega) = k/2^n, \quad \text{if } \sigma(\omega) \in [(k-1)/2^n, k/2^n),$$

and define τ_n similarly. Then σ_n and τ_n are stopping times, and $\sigma_n \leq \tau_n$. We can apply the optional stopping theorem for the submartingale $(X_{k/2^n}, \mathcal{F}_{k/2^n})$, and stopping times σ_n, τ_n. Then

$$\mathbf{E}[X_{\tau_n}|\mathcal{F}_{\sigma_n}] \geq X_{\sigma_n},$$

that is for $A \in \mathcal{F}_{\sigma_n}$

$$\int_A X_{\tau_n} \mathrm{d}\mathbf{P} \geq \int_A X_{\sigma_n} \mathrm{d}\mathbf{P}.$$

Since $\sigma_n \geq \sigma$ for each n, $\mathcal{F}_{\sigma_n} \supset \mathcal{F}_\sigma$. Therefore, for $A \in \mathcal{F}_\sigma$

$$\int_A X_{\tau_n} \mathrm{d}\mathbf{P} \geq \int_A X_{\sigma_n} \mathrm{d}\mathbf{P}.$$

By the right-continuity $X_{\tau_n} \to X_\tau$ and $X_{\sigma_n} \to X_\sigma$ a.s. This combined with the uniform integrability implies

$$\int_A X_\tau \mathrm{d}\mathbf{P} \geq \int_A X_\sigma \mathrm{d}\mathbf{P},$$

proving the result. □

Exercise 11. Prove that σ_n, τ_n are indeed stopping times.

2.5 Doob-Meyer decomposition

The Doob-Meyer decomposition is the continuous time analogue of the Doob's decomposition of submartingales. While the latter is basically trivial, the Doob-Meyer decomposition is highly nontrivial, and needs further assumptions.

Recall that a class $\mathcal{D}$ of random variables are *uniformly integrable*, if for any $\varepsilon > 0$ there exists $K > 0$ such that for all $X \in \mathcal{D}$

$$\int_{|X|>K} |X| \mathrm{d}\mathbf{P} < \varepsilon.$$

Put

$$\mathcal{S}_a = \{\tau : \tau \text{ stopping time }, \tau \leq a\}.$$

The adapted process (X_t) belongs to the class DL is for any $a > 0$ the class $\{X_\tau\}_{\tau \in \mathcal{S}_a}$ of random variables is uniformly integrable.

Theorem 16 (Doob-Meyer decomposition)**.** *Let the filtration $\mathcal{F}_t$ satisfy the usual conditions, and let $(X_t)_t$ be a right-continuous submartingale in DL. Then there exist (M_t) and (A_t) such that (M_t) is a martingale, (A_t) is an adapted nondecreasing right-continuous process with $A_0 \equiv 0$, and*

$$X_t = M_t + A_t, \quad t \geq 0.$$

Furthermore, the decomposition is unique.

Example 8. If (N_t) is a Poisson process with intensity $\lambda > 0$, then it is a submartingale. Its Doob-Meyer decomposition is

$$N_t = (N_t - \lambda t) + \lambda t.$$

If (W_t) is a standard Brownian motion, then (W_t^2) is a submartingale and its Doob-Meyer decomposition is

$$W_t^2 = (W_t^2 - t) + t.$$

3 Wiener process

This part is from Karatzas and Shreve [5].

3.1 First properties and existence

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space. Then $W = (W_t, \mathcal{F}_t)_{t \geq 0}$ is a *Wiener process* or *standard Brownian motion* if

(W1) $W_0 = 0$ a.s.,

(W2) W has independent increments, that is $W_t - W_s$ is independent of $\mathcal{F}_s$ for any $s < t$,

(W3) $W_t - W_s \sim \mathrm{N}(0, t - s)$,

(W4) W_t has continuous sample path.

Exercise 12. Show that (W2) and (W3) with $s = 0$ (i.e. $W_t \sim \mathrm{N}(0, t)$) implies (W3).

Proposition 7. *(i)* $\mathbf{E}(W_t) = 0$ *for all* t.

(ii) $\mathbf{Cov}(W_s, W_t) = \mathbf{E}(W_s W_t) = \min(s, t) =: s \wedge t$, $s, t \geq 0$.

(iii) For any $k \in \mathbb{N}$ *and* $0 \leq t_1 < \cdots < t_k$, *the random vector* $(W_{t_1}, \ldots, W_{t_k})$ *has a multivariate normal distribution with mean 0 and covariance*

$$\Sigma = \Sigma_{t_1, \ldots, t_k} = \begin{pmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_k \end{pmatrix}.$$

Proof. Part (i) and (ii) are trivial. For part (iii) note that by the independent increment property the components of

$$X = (W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_k} - W_{t_{k-1}})^\top$$

are independent normal random variables. Therefore X is a multivariate normal. Since

$$(W_{t_1}, \dots, W_{t_k})^\top = AX,$$

the statement follows from the fact that a linear transformation of a multivariate normal is normal with covariance matrix $A\mathbf{Cov}(X)A^\top$. $\square$

Let (X_t) be a stochastic process with finite second moment. Then $m(t) = \mathbf{E}X_t$ is the mean value and $r(s,t) = \mathbf{Cov}(X_s, X_t) = \mathbf{E}([X_s - m(s)][X_t - m(t)])$, is the covariance function.

Clearly r is symmetric, and nonnegative definite, i.e.

$$\sum_{j=1}^{k}\sum_{\ell=1}^{k} c_j c_\ell \, r(t_j, t_\ell) \geq 0\,, \quad k \in \mathbb{N}\,,\ t_1, \dots, t_k \in T\,,\ c_1, \dots, c_k \in \mathbb{R}\,.$$

Definition 1. The stochastic process (X_t) is a Gaussian process with mean function $m(t)$ and covariance function $r(t,s)$ if for any $k \in \mathbb{N}$ and $t_1, \dots, t_k$ the random vector $(X_{t_1}, \dots, X_{t_k})$ has multivariate normal distribution with mean $(m(t_1), \dots, m(t_k))$ and covariance $(r(t_j, t_\ell))_{j,\ell=1}^k$.

A simple, but not very interesting example to a Gaussian process is $X_t = a(t)Z + b(t)$, where $Z \sim N(0,1)$.

We proved that the Wiener process (W_t) is a Gaussian process with mean $m(t) \equiv 0$ and covariance function $r(s,t) = \min(s,t)$. This could be the definition of the Wiener process.

Proposition 8. *Let (W_t) be a continuous Gaussian process with mean 0 and covariance function $r(s,t) = \min(s,t)$. Then (W_t) is a Wiener process.*

Exercise 13. Prove the statement.

Exercise 14. Let $(W(t))$ be SBM. Show that

(i) $W_1(t) = W(c+t) - W(c)$, $t \geq 0$;

(ii) $W_2(t) = \sqrt{c}\, W(t/c)$, $t \geq 0$;

(iii) $W_3(t) = tW(1/t)$

are SBM.

Kolmogorov's consistency theorem yields the the existence of Gaussian processes.

Theorem 17. *Let $\mathbb{T} \subset \mathbb{R}$, and let $m(t)$ be an arbitrary function and $r(s,t)$ a nonnegative definite function. Then there exists a Gaussian process $(X_t)_{t\in\mathbb{T}}$ with mean function m and covariance function r.*

Therefore, *apart from continuity*, we have a Wiener process. That is, we have a probability space $(\mathbb{R}^{[0,\infty)}, \mathcal{B}^{[0,\infty)}, \mathbf{P})$ and a stochastic process $(\widetilde{W}_t(\omega) = \omega_t)_{t\geq 0}$, which satisfies (W1)–(W3).

Let $C = C[0,\infty)$ be the space of continuous function on $[0,\infty)$. We have to show that $\mathbf{P}(\widetilde{W} \in C) = 1$. *The problem is that C does not belong to the product σ-algebra $\mathcal{B}^{[0,\infty)}$.* Indeed, it can be shown that

$$\mathcal{B}^{[0,\infty)} = \cup\{\pi_K^{-1}(\mathcal{B}^K) : K \subset [0,\infty), K \text{ countable}\}.$$

Therefore, if $C \in \mathcal{B}^{[0,\infty)}$, then $C = \pi_K^{-1}(\mathcal{B}^K)$ for some countable $K \subset [0,\infty)$. But continuity cannot be determined from the values of the function on a countable set. Similarly,

$$\left\{\omega \in \mathbb{R}^{[0,\infty)} : \sup_{0\leq t\leq 1} \omega_t \leq x\right\}, \quad x \in \mathbb{R},$$

is not $\mathcal{B}^{[0,\infty)}$-measurable, so we cannot define $\sup_{t\in[0,1]} \widetilde{W}_t$.

Thus the setup in Kolmogorov's consistency theorem cannot deal with continuous processes. We need a different approach.

Recall that Y is a *modification* of X if $X_t = Y_t$ a.s. for any fix t, i.e. $\mathbf{P}(X_t = Y_t) = 1$ for each $t \geq 0$.

Theorem 18 (Kolmogorov continuity theorem)**.** *Let $(X_t)_{t\in[0,T]}$ be a stochastic process on $(\Omega, \mathcal{A}, \mathbf{P})$, such that for some positive constants α, β, C*

$$\mathbf{E}|X_t - X_s|^\alpha \leq C|t-s|^{1+\beta}, \quad 0 \leq s,t \leq T.$$

Then X has a continuous modification $\widetilde{X}$ which is Hölder continuous with exponent γ for every $\gamma \in (0, \beta/\alpha)$, that is for some $h(\omega)$ a.s. positive random variable and $\delta > 0$

$$\mathbf{P}\left(\omega : \sup_{0<t-s<h(\omega)} \frac{\widetilde{X}_t(\omega) - \widetilde{X}_s(\omega)}{|t-s|^\gamma} \leq \delta\right) = 1.$$

Proof. We can assume that $T = 1$. By Chebyshev

$$\mathbf{P}(|X_t - X_s| > \varepsilon) \leq \varepsilon^{-\alpha}\mathbf{E}|X_t - X_s|^\alpha \leq C\varepsilon^{-\alpha}|t-s|^{1+\beta},$$

in particular $X_t \to X_s$ in probability as $t \to s$. Fix $\gamma \in (0, \beta/\alpha)$. Then

$$\begin{aligned}
&\mathbf{P}\left(\max_{1 \leq k \leq 2^n} |X_{k2^{-n}} - X_{(k-1)2^{-n}}| > 2^{-\gamma n}\right) \\
&\leq 2^n \mathbf{P}\left(|X_{k2^{-n}} - X_{(k-1)2^{-n}}| > 2^{-\gamma n}\right) \\
&\leq 2^n \, C 2^{-\alpha\gamma n} 2^{-n(1+\beta)} \\
&= C 2^{-n(\beta - \alpha\gamma)}.
\end{aligned}$$

By the first Borel–Cantelli lemma with probability 1 only finitely many of the events

$$\max_{1 \leq k \leq 2^n} |X_{k2^{-n}} - X_{(k-1)2^{-n}}| > 2^{-\gamma n}$$

occur. That is, there is a set Ω_0 with $\mathbf{P}(\Omega_0) = 1$, and a threshold $n_0(\omega)$ (depending on ω!) such that for $\omega \in \Omega_0$

$$\max_{1 \leq k \leq 2^n} |X_{k2^{-n}} - X_{(k-1)2^{-n}}| \leq 2^{-\gamma n}, \quad n \geq n_0(\omega).$$

Fix $\omega \in \Omega_0$. Put $D_n = \{k2^{-n} : k = 0, 1, \ldots, 2^n\}$, and $D = \cup_n D_n$. Then for $n \geq n_0(\omega)$ and $m > n$ induction gives that

$$|X_t(\omega) - X_s(\omega)| \leq 2 \sum_{j=n+1}^{m} 2^{-\gamma j}, \quad t, s \in D_m, \ |t - s| \leq 2^{-n}.$$

This implies that $(X_t(\omega))_{t \in D}$ is uniformly continuous in $t \in D$. Indeed, for any $t, s \in D$ with $0 < t - s < h(\omega) = 2^{-n_0(\omega)}$ there is an $n \geq n_0$ such that $2^{-n-1} \leq t - s < 2^{-n}$, thus

$$|X_t(\omega) - X_s(\omega)| \leq 2 \sum_{j=n+1}^{\infty} 2^{-\gamma j} = 2^{-\gamma(n+1)} \frac{2}{1 - 2^{-\gamma}} \leq |t - s|^\gamma \frac{2}{1 - 2^{-\gamma}}.$$

Informally, we proved that (X_t) behaves regularly on D. We define $\widetilde{X}$. If $\omega \notin \Omega_0$ let $\widetilde{X}(\omega) = 0$, (or anything). If $\omega \in \Omega_0$ and $t \in D$ let $\widetilde{X}_t(\omega) = X_t(\omega)$, while if $t \notin D$ choose a sequence $s_n \in D$ such that $s_n \to t$ and let

$$\widetilde{X}_t(\omega) = \lim_{n \to \infty} X_{s_n}(\omega).$$

By the uniform continuity and the Cauchy criteria the limit on the right-hand side exist.

The a.s. uniqueness of the stochastic limit together with the stochastic continuity of X implies that $\widetilde{X}$ is a modification of X. $\square$

Exercise 15 (Random fields)**.** A *random field* is a collection of random variables indexed by a partially ordered set. Let $(X_t)_{t\in[0,T]^d}$ be a random field satisfying

$$\mathbf{E}|X_t - X_s|^\alpha \leq C\|t-s\|^{d+\beta},$$

for some positive constants. Show that there exists a continuous modification $\widetilde{X}$ which is Hölder continuous with exponent γ for every $\gamma \in (0, \beta/\alpha)$, that is for some $h(\omega)$ a.s. positive random variable and $\delta > 0$

$$\mathbf{P}\left(\omega : \sup_{0<\|t-s\|<h(\omega)} \frac{\widetilde{X}_t(\omega) - \widetilde{X}_s(\omega)}{\|t-s\|^\gamma} \leq \delta\right) = 1.$$

Exercise 16. Show that if $W_t - W_s \sim N(0, t-s)$ then for any $n > 0$

$$\mathbf{E}|W_t - W_s|^{2n} = C_n|t-s|^n,$$

where $C_n = \mathbf{E}|Z|^n$, $Z \sim N(0,1)$.

Corollary 6. *Wiener process exists.*

Proof. We need only the continuity part. The condition of Kolmogorov continuity theorem holds with $\alpha = 2n$ and $\beta = n-1$ for any $n > 1$. Thus there exists a continuous modification on $[0, N]$, for any $N \in \mathbb{N}$. Necessarily, X^{N_1} and X^{N_2} agrees a.s. for any fix $t \in [0, N_1 \wedge N_2]$, which allows us to extend the process to $[0, \infty)$. $\square$

In fact, we proved that the Wiener process is locally γ-Hölder continuous for any $\gamma < 1/2$.

Exercise 17. Let (N_t) be a Poisson process with intensity 1. Compute the order $\mathbf{E}|N_t - N_s|^\alpha$ for $t-s$ small. (Thus the condition in the continuity theorem holds for $\beta = 0$. Well, of course, Poisson processes are not continuous.)

More generally, we obtain a result on continuity of Gaussian processes.

Theorem 19. *Let (X_t) be a Gaussian process with continuous mean function m, and covariance function r. If there exist positive constants δ, C such that for all s,t*

$$r(t,t) - 2r(s,t) + r(s,s) \leq C|t-s|^{\delta},$$

then (X_t) has a continuous modification which is locally γ-Hölder continuous for any $\gamma \in (0, \delta/2)$.

Proof. Subtracting the mean function we may and do assume that $m(t) \equiv 0$. Simply

$$\mathbf{Var}(X_t - X_s) = r(t,t) - 2r(s,t) + r(s,s) = \sigma^2(s,t),$$

therefore

$$\mathbf{E}|X_t - X_s|^{\alpha} = \mathbf{E}|Z|^{\alpha}\sigma(s,t)^{\alpha},$$

with $Z \sim N(0,1)$. Thus

$$\mathbf{E}|X_t - X_s|^{\alpha} \leq C|t-s|^{\delta\alpha/2},$$

which implies that the condition of the continuity theorem holds with $\alpha > 0$, $\beta = \delta\alpha/2 - 1$. Letting $\alpha \to \infty$ the result follows. $\square$

Exercise 18 (Fractional Brownian motion)**.** Fractional Brownian motion with Hurst index $H \in (0,1)$ is a Gaussian process $(B(t))$ with mean function $m(t) \equiv 0$ and covariance function

$$r(s,t) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right).$$

Note that $H = 1/2$ corresponds to the usual Brownian motion.

(i) Show that it is self-similar, i.e. $B(at) \sim a^H B(t)$.

(ii) Show that it has stationary increments: $B(t) - B(s) \sim B(t-s)$.

(iii) Prove that a continuous modification exists, which is γ-Hölder for any $\gamma < H$. (That is H is the ‘roughness parameter’: for small H the process strongly oscillates, while for H close to 1 the paths are almost smooth.)

(iv) Are the increments independent?

Exercise 19. Let $(X_t)_{t\in[0,1]}$ be a continuous Gaussian process with mean 0 and covariance function $r(s,t)$. Show that $Y = \int_0^1 X_t \mathrm{d}t \sim N(0,\sigma^2)$, where

$$\sigma^2 = \int_0^1 \int_0^1 r(s,t)\,\mathrm{d}s\,\mathrm{d}t.$$

Show that $Y_t = \int_0^t X_s \mathrm{d}s$ is a Gaussian process. Determine its covariance function.

A version of the continuity theorem is the following.

Theorem 20. *Let $T \subset \mathbb{R}$ finite or infinite interval, and $(X_t)_{t\in T}$ a stochastic process such that for $\delta > 0$ small enough*

$$\mathbf{P}\left(|X_t - X_s| \geq g(\delta)\right) \leq h(\delta) \quad \textit{whenever } |s-t| < \delta\,,\ s,t \in T,$$

where g and h are continuous function such that

$$\sum_{n=1}^{\infty} g\left(2^{-n}\right) < \infty\,, \quad \sum_{n=1}^{\infty} 2^n h\left(2^{-n}\right) < \infty\,,$$

Then X has a continuous modification.

Recall that

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

is the standard normal density function, and

$$\Phi(x) = \int_{-\infty}^{x} \varphi(y)\mathrm{d}y$$

is the standard normal distribution function.

Lemma 5. *For any $x > 0$*

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\varphi(x) \leq 1 - \Phi(x) \leq \frac{1}{x}\varphi(x)$$

and

$$\lim_{x\to\infty} \frac{1-\Phi(x)}{\frac{1}{x}\varphi(x)} = 1.$$

Proof. The first follows from integrating the inequality

$$\left(1 - \frac{3}{y^4}\right)\varphi(y) \leq \varphi(y) \leq \left(1 + \frac{1}{y^2}\right)\varphi(y),$$

on (x,∞). The second is immediate from the first. $\square$

Using Theorem 20 we obtain a better criteria for continuity.

