Stochastic processes

Péter Kevei

February 24, 2025

Contents

1 Kolmogorov consistency theorem **1**
- 1.1 Arbitrary product of measurable spaces 1
- 1.2 Consistency theorem 2

2 Conditional expectation **7**
- 2.1 Existence and uniqueness 8
- 2.2 Conditional expectation with respect to random variable . . . 10
- 2.3 Elementary setup 12
  - 2.3.1 Discrete case 12
  - 2.3.2 Continuous case 13
- 2.4 Properties 14
- 2.5 Regular conditional probabilities 17

3 Discrete time martingales **19**
- 3.1 Definition, properties 20
- 3.2 Martingale convergence theorem 21
- 3.3 Examples 23
  - 3.3.1 Bounded increments 23
  - 3.3.2 Pólya urns 23
  - 3.3.3 Galton–Watson process 25
- 3.4 Doob's decomposition 26
- 3.5 Doob's maximal inequality 27
- 3.6 Optional stopping theorem 30

4 Continuous time martingales **36**
- 4.1 Definitions and simple properties 36
- 4.2 Martingale convergence theorem 39
- 4.3 Inequalities 40
- 4.4 Optional stopping 42
- 4.5 Doob-Meyer decomposition 43

5 Wiener process **44**
- 5.1 First properties and existence 44
- 5.2 The space $C[0,\infty)$ 51
- 5.3 Donsker theorem 54
- 5.4 Path properties 57

6 General Markov processes **64**
6.1 Transition probabilities and Chapman–Kolmogorov equations 64
6.2 Infinitesimal generator 67
6.3 Kolmogorov equations 68
6.4 Diffusion processes 72

1 Kolmogorov consistency theorem

This part is from the Csörgő notes [2].

1.1 Arbitrary product of measurable spaces

Let I be an arbitrary index set, and $\mathcal{X}_i$, $i \in I$, arbitrary sets. The product

$$\mathcal{X}_I = \{x \mid x : I \mapsto \cup_{i\in I}\mathcal{X}_i\,, \quad x(i) \in \mathcal{X}_i\,, \, i \in I\}\,,$$

that is the set of functions defined on I, such that $x(i) \in \mathcal{X}_i$.

Typical examples are $I = \{0, 1, \ldots\}$ (Markov chains, or iid random variables), $I = [0, \infty)$ (Poisson process, continuous time Markov chains), and $\mathcal{X}_i = \mathcal{X} = \mathbb{R}$.

For $I \supset J \supset K$ define the projection

$$\pi_{J,K} : \mathcal{X}_J \to \mathcal{X}_K$$

such that we forget the coordinates in $J - K$, that is

$$\pi_{J,K}(x) = x|_K\,, \quad x \in \mathcal{X}_J.$$

Put $\pi_K = \pi_{I,K}$ and $\pi_i = \pi_{\{i\}}$.

In what follows let $(\mathcal{X}_i, \mathcal{F}_i)$ be measurable spaces.

The sets

$$R = \{x : x(i) \in B_i\,, \, i \in J\} = \pi_J^{-1}\left(\prod_{i\in J} B_i\right)$$

for $B_i \in \mathcal{F}_i$, $i \in J$, are measurable rectangles. For fix J the set of all measurable rectangles are

$$\mathcal{R}_J = \left\{\pi_J^{-1}\left(\prod_{i\in J} B_i\right) : B_i \in \mathcal{F}_i\,, \, i \in J\right\}$$

and the set of all rectangles

$$\mathcal{R} = \cup_{J\,\text{finite}}\mathcal{R}_J.$$

Recall that for $J \subset I$ finite, the product σ-algebra $\mathcal{F}_J := \prod_{i\in J}\mathcal{F}_i = \sigma(\mathcal{R}_J)$, is the σ-algebra generated by the measurable rectangles. In what follows we define the product σ-algebra for arbitrary products.

Sets of the form

$$\pi_J^{-1}(B) \subset \mathcal{X}_I\,, \quad \text{where} \quad B \in \mathcal{F}_J = \prod_{i \in J} \mathcal{F}_i$$

are cylinders. Put

$$\mathcal{C}_J = \left\{ \pi_J^{-1}(B) : B \in \mathcal{F}_J \right\}$$

and

$$\mathcal{C} = \cup \{ \mathcal{C}_J : J \subset I, \ J \,\text{finite} \}.$$

Recall that a $\mathcal{A}$ is *semialgebra* on $\mathcal{X}$ if (1) $\mathcal{X} \in \mathcal{A}$, (2) is closed under intersection, and (3) if $A \in \mathcal{A}$ then there exists $A_1, \ldots, A_n \in \mathcal{A}$ disjoint, such that $A^c = \cup_{i=1}^n A_i$. It is an algebra, if (1), (2) hold, if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$.