Corollary 7. *Let $T \subset \mathbb{R}$ be a finite or infinite interval and let $(X_t)_{t \in T}$ be a Gaussian process with continuous mean function m, and covariance function r such that for δ small enough*

$$\sup_{|s-t| \leq \delta} \left(r(t,t) - 2r(s,t) + r(s,s) \right) \leq C \left(-\log \delta \right)^{-3(1+\alpha)}$$

for some $C > 0$, $\alpha > 0$. Then (X_t) has a continuous modification.

3.2 The space $C[0, \infty)$

As SBM is continuous, its natural space is the space of continuous functions. Instead of a collection of random variables a stochastic process (W_t) can be understood as a random element of a function space.

Recall that ρ is a metric if on S

(i) $\rho \geq 0$, $\rho(\omega_1, \omega_2) = 0$ iff $\omega_1 = \omega_2$;

(ii) symmetric;

(iii) the triangle inequality holds, i.e.

$$\rho(\omega_1, \omega_2) \leq \rho(\omega_1, \omega_3) + \rho(\omega_2, \omega_3).$$

Then (S, ρ) is a metric space.

The sequence (x_n) is *Cauchy* if for each $\varepsilon > 0$ there exist $n_0(\varepsilon)$ such that $\rho(x_m, x_n) \leq \varepsilon$ for all $m, n \geq n_0$. The space (S, ρ) is *complete* if every Cauchy sequence converges. A set $A \subset S$ is dense, if for any $x \in S$ there exists a sequence $(x_n) \subset A$ such that $x_n \to x$. The space (S, ρ) is *separable* if there exists a countable dense subset.

Let $C[0, \infty)$ denote the space of continuous real functions on $[0, \infty)$ with metric

$$\rho(\omega_1, \omega_2) = \sum_{n=1}^{\infty} \frac{1}{2^n} \max_{t \in [0,n]} \left(|\omega_1(t) - \omega_2(t)| \wedge 1 \right).$$

Proposition 9. *ρ is a metric, and $(C[0, \infty), \rho)$ is a complete separable metric space.*

Proof. It is clear that ρ is a metric. Fix a Cauchy sequence (x_n). For any fix $N \in \mathbb{N}$ the limit $\lim_{n\to\infty} x_n(t) = x_\infty(t)$ exists for $t \in [0, N]$, and it is continuous. Thus x_∞ exists and continuous.

To find a countable dense subset consider functions which are 0 for $t \geq n$, and it is rational at k/n for $k = 0, 1, \dots, n^2 - 1$. $\square$

If (S, ρ) is a metric space we can define open sets. The σ-algebra generated by open sets is the Borel-σ-algebra $\mathcal{B}(S)$. With this $(S, \mathcal{B}(S))$ is a measurable space.

If $(\Omega, \mathcal{A}, \mathbf{P})$ is a probability space and $(S, \mathcal{B}(S))$ is a measurable space then a measurable $X : \Omega \to S$ is a *random variable / random element* in S. It induces a probability measure $\mathbf{P} \circ X^{-1}$ on S as

$$\mathbf{P} \circ X^{-1}(B) = \mathbf{P}(X \in B) = \mathbf{P}(\{\omega : X(\omega) \in B\}).$$

Let (P_n) be a sequence of probability measure on $(S, \mathcal{B}(S))$ and P another measure on it. Then P_n *converges weakly to* P, $P_n \stackrel{w}{\to} P$, if

$$\lim_{n\to\infty} \int_S f(s) \mathrm{d}P_n(s) = \int_S f(s) \mathrm{d}P(s)$$

for every continuous real function f. Note that the limit measure is necessarily a probability measure.

Let X_n and X be random elements in S, defined possibly on different probability spaces. The sequence (X_n) *converges in distribution to* X if the corresponding induced measures converge weakly. Equivalently,

$$\mathbf{E} f(X_n) \to \mathbf{E} f(X)$$

for all continuous and bounded f.

Assume that $X_n \to X$ in distribution. For any $0 \leq t_1 < \ldots < t_k$ consider the projection $\pi_{t_1,\ldots,t_k} : C[0,\infty) \to \mathbb{R}^k$

$$\pi_{t_1,\ldots,t_k}(\omega) = (\omega(t_1), \ldots, \omega(t_k)).$$

This is clearly continuous. For a continuous bounded function $f : \mathbb{R}^k \to \mathbb{R}$ the composite function $f(\pi_{t_1,\ldots,t_k})$ is bounded and continuous. Therefore, by the definition of convergence in distribution

$$\mathbf{E} f(\pi_{t_1,\ldots,t_k}(X_n)) \to \mathbf{E} f(\pi_{t_1,\ldots,t_k}(X))$$

that is

$$\mathbf{E}f(X_n(t_1),\ldots,X_n(t_k)) \to \mathbf{E}f(X(t_1),\ldots,X(t_k)).$$

That is, for any $0 \le t_1 < \ldots < t_k$

$$(X_n(t_1),\ldots,X_n(t_k)) \xrightarrow{\mathcal{D}} (X(t_1),\ldots,X(t_k)).$$

This means that the finite dimensional distributions converge.

We proved the following.

Proposition 10. *If (X_n) converges in distribution to X then all finite dimensional distributions converge.*

The converse is not true in general.

Example 9. Let

$$X_n(t) = nt\mathbf{I}_{[0,(2n)^{-1}]}(t) + (1-nt)\mathbf{I}_{((2n)^{-1},n^{-1}]}(t), \quad t \ge 0.$$

Then all finite dimensional distributions converge to the corresponding finite dimensional distributions of $X \equiv 0$. However, convergence as a process does not hold.

In what follows we try to understand what goes wrong in the example above, and state a converse of the Proposition above.

A family of probability measures Π on $(S, \mathcal{B}(S))$ is *tight* if for every $\varepsilon > 0$ there exists a compact set $K \subset S$ such that $P(K) \ge 1-\varepsilon$ for all $P \in \Pi$. The family Π is *relatively compact* if each sequence of elements from Π contains a convergent subsequence. A family of random elements is tight (relatively compact) if the family of induced measures is tight (relatively compact).

Theorem 21 (Prohorov)**.** *Let Π be a family of probability measures on a complete separable metric space S. Then Π is tight if and only if it is relatively compact.*

The modulus of continuity plays an important role in characterization of tightness on C. Fix $T > 0$ and $\delta > 0$, and let $\omega \in C[0,\infty)$. The modulus of continuity on $[0,T]$

$$m^T(\omega,\delta) = \max\left\{|\omega(s)-\omega(t)| : \ |s-t| \le \delta, 0 \le s,t \le T\right\}.$$

Exercise 20. Show that m^T is continuous in $\omega \in C[0,\infty)$ under the metric ρ, is nondecreasing in δ, and $\lim_{\delta\downarrow 0} m^T(\omega,\delta) = 0$ for each $\omega \in C[0,T)$.

Theorem 22 (Arselà–Ascoli). *A set $A \subset C[0,\infty)$ has compact closure if and only if the following two conditions hold:*

(i) $\sup_{\omega \in A} |\omega(0)| < \infty$*;*

(ii) for every $T > 0$

$$\lim_{\delta \downarrow 0} \sup_{\omega \in A} m^T(\omega, \delta) = 0.$$

Now we can characterize tightness of probability measures.

Theorem 23. *A sequence (P_n) of probability measures on $(C[0,\infty), \mathcal{B})$ is tight if and only if the following two conditions hold:*

(i) $\lim_{\lambda \uparrow \infty} \sup_{n \geq 1} P_n(\omega : \, |\omega(0)| > \lambda) = 0$*;*

(ii) for all $T > 0$ and $\varepsilon > 0$

$$\lim_{\delta \downarrow 0} \sup_{n \geq 1} P_n(\omega : \, m^T(\omega, \delta) > \varepsilon) = 0.$$

Theorem 24. *Let (X_n) be a tight sequence of continuous processes such that its finite dimensional distributions converge. Then the sequence of induced measures (P_n) converge weakly to a measure P such that the coordinate mapping $W_t(\omega) = \omega_t$ on $C[0,\infty)$ satisfies*

$$(X_n(t_1), \ldots, X_n(t_k)) \xrightarrow{\mathcal{D}} (W(t_1), \ldots, W(t_k)),$$

for any $0 \leq t_1 < \ldots < t_k < \infty$, $k \geq 1$.

Proof. Tightness is the same as relative compactness. Therefore, every subsequence contains a further convergent subsequence. Because of the convergence of finite dimensional distributions any two limit measure has the same finite dimensional distributions. But finite dimensional distributions determine the measure. □

3.3 Donsker theorem

Let $\xi, \xi_1, \xi_2, \ldots$ be iid random variables with $\mathbf{E}\xi = 0$, $\mathbf{E}\xi^2 = \sigma^2 \in (0,\infty)$, and let $S_n = \sum_{i=1}^n \xi_i$ denote the partial sum. Define the continuous time process $(Y_t)_{t \geq 0}$ as

$$Y_t = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)\xi_{\lfloor t \rfloor + 1},$$

where $\lfloor \cdot \rfloor$ stands for the usual integer part. For $n \in \mathbb{N}$ define the scaled process

$$X_t^{(n)} = \frac{1}{\sigma\sqrt{n}} Y_{nt}, \quad t \geq 0.$$

Then $X_t^{(n)} - X_s^{(n)}$ for $s, t \in \mathbb{N}/n$ is independent of $\sigma(\xi_1, \ldots, \xi_{sn})$, and by the CLT its distribution tends to $N(0, t-s)$.

Theorem 25 (Invariance principle of Donsker)**.** *Let P_n denote the measure on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$ induced by $X^{(n)}$. Then P_n converges weakly to a measure $P_\star$. Under $P_\star$ the coordinate mapping $W_t(\omega) = \omega(t)$, $\omega \in C[0,\infty)$ is SBM.*

Proof. According to Theorem 24 we have to show that $(X^{(n)})$ is tight and the finite dimensional distributions converge to those of a SBM.

To prove tightness we have to show that the conditions of Theorem 23 hold for P_n. This can be done by proving some maximal inequalities. We skip this part.

We prove the convergence of finite dimensional distributions. Fix $d \in \mathbb{N}$ and $0 \leq t_1 < \ldots < t_d < \infty$. We have to show that

$$\left(X_{t_1}^{(n)}, \ldots, X_{t_d}^{(n)} \right) \xrightarrow{\mathcal{D}} (W_{t_1}, \ldots, W_{t_d}).$$

To ease notation let $d = 2$ and $(t_1, t_2) = (s, t)$. We want to show that

$$(X_s^{(n)}, X_t^{(n)}) \xrightarrow{\mathcal{D}} (W_s, W_t).$$

By the definition of $X^{(n)}$

$$\left\| (X_s^{(n)}, X_t^{(n)}) - \frac{1}{\sigma\sqrt{n}} (S_{\lfloor sn \rfloor}, S_{\lfloor tn \rfloor}) \right\| \xrightarrow{\mathbf{P}} 0,$$

therefore it is enough to show that

$$\frac{1}{\sigma\sqrt{n}} (S_{\lfloor sn \rfloor}, S_{\lfloor tn \rfloor}) \xrightarrow{\mathcal{D}} (W_s, W_t).$$

By Lévy's CLT

$$\frac{1}{\sigma\sqrt{n}} (S_{\lfloor sn \rfloor}, S_{\lfloor tn \rfloor} - S_{\lfloor sn \rfloor}) \xrightarrow{\mathcal{D}} (\sqrt{s}Z, \sqrt{t-s}Z'),$$

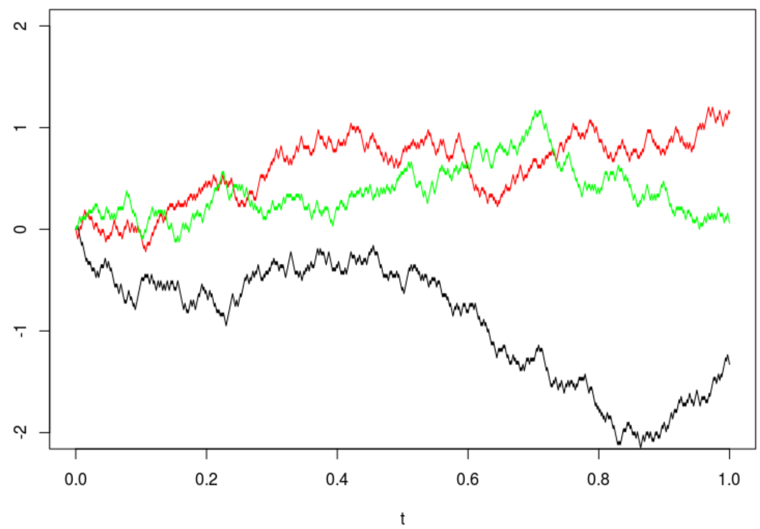

Figure 1: Simulation of 3 independent SBM

where Z, Z' are independent $N(0,1)$. Therefore

$$\frac{1}{\sigma\sqrt{n}}(S_{\lfloor sn \rfloor}, S_{\lfloor tn \rfloor}) \xrightarrow{\mathcal{D}} (\sqrt{s}Z, \sqrt{s}Z + \sqrt{t-s}Z') \stackrel{\mathcal{D}}{=} (W_s, W_t),$$

as claimed. □

In the proof above we used the following simple statements.

Exercise 21. Let (X_n) be a sequence of random elements in the metric space (S_1, ρ_1) converging in distribution to X. Let $\varphi : S_1 \to S_2$ be continuous, where (S_2, ρ_2) is another metric space. Show that $\varphi(X_n)$ converges in distribution to $\varphi(X)$.

Exercise 22. Let (X_n), (Y_n) be random elements in the separable metric space (S, ρ) defined on the same probability space. Show that if X_n converges in distribution to X and $\rho(X_n, Y_n) \to 0$ in probability then Y_n converges in distribution to X.

As a consequence of Donsker's invariance principle we obtain limit result for the path of random walks. Let us restrict to the interval $[0,1]$ and consider the space $C[0,1]$ with the supremum norm. Consider the continuous functional

$$f : C[0,1] \to \mathbb{R};\ \omega \mapsto \max_{t\in[0,1]} \omega(t).$$

Since $X^{(n)} \to W$ in distribution (in $C[0,1]$) we have that $f(X^{(n)}) \to f(W)$ in distribution (in $\mathbb{R}$!). That is

$$\mathbf{P}(\max_{t\in[0,1]} X_t^{(n)} \le x) \to \mathbf{P}(\max_{t\in[0,1]} W_t \le x),$$

for each $x \in \mathbb{R}$ (well, only for continuity point of the limit, but it is continuous). By the definition of $X^{(n)}$ we can rewrite the RHS to get

$$\mathbf{P}\left(\max_{k\le n} S_k \le \sqrt{n}\sigma x\right) \to \mathbf{P}(\max_{t\in[0,1]} W_t \le x).$$

Next we determine the LHS. Using the reflection principle

$$\begin{aligned}
&\mathbf{P}\left(\max_{t\in[0,1]} W_t > x\right)\\
&= \mathbf{P}\left(\max_{t\in[0,1]} W_t > x,\ W_1 > x\right) + \mathbf{P}\left(\max_{t\in[0,1]} W_t > x,\ W_1 < x\right)\\
&= 2\mathbf{P}\left(\max_{t\in[0,1]} W_t > x,\ W_1 > x\right)\\
&= 2\mathbf{P}\left(W_1 > x\right) = 2\left(1 - \Phi(x)\right).
\end{aligned}$$

Summarizing

$$\lim_{n\to\infty} \mathbf{P}\left(\max_{k\le n} S_k \le \sqrt{n}\sigma x\right) = 2\Phi(x) - 1.$$

3.4 Markov property

Assume that we have a SBM (W_t) and we know everything up to time s. Conditioned on that information, what is the distribution of W_t, $t > s$?

Formally, $(W_t, \mathcal{F}_t)$ is a SBM, and we are interested in the conditional probabilities

$$\mathbf{P}(W_t \in A | \mathcal{F}_s).$$

Since $W_t = W_s + W_t - W_s$, where W_s is $\mathcal{F}_s$-measurable and $W_t - W_s$ is independent of $\mathcal{F}_s$, we obtain

$$\mathbf{P}(W_t \in A|\mathcal{F}_s) = \mathbf{P}(W_t \in A|W_s) = \mathbf{P}_{W_s}(W_{t-s} \in A),$$

where $\mathbf{P}_x(W_u \in A) = \mathbf{P}(W_u \in A | W_0 = x)$, that is under $\mathbf{P}_x$ W is a SBM starting at x. That is knowing the whole past up to s gives no more information than knowing only W_s. This is the Markov property.

To make the previous argument formal we need the following.

Exercise 23. Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space, $\mathcal{G} \subset \mathcal{A}$ a sub-σ-algebra, and X, Y random variables such that X is independent of $\mathcal{G}$ and Y is $\mathcal{G}$-measurable. Then

$$\mathbf{P}(X + Y \in A | \mathcal{G}) = \mathbf{P}(X + Y \in A | Y) \quad \mathbf{P} - \text{a.s.}$$

and

$$\mathbf{P}(X + Y \in A | Y = y) = \mathbf{P}(X + y \in A) \quad \mathbf{P}Y^{-1} - \text{a.s.}$$

For the latter note that for some $\sigma(Y)/\mathcal{B}(\mathbb{R})$-measurable h

$$\mathbf{P}(X + Y \in A | Y) = h(Y).$$

So the latter statement claims that $h(y) = \mathbf{P}(X + y \in A)$ a.s. with respect to the induced measure $\mathbf{P}Y^{-1}$.

A (d-dimensional) adapted process (X_t) is *Markov process with initial distribution* μ if

(i) $\mathbf{P}(X_0 \in A) = \mu(A)$;

(ii) $\mathbf{P}(X_{t+s} \in A | \mathcal{F}_s) = \mathbf{P}(X_{t+s} \in A | X_s)$, for all A and $t, s > 0$.

Sometimes it is more convenient to work with various initial distributions. A *Markov family* is an adapted process (X_t) together with a family of probability measures $(\mathbf{P}_x)$ such that

(i) $\mathbf{P}_x(X_0 = x) = 1$;

(ii) $\mathbf{P}_x(X_{t+s} \in A | \mathcal{F}_s) = \mathbf{P}_x(X_{t+s} \in A | X_s)$;

(iii) $\mathbf{P}_x(X_{t+s} \in A | X_s = y) = \mathbf{P}_y(X_t \in A)$ $\mathbf{P}_x X_s^{-1}$-a.s.

For a given $\omega \in \Omega$ denote $X_{s+\cdot}$ the function X_{s+t}, that is we shift the path by s. The property in the definition of Markov process easily extends to path.

Proposition 11. *For a Markov family for any* $F \in \mathcal{B}(\mathbb{R}^{[0,\infty)})$

(i) $\mathbf{P}_x(X_{s+\cdot} \in F | \mathcal{F}_s) = \mathbf{P}_x(X_{s+\cdot} \in F | X_s)$*;*

(ii) $\mathbf{P}_x(X_{s+\cdot} \in F | X_s = y) = \mathbf{P}_y(X_\cdot \in F)$ $\mathbf{P}_x X_s^{-1}$*-a.s.*

The proof goes by the usual technical machinery. The sets F satisfying the above properties forms a λ-system and it contains the finite dimensional cylinders.

Markov property states that the process restarts at fixed times t. Sometimes we need to restart the process at stopping times τ. This property is the *strong Markov property.*

A (d-dimensional) adapted process (X_t) is *strong Markov process with initial distribution* μ if

(i) $\mathbf{P}(X_0 \in A) = \mu(A)$;

(ii) $\mathbf{P}(X_{\tau+t} \in A|\mathcal{F}_\tau) = \mathbf{P}(X_t \in A|X_\tau)$, for all A and stopping time τ.

Similarly, a *strong Markov family* is an adapted process (X_t) together with a family of probability measures $\mathbf{P}_x$ such that

(i) $\mathbf{P}_x(X_0 = x) = 1$;

(ii) $\mathbf{P}_x(X_{\tau+t} \in A|\mathcal{F}_\tau) = \mathbf{P}_x(X_{\tau+t} \in A|X_\tau)$ for all A and stopping time τ;

(iii) $\mathbf{P}_x(X_{\tau+t} \in A|X_\tau = y) = \mathbf{P}_y(X_t \in A)$ $\mathbf{P}_x X_\tau^{-1}$-a.s. for all A and stopping time τ;

Proposition 12. *For a strong Markov family for any* $F \in \mathcal{B}((\mathbb{R})^{[0,\infty)})$

(i) $\mathbf{P}_x(X_{\tau+\cdot} \in F|\mathcal{F}_\tau) = \mathbf{P}_x(X_{\tau+\cdot} \in F|X_\tau)$*;*

(ii) $\mathbf{P}_x(X_{\tau+\cdot} \in F|X_\tau = x) = \mathbf{P}_x(X_\cdot \in F)$ $\mathbf{P}_x X_\tau^{-1}$*-a.s.*

We proved that SBM is Markov. In fact, it is strong Markov.

Theorem 26. *SBM is a strong Markov process.*

3.5 Path properties

Theorem 27. *Almost surely the sample path of a SBM is not monotone in any interval.*

Proof. Let

$$A = \{\omega : W(\cdot, \omega) \text{ is monotone on some interval}\}.$$

Clearly

$$A = \cup_{r,s\in\mathbf{Q}} \{\omega : W(\cdot, \omega) \text{ is monotone on } [r, s]\}.$$

Since this is a countable union it is enough to prove that each event has probability zero. To ease notation choose $r = 0$, $s = 1$, and put

$$B = \{\omega : W(\cdot, \omega) \text{ is nondecreasing on } [0,1]\}.$$

We have

$$B = \cap_{n=1}^{\infty} \{\omega : W((i+1)/n, \omega) \geq W(i/n, \omega),\ i = 0, 1, \ldots, n-1\} =: \cap_{n=1}^{\infty} B_n.$$

By the independent increment property

$$\mathbf{P}(B_n) = \prod_{i=0}^{n-1} \mathbf{P}(W((i+1)/n) \geq W(i/n)) = 2^{-n},$$

which implies that $\mathbf{P}(B) = 0$ as claimed. $\square$

For any interval $[a,b]$ let $\Pi_n = \{a = t_0 < t_1 < \ldots < t_n = b\}$ a partition with mesh

$$\|\Pi_n\| = \max\{t_i - t_{i-1} \ :\ i = 1, 2, \ldots, n\}.$$

We determine the quadratic variation of the Wiener process.

Theorem 28. *Let* $\Pi_n = \{a = t_0 < t_1 < \ldots < t_n = b\}$, $n = 1, 2, \ldots$, *a sequence of partitions of* $[a,b]$ *such that* $\|\Pi_n\| \to 0$. *Then*

$$\sum_{i=1}^{n} (W_{t_i} - W_{t_{i-1}})^2 \xrightarrow{L^2} b - a.$$

Proof. Assume that $[a,b] = [0,1]$. We have to show that

$$\mathbf{E}\left(\sum_{i=1}^{n} (W_{t_i} - W_{t_{i-1}})^2 - 1\right)^2 \longrightarrow 0.$$

Using $1 = \sum_{i=1}^{n} (t_i - t_{i-1})$ we have

$$\begin{aligned} &\mathbf{E}\left(\sum_{i=1}^{n} (W_{t_i} - W_{t_{i-1}})^2 - 1\right)^2 = \\ &\sum_{i,j=1}^{n} \mathbf{E}\left(\left[(W_{t_i} - W_{t_{i-1}})^2 - (t_i - t_{i-1})\right]\left[(W_{t_j} - W_{t_{j-1}})^2 - (t_j - t_{j-1})\right]\right). \end{aligned} \tag{9}$$

If $i \neq j$ then $W_{t_i} - W_{t_{i-1}}$ and $W_{t_j} - W_{t_{j-1}}$ are independent. Therefore

$$\mathbf{E}\left[(W_{t_i} - W_{t_{i-1}})^2 - (t_i - t_{i-1})\right] = 0,$$

so the mixed products in (9) are 0. Using that $W_t - W_s \sim N(0, t-s)$ we obtain

$$\begin{aligned}
\mathbf{E}\left(\sum_{i=1}^{n}(W_{t_i} - W_{t_{i-1}})^2 - 1\right)^2 &= \sum_{i=1}^{n} \mathbf{E}\left[(W_{t_i} - W_{t_{i-1}})^2 - (t_i - t_{i-1})\right]^2 \\
&= \sum_{i=1}^{n}(t_i - t_{i-1})^2 \mathbf{E}\left[\left(\frac{W_{t_i} - W_{t_{i-1}}}{\sqrt{t_i - t_{i-1}}}\right)^2 - 1\right]^2 \\
&= \mathbf{E}(Z^2 - 1)^2 \sum_{i=1}^{n}(t_i - t_{i-1})^2,
\end{aligned}$$

where $Z \sim N(0,1)$. Since

$$\sum_{i=1}^{n}(t_i - t_{i-1})^2 \leq \|\Pi_n\| \sum_{i=1}^{n}(t_i - t_{i-1}) = \|\Pi_n\| \to 0,$$

the proof is ready. $\square$

Under some extra conditions a.s. convergence hold. Recall that in general neither L^2 convergence nor a.s. convergence implies the other. Moreover, L^2 convergence implies a.s. convergence on a subsequence. However, if $\sum_{n=1}^{\infty} \|\Pi_n\| < \infty$ then the Borel–Cantelli lemma implies a.s. convergence.

Exercise 24. Let $\Pi_n = \{a = t_0 < t_1 < \ldots < t_n = b\}$, $n = 1, 2, \ldots$, a sequence of partitions of $[a,b]$ such that $\sum_{n=1}^{\infty} \|\Pi_n\| < \infty$. Then a.s.

$$\sum_{i=1}^{n}(W_{t_i} - W_{t_{i-1}})^2 \longrightarrow b - a.$$

Corollary 8. *Let (Π_n) be a sequence of partitions of the interval $[a,b]$ such that $\sum_{n=1}^{\infty} \|\Pi_n\| < \infty$. Then $\sum_{i=1}^{n} |W_{t_i} - W_{t_{i-1}}| \to \infty$ a.s.*

Proof. Clearly,

$$\sum_{i=1}^{n}(W_{t_i} - W_{t_{i-1}})^2 \leq \sup_{1 \leq i \leq n} |W_{t_i} - W_{t_{i-1}}| \sum_{i=1}^{n} |W_{t_i} - W_{t_{i-1}}|.$$

The left-hand side converges to $b-a$ a.s. on a subsequence. On the right-hand side the first factor goes to 0 a.s. by the continuity of the Wiener process. (Recall that continuous function is uniformly continuous on compacts.) Therefore the second term necessarily tends to infinity. $\square$

We proved that the sample path of W are Hölder continuous with exponent $< 1/2$, and that the sample path are not of bounded variation. These results suggest that the trajectories are quite irregular. In fact, they are a.s. nowhere differentiable.

Theorem 29 (Paley, Wiener, Zygmund (1933))**.** *Almost surely the path $W(\cdot,\omega)$ is nowhere differentiable.*

Proof. For $n,k \in \mathbb{N}$ consider

$$\begin{aligned} X_{nk} = \max\Big\{ &\left|W\left(k2^{-n}\right) - W\left((k-1)2^{-n}\right)\right|, \left|W\left((k+1)2^{-n}\right) - W\left(k2^{-n}\right)\right|, \\ &\left|W\left((k+2)2^{-n}\right) - W\left((k+1)2^{-n}\right)\right| \Big\}. \end{aligned}$$

Using the independent increment property and the scale invariance

$$\mathbf{P}(X_{nk} \le \varepsilon) = \left(\mathbf{P}(|W(1/2^n)| \le \varepsilon)\right)^3 \le \left(2 \cdot 2^{n/2}\varepsilon\right)^3.$$

Putting $Y_n = \min_{1 \le k \le n2^n} X_{nk}$ we obtained

$$\mathbf{P}(Y_n \le \varepsilon) \le \sum_{k=1}^{n2^n} \mathbf{P}(X_{nk} \le \varepsilon) < n\,2^n \left(2 \cdot 2^{n/2}\,\varepsilon\right)^3.$$

Introduce the event

$$A = \{\omega : W(\cdot\,,\omega) \text{ is somewhere differentiable}\}.$$

If $\omega \in A$ then there exist $t = t(\omega)$ such that $W'(t,\omega) = D(\omega) \in \mathbb{R}$. Thus

$$\lim_{s\to t}\left|\frac{W(s,\omega) - W(t,\omega)}{s-t}\right| = |D(\omega)| < \infty\,.$$

Therefore there exists $\delta(\omega) = \delta(\omega,t) > 0$ such that for $|s-t| < \delta(\omega)$

$$|W(s,\omega) - W(t,\omega)| \le (|D(\omega)|+1)|s-t|\,.$$

Let $n_0(\omega) = n_0(\omega, t)$ so large that

$$2^{-n_0(\omega)} < \frac{\delta(\omega)}{2}\,, \quad n_0(\omega) \geq \max\{4(|D(\omega)|+1), t+1\}.$$

Fix $n \geq n_0(\omega)$ and let

$$\frac{k(\omega)}{2^n} \leq t < \frac{k(\omega)+1}{2^n}\,.$$

Then

$$\max\left\{\left|t - \frac{j}{2^n}\right| \;:\; j = k(\omega)-1,\ k(\omega),\ k(\omega)+1,\ k(\omega)+2\right\} \leq \frac{2}{2^n} < \delta(\omega)\,,$$

thus

$$\begin{aligned} X_{n,\,k(\omega)}(\omega) &\leq \max\left\{\left|W\left(\frac{j}{2^n},\omega\right) - W(t,\omega)\right| + \left|W\left(\frac{j-1}{2^n},\omega\right) - W(t,\omega)\right|\right\} \\ &\leq 2\big(|D(\omega)|+1\big)\,\frac{2}{2^n} = 4\Big(|D(\omega)|+1\Big)\,\frac{1}{2^n} \leq \frac{n}{2^n}\,, \end{aligned}$$

where the max is taken on the set $j \in \{k(\omega), k(\omega)+1, k(\omega)+2\}$.

Since $k(\omega) \leq n\,2^n$, we obtained

$$Y_n(\omega) = \min_{1\leq k\leq n2^n} X_{nk}(\omega) \leq n/2^n.$$

Thus $\omega \in A$ implies $\omega \in A_n = \{\omega : Y_n(\omega) \leq n/2^n\}$ for all $n \geq n_0(\omega)$ so

$$\begin{aligned} \omega \in \liminf_{n\to\infty} A_n &= \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} A_m \\ &= \{\omega : \omega \in A_k \text{ except finitely many } k\}. \end{aligned}$$

That is $A \subset B := \liminf_{n\to\infty} A_n$. Using the Fatou lemma

$$\begin{aligned} \mathbf{P}(B) &\leq \liminf_{n\to\infty} \mathbf{P}(A_n) \leq \liminf_{n\to\infty} \mathbf{P}\left(Y_n \leq \frac{n}{2^n}\right) \\ &\leq \liminf_{n\to\infty} n\,2^n\left(2\cdot 2^{n/2}\,\frac{n}{2^n}\right)^3 = 0. \end{aligned}$$

So $A \subset B$ and $\mathbf{P}(B) = 0$ as claimed. $\square$

Note that we don't claim that $A \in \mathcal{A}$. Now we see the usefulness of the *usual conditions*. The usual conditions include that $\mathcal{F}_0$ contains the null-sets of $\mathcal{A}$.

Let

$$Z(\omega) = \{t : W(t,\omega) = 0\}$$

denote the set of zeros. Let λ be the Lebesgue measure. By Fubini

$$\begin{aligned}
\mathbf{E}\lambda(Z) &= \int_\Omega \lambda(Z(\omega))\mathbf{P}(\mathrm{d}\omega) \\
&= \int_\Omega \int_\mathbb{R} \mathbf{I}(W(t,\omega) = 0)\, \mathrm{d}t\mathbf{P}(\mathrm{d}\omega) \\
&= \int_\mathbb{R} \mathbf{P}(W(t,\omega) = 0)\mathrm{d}t = 0.
\end{aligned}$$

Since $\lambda(Z) \geq 0$ this implies $\lambda(Z) = 0$ a.s.

Theorem 30 (Khinchin, 1933)**.** *For almost every* ω

$$\limsup_{t\downarrow 0} \frac{W_t(\omega)}{\sqrt{2t\log\log 1/t}} = 1 \quad \textit{and} \quad \liminf_{t\downarrow 0} \frac{W_t(\omega)}{\sqrt{2t\log\log 1/t}} = -1,$$

and

$$\limsup_{t\to\infty} \frac{W_t(\omega)}{\sqrt{2t\log\log t}} = 1 \quad \textit{and} \quad \liminf_{t\to\infty} \frac{W_t(\omega)}{\sqrt{2t\log\log t}} = -1.$$

Proof. By symmetry it is enough to prove the limsup results, and since $(tW_{1/t})$ is SBM it is enough to prove at 0.

Let

$$X_t = \exp\left\{\lambda W_t - \frac{\lambda^2}{2}t\right\}.$$

This is a martingale, therefore by the maximal inequality

$$\mathbf{P}\left(\max_{s\in[0,t]}\left(W_s - \frac{\lambda}{2}s\right) \geq \beta\right) = \mathbf{P}\left(\max_{s\in[0,t]} X_s \geq e^{\lambda\beta}\right) \leq e^{-\lambda\beta}.$$

Put $h(t) = \sqrt{2t\log\log(1/t)}$. Fix $\theta, \delta \in (0,1)$. Choose $\lambda = (1+\delta)\theta^{-n}h(\theta^n)$, $\beta = h(\theta^n)/2$, and $t = \theta^n$. Then

$$\mathbf{P}\left(\max_{s\in[0,t]}\left(W_s - \frac{\lambda}{2}s\right) \geq \beta\right) \leq e^{-\lambda\beta} = (n\log 1/\theta)^{-(1+\delta)}.$$

This is summable, therefore by the Borel–Cantelli lemma there exists $N(\omega)$, and $\Omega_{\delta,\theta}$ with $\mathbf{P}(\Omega_{\delta,\theta}) = 1$ such that

$$\max_{s\in[0,\theta^n]}\left(W_s - \frac{1+\delta}{2}s\theta^{-n}h(\theta^n)\right) \le \frac{1}{2}h(\theta^n) \quad \text{for } n \ge N(\omega).$$

Thus for $t \in (\theta^{n+1}, \theta^n]$

$$W_t(\omega) \le \max_{s\in[0,\theta^n]} W_s(\omega) \le (1+\delta/2)\,h(\theta^n) \le (1+\delta/2)\,\theta^{-1/2}\,h(t).$$

Therefore for $n \ge N(\omega)$

$$\sup_{t\in(\theta^{n+1},\theta^n]} \frac{W_t(\omega)}{h(t)} \le (1+\delta/2)\,\theta^{-1/2},$$

which implies as $n \to \infty$

$$\limsup_{t\downarrow 0} \frac{W_t(\omega)}{h(t)} \le (1+\delta/2)\,\theta^{-1/2}.$$

Letting $\delta \downarrow 0$ and $\theta \uparrow 1$ through rationals we obtain

$$\limsup_{t\downarrow 0} \frac{W_t(\omega)}{h(t)} \le 1. \tag{10}$$

For the opposite direction we need the second Borel–Cantelli lemma, which requires independence. Fix $\theta \in (0,1)$ and let

$$A_n = \{W_{\theta^n} - W_{\theta^{n+1}} \ge \sqrt{1-\theta}h(\theta^n)\}.$$

Putting $x = \sqrt{2\log n + 2\log\log 1/\theta}$

$$\mathbf{P}(A_n) = \mathbf{P}\left(\frac{W_{\theta^n} - W_{\theta^{n+1}}}{\sqrt{\theta^n - \theta^{n+1}}} \ge x\right) \ge Cx^{-1}e^{-\frac{x^2}{2}} \ge C'\frac{1}{n\sqrt{\log n}},$$

where we use Lemma 5. The lower bound is a divergent series in n, therefore the event A_n occur infinitely often. On the other hand by (10) (for $-W_t$)

$$-W_{\theta^{n+1}} \le 2h(\theta^{n+1}) \le 4\theta^{1/2}h(\theta^n)$$

for all $n \geq N(\omega)$. Therefore whenever A_n occur

$$\frac{W_{\theta^n}(\omega)}{h(\theta^n)} \geq \sqrt{1-\theta} - 4\sqrt{\theta}.$$

Letting $n \to \infty$ we have

$$\limsup_{t \downarrow 0} \frac{W_t}{h(t)} \geq \sqrt{1-\theta} - 4\sqrt{\theta},$$

and the result follows by letting $\theta \downarrow 0$. □

Exercise 25. Show that if W is SBM then for any λ

$$X_t = \exp\left\{\lambda W_t - \frac{\lambda^2}{2} t\right\}$$

is a martingale.

4 Stochastic integral

This part is from Karatzas and Shreve [5], in a rather simplified way. Stochastic integration is only worked out in detail with respect to SBM, and not with respect to a continuous martingales. A lot of technical detailes are omitted.

Here we define the integration with respect to the Brownian motion. Note that SBM is not of bounded variation, therefore we cannot define the integral pathwise. This is the major difficulty in the theory.

4.1 Integration of simple processes

In what follows we work on $[0, T]$, for $T < \infty$. Let $(W_t, \mathcal{F}_t)$ be SBM.

The process (X_t) is a *simple process*, if

$$X_t(\omega) = \xi_0(\omega)\mathbf{I}_{\{0\}}(t) + \sum_{i=1}^{n-1} \xi_i(\omega)\mathbf{I}_{(t_i, t_{i+1}]}(t),$$

where $0 = t_0 < t_1 < \ldots < t_n = T$ is a partition of $[0, T]$, and ξ_i is $\mathcal{F}_{t_i}$-measurable.

That is $(X_t(\omega))$ is a step function for each $\omega \in \Omega$, where the step sizes are random. Note that ξ_i is measurable with respect to the σ-algebra corresponding to the left end point of the interval.

Exercise 26. Show that a simple process is adapted.

The definition of the integral of simple processes is straightforward. Let k be such that $t \in (t_k, t_{k+1}]$. Then

$$I_t(X) = \int_0^t X_s \mathrm{d}W_s = \sum_{i=0}^{k-1} \xi_i (W_{t_{i+1}} - W_{t_i}) + \xi_k (W_t - W_{t_k}), \quad t \in [0,T].$$

Note that we defined the process for each $t \in [0,T]$.

Theorem 31. *Let X, Y be simple processes with square integrable coefficients.*

(i) $I_t(X)$ is a continuous martingale, $I_0(X) = 0$ a.s.

(ii) For $t > s$

$$\mathbf{E}\left[\left(\int_s^t X_u \mathrm{d}W_u \right)^2 \Big| \mathcal{F}_s \right] = \mathbf{E}\left[\int_s^t X_u^2 \mathrm{d}u \Big| \mathcal{F}_s \right];$$

in particular $\mathbf{E} I_t(X)^2 = \mathbf{E} \int_0^t X_u^2 \mathrm{d}u$.

(iii) The integral is linear, that is

$$I(\alpha X + \beta Y) = \alpha I(X) + \beta I(Y), \quad \alpha, \beta \in \mathbb{R}.$$

(iv) $\mathbf{E} \sup_{0 \le t \le T} \left(\int_0^t X_u \mathrm{d}W_u \right)^2 \le 4 \mathbf{E} \int_0^T X_u^2 \mathrm{d}u$.

Proof. (iii) is clear. (iv) follows from Doob's maximal inequality.

(i) The continuity is obvious and $I_0(X) = 0$. We prove that (I_t) is martingale. Let $s < t$ and $s \in (t_k, t_{k+1}], t \in (t_m, t_{m+1}]$. Then

$$\begin{aligned} \int_0^t X_u \mathrm{d}W_u &= \sum_{i=0}^{k-1} \xi_i (W_{t_{i+1}} - W_{t_i}) + \xi_k (W_s - W_{t_k}) \\ &\quad + \xi_k (W_{t_{k+1}} - W_s) + \sum_{i=k+1}^{m-1} \xi_i (W_{t_{i+1}} - W_{t_i}) + \xi_m (W_t - W_{t_m}). \end{aligned}$$

By the tower rule

$$\begin{aligned} \mathbf{E}[\xi_i (W_{t_{i+1}} - W_{t_i}) | \mathcal{F}_s] &= \mathbf{E}\left[\mathbf{E}[\xi_i (W_{t_{i+1}} - W_{t_i}) | \mathcal{F}_{t_i}] | \mathcal{F}_s \right] \\ &= \mathbf{E}\left[\xi_i \mathbf{E}[W_{t_{i+1}} - W_{t_i} | \mathcal{F}_{t_i}] | \mathcal{F}_s \right] \\ &= \mathbf{E}[\xi_i \cdot 0 | \mathcal{F}_s] = 0. \end{aligned}$$

The first and last term can be handled similarly.

(ii) We showed that

$$\int_s^t X_u \mathrm{d}W_u = \xi_k(W_{t_{k+1}} - W_s) + \sum_{i=k+1}^{m-1} \xi_i(W_{t_{i+1}} - W_{t_i}) + \xi_m(W_t - W_{t_m}).$$

Taking square and conditional expectation we end up with sum of terms

$$\mathbf{E}[\xi_i(W_{t_{i+1}} - W_{t_i})\xi_j(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_s]$$

We show that this equals 0, whenever $i \neq j$. Indeed,

$$\begin{aligned}
&\mathbf{E}[\xi_i(W_{t_{i+1}} - W_{t_i})\xi_j(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_s] \\
&= \mathbf{E}\left[\mathbf{E}[\xi_i(W_{t_{i+1}} - W_{t_i})\xi_j(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_{t_j}]|\mathcal{F}_s\right] = 0.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathbf{E}\left[\left(\int_s^t X_u \mathrm{d}W_u\right)^2 |\mathcal{F}_s\right] \\
&= \mathbf{E}\left[\xi_k^2(W_{t_{k+1}} - W_s)^2 + \sum_{i=k+1}^{m-1} \xi_i^2(W_{t_{i+1}} - W_{t_i})^2 + \xi_m^2(W_t - W_{t_m})^2|\mathcal{F}_s\right].
\end{aligned}$$

By the tower rule again

$$\begin{aligned}
\mathbf{E}[\xi_i^2(W_{t_{i+1}} - W_{t_i})^2|\mathcal{F}_s] &= \mathbf{E}\left[\mathbf{E}[\xi_i^2(W_{t_{i+1}} - W_{t_i})^2|\mathcal{F}_{t_i}]\mathcal{F}_s\right] \\
&= \mathbf{E}[\xi_i^2(t_{i+1} - t_i)|\mathcal{F}_s] \\
&= \mathbf{E}\left[\int_{t_i}^{t_{i+1}} X_u^2 \mathrm{d}u|\mathcal{F}_s\right].
\end{aligned}$$

Summing we obtain the result. □

4.2 Extending the definition

The idea is the following. We defined the integral for simple processes. Adapted processes can be approximated by simple processes, so we can define the integral of adapted process as a limit and hope for the best. This was the method at the definition of both Riemann and Lebesgue integral.

Let

$$\mathcal{H} = \left\{(X_t) \ : \mathcal{F}_t\text{-adapted and } \mathbf{E}\int_0^T X_u^2 \mathrm{d}u < \infty\right\}.$$

We extend the definition to the class $\mathcal{H}$.

Lemma 6. *Let* $(X_t) \in \mathcal{H}$. *There exists a sequence of simple processes* $\{(X^n_t)\}_n$ *such that*

$$\lim_{n\to\infty} \mathbf{E} \int_0^T (X_s - X^n_s)^2 \, \mathrm{d}s = 0.$$

Proof. We only prove in the special case when X is bounded and continuous. Let

$$X^n_t(\omega) = X_0(\omega)\mathbf{I}_{\{0\}}(t) + \sum_{k=0}^{2^n-1} X_{\frac{kT}{2^n}}(\omega)\mathbf{I}_{(\frac{kT}{2^n}, \frac{(k+1)T}{2^n}]}(t).$$

These are simple processes. Since continuous function is uniformly continuous on compacts, almost surely

$$\int_0^T |X^n_u - X_u|^2 \, \mathrm{d}t \to 0.$$

Lebesgue's dominated convergence gives the proof. □

Let $X \in \mathcal{H}$ and $\{X^n\}_n$ given in the lemma. By Theorem 31 (iv)

$$\mathbf{E} \sup_{t\in[0,T]} \left(\int_0^t (X^n_u - X^m_u) \mathrm{d}W_u \right)^2 \leq 4\mathbf{E} \int_0^T (X^n_u - X^m_u)^2 \mathrm{d}u.$$

The right-hand side tends to 0 by the lemma above, therefore the left-hand side too. Thus there exists a sequence $\{n_k\}$ such that

$$\mathbf{E} \sup_{t\in[0,T]} \left(\int_0^t (X^{n_{k+1}}_u - X^{n_k}_u) \mathrm{d}W_u \right)^2 \leq 2^{-k}. \tag{11}$$

The first Borel–Cantelli lemma implies

$$I(X^{n_k}) \to I(X), \text{ uniformly on } [0,T]\text{-n a.s.}$$

As $I(X^{n_k})$ is continuous, so is $I(X)$. We have to show that $I(X)$ does not depend on the subsequence. In (11) letting $m \to \infty$

$$\mathbf{E} \sup_{t\in[0,T]} (I_t(X) - I_t(X^n))^2 \leq 4\mathbf{E} \int_0^T (X_u - X^n_u)^2 \mathrm{d}u,$$

so $I(X)$ does not depend on the subsequence.

Next we show that $I(X)$ is martingale, i.e. for any $s < t$

$$\mathbf{E}[I_t(X)|\mathcal{F}_s] = I_s(X).$$

For any n

$$\begin{aligned}\|\mathbf{E}[I_t(X)|\mathcal{F}_s] - I_s(X)\|_{L^2} &\leq \|\mathbf{E}[I_t(X) - I_t(X^n)|\mathcal{F}_s]\|_{L^2} \\ &+ \|\mathbf{E}[I_t(X^n) - I_s(X^n)|\mathcal{F}_s]\|_{L^2} + \|I_s(X^n) - I_s(X)\|_{L^2},\end{aligned}$$

where $\|X\|_{L^2} = \sqrt{\mathbf{E}X^2}$. The second term on the RHS equals 0, since $I(X^n)$ is martingale, while the first and third term can be arbitrarily small. So $I(X)$ is indeed a martingale.

Summarizing, for $X \in \mathcal{H}$ we defined the stochastic integral

$$I_t(X) = \int_0^t X_u \mathrm{d}W_u$$

and showed that it satisfies the properties of Theorem 31.

We note that the definition of the integral can be further extended from $\mathcal{H}$ to the larger class

$$\mathcal{H}' = \left\{ (X_t) \ : \mathcal{F}_t\text{-adapted and } \int_0^T X_u^2 \,\mathrm{d}u < \infty \text{ a.s.} \right\}$$

such that Theorem 31 remains true.

Example 10 (Approximation of $\int_0^t W_s \mathrm{d}W_s$)**.** Fix $\varepsilon \in [0, 1]$ and consider

$$S_\varepsilon(\Pi) = \sum_{i=0}^{n-1} \left(\varepsilon W_{t_{i+1}} + (1-\varepsilon) W_{t_i}\right) \left(W_{t_{i+1}} - W_{t_i}\right).$$

We prove that

$$\lim_{\|\Pi\| \to 0} S_\varepsilon(\Pi) \overset{L^2}{=} \frac{1}{2} W_t^2 + \left(\varepsilon - \frac{1}{2}\right) t. \tag{12}$$

We know that $(W_t^2 - t)$ is martingale, thus the limit above is martingale iff $\varepsilon = 0$, which corresponds to the definition of Itô stochastic integral. There are other stochastic integrals: $\varepsilon = 1/2$ corresponds to the *Fisk–Stratonovich integral*, and $\varepsilon = 1$ corresponds to the *backward Itô integral*.

By (12)

$$\int_0^t W_s \mathrm{d}W_s = \frac{W_t^2 - t}{2}.$$

Next we prove (12). Since

$$\varepsilon W_{t_{i+1}} + (1-\varepsilon) W_{t_i} = \frac{W_{t_{i+1}} + W_{t_i}}{2} + \left(\varepsilon - \frac{1}{2}\right) \left(W_{t_{i+1}} - W_{t_i}\right),$$

we have to determine the limits

$$\sum_{i=0}^{n-1} (W_{t_{i+1}} - W_{t_i})^2, \quad \sum_{i=0}^{n-1} (W_{t_{i+1}}^2 - W_{t_i}^2).$$

The first is exactly the quadratic variation of SBM, therefore converges to t in L^2, while the second is a telescopic sum, giving W_t^2.

Example 11. Let X be simple process and W SBM. Let

$$\zeta_t^s(X) = \int_s^t X_u \mathrm{d}W_u - \frac{1}{2} \int_s^t X_u^2 \mathrm{d}u, \quad \zeta_t = \zeta_t^0.$$

We show that $(Y_t = e^{\zeta_t})$ is martingale.

Since X is simple, we have

$$X_t = \xi_0 \mathbf{I}_{\{0\}}(t) + \sum_{i=0}^{n-1} \xi_i \mathbf{I}_{(t_i, t_{i+1}]}(t),$$

where ξ_i is $\mathcal{F}_{t_i}$-measurable. Thus if $s \in (t_k, t_{k+1}]$, $t \in (t_m, t_{m+1}]$, then

$$\begin{aligned} \zeta_t^s = & \xi_k (W_{t_{k+1}} - W_s) - \frac{\xi_k^2}{2} (t_{k+1} - s) + \sum_{i=k+1}^{m-1} \left[\xi_i (W_{t_{i+1}} - W_{t_i}) - \frac{\xi_i^2}{2} (t_{i+1} - t_i) \right] \\ & + \xi_m (W_t - W_{t_m}) - \frac{\xi_m^2}{2} (t - t_m). \end{aligned} \tag{13}$$

Since ζ_s is $\mathcal{F}_s$-measurable we obtain

$$\mathbf{E}[e^{\zeta_t} | \mathcal{F}_s] = e^{\zeta_s} \mathbf{E}[e^{\zeta_t^s} | \mathcal{F}_s].$$

We only have to show that

$$\mathbf{E}[e^{\zeta_t^s} | \mathcal{F}_s] = 1.$$

This can be done by a repeated application of the tower rule. In (13) all terms but the last are $\mathcal{F}_{t_m}$-measurable and

$$\begin{aligned}&\mathbf{E}\left[\exp\left\{\xi_m(W_t - W_{t_m}) - \frac{\xi_m^2}{2}(t-t_m)\right\}|\mathcal{F}_{t_m}\right]\\&= e^{-\frac{\xi_m^2}{2}(t-t_m)}\mathbf{E}\left[\exp\{\xi_m(W_t - W_{t_m})\}|\mathcal{F}_{t_m}\right].\end{aligned}$$

In the exponent of the RHS ξ_m is $\mathcal{F}_{t_m}$-measurable and $W_t - W_{t_m}$ is independent of $\mathcal{F}_{t_m}$, therefore (by the next exercise) ξ_m can be handled as a constant. We have

$$\mathbf{E}e^{\lambda Z} = e^{\frac{\lambda^2}{2}},$$

therefore

$$\mathbf{E}\left[\exp\{\xi_m(W_t - W_{t_m})\}|\mathcal{F}_{t_m}\right] = e^{\frac{\xi_m^2}{2}(t-t_m)}.$$

Summarizing

$$\mathbf{E}\left[\exp\{\xi_m(W_t - W_{t_m}) - \frac{\xi_m^2}{2}(t-t_m)\}|\mathcal{F}_{t_m}\right] = 1.$$

Applying repeatedly the tower rule first to the σ-algebra $\mathcal{F}_{t_{m-1}}$, then to $\mathcal{F}_{t_{m-2}}$, $\ldots$, we obtain that each factor equals 1.

Using the Itô formula we show that Y is martingale for more general processes and it satisfies a certain stochastic differential equation.

Exercise 27. Let X, Y be random variables, X is $\mathcal{G}$-measurable, and Y is independent of $\mathcal{G}$. Then

$$\mathbf{E}[h(X,Y)|\mathcal{G}] = \int h(X,y)\mathrm{d}F(y),$$

where $F(y) = \mathbf{P}(Y \leq y)$ is the distribution function of Y.

4.3 Itô's formula

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $(\mathcal{F}_t)$ a filtration, and (W_t) SBM for this filtration. Then (X_t) *is Itô process* if

$$X_t = X_0 + \int_0^t K_s \mathrm{d}s + \int_0^t H_s \mathrm{d}W_s, \tag{14}$$

where

- X_0 $\mathcal{F}_0$-measurable;
- K, H are $\mathcal{F}_t$-adapted processes;
- $\int_0^T |K_u| \mathrm{d}u < \infty$, $\int_0^T H_s^2 \mathrm{d}s < \infty$ a.s.

The part $\int_0^t K_s \mathrm{d}s$ is the bounded variation part of the process, while $\int_0^t H_s \mathrm{d}W_s$ is the martingale part.

Lemma 7. *If $M_t = \int_0^t K_s \mathrm{d}s$ is a continuous martingale and $\int_0^T |K_s| \mathrm{d}s < \infty$ almost surely then $M_t \equiv 0$.*

Proof. Assume that $\int_0^T |K_s| \mathrm{d}s \leq C$ for some $C < \infty$. Then for a sequence of partitions $(\Pi_n = \{0 = t_0 < t_1 < \ldots < t_n = T\})$ of $[0, T]$

$$\begin{aligned}
\mathbf{E} \sum_{i=0}^{n-1} (M_{t_{i+1}} - M_{t_i})^2 &\leq \mathbf{E} \sup_{0 \leq i \leq n-1} |M_{t_{i+1}} - M_{t_i}| \int_0^T |K_s| \mathrm{d}s \\
&\leq C \mathbf{E} \sup_{0 \leq i \leq n-1} |M_{t_{i+1}} - M_{t_i}| \to 0,
\end{aligned}$$

as $\|\Pi_n\| \to 0$. We used that continuous function is uniformly continuous on compacts and Lebesgue's dominated convergence can be used because of the boundedness.

Furthermore,

$$\begin{aligned}
\mathbf{E}(M_t - M_s)^2 &= \mathbf{E} M_t^2 + \mathbf{E} M_s^2 - 2 \mathbf{E} \left(\mathbf{E}[M_t M_s | \mathcal{F}_s] \right) \\
&= \mathbf{E} M_t^2 - \mathbf{E} M_s^2,
\end{aligned}$$

for $s < t$, thus

$$\mathbf{E} \sum_{i=0}^{n-1} (M_{t_{i+1}} - M_{t_i})^2 = \mathbf{E}(M_t^2 - M_0^2) = \mathbf{E} M_t^2.$$

Therefore $\mathbf{E} M_t^2 = 0$ for all t, and the statement follows. □

Corollary 9. *Representation* (14) *is unique.*

Proof. Indeed, if

$$\int_0^t K_s \mathrm{d}s + \int_0^t H_s \mathrm{d}W_s = \int_0^t L_s \mathrm{d}s + \int_0^t G_s \mathrm{d}W_s,$$

then
$$\int_0^t (K_s - L_s)\mathrm{d}s = \int_0^t (G_s - H_s)\mathrm{d}W_s.$$
The RHS is a continuous martingale, therefore by the previous lemma it has to be constant 0. $\square$

In what follows we use the *notation*
$$\mathrm{d}X_t = K_t\mathrm{d}t + H_t\mathrm{d}W_t.$$

Theorem 32 (Itô formula (1944))**.** *Let $X_t = X_0 + \int_0^t K_s\mathrm{d}s + \int_0^t H_s\mathrm{d}W_s$ be an Itô process and $f \in C^2$. Then*
$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)H_s^2\mathrm{d}s.$$

That is $(f(X_t))$ is an Itô process too, with representation (14)
$$f(X_t) = f(X_0) + \int_0^t \left(f'(X_s)K_s + \frac{1}{2}f''(X_s)H_s^2\right)\mathrm{d}s + \int_0^t f'(X_s)H_s\mathrm{d}W_s.$$

Example 12. We already calculated the stochastic integral $\int W_s\mathrm{d}W_s$ in Example 10. Now we determine it again.

The SBM as an Itô process can be represented with $K_s \equiv 0$, $H_s \equiv 1$. Let $f(x) = x^2$. Then
$$W_t^2 = W_0^2 + \int_0^t 2W_s\mathrm{d}W_s + \frac{1}{2}\int_0^t 2\mathrm{d}s.$$
From this we obtain
$$\int_0^t W_s\mathrm{d}W_s = \frac{W_t^2 - t}{2}.$$
We see immediately that $W_t^2 - t$ is martingale.

Proof. We only prove under the following extra assumptions: f is compactly supported; $\sup_{s,\omega}|K_s(\omega)| < K$, $\sup_{s,\omega}|H_s(\omega)| < K$ for some $K < \infty$. (This is not an essential restriction.)

Take $\Pi = \{0 = t_0 < t_1 < \ldots < t_m = T\}$. Using the Taylor formula

$$\begin{aligned}
f(X_t) - f(X_0) &= \sum_{k=1}^{m} \left[f(X_{t_k}) - f(X_{t_{k-1}}) \right] \\
&= \sum_{k=1}^{m} f'(X_{t_{k-1}})(X_{t_k} - X_{t_{k-1}}) + \frac{1}{2} \sum_{k=1}^{m} f''(\eta_k)(X_{t_k} - X_{t_{k-1}})^2 \\
&= \sum_{k=1}^{m} f'(X_{t_{k-1}}) \int_{t_{k-1}}^{t_k} K_s \mathrm{d}s + \sum_{k=1}^{m} f'(X_{t_{k-1}}) \int_{t_{k-1}}^{t_k} H_s \mathrm{d}W_s \\
&\quad + \frac{1}{2} \sum_{k=1}^{m} f''(\eta_k)(X_{t_k} - X_{t_{k-1}})^2 \\
&= I_1 + I_2 + I_3,
\end{aligned}$$

where $\eta_k(\omega)$ is between $X_{t_{k-1}}(\omega)$ and $X_{t_k}(\omega)$.

It is easy to handle I_1. As f' and X_t are continuous

$$I_1 = \sum_{k=1}^{m} f'(X_{t_{k-1}}) \int_{t_{k-1}}^{t_k} K_s \mathrm{d}s \longrightarrow \int_0^t f'(X_s) K_s \mathrm{d}s \quad \text{a.s.}, \tag{15}$$

as $\|\Pi\| \to 0$.

Rewrite I_2 as

$$I_2 = \sum_{k=1}^{m} f'(X_{t_{k-1}}) \int_{t_{k-1}}^{t_k} H_s \mathrm{d}W_s = \int_0^t \sum_{k=1}^{m} f'(X_{t_{k-1}}) \mathbf{I}_{(t_{k-1}, t_k]}(s) H_s \mathrm{d}W_s.$$

As $\|\Pi\| \to 0$

$$\mathbf{E} \int_0^t \left(f'(X_s) H_s - \sum_{k=1}^{m} f'(X_{t_{k-1}}) \mathbf{I}_{(t_{k-1}, t_k]}(s) H_s \right)^2 \mathrm{d}s \to 0.$$

Indeed, for any $\omega \in \Omega$ fix the integrand is bounded and by continuity goes to 0, therefore the dominated Lebesgue convergence theorem applies. Theorem 31 (ii) implies

$$I_2 = \int_0^t \sum_{k=1}^{m} f'(X_{t_{k-1}}) I_{(t_{k-1}, t_k]}(s) H_s \mathrm{d}W_s \xrightarrow{L^2} \int_0^t f'(X_s) H_s \mathrm{d}W_s. \tag{16}$$

Next comes I_3, the difficult part. We have to show that

$$I_3 \to \frac{1}{2}\int_0^t f''(X_s)H_s^2 \mathrm{d}s.$$

Write

$$\begin{aligned}
(X_{t_k} - X_{t_{k-1}})^2 &= \left(\int_{t_{k-1}}^{t_k} K_s \mathrm{d}s + \int_{t_{k-1}}^{t_k} H_s \mathrm{d}W_s\right)^2 \\
&= \left(\int_{t_{k-1}}^{t_k} K_s \mathrm{d}s\right)^2 + 2\int_{t_{k-1}}^{t_k} K_s \mathrm{d}s \cdot \int_{t_{k-1}}^{t_k} H_s \mathrm{d}W_s \\
&\quad + \left(\int_{t_{k-1}}^{t_k} H_s \mathrm{d}W_s\right)^2.
\end{aligned}$$

We show that the contribution of the first two terms is negligible to the whole sum. For the first

$$\left|\sum_{k=1}^m f''(\eta_k)\left(\int_{t_{k-1}}^{t_k} K_s \mathrm{d}s\right)^2\right| \le \|f''\|_\infty \cdot K^2 \sum_{k=1}^m (t_k - t_{k-1})^2 \to 0 \quad \text{a.s.} \tag{17}$$

To handle the second introduce $M_t = \int_0^t H_s \mathrm{d}W_s$. Then

$$\begin{aligned}
&\left|\sum_{k=1}^m f''(\eta_k)\int_{t_{k-1}}^{t_k} K_s \mathrm{d}s \cdot \int_{t_{k-1}}^{t_k} H_s \mathrm{d}W_s\right| \\
&\le \|f''\|_\infty \cdot K \sup_{1\le k\le m} |M_{t_k} - M_{t_{k-1}}| \cdot \sum_{k=1}^m (t_k - t_{k-1}) \\
&= \|f''\|_\infty \cdot K\, t \sup_{1\le k\le m} |M_{t_k} - M_{t_{k-1}}| \to 0, \quad \text{a.s.},
\end{aligned} \tag{18}$$

since $M_t = \int_0^t H_s \mathrm{d}W_s$ is a continuous martingale.

We have to deal with the sum

$$\sum_{k=1}^m f''(\eta_k)\left(\int_{t_{k-1}}^{t_k} H_s \mathrm{d}W_s\right)^2.$$

First we change η_k to $X_{t_{k-1}}$. Taking the difference

$$\begin{aligned}
&\sum_{k=1}^{m}[f''(\eta_k) - f''(X_{t_{k-1}})](M_{t_k} - M_{t_{k-1}})^2 \\
&\leq \sup_{1\leq k\leq m} |f''(\eta_k) - f''(X_{t_{k-1}})| \cdot \sum_{k=1}^{m}(M_{t_k} - M_{t_{k-1}})^2.
\end{aligned}$$

By the Cauchy–Schwarz inequality

$$\begin{aligned}
&\left|\mathbf{E}\sum_{k=1}^{m}[f''(\eta_k) - f''(X_{t_{k-1}})](M_{t_k} - M_{t_{k-1}})^2\right| \\
&\leq \sqrt{\mathbf{E}\sup_{1\leq k\leq m}(f''(\eta_k) - f''(X_{t_{k-1}}))^2}\sqrt{\mathbf{E}\left(\sum_{k=1}^{m}(M_{t_k} - M_{t_{k-1}})^2\right)^2}.
\end{aligned} \tag{19}$$

The first term tends to 0 because (X_t) is continuous and f'' is bounded. The second is bounded by the following lemma.

Lemma 8. *Let (M_t) be a continuous bounded martingale on $[0,t]$, that is $\sup_{s,\omega}|M_s(\omega)| \leq K$, and let $\Pi = \{0 = t_0 < t_1 < \ldots < t_m = t\}$ be a partition. Then*

$$\mathbf{E}\left(\sum_{i=1}^{m}(M_{t_i} - M_{t_{i-1}})^2\right)^2 \leq 6K^4.$$

Proof. Expanding the square

$$\begin{aligned}
&\mathbf{E}\left(\sum_{i=1}^{m}(M_{t_i} - M_{t_{i-1}})^2\right)^2 \\
&= \sum_{i=1}^{m}\mathbf{E}(M_{t_i} - M_{t_{i-1}})^4 + \sum_{i\neq j}\mathbf{E}(M_{t_i} - M_{t_{i-1}})^2(M_{t_j} - M_{t_{j-1}})^2.
\end{aligned}$$

Using several times that

$$\mathbf{E}[(M_t - M_s)^2|\mathcal{F}_s] = \mathbf{E}[M_t^2 - M_s^2|\mathcal{F}_s], \quad s < t,$$

we obtain

$$
\begin{aligned}
&\sum_{i\neq j}\mathbf{E}(M_{t_i}-M_{t_{i-1}})^2(M_{t_j}-M_{t_{j-1}})^2\\
&=2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\mathbf{E}(M_{t_i}-M_{t_{i-1}})^2(M_{t_j}-M_{t_{j-1}})^2\\
&=2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\mathbf{E}\left[\mathbf{E}[(M_{t_i}-M_{t_{i-1}})^2(M_{t_j}-M_{t_{j-1}})^2|\mathcal{F}_{t_{j-1}}]\right]\\
&=2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\mathbf{E}(M_{t_i}-M_{t_{i-1}})^2(M_{t_j}^2-M_{t_{j-1}}^2)\\
&=2\sum_{i=1}^{m-1}\mathbf{E}(M_{t_i}-M_{t_{i-1}})^2(M_t^2-M_{t_i}^2)\\
&\leq 2K^2\sum_{i=1}^{m-1}\mathbf{E}(M_{t_i}-M_{t_{i-1}})^2\\
&=2K^2\sum_{i=1}^{m-1}\mathbf{E}(M_{t_i}^2-M_{t_{i-1}}^2)\leq 2K^4.
\end{aligned}
$$

While, for the sum of 4th powers

$$
\begin{aligned}
\sum_{i=1}^{m}\mathbf{E}(M_{t_i}-M_{t_{i-1}})^4 &\leq 4K^2\mathbf{E}\sum_{i=1}^{m}\mathbf{E}(M_{t_i}-M_{t_{i-1}})^2\\
&=4K^2\mathbf{E}(M_t^2-M_0^2)\leq 4K^4.
\end{aligned}
$$

□

Summarizing from I_3 we have the sum

$$\sum_{k=1}^{m}f''(X_{t_{k-1}})(M_{t_k}-M_{t_{k-1}})^2.$$

We claim that

$$\sum_{k=1}^{m}f''(X_{t_{k-1}})(M_{t_k}-M_{t_{k-1}})^2\xrightarrow{L^1}\int_0^t f''(X_s)H_s^2\mathrm{d}s. \tag{20}$$

Since X and f'' are continuous

$$\sum_{k=1}^{m} f''(X_{t_{k-1}}) \int_{t_{k-1}}^{t_k} H_s^2 \mathrm{d}s \to \int_0^t f''(X_s) H_s^2 \mathrm{d}s \quad \text{a.s.}$$

Thus it is enough to show that

$$\sum_{k=1}^{m} f''(X_{t_{k-1}}) \left((M_{t_k} - M_{t_{k-1}})^2 - \int_{t_{k-1}}^{t_k} H_s^2 \mathrm{d}s \right) \xrightarrow{L^2} 0.$$

Theorem 31 (ii) implies

$$\begin{aligned}
\mathbf{E}\left[(M_{t_k} - M_{t_{k-1}})^2 | \mathcal{F}_{t_{k-1}}\right] &= \mathbf{E}\left[\left(\int_{t_{k-1}}^{t_k} H_s \,\mathrm{d}W_s \right)^2 | \mathcal{F}_{t_{k-1}} \right] \\
&= \mathbf{E}\left[\int_{t_{k-1}}^{t_k} H_s^2 \,\mathrm{d}s | \mathcal{F}_{t_{k-1}} \right],
\end{aligned}$$

so in

$$\mathbf{E}\left(\sum_{k=1}^{m} f''(X_{t_{k-1}}) \left((M_{t_k} - M_{t_{k-1}})^2 - \int_{t_{k-1}}^{t_k} H_s^2 \mathrm{d}s \right) \right)^2$$

the expectation of the mixed term is 0. Thus this equals

$$\begin{aligned}
&= \mathbf{E} \sum_{k=1}^{m} f''(X_{t_{k-1}})^2 \left((M_{t_k} - M_{t_{k-1}})^2 - \int_{t_{k-1}}^{t_k} H_s^2 \mathrm{d}s \right)^2 \\
&\leq \|f\|_\infty^2 \Bigg[\mathbf{E} \sum_{k=1}^{m} (M_{t_k} - M_{t_{k-1}})^4 + 2\mathbf{E} \sum_{k=1}^{m} (M_{t_k} - M_{t_{k-1}})^2 \int_{t_{k-1}}^{t_k} H_s^2 \mathrm{d}s \\
&\qquad + \mathbf{E} \sum_{k=1}^{m} \left(\int_{t_{k-1}}^{t_k} H_s^2 \mathrm{d}s \right)^2 \Bigg] \\
&\leq \|f\|_\infty^2 \left[\mathbf{E} \sum_{k=1}^{m} (M_{t_k} - M_{t_{k-1}})^4 + 2K^2 t \mathbf{E} \sup_{1 \leq k \leq m} (M_{t_k} - M_{t_{k-1}})^2 + K^4 t \|\Pi\| \right].
\end{aligned}$$

The second and third term tend to 0, and for the first

$$\begin{aligned}
\mathbf{E}\sum_{k=1}^{m}(M_{t_k}-M_{t_{k-1}})^4 &\leq \mathbf{E}\left[\sum_{k=1}^{m}(M_{t_k}-M_{t_{k-1}})^2\cdot\sup_{1\leq k\leq m}|M_{t_k}-M_{t_{k-1}}|^2\right] \\
&\leq \sqrt{\mathbf{E}\left[\sum_{k=1}^{m}(M_{t_k}-M_{t_{k-1}})^2\right]^2}\sqrt{\mathbf{E}\sup_{1\leq k\leq m}|M_{t_k}-M_{t_{k-1}}|^4} \\
&\leq \sqrt{6}K^2\sqrt{\mathbf{E}\sup_{1\leq k\leq m}|M_{t_k}-M_{t_{k-1}}|^4}\to 0.
\end{aligned}$$

Summarizing we obtained L^1, L^2 and almost sure convergence in (15)–(20). Since everything is bounded, L^1 convergence follows in each case, that is

$$\begin{aligned}
f(X_t)-f(X_0) &= \sum_{k=1}^{m}[f(X_{t_k})-f(X_{t_{k-1}})] \\
&\xrightarrow{L^1}\int_0^t f'(X_s)\mathrm{d}X_s+\frac{1}{2}\int_0^t f''(X_s)H_s^2\mathrm{d}s.
\end{aligned}$$

Convergence in L^1 implies a.s. convergence on a subsequence. As both sides are continuous we obtained that the two process are indistinguishable. □

Example 13 (Continuation of Example 11)**.** Let

$$\zeta_t^s=\int_s^t X_u\mathrm{d}W_u-\frac{1}{2}\int_s^t X_u^2\mathrm{d}u,\quad \zeta_t=\zeta_t^0,$$

where X_t is an adapted process. Then $Z_t=e^{\zeta_t}$ satisfies the stochastic differential equation

$$Z_t=1+\int_0^t Z_sX_s\mathrm{d}W_s,$$

or with a common notation

$$\mathrm{d}Z_t=Z_tX_t\mathrm{d}W_t,\quad Z_0=1.$$

Writing ζ as an Itô process

$$\zeta_t=\int_0^t -\frac{1}{2}X_u^2\mathrm{d}u+\int_0^t X_u\mathrm{d}W_u.$$

Using Itô's formula with $f(x) = e^x$

$$\begin{aligned}
Z_t = e^{\zeta_t} &= 1 + \int_0^t e^{\zeta_s} \mathrm{d}\zeta_s + \frac{1}{2} \int_0^t e^{\zeta_s} X_s^2 \mathrm{d}s \\
&= 1 + \int_0^t e^{\zeta_s} \left(-\frac{1}{2} X_s^2 \mathrm{d}s + X_s \mathrm{d}W_s \right) + \frac{1}{2} \int_0^t e^{\zeta_s} X_s^2 \mathrm{d}s \\
&= 1 + \int_0^t e^{\zeta_s} X_s \mathrm{d}W_s \\
&= 1 + \int_0^t Z_s X_s \mathrm{d}W_s,
\end{aligned}$$

as claimed. We see that Z_t is martingale.

Exercise 28. Let ζ_t be as above. Show that $Y_t = e^{-\zeta_t}$ satisfies the SDE

$$\mathrm{d}Y_t = Y_t X_t^2 \mathrm{d}t - X_t Y_t \mathrm{d}W_t, \quad Y_0 = 1.$$

Similarly, one can show a more general version, where f depends on the time variable t.

Theorem 33 (More general Itô formula)**.** *Let X_t be an Itô process and $f \in C^{1,2}$. Then*

$$\begin{aligned}
f(t, X_t) = f(0, X_0) &+ \int_0^t \frac{\partial}{\partial s} f(s, X_s) \mathrm{d}s + \int_0^t \frac{\partial}{\partial x} f(s, X_s) \mathrm{d}X_s \\
&+ \frac{1}{2} \int_0^t \frac{\partial^2}{\partial x^2} f(s, X_s) H_s^2 \mathrm{d}s.
\end{aligned}$$

4.4 Multidimensional Itô processes

Let $W = (W^1, W^2, \dots, W^r)$ be an r-dimensional SBM, that is its component are iid SBM's. Then (X_t) is a *d-dimensional Itô process*, if

$$X_t^i = X_0^i + \int_0^t K_s^i \mathrm{d}s + \sum_{j=1}^r \int_0^t H_s^{i,j} \mathrm{d}W_s^j, \tag{21}$$

where $\int_0^T |K_s^i| \mathrm{d}s < \infty, \int_0^T (H_s^{i,j})^2 \mathrm{d}s < \infty$ a.s., and $K^i, H^{i,j}$ are $\mathcal{F}_t$-adapted, $i = 1, 2, \dots, d$, $j = 1, 2, \dots, r$.

Theorem 34 (Multidimensional Itô formula)**.** *Let (X_t) be a multidimensional Itô process and $f : \mathbb{R}^{1+d} \to \mathbb{R}$, $f \in C^{1,2}$. Then*

$$
\begin{aligned}
f(t, X_t^1, \ldots, X_t^d) = f(0, X_0^1, \ldots, X_0^d) &+ \int_0^t \frac{\partial}{\partial s} f(s, X_s^1, \ldots, X_s^d)\,\mathrm{d}s \\
&+ \sum_{i=1}^d \int_0^t \frac{\partial}{\partial x_i} f(s, X_s^1, \ldots, X_s^d)\,\mathrm{d}X_s^i \\
&+ \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2}{\partial x_i \partial x_j} f(s, X_s^1, \ldots, X_s^d) \sum_{k=1}^r H_s^{i,k} H_s^{j,k}\,\mathrm{d}s.
\end{aligned}
$$

4.5 Applications

Example 14 (Integration by parts I)**.** Let (X, Y) be a two-dimensional Itô process with representation

$$
\begin{aligned}
X_t &= X_0 + \int_0^t K_s\,\mathrm{d}s + \int_0^t H_s\,\mathrm{d}W_s \\
Y_t &= Y_0 + \int_0^t L_s\,\mathrm{d}s + \int_0^t G_s\,\mathrm{d}W_s,
\end{aligned}
$$

where K, L, H, G are as usual. Then

$$\int_0^t X_s \mathrm{d}Y_s = X_t Y_t - X_0 Y_0 - \int_0^t Y_s \mathrm{d}X_s - \int_0^t H_s G_s \mathrm{d}s.$$

Note that in the deterministic integration by parts formula the last term is missing.

For the proof apply Itô's formula for (X, Y) and $f(x, y) = xy$. Then

$$r = 1,\ d = 2,\ K_s^1 = K_s,\ K_s^2 = L_s,\ H_s^{1,1} = H_s,\ H_s^{2,1} = G_s.$$

Since $\frac{\partial f}{\partial x} = y$, $\frac{\partial f}{\partial y} = x$, $\frac{\partial^2 f}{\partial^2 x} = \frac{\partial^2 f}{\partial^2 y} = 0$, and $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 1$, we obtain

$$X_t Y_t = X_0 Y_0 + \int_0^t Y_s \mathrm{d}X_s + \int_0^t X_s \mathrm{d}Y_s + \frac{1}{2} 2 \int_0^t H_s G_s \mathrm{d}s,$$

as claimed.

Example 15 (Integration by parts II)**.** To change a bit let $\widetilde{W}$ be another SBM independent of W and (X, Y)

$$
\begin{aligned}
X_t &= X_0 + \int_0^t K_s \,\mathrm{d}s + \int_0^t H_s \,\mathrm{d}W_s \\
Y_t &= Y_0 + \int_0^t L_s \,\mathrm{d}s + \int_0^t G_s \,\mathrm{d}\widetilde{W}_s.
\end{aligned}
$$

Then

$$\int_0^t X_s \mathrm{d}Y_s = X_t Y_t - X_0 Y_0 - \int_0^t Y_s \mathrm{d}X_s.$$

The proof is the same but here $d = r = 2$, and no extra term appears.

Example 16 (Geometric Brownian motion)**.** Let $\mu \in \mathbb{R}$, $\sigma > 0$. Solve the SDE

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t. \tag{22}$$

We have

$$X_t = X_0 + \int_0^t \mu X_s \mathrm{d}s + \int_0^t \sigma X_s \mathrm{d}W_s.$$

Applying Itô's formula with $f(x) = \log x$

$$
\begin{aligned}
\log X_t &= \log X_0 + \int_0^t \frac{1}{X_s} \left(\mu X_s \mathrm{d}s + \sigma X_s \mathrm{d}W_s \right) + \frac{1}{2} \int_0^t -\frac{1}{X_s^2} \sigma^2 X_s^2 \mathrm{d}s \\
&= \log X_0 + \sigma W_t + \left(\mu - \frac{\sigma^2}{2} \right) t.
\end{aligned}
$$

Thus

$$X_t = X_0 \cdot e^{\sigma W_t + \left(\mu - \frac{\sigma^2}{2} \right) t}. \tag{23}$$

This is martingale iff $\mu = 0$.

Note that $\log x$ is not defined at 0, so the proof is not complete. It only gives us a potential solution.

Exercise 29. Show that X_t in (23) is indeed a solution to the SDE (22).

A more constructive solution is to apply Itô's formula with a general f, and then choose f to obtain a simple equation. With $f(x) = \log x$ the integrand in the martingale part is constant.

Exercise 30. Show that $Y(t) = e^{t/2} \cos W_t$ is martingale.

Exercise 31. Show that

$$\int_0^t W_s^2 \mathrm{d}W_s = \frac{1}{3}W_t^3 - \int_0^t W_s \mathrm{d}s,$$

and

$$\int_0^t W_s^3 \mathrm{d}W_s = \frac{1}{4}W_t^4 - \frac{3}{2}\int_0^t W_s^2 \mathrm{d}s.$$

Exercise 32. Let $\mathbf{W} = (W^1, \ldots, W^r)$ be an r-dimensional SBM, $r \geq 2$, and let

$$R_t = \sqrt{\sum_{i=1}^r (W_t^i)^2}.$$

Show that R satisfies the SDE

$$\mathrm{d}R_t = \frac{r-1}{2R_t}\mathrm{d}t + \sum_{i=1}^r \frac{W_t^i}{R_t}\mathrm{d}W_t^i.$$

This is the Bessel equation and R is the *Bessel process*.

4.6 Quadratic variation and the Doob–Meyer decomposition

We proved that

$$\mathbf{E}\left[\left(\int_s^t X_u \mathrm{d}W_u\right)^2 |\mathcal{F}_s\right] = \mathbf{E}\left[\int_s^t X_u^2 \,\mathrm{d}u \big| \mathcal{F}_s\right],$$

which means that the process

$$\left(\int_0^t X_u \,\mathrm{d}W_u\right)^2 - \int_0^t X_u^2 \,\mathrm{d}u \tag{24}$$

is a continuous martingale. In the decomposition

$$\left(\int_0^t X_u \mathrm{d}W_u\right)^2 = \int_0^t X_u^2 \,\mathrm{d}u + \left(\int_0^t X_u \mathrm{d}W_u\right)^2 - \int_0^t X_u^2 \,\mathrm{d}u$$

the first term is an increasing process and the second term is a martingale, that is we obtained the Doob–Meyer decomposition of $I_t(X)^2$.

On the other hand, at the proof of Itô's formula we showed (see (20)) that

$$\sum_{i=1}^{n} \left(\int_{t_{i-1}}^{t_i} X_u \mathrm{d}W_u \right)^2 \xrightarrow{L^1} \int_0^t X_u^2 \mathrm{d}u, \quad \text{as} \quad \|\Pi_n\| \to 0.$$

The left-hand side is exactly the *quadratic variation process* of the martingale $I_t(X)$.

Summarizing, we proved the following.

Theorem 35. *For any Itô process X_t, the quadratic variation of $I_t(X)$ and the increasing process in the Doob–Meyer decomposition of $I_t(X)^2$ are the same.*

This result holds in a more general setup.

Let (X_t) be a (continuous) square integrable martingale, $X \in \mathcal{M}_2$ (or $X \in \mathcal{M}_2^c$). Then X_t^2 is a submartingale, so by the Doob–Meyer decomposition there exists a unique (up to indistinguishibility) adapted increasing process A_t, such that $A_0 = 0$ a.s. and $X_t^2 - A_t$ is a martingale. The process $\langle X \rangle_t = A_t$ is the *quadratic variation of* X.

With this notation, Theorem 35 states that

$$\left\langle \int_0^{\cdot} X_u \mathrm{d}W_u \right\rangle_t = \langle I(X) \rangle_t = \int_0^t X_u^2 \, \mathrm{d}u.$$

Without proof we mention that Theorem 35 holds not only for Itô processes but for *continuous square integrable martingales*.

Theorem 36. *Let $X \in \mathcal{M}_2^c$. For partition Π of $[0,t]$ we have*

$$V_t^{(2)}(\Pi) := \sum_{k=1}^{n} (X_{t_k} - X_{t_{k-1}})^2 \xrightarrow{\mathbf{P}} \langle X \rangle_t \quad \textit{as} \quad \|\Pi\| \to 0.$$

For square integrable martingales X, Y the *crossvariation process of X and Y* is

$$\langle X, Y \rangle_t = \frac{1}{4} \left(\langle X + Y \rangle_t - \langle X - Y \rangle_t \right).$$

The processes X and Y are *orthogonal* if $\langle X, Y \rangle_t = 0$ a.s. for any t.

Exercise 33. Show that if $X, Y \in \mathcal{M}_2$, then $XY - \langle X, Y \rangle$ is a martingale.

One can define stochastic integral with respect to more general processes. The process (X_t) is a continuous *semimartingale* if

$$X_t = M_t + A_t,$$

where M_t is a continuous martingale and A_t is of bounded variation, and both are adapted. As in Lemma 7 it can be shown that this decomposition is essentially unique.

We can define stochastic integral with respect to semimartingales. Indeed, integral with respect to A_t can be defined pathwise, since A is of bounded variation, and integration with respect to continuous M_t can be defined similarly as for SBM.

The following version of Itô's formula holds.

Theorem 37 (Itô formula for semimartingales)**.** *Let $X_t = M_t + A_t$ be a continuous semimartingale, and let $f \in C^2$. Then*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)\mathrm{d}\langle M\rangle_s.$$

5 Stochastic differential equations

5.1 Existence and uniqueness

We define the strong solution of SDEs and obtain existence and uniqueness results.

The followings are given:

- probability space $(\Omega, \mathcal{A}, \mathbf{P})$;
- with a filtration $(\mathcal{F}_t)_{t\in[0,T]}$;
- a d-dimensional SBM $W_t = (W_t^1, \ldots, W_t^r)$ with respect to the filtration $(\mathcal{F}_t)$;
- measurable functions $f : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times [0,T] \to \mathbb{R}^{d\times r}$;
- $\mathcal{F}_0$-measurable rv $\xi : \Omega \to \mathbb{R}^d$.

The (d-dimensional) process (X_t) is *strong solution to the SDE*

$$\begin{aligned} \mathrm{d}X_t &= f(X_t,t)\,\mathrm{d}t + \sigma(X_t,t)\,\mathrm{d}W_t, \\ X_0 &= \xi, \end{aligned} \tag{25}$$

if $\int_0^t f(X_s, s)\mathrm{d}s$ are $\int_0^t \sigma(X_s, s)\mathrm{d}W_s$ well-defined for all $t \in [0, T]$ and the integral version of (25) holds, i.e.

$$X_t = \xi + \int_0^t f(X_s, s)\,\mathrm{d}s + \int_0^t \sigma(X_s, s)\,\mathrm{d}W_s, \quad \text{for all } t \in [0, T] \text{ a.s.}$$

Written coordinatewise

$$X_t^i = \xi^i + \int_0^t f^i(X_s, s)\,\mathrm{d}s + \int_0^t \sum_{j=1}^r \sigma_{i,j}(X_s, s)\,\mathrm{d}W_s^j, \quad i = 1, 2, \ldots, d.$$

It is important to emphasize that with strong solutions not only the SDE (25) is given, but the driving SBM, the initial condition (not just distribution!) ξ and the filtration.

For d-dimensional vectors $|x| = \sqrt{x_1^2 + \ldots + x_d^2}$ stands for the usual Euclidean norm, and for a matrix $\sigma \in \mathbb{R}^{d\times r}$, define $|\sigma| = \sqrt{\sum_{i,j} \sigma_{ij}^2}$,

Theorem 38. *Assume that for the functions in* (25) *the following hold:*

$$\begin{aligned} &|f(x,t) - f(y,t)| + |\sigma(x,t) - \sigma(y,t)| \le K|x-y|, \\ &|f(x,t)|^2 + |\sigma(x,t)|^2 \le K_0(1 + |x|^2), \\ &\mathbf{E}|\xi|^2 < \infty. \end{aligned}$$

Then (25) *has a unique strong solution X, and*

$$\mathbf{E} \sup_{0\le t\le T} |X_t|^2 \le C(1 + \mathbf{E}|\xi|^2).$$

Proof. We only prove for $d = r = 1$. The general case is similar, but notationally messy.

Recall the following statement from the theory of ordinary differential equations.

Lemma 9 (Gronwall–Bellman)**.** *Let α, β be integrable functions for which*

$$\alpha(t) \le \beta(t) + H \int_a^t \alpha(s)\,\mathrm{d}s, \quad t \in [a, b],$$

for some $H \ge 0$. Then

$$\alpha(t) \le \beta(t) + H \int_a^t e^{H(t-s)}\beta(s)\,\mathrm{d}s.$$

Uniqueness. Let X_t, Y_t be solutions. Then

$$X_t - Y_t = \int_0^t (f(X_s,s) - f(Y_s,s))\,\mathrm{d}s + \int_0^t (\sigma(X_s,s) - \sigma(Y_s,s))\,\mathrm{d}W_s.$$

Since $(a+b)^2 \leq 2a^2 + 2b^2$, by Theorem 31 (ii) and the Cauchy–Schwarz inequality

$$\begin{aligned}
\mathbf{E}(X_t - Y_t)^2 &\leq 2\,\mathbf{E}\left(\int_0^t (f(X_s,s) - f(Y_s,s))\mathrm{d}s\right)^2 \\
&\quad + 2\,\mathbf{E}\int_0^t (\sigma(X_s,s) - \sigma(Y_s,s))^2\mathrm{d}s \\
&\leq 2(T+1)K^2 \int_0^t \mathbf{E}(X_s - Y_s)^2\,\mathrm{d}s.
\end{aligned}$$

With the notation $\varphi(t) = \mathbf{E}(X_t - Y_t)^2$ we obtained

$$\varphi(t) \leq 2(T+1)K^2 \int_0^t \varphi(s)\,\mathrm{d}s.$$

By the Gronwall–Bellman lemma $\varphi(t) \equiv 0$, i.e. $X_t = Y_t$ a.s. Since $X_t - Y_t$ is continuous, the two processes are indistinguishable, meaning

$$\mathbf{P}(X_t = Y_t,\ \forall t \in [0,T]) = 1.$$

Thus the uniqueness is proved.
Existence. Sketch. The proof goes similarly as the proof of the Picard–Lindelöf theorem for ODEs. We do Picard iteration. Let $X_t^{(0)} \equiv \xi$, and if $X_t^{(n)}$ is given, let

$$X_t^{(n+1)} = \xi + \int_0^t f(X_s^{(n)}, s)\mathrm{d}s + \int_0^t \sigma(X_s^{(n)}, s)\mathrm{d}W_s.$$

Write

$$\begin{aligned}
X_t^{(n+1)} - X_t^{(n)} &= \int_0^t \big(f(X_s^{(n)}, s) - f(X_s^{(n-1)}, s)\big)\,\mathrm{d}s \\
&\quad + \int_0^t \big(\sigma(X_s^{(n)}, s) - \sigma(X_s^{(n-1)}, s)\big)\,\mathrm{d}W_s \\
&=: B_t^{(n)} + M_t^{(n)}.
\end{aligned}$$

By Doob's maximal inequality, as in the proof of uniqueness

$$\mathbf{E}\left(\sup_{s\in[0,t]}(M_s^{(n)})^2\right) \leq 4\mathbf{E}\int_0^t \left(\sigma(X_s^{(n)},s)-\sigma(X_s^{(n-1)},s)\right)^2 \mathrm{d}s$$

$$\leq 4K^2\int_0^t \mathbf{E}(X_s^{(n)}-X_s^{(n-1)})^2\,\mathrm{d}s.$$

On the other hand, by Cauchy–Schwarz

$$\mathbf{E}\left(\sup_{s\in[0,t]}(B_s^{(n)})^2\right) \leq tK^2\,\mathbf{E}\int_0^t \left(X_s^{(n)}-X_s^{(n-1)}\right)^2 \mathrm{d}s.$$

This implies

$$\mathbf{E}\left(\sup_{s\in[0,t]}(X_s^{(n+1)}-X_s^{(n)})^2\right) \leq L\int_0^t \mathbf{E}(X_s^{(n)}-X_s^{(n-1)})^2\mathrm{d}s,$$

with $L=2(T+4)K^2$. Iterating and changing the order of integration

$$\begin{aligned}\mathbf{E}\left(\sup_{s\in[0,t]}(X_s^{(n+1)}-X_s^{(n)})^2\right) &\leq L\int_0^t \mathbf{E}(X_s^{(n)}-X_s^{(n-1)})^2\,\mathrm{d}s\\ &\leq L^2\int_0^t\int_0^s \mathbf{E}(X_u^{(n-1)}-X_u^{(n-2)})^2\,\mathrm{d}u\,\mathrm{d}s\\ &\leq L^2\int_0^t (t-s)\mathbf{E}(X_s^{(n-1)}-X_s^{(n-2)})^2\,\mathrm{d}s.\end{aligned}$$

Continuing, and using the assumption on ξ we obtain

$$\begin{aligned}&\mathbf{E}\left(\sup_{s\in[0,t]}(X_s^{(n+1)}-X_s^{(n)})^2\right)\\ &\leq L^n\int_0^t \frac{(t-s)^{n-1}}{(n-1)!}\mathbf{E}(X_s^1-\xi)^2\,\mathrm{d}s \leq C\frac{(LT)^n}{n!}.\end{aligned}$$

By Chebyshev

$$\sum_{n=1}^{\infty}\mathbf{P}\left(\sup_{0\leq t\leq T}|X_t^{(n+1)}-X_t^n|>n^{-2}\right) \leq \sum_{n=1}^{\infty} C'n^4\frac{(LT)^n}{n!}<\infty.$$

Therefore, applying the first Borel–Cantelli lemma the infinite sum

$$\sum_{n=0}^{\infty}(X_t^{(n+1)} - X_t^n)$$

converges a.s. Clearly the sum is a solution to the SDE (25). $\square$

5.2 Examples

Most of the examples and exercises are from Evans [4].

Example 17. Let g be a continuous function, and consider the SDE

$$\begin{cases} \mathrm{d}X_t = g(t)X_t \mathrm{d}W_t \\ X_0 = 1. \end{cases}$$

Show that the unique solution is

$$X_t = \exp\left\{-\frac{1}{2}\int_0^t g(s)^2 \mathrm{d}s + \int_0^t g(s)\mathrm{d}W_s\right\}.$$

The uniqueness follows from Theorem 38, assuming g is nice enough. To check that X_t is indeed a solution, we use Itô's formula. Let

$$Y_t = -\frac{1}{2}\int_0^t g(s)^2 \mathrm{d}s + \int_0^t g(s)\mathrm{d}W_s.$$

With $f(x) = e^x$, we have

$$\begin{aligned} X_t = e^{Y_t} &= 1 + \int_0^t e^{Y_s}\mathrm{d}Y_s + \frac{1}{2}\int_0^t e^{Y_s} g^2(s)\mathrm{d}s \\ &= 1 + \int_0^t X_s g(s)\mathrm{d}W_s, \end{aligned}$$

as claimed.

Exercise 34. Let f and g be continuous functions, and consider the SDE

$$\begin{cases} \mathrm{d}X_t = f(t)X_t \mathrm{d}t + g(t)X_t \mathrm{d}W_t \\ X_0 = 1. \end{cases}$$

Show that the unique solution is

$$X_t = \exp\left\{\int_0^t \left[f(s) - \frac{1}{2}g(s)^2\right]\mathrm{d}s + \int_0^t g(s)\mathrm{d}W_s\right\}.$$

Exercise 35 (Brownian bridge)**.** Show that

$$B_t = (1-t)\int_0^t \frac{1}{1-s}\mathrm{d}W_s$$

is the unique solution of the SDE

$$\begin{cases} \mathrm{d}B_t = -\frac{B_t}{1-t}\mathrm{d}t + \mathrm{d}W_t \\ B_0 = 0. \end{cases}$$

Calculate the mean and covariance function of B.

A mean zero Gaussian process B_t on $[0,1]$ is called *Brownian bridge* if its covariance function is

$$\mathbf{Cov}(B_s, B_t) = \min(s,t) - st.$$

Exercise 36. Show that if W is SBM then $B_t = W_t - tW_1$ is Brownian bridge.

Exercise 37. Solve the SDE

$$\begin{cases} \mathrm{d}X_t = -\frac{1}{2}e^{-2X_t}\mathrm{d}t + e^{-X_t}\mathrm{d}W_t \\ X(0) = 0 \end{cases}$$

and show that it explodes in a finite random time. *Hint: Look for a solution $X_t = u(W_t)$.*

Exercise 38. Solve the SDE

$$\mathrm{d}X_t = -X_t\mathrm{d}t + e^{-t}\mathrm{d}W_t.$$

Exercise 39. Show that $(X_t, Y_t) = (\cos W_t, \sin W_t)$ is a solution to the SDE

$$\begin{cases} \mathrm{d}X_t = -\frac{1}{2}X_t\mathrm{d}t - Y_t\mathrm{d}W_t \\ \mathrm{d}Y_t = -\frac{1}{2}Y_t\mathrm{d}t + X_t\mathrm{d}W_t. \end{cases}$$

Show that $\sqrt{X_t^2 + Y_t^2}$ is a constant for any solution (X, Y)!

Exercise 40. Solve the SDE

$$\begin{cases} \mathrm{d}X_t = \mathrm{d}t + \mathrm{d}W_t^{(1)} \\ \mathrm{d}Y_t = X_t \mathrm{d}W_t^{(2)}, \end{cases}$$

where $W^{(1)}$ and $W^{(2)}$ are independent SBMs.

Exercise 41. Solve the SDE

$$\begin{cases} \mathrm{d}X_t = Y_t \mathrm{d}t + \mathrm{d}W_t^{(1)} \\ \mathrm{d}Y_t = X_t \mathrm{d}t + \mathrm{d}W_t^{(2)}, \end{cases}$$

where $W^{(1)}$ and $W^{(2)}$ are independent SBMs.

6 General Markov processes

This part is from Breiman [1].

6.1 Transition probabilities and Chapman–Kolmogorov equations

The process (X_t) is a *Markov process*, if for each Borel set $B \in \mathcal{B}(\mathbb{R})$, and $t, \tau \mathbb{R}$

$$\mathbf{P}(X_{t+\tau} \in B | X_s, s \leq t) = \mathbf{P}(X_{t+\tau} \in B | X_t).$$

Choosing natural filtration $\mathcal{F}_t = \sigma(X_s, s \leq t)$, the definition is the same as in Subsection 3.4

Since regular conditional distributions exist, we may choose the probabilities

$$p_{t_2,t_1}(B|x) = \mathbf{P}(X_{t_2} \in B | X_{t_1} = x), \quad t_2 > t_1, B \in \mathcal{B},$$

such that

- for x fixed, $p_{t_2,t_1}(\cdot|x)$ is a probability measure;
- for $B \in \mathcal{B}$ fixed, $p_{t_2,t_1}(B|\cdot)$ is measurable.

These probabilities are the *transition probabilities* of the Markov process (X_t).

Let $\tau < s < t$, $B \in \mathcal{B}$. By the tower rule, the Markov property, and the properties of regular conditional distribution

$$\begin{aligned}\mathbf{P}(X_t \in B | X_\tau) &= \mathbf{E}\left[\mathbf{P}(X_t \in B | X_\tau, X_s) | X_\tau\right] \\ &= \mathbf{E}\left[\mathbf{P}(X_t \in B | X_s) | X_\tau\right] \\ &= \mathbf{E}\left[h(X_s) | X_\tau\right] \\ &= \int h(y) \mathrm{d}\mathbf{P}(X_s \in \mathrm{d}y | X_\tau) \\ &= \int \mathbf{P}(X_t \in B | X_s = y) \mathbf{P}\left(X_s \in \mathrm{d}y | X_\tau\right) \\ &= \int_{\mathbb{R}} p_{t,s}(B|y) p_{s,\tau}(\mathrm{d}y | X_\tau).\end{aligned}$$

That is

$$p_{t,\tau}(B|x) = \int p_{t,s}(B|y) p_{s,\tau}(\mathrm{d}y|x).$$

We proved the following.

Theorem 39 (Chapman–Kolmogorov equations)**.** *The transition probabilities of a Markov process satisfies the equations*

$$p_{t,\tau}(B|x) = \int p_{t,s}(B|y) p_{s,\tau}(\mathrm{d}y|x), \quad \tau < s < t, B \in \mathcal{B}. \tag{26}$$

The expression $p_{t,\tau}(B|x)$ is the probability that starting from x in time τ we end up in B at time t. Consider any s between τ and t. The distribution of X_s given $X_\tau = x$ is $p_{s,\tau}(\cdot|x)$, that is the probability being in y is $p_{s,\tau}(\mathrm{d}y|x)$. Therefore, the Chapman–Kolmogorov equation is the law of total probability plus Markov property.

We are cheating again a bit. What we proved is that (26) holds for fixed $\tau < s < t$ *almost surely* with respect to the probability $\mathbf{P}(X_\tau \in \cdot)$. Indeed, in the proof we calculated conditional probabilities, where each equality is only an almost sure equality. In what follows *we assume that* (26) *holds for every* x.

The Markov process (X_t) is *stationary* if the transition probabilities depend only on the time increment, i.e. $p_{t,\tau}(B|x) = p_{t-\tau}(B|x)$. Then $p_t(B|x) = p_{t,0}(B|x)$, and the Chapman–Kolmogorov equations simplify to

$$p_{t+s}(B|x) = \int p_t(B|y) p_s(\mathrm{d}y|x). \tag{27}$$

Assume that (X_t) is stochastically continuous at 0, that is

$$X_t \xrightarrow{\mathbf{P}} X_0, \quad t \to 0.$$

If (X_t) starts at x then its distribution is denoted by $\mathbf{P}_x$, and the corresponding expectation is $\mathbf{E}_x$, that is

$$\mathbf{P}_x(X_t \in B) = \mathbf{P}(X_t \in B | X_0 = x), \quad \mathbf{E}_x f(X_t) = \mathbf{E}\left[f(X_t) | X_0 = x \right].$$

Example 18 (Poisson process)**.** Let N_t be a standard Poisson process. Then $N_t - N_s \sim \text{Poisson}(t-s)$, so

$$\mathbf{P}_x(N_t = x + k) = p_t(\{x+k\}|x) = \frac{t^k}{k!} e^{-t},$$

or, what is the same

$$p_t(B|x) = \sum_{k: x+k \in B} \frac{t^k}{k!} e^{-t}.$$

The Chapman–Kolmogorov equation (27) become

$$p_{t+s}(\{k\}|0) = \sum_{\ell=0}^{\infty} p_t(\{k\}|\ell) p_s(\{\ell\}|0),$$

which is just a reformulation of the fact that the sum of two independent Poisson random variables is Poisson, and the parameter is the sum of the parameters.

Example 19 (Wiener process)**.** Let W_t be SBM. Then

$$\begin{aligned} p_t(B|x) = \mathbf{P}_x(W_t \in B) &= \mathbf{P}_0(x + W_t \in B) = \mathbf{P}_0(W_t \in B - x) \\ &= \int_{B-x} \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} \mathrm{d}y \\ &= \int_B \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} \mathrm{d}y. \end{aligned}$$

That is $p_t(B|x)$ is absolutely continuous with *transition density* $p_t(\mathrm{d}y|x) = \rho_t(y|x)\mathrm{d}y$

$$\rho_t(y|x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}}.$$

The Chapman–Kolmogorov equation (27) become

$$p_{t+s}(B|x) = \int_{\mathbb{R}} p_t(B|y)\rho_s(y|x)\mathrm{d}y,$$

or for the densities

$$\rho_{t+s}(z|x) = \int_{\mathbb{R}} \rho_t(z|y)\rho_s(y|x)\mathrm{d}y.$$

This is a reformulation of the fact that the sum of independent normals is normal. Recall the convolution formula for densities.

6.2 Infinitesimal generator

The *infinitesimal generator* of X an operator defined by

$$f \mapsto Sf : Sf(x) = \lim_{t\to 0+} \frac{1}{t}\mathbf{E}_x \left[f(X_t) - f(x)\right], \tag{28}$$

whenever the limit exists. Its domain is denoted by $\mathcal{D}(S)$.

We determine the infinitesimal generator of the Poisson process and the Wiener process.

Example 20 (Poisson process)**.** Let (N_t) be a Poisson process with intensity 1, and let f be a bounded measurable function. By definition $N_t - N_0 \sim$ Poisson(t), thus

$$\mathbf{E}_x f(N_t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} e^{-t} f(k+x).$$

Since f is bounded the sum is finite, and as $t \downarrow 0$

$$\mathbf{E}_x f(N_t) = f(x)e^{-t} + f(x+1)te^{-t} + O(t^2).$$

Thus

$$\begin{aligned}
Sf(x) &= \lim_{t\to 0} \frac{1}{t}\mathbf{E}_x \left[f(N_t) - f(x)\right] \\
&= \lim_{t\to 0} \left(f(x)\frac{e^{-t}-1}{t} + f(x+1)e^{-t} \right) \\
&= f(x+1) - f(x).
\end{aligned}$$

The limit exists for any bounded measurable function.

Example 21 (Wiener process)**.** Let (W_t) be SBM and $f \in C_c^2$ twice continuously differentiable function with compact support. Using Taylor expansion

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x) + o(h^2),$$

and since $\mathbf{E}_0 W_t = 0$, $\mathbf{E}_0 W_t^2 = t$, we have

$$\begin{aligned}
\mathbf{E}_x f(W_t) &= \mathbf{E}_0 f(x + W_t) \\
&= \mathbf{E}_0 \left[f(x) + W_t f'(x) + \frac{W_t^2}{2} f''(x) + o(W_t^2) \right] \\
&= f(x) + \frac{t}{2} f''(x) + o(t).
\end{aligned}$$

Thus

$$Sf(x) = \lim_{t \to 0} \frac{1}{t} \mathbf{E}_x \left[f(W_t) - f(x) \right] = \frac{f''(x)}{2}.$$

We see that $C_c^2 \subset \mathcal{D}(S)$.

6.3 Kolmogorov equations

Backward. Let $t > 0$ fix, $B \in \mathcal{B}(\mathbb{R})$, $\tau > 0$ small. By the tower rule and the Markov property

$$\mathbf{P}(X_{t+\tau} \in B | X_0 = x) = \mathbf{E} \left[\mathbf{P}(X_{t+\tau} \in B | X_\tau) | X_0 = x \right].$$

With the notation $\varphi_t(x) = p_t(B|x)$

$$\varphi_{t+\tau}(x) = \mathbf{E}_x \varphi_t(X_\tau),$$

which reads as

$$\frac{1}{\tau} \left[\varphi_{t+\tau}(x) - \varphi_t(x) \right] = \frac{1}{\tau} \mathbf{E}_x \left[\varphi_t(X_\tau) - \varphi_t(x) \right].$$

Letting τ tend to 0, we obtain

$$\frac{\partial}{\partial t} \varphi_t(x) = (S\varphi_t)(x).$$

Substituting back the definition of φ, we obtain *Kolmogorov's backward equation*

$$\frac{\partial}{\partial t} p_t(B|x) = (S p_t(B|\cdot))(x). \tag{29}$$

Forward. Let $t > 0$ fix, $f \in \mathcal{D}(S)$. By the tower rule and the Markov property

$$\mathbf{E}_x f(X_{t+\tau}) = \mathbf{E}_x \left[\mathbf{E}_x[f(X_{t+\tau})|X_t]\right],$$

which can be rewritten as

$$\int f(y)p_{t+\tau}(\mathrm{d}y|x) = \int\int f(z)p_\tau(\mathrm{d}z|y)p_t(\mathrm{d}y|x) = \int \mathbf{E}_y f(X_\tau)p_t(\mathrm{d}y|x).$$

Subtracting

$$\mathbf{E}_x f(X_t) = \int f(y)p_t(\mathrm{d}y|x)$$

and dividing by τ

$$\int f(y)\frac{p_{t+\tau}(\mathrm{d}y|x) - p_t(\mathrm{d}y|x)}{\tau} = \int \frac{1}{\tau}\left[\mathbf{E}_y f(X_\tau) - f(y)\right]p_t(\mathrm{d}y|x).$$

Letting $\tau \downarrow 0$

$$\int f(y)\frac{\partial}{\partial t}p_t(\mathrm{d}y|x) = \int (Sf)(y)p_t(\mathrm{d}y|x). \tag{30}$$

The adjoint of the operator S is an operator S^* on the space of measures such that

$$\int (Sf)(y)\mu(\mathrm{d}y) = \int f(y)(S^*\mu)(\mathrm{d}y).$$

If this holds for sufficiently many f and μ, then it is unique.

Using the definition of adjoint in (30)

$$\int f(y)\frac{\partial}{\partial t}p_t(\mathrm{d}y|x) = \int f(y)\left(S^*p_t(\cdot|x)\right)(\mathrm{d}y),$$

from which we get *Kolmogorov's forward equation*

$$\frac{\partial}{\partial t}p_t(B|x) = \left(S^*p_t(\cdot|x)\right)(B). \tag{31}$$

Remark 2. The derivation of the forward equation is rather intuitive. What kind of space is the domain $\mathcal{D}(S)$, and how the adjoint operator defined? Furthermore, in (30)) we differentiated a family of measures with respect to t. If the measure are absolutely continuous, i.e.

$$p_t(\mathrm{d}y|x) = \rho_t(y|x)\mathrm{d}y,$$

then

$$\lim_{\tau\to 0}\frac{\rho_{t+\tau}(y|x)-\rho_t(y|x)}{\tau}=\frac{\partial}{\partial t}\rho_t(y|x).$$

In general, both for the backward and for the forward equations extra conditions are needed. As it can be guessed from the derivation, for the forward equation more restrictive conditions are needed.

The importance of the Kolmogorov equations (29) and (31) is that from infinitesimal conditions (from the generator S) one can determine the evolution of the whole process, that is the transition probabilities. In most of the cases the solution cannot be determined explicitly, only by simulation.

Example 22 (Poisson process)**.** Let (N_t) be a Poisson process with intensity 1. We proved that

$$(Sf)(x)=f(x+1)-f(x).$$

Therefore, the backward equation reads as

$$\frac{\partial}{\partial t}p_t(B|x)=p_t(B|x+1)-p_t(B|x). \tag{32}$$

For the forward equation we determine the adjoint of S. We need an $S^*\mu$ such that

$$\int[f(x+1)-f(x)]\mu(\mathrm{d}x)=\int f(x)(S^*\mu)(\mathrm{d}x).$$

From this form we can guess that

$$S^*\mu(A)=\mu(A-1)-\mu(A),$$

should work, where $A-1=\{a-1\,:\,a\in A\}$. This indeed holds, therefore the forward equation reads as

$$\frac{\partial}{\partial t}p_t(B|x)=p_t(B-1|x)-p_t(B|x).$$

The initial condition in both cases is

$$p_0(B|x)=\delta_x(B)=\begin{cases}1, & \text{if } x\in B,\\ 0, & \text{otherwise.}\end{cases}$$

In this special case we can solve the equation (32). Let $x=0$ and $B=\{0\}$. Since the process have only upwards jumps $p_t(\{0\}|1)=0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}p_t(\{0\}|0)=-p_t(\{0\}|0),$$

which together with the initial condition $p_0 = 1$ gives

$$p_t(\{0\}|0) = e^{-t}.$$

Now $B = \{1\}$ gives

$$\frac{\mathrm{d}}{\mathrm{d}t} p_t(\{1\}|0) = e^{-t} - p_t(\{1\}|0).$$

Multiplying by e^t

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^t p_t(\{1\}|0)\right) = 1,$$

which with the initial condition $p_0(\{1\}|0) = 0$ gives

$$p_t(\{1\}|0) = te^{-t}.$$

In general, induction gives that

$$p_t(\{k\}|0) = \frac{t^k}{k!} e^{-t}.$$

Example 23 (Wiener process)**.** Let (W_t) be SBM. Since $(Sf)(x) = f''(x)/2$, the backward equation is

$$\frac{\partial}{\partial t} p_t(B|x) = \frac{1}{2}\frac{\partial^2}{\partial x^2} p_t(B|x).$$

For the density $p_t(\mathrm{d}y|x) = \rho_t(y|x)\mathrm{d}y$ we get

$$\frac{\partial}{\partial t}\rho_t(y|x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\rho_t(y|x).$$

This is the heat equation.

For the forward equation we need again the adjoint of S. Let μ be absolutely continuous with respect to the Lebesgue measure, $\mu(\mathrm{d}y) = g(y)\mathrm{d}y$, and let $f \in C_c^2$. Integration by parts twice gives

$$\int f''(y)g(y)\mathrm{d}y = \int f(y)g''(y)\mathrm{d}y.$$

That is $(S^*\mu)(\mathrm{d}y) = \frac{1}{2}g''(y)\mathrm{d}y$. The forward equation is

$$\frac{\partial}{\partial t} p_t(y|x)\mathrm{d}y = \frac{1}{2}\frac{\partial^2}{\partial y^2} p_t(y|x)\mathrm{d}y,$$

which for the densities gives

$$\frac{\partial}{\partial t}\rho_t(y|x) = \frac{1}{2}\frac{\partial^2}{\partial y^2}\rho_t(y|x),$$

again the heat equation.

Recall that the *fundamental solution* to the heat equation

$$\frac{\partial}{\partial t}u(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}u(t,x)$$

is

$$F(t,x) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}},$$

which is exactly the transition density of the SBM.

6.4 Diffusion processes

Diffusions can be handled as solution to SDEs. We showed that under general conditions unique strong solution to SDEs exists, implying the existence of diffusion processes. This is the probabilistic approach due to Lévy and Itô. Another more analytical approach to such processes was applied by Kolmogorov and Feller. They treated diffusions as general Markov processes and using tools from the theory of partial differential equations, they showed that under suitable conditions the Kolmogorov backward and forward equations have a unique solution. Then the existence of a desired Markov process follows from Kolmogorov's consistency theorem, and the continuity property of the process can be treated by Kolmogorov's continuity theorem (Theorem 18). Here we look a bit into the latter approach.

A diffusion process *locally* behaves as a Wiener process, in the sense that it satisfies the SDE

$$\mathrm{d}Y_t = \mu(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}W_t.$$

That is, for $h > 0$

$$\begin{aligned}\Delta Y_t = Y_{t+h} - Y_t &= \int_t^{t+h} \mu(Y_s)\mathrm{d}s + \int_t^{t+h} \sigma(Y_s)\mathrm{d}W_s \\ &\approx h\mu(Y_t) + \sigma(Y_t)(W_{t+h} - W_t),\end{aligned}$$

thus

$$\mathbf{E}\left[\Delta Y_t | Y_t = y\right] = \mu(y)h + o(h),$$
$$\mathbf{E}\left[(\Delta Y_t)^2 | Y_t = y\right] = \sigma^2(y)h + o(h).$$

A *diffusion process* (Y_t) is a continuous Markov process satisfying as $h \downarrow 0$

(i) $\mathbf{P}(|\Delta Y_t| > \varepsilon | Y_t = y) = o(h)$;

(ii) $\mathbf{E}\left(\Delta_\varepsilon Y_t | Y_t = y\right) = \mu(y)h + o(h)$;

(iii) $\mathbf{E}\left((\Delta_\varepsilon Y_t)^2 | Y_t = y\right) = \sigma^2(y)h + o(h)$,

where $\Delta Y_t = Y_{t+h} - Y_t$, and

$$\Delta_\varepsilon Y_t = \begin{cases} \Delta Y_t, & \text{if } |\Delta Y_t| \le \varepsilon, \\ 0, & \text{otherwise.} \end{cases}$$

The definition determines the infinitesimal generator of the process. For $f \in C^2$

$$\begin{aligned} \mathbf{E}_x f(Y_t) &= \mathbf{E}_x \left[f(x) + (Y_t - x)f'(x) + (Y_t - x)^2 \frac{f''(x)}{2} + o((Y_t - x)^2) \right] \\ &= f(x) + t\mu(x)f'(x) + t\sigma^2(x)\frac{f''(x)}{2} + o(t). \end{aligned}$$

Therefore,

$$(Sf)(x) = \lim_{t\to 0} \frac{1}{t} \mathbf{E}_x \left[f(Y_t) - f(x)\right] = \mu(x)f'(x) + \sigma^2(x)\frac{f''(x)}{2}.$$

Kolmogorov backward equation is

$$\frac{\partial}{\partial t} p_t(y|x) = \mu(x)\frac{\partial}{\partial x} p_t(y|x) + \frac{\sigma^2(x)}{2}\frac{\partial^2}{\partial x^2} p_t(y|x).$$

For the forward equation we need the adjoint of S. This can be determined as for the SBM. Let $\rho_t(y|x)$ denote the density of the process, i.e. $p_t(\mathrm{d}y|x) = \rho_t(y|x)\mathrm{d}y$. Let $\mu(\mathrm{d}y) = g(y)\mathrm{d}y$. If f has compact support then in the integration by parts formula the increment disappears and we get

$$\begin{aligned} \int (Sf)(y)g(y)\mathrm{d}y &= \int \left[\mu(y)f'(y) + \frac{\sigma^2(y)}{2} f''(y) \right] g(y)\mathrm{d}y \\ &= \int f(y) \left[-\frac{\mathrm{d}}{\mathrm{d}y} \left(\mu(y)g(y)\right) + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}y^2} \left(\sigma^2(y)g(y)\right) \right] \mathrm{d}y. \end{aligned}$$

Thus

$$(S^* p_t(\cdot|x))\,(\mathrm{d}y) = \left[-\frac{\mathrm{d}}{\mathrm{d}y}\left(\mu(y)\rho_t(y|x)\right) + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}y^2}\left(\sigma^2(y)\rho_t(y|x)\right)\right]\mathrm{d}y,$$

and the forward equation is

$$\frac{\partial}{\partial t}\rho_t(y|x) = -\frac{\partial}{\partial y}\left(\mu(y)\rho_t(y|x)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y)\rho_t(y|x)\right).$$

Example 24 (Ornstein–Uhlenbeck process)**.** Consider the *Langevin equation*

$$\mathrm{d}Y_t = -\mu Y_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t,$$

where $\mu > 0$, $\sigma > 0$, and Y_0 is independent of $\sigma(W_s : s \geq 0)$.

The solution of the homogeneous equation is $e^{-\mu t}$. Taking the derivative of $e^{\mu t}Y_t$ we obtain

$$\mathrm{d}\left(e^{\mu t}Y_t\right) = e^{\mu t}\,\mathrm{d}Y_t + \mu\, e^{\mu t}Y_t\,\mathrm{d}t = e^{\mu t}\,\sigma\,\mathrm{d}W_t,$$

which gives

$$Y_t = e^{-\mu t}\left(Y_0 + \int_0^t e^{\mu s}\,\sigma\,\mathrm{d}W_s\right).$$

This is the *Ornstein–Uhlenbeck process*. The integral of a deterministic function with respect to SBM is Gaussian, thus

$$Y_t - e^{-\mu t}Y_0$$

is normal with mean and variance

$$\begin{aligned}
&\mathbf{E}Y_t = e^{-\mu t}\,\mathbf{E}Y_0,\\
&\mathbf{E}Y_t^2 = e^{-2\mu t}\,\mathbf{E}Y_0^2 + e^{-2\mu t}\int_0^t \sigma^2\, e^{2\mu s}\,\mathrm{d}s = e^{-2\mu t}\,\mathbf{E}Y_0^2 + \frac{\sigma^2}{2\mu}\,(1 - e^{-2\mu t}).
\end{aligned}$$

We see that as $t \to \infty$

$$Y_t \xrightarrow{\mathcal{D}} N(0, \sigma^2/(2\mu)).$$

Taking the limit for the initial distribution Y_0 we see that (Y_t) is Gaussian and

$$Y_t \sim \mathrm{N}\left(0, \frac{\sigma^2}{2\mu}\right).$$

Next we determine the covariance function of Y. Since

$$Y_t = e^{-\mu t}\left(Y_0 + \int_0^t \sigma\, e^{\mu u}\, \mathrm{d}W_u\right)$$

we get

$$Y_t - e^{-\mu(t-s)}Y_s = e^{-\mu t}\int_s^t \sigma\, e^{\mu u}\, \mathrm{d}W_u,\ t > s, \tag{33}$$

which is independent of $\sigma(W_u : u \leq s)\ \sigma$. Therefore,

$$\begin{aligned}\mathbf{Cov}(Y_t, Y_s) &= \mathbf{E}Y_tY_s = \mathbf{E}\left(Y_t - e^{-\mu(t-s)}Y_s + e^{-\mu(t-s)}Y_s\right)Y_s \\ &= e^{-\mu(t-s)}\,\mathbf{E}Y_s^2 = \frac{\sigma^2}{2\mu}e^{-\mu(t-s)},\end{aligned}$$

which depends only on $t-s$. That is (Y_t) is stationary.

Using formula (33) for $A \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned}&\mathbf{P}(Y_t \in A | Y_u : u \leq s, Y_s = x) \\ &= \mathbf{P}(Y_t - e^{-\mu(t-s)}Y_s \in A - e^{-\mu(t-s)}x | Y_u : u \leq s, Y_s = x) \\ &= \mathbf{P}(Y_t - e^{-\mu(t-s)}Y_s \in A - e^{-\mu(t-s)}x).\end{aligned}$$

The variable $Y_t - e^{-\mu(t-s)}Y_s$ is mean zero Gaussian with variance

$$\mathbf{E}\left(Y_t - e^{-\mu(t-s)}Y_s\right)^2 = e^{-2\mu t}\int_s^t \sigma^2 e^{2\mu u}\mathrm{d}u = \frac{\sigma^2}{2\mu}\left(1 - e^{-2\mu(t-s)}\right).$$

Substituting $s = 0$

$$p_t(\cdot|x) \sim \mathrm{N}\left(e^{-\mu t}x, \frac{\sigma^2}{2\mu}\left(1 - e^{-2\mu t}\right)\right),$$

that is, the transition density

$$\rho_t(y|x) = \sqrt{\frac{\mu}{\pi\sigma^2(1 - e^{-2\mu t})}}\ \exp\left\{-\frac{\mu(y - e^{-\mu t}x)^2}{\sigma^2(1 - e^{-2\mu t})}\right\}.$$

We proved that (Y_t) is a continuous stationary Markov process. It can be shown that this characterizes the OU process.

Finally, we spell out the Kolmogorov equations. The backward is

$$\frac{\partial}{\partial t}\rho_t(y|x) = -\mu x\frac{\partial}{\partial x}\rho_t(y|x) + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}\rho_t(y|x),$$

which is called *Fokker–Planck equation.* The forward is

$$\frac{\partial}{\partial t}\rho_t(y|x) = -\frac{\partial}{\partial y}\left(-\mu y \rho_t(y|x)\right) + \frac{\sigma^2}{2}\frac{\partial^2}{\partial y^2}\rho_t(y|x).$$

It is important to emphasize that in general explicit formulas for the transition densities cannot be obtained. For simulation results the Kolmogorov equations are important, because solutions can be approximated numerically.

7 Brownian motion and PDEs

This part is from Karatzas and Shreve [5].

We showed that the infinitesimal generator of the SBM is the Laplacian operator Δ. Furthermore the transition density of SBM is the fundamental solution to the heat equation. These facts already show the intrinsic connection between Brownian motion and partial differential equations. Here we spell out this connection a bit more.

7.1 Harmonic functions and the Dirichlet problem

Let D be an open subset of $\mathbb{R}^d$. Let W be a d-dimensional standard Brownian motion, and let

$$\tau_D = \inf\{t \geq 0 : W_t \in D^c\}$$

the first exit time from D. Let B_r be the open ball centered at the origin, V_r its volume and S_r its surface. The normalized surface measure on B_r is μ_r

$$\mu_r(\mathrm{d}x) = \mathbf{P}_0(W_{\tau_{B_r}} \in \mathrm{d}x).$$

Then

$$\int_{B_r} f(x)\mathrm{d}x = \int_0^r S_\rho \int_{\partial B_\rho} f(x)\mu_\rho(\mathrm{d}x)\mathrm{d}\rho. \tag{34}$$

A function u is *harmonic* in D if

$$\Delta u = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} u = 0$$

in D. A function $u : D \to \mathbb{R}$ satisfies the *mean-value property*, if for every $a \in D$ and $r > 0$ such that $a + \overline{B}_r \subset D$,

$$u(a) = \int_{\partial B_r} u(a + x)\mu_r(\mathrm{d}x).$$

We know that u is harmonic if and only if it satisfies the mean-value property. We give a simple proof to one direction using Itô formula.

Proposition 13. *If u is harmonic in D, then it satisfies the mean-value property there.*

Proof. By Itô's formula

$$u(W_{t\wedge\tau_{a+B_r}}) = u(W_0) + \sum_{i=1}^{d} \int_0^{t\wedge\tau_{a+B_r}} \frac{\partial u}{\partial x_i}(W_s)\mathrm{d}W_s^{(i)} + \frac{1}{2}\int_0^{t\wedge\tau_{a+B_r}} \Delta u(W_s)\mathrm{d}s.$$

Taking expectation $\mathbf{E}_a$ and letting $t\to\infty$

$$\mathbf{E}_a u(W_{\tau_{a+B_r}}) = u(a),$$

as stated. $\square$

Let D be an open set of $\mathbb{R}^d$ and $f:\partial D\to\mathbb{R}$ be a continuous function. Consider the Dirichlet problem

$$\begin{aligned} \Delta u &= 0, \quad \text{in } D, \\ u &= f, \quad \text{on } \partial D. \end{aligned} \tag{35}$$

A solution to the Dirichlet problem is a continuous function $u:\overline{D}\to\mathbb{R}$ which satisfies the equation above.

Then one can guess that

$$u(x) = \mathbf{E}_x f(W_{\tau_D}) \tag{36}$$

should be a solution, provided that the expectation exists.

Indeed, the boundary condition holds by the definition of τ_D. Using the strong Markov property

$$\begin{aligned} u(a) = \mathbf{E}_a f(W_{\tau_D}) &= \mathbf{E}_a\left[\mathbf{E}_a[f(W_{\tau_D})|\mathcal{F}_{\tau_{a+B_r}}]\right] \\ &= \mathbf{E}_a u(W_{\tau_{a+B_r}}) = \int_{\partial B_r} u(a+x)\mu_r(\mathrm{d}x), \end{aligned}$$

that is the mean-value property holds, which means that u is indeed harmonic.

We proved the following.

Proposition 14. *If u in* (36) *is well-defined then it is harmonic.*

The proof of Proposition 13 shows in fact uniqueness.

Proposition 15. *If f is bounded and $\mathbf{P}_a(\tau_D < \infty) = 1$ for all $a \in D$, then any bounded solution to* (35) *has the form* (36)*.*

Proof. Consider a bounded solution u. By Itô's formula

$$u(W_{t\wedge\tau_D}) = u(W_0) + \sum_{i=1}^{d} \int_0^{t\wedge\tau_{a+B_r}} \frac{\partial u}{\partial x_i}(W_s)\mathrm{d}W_s^{(i)}.$$

Taking expectation $\mathbf{E}_a$ and letting $t \to \infty$

$$\mathbf{E}_a u(W_{\tau_D}) = u(a),$$

as stated. □

Note that a solution to the Dirichlet problem (35) is necessarily continuous. Therefore, we need conditions characterizing the points $a \in \partial D$ for which

$$\lim_{x\to a, x\in D} \mathbf{E}_x f(W_{\tau_D}) = f(a) \tag{37}$$

holds for any bounded measurable function, which is continuous at a.

Define the stopping time $\sigma_D = \inf\{t > 0 : W_t \in D^c\}$. Note the $>$ compared to $\geq$ in τ_D. A point $a \in \partial D$ is *regular* for D is $\mathbf{P}_a(\sigma_D = 0) = 1$.

Without proof we state the result on regularity.

Theorem 40. *Let $d \geq 2$ and fix $a \in \partial D$. The following are equivalent:*

(i) (37) *holds for every bounded, measurable function which continuous at a;*

(ii) a is regular for D;

(iii) for all $\varepsilon > 0$ we have

$$\lim_{x\to a, x\in D} \mathbf{P}_x(\tau_D > \varepsilon) = 0.$$

For $d = 1$ every point of ∂D is regular. The Dirichlet problem is always solvable, the solution is piecewise linear. For $d \geq 2$ consider the punctured unit ball $D = \{x \in \mathbb{R}^d : 0 < \|x\| < 1\}$. Clearly, the origin is irregular for D. For any $x \in D$ the SBM exits D on its outer boundary, therefore we do not see the value of f at 0. For this D the Dirichlet problem has a solution only if $f(0) = \widetilde{u}(0)$, where $\widetilde{u}$ is the solution for B_1.

7.2 Feynman–Kac formula

Consider the heat equation

$$\frac{\partial u}{\partial t} = \frac{1}{2}\Delta u, \tag{38}$$

with initial condition $u(0,x) = f(x)$.

The fundamental solution to the heat equation is in fact the transition probabilities of the d-dimensional SBM

$$\rho_t(y|x) = \frac{1}{(2\pi t)^{d/2}} e^{-\frac{\|x-y\|^2}{2t}}.$$

Under some growth condition on f, the unique solution to (38) has the form

$$u(t,x) = \mathbf{E}_x f(W_t) = \int f(y)\rho_t(y|x)\mathrm{d}y.$$

The probabilistic representation of the solution to certain PDEs holds in a more general setup.

Consider the equation

$$\begin{aligned} -\frac{\partial v}{\partial t} + kv &= \frac{1}{2}\Delta v + g \quad \text{on } [0,T) \times \mathbb{R}^d, \\ v(T,x) &= f(x), \quad x \in \mathbb{R}^d, \end{aligned} \tag{39}$$

where $f : \mathbb{R}^d \to \mathbb{R}$, $k : \mathbb{R}^d \to [0,\infty)$, and $g : [0,T] \times \mathbb{R}^d \to \mathbb{R}$.

Theorem 41 (Feynman–Kac formula)**.** *Assume that* (39) *has a solution and*

$$\max_{0\le t\le T} |v(t,x)| + \max_{0\le t\le T} |g(t,x)| \le K e^{a\|x\|^2}, \quad \forall x \in \mathbb{R}^d,$$

for some constant $K > 0$, $0 < a < 1/(2Td)$. Then v admits the stochastic representation

$$v(t,x) = \mathbf{E}_x \left[f(W_{T-t}) e^{-\int_0^{T-t} k(W_s)\mathrm{d}s} + \int_0^{T-t} g(t+\theta, W_\theta) e^{-\int_0^\theta k(W_s)\mathrm{d}s} \mathrm{d}\theta \right].$$

Proof. Consider the case $d = 1$. The general case is the same. Fix t. Let

$$\begin{aligned} X_\theta &= v(t+\theta, W_\theta) \\ Y_\theta &= e^{-\int_0^\theta k(W_s)\mathrm{d}s}. \end{aligned}$$

Then, by Itô's formula

$$\mathrm{d}X_\theta = \frac{\partial}{\partial t}v(t+\theta, W_\theta)\mathrm{d}\theta + \frac{\partial}{\partial x}v(t+\theta, W_\theta)\mathrm{d}W_\theta + \frac{1}{2}\Delta v(t+\theta, W_\theta)\mathrm{d}\theta.$$

and, using also (39)

$$\begin{aligned}\mathrm{d}(X_\theta Y_\theta) &= X_\theta \mathrm{d}Y_\theta + Y_\theta \mathrm{d}X_\theta \\ &= -g(t+\theta, W_\theta)e^{-\int_0^\theta k(W_s)\mathrm{d}s}\mathrm{d}\theta + \frac{\partial}{\partial x}v(t+\theta, W_\theta)\mathrm{d}W_\theta.\end{aligned}$$

Taking expectation $\mathbf{E}_x$ and integrating on $[0, T-t]$, and using the terminal condition we obtain the desired form. $\square$

As a consequence we obtain the representation of the solution to the *parabolic equation*

$$\begin{aligned}&\frac{\partial u}{\partial t}u + ku = \frac{1}{2}\Delta u + g, \quad t \in (0,\infty), x \in \mathbb{R}^d, \\ &u(0,x) = f(x), \quad x \in \mathbb{R}^d,\end{aligned} \tag{40}$$

where $k : \mathbb{R}^d \to [0,\infty)$, $g : (0,\infty) \times \mathbb{R}^d \to \mathbb{R}$, $f : \mathbb{R}^d \to \mathbb{R}$.

Corollary 10. *Assume that f, k, and g are continuous, $u : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ is continuous, on $(0,\infty) \times \mathbb{R}^d$ it is $C^{1,2}$, and satisfies (40). Further assume that for each $T > 0$ there exist $K > 0$ and $0 < a < 1/(2Td)$ such that*

$$\max_{0 \le t \le T} |u(t,x)| + \max_{0 \le t \le T} |g(t,x)| \le K e^{a\|x\|^2}, \quad \forall x \in \mathbb{R}^d.$$

Then u admits the stochastic representation

$$u(t,x) = \mathbf{E}_x f(W_t) e^{-\int_0^t k(W_s)\mathrm{d}s} + \int_0^t g(t-\theta, W_\theta) e^{-\int_0^\theta k(W_s)\mathrm{d}s}\mathrm{d}\theta.$$

Proof. Fix $T > 0$ and consider the PDE (39) with $g_{v,T}(t,x) = g(T-t,x)$. Then $v_T(t,x) = u(T-t,x)$ satisfies the conditions of Theorem 41 with $g_{v,T}$ instead of g. Therefore v_T has a Feynman–Kac representation, which, rewriting to u gives the statement. $\square$

References

[1] Leo Breiman. *Probability*. Addison-Wesley Publishing Company, Reading, Mass.-London-Don Mills, Ont., 1968.

[2] Sándor Csörgő. *Fejezetek a valószínűségelméletből*. Polygon, 2010.

[3] Rick Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fourth edition, 2010.

[4] Lawrence C. Evans. *An introduction to stochastic differential equations*. American Mathematical Society, Providence, RI, 2013.

[5] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.