Lemma 1.1. *(i) $\mathcal{R}$ is semialgebra.*

(ii) $\mathcal{C}_J$ is σ-algebra for any finite J.

(iii) $\mathcal{C}$ is algebra.

Proof. It is easy. Recall that the inverse image of a σ-algebra is σ-algebra. □

Now we can define the product σ-algebra. Let

$$\mathcal{F}_I = \prod_{i \in I} \mathcal{F}_i = \sigma(\mathcal{C}).$$

1.2 Consistency theorem

Let $(\Omega, \mathcal{A}, \mathbf{P})$ a probability space, and $(\mathcal{X}_i, \mathcal{F}_i)$, $i \in I$, measurable spaces, and $X_i : \Omega \to \mathcal{X}_i$ a *random element* in $\mathcal{X}_i$ (that is, it is $\mathcal{F}_i$–$\mathcal{A}$-measurable). Then

$$X : \Omega \to \mathcal{X}_I, \quad X(\omega)(i) = X_i(\omega), \quad \omega \in \Omega, \ i \in I,$$

is a random element in $\mathcal{X}_I$.

Indeed, it is enough to check measurability on a generating system, and it is easy to see that $\mathcal{F}_I = \sigma(\mathcal{R})$. For $\pi_J^{-1}\left(\prod_{i \in J} B_i\right) \in \mathcal{R}$

$$X^{-1}\left(\pi_J^{-1} \left(\prod_{i \in J} B_i \right) \right) = \cap_{i \in J} X_i^{-1}(B_i) \in \mathcal{A},$$

since X_i is measurable.

The random element X is a *stochastic process indexed by* I. Its *distribution* is

$$Q = \mathbf{P} \circ X^{-1}.$$

For any $B \in \mathcal{F}_I$

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

If $J \subset I$, $C \in \mathcal{F}_J$

$$Q \circ \pi_J^{-1}(C) = \mathbf{P}(X \in \pi_J^{-1}(C)) = \mathbf{P}(X|_J \in C).$$

Note that Q is a measure on $\mathcal{F}_I$, while $Q \circ \pi_J^{-1}$, $J \subset I$ finite, is a measure on the finite product space $(\mathcal{X}_J, \mathcal{F}_J)$. The set of measures

$$\{Q \circ \pi_J^{-1} : J \subset I, J \,\text{finite}\}$$

are the *finite dimensional distributions of* X. For $K \subset J \subset I$

$$(Q \circ \pi_J^{-1}) \circ \pi_{J,K}^{-1} = Q \circ \pi_J^{-1} \circ \pi_{J,K}^{-1} = Q \circ (\pi_{J,K} \circ \pi_J)^{-1} = Q \circ \pi_K^{-1}.$$

With the notation $Q_J = Q \circ \pi_J^{-1}$ we obtained the *consistency condition*

$$Q_J \circ \pi_{J,K}^{-1} = Q_K.$$

In probabilistic terms $\mathbf{P}(X|_J|_K \in B) = \mathbf{P}(X|_K \in B)$ for each $B \in \mathcal{F}_K$.

Exercise 1.2. Show that $\sigma(\mathcal{R}) = \mathcal{F}_I$.

The measurable space $(\mathcal{X}, \mathcal{F})$ is an *Euclidean space* if $\mathcal{X} \in \mathcal{B}(\mathbb{R}^n)$, and $\mathcal{F} = \mathcal{X} \cap \mathcal{B}(\mathbb{R}^n)$ for some n.

Theorem 1.3 (Kolmogorov's consistency theorem)**.** *Assume that* $(\mathcal{X}_i, \mathcal{F}_i)$, $i \in I$, *are Euclidean spaces. For* $J \subset I$ *finite, let* Q_J *be a measure on the (finite) product* σ*-algebra* $\mathcal{F}_J$. *If the set of measures* $\{Q_J : J \subset I, J \textit{ finite}\}$ *satisfies the consistency condition*

$$Q_J \circ \pi_{J,K}^{-1} = Q_K, \quad \forall K \subset J \subset I, \;\; J, K \;\; \textit{finite}, \tag{1}$$

then there exists a unique measure Q *on the product* σ*-algebra* $\mathcal{F}_I$, *for which*

$$Q \circ \pi_J^{-1} = Q_J \quad \forall J \textit{ finite}.$$

Before the proof we recall some results from measure theory.

Lemma 1.4 (Regularity)**.** *Let $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mu)$ be a finite measure space. For any $B \in \mathcal{B}(\mathbb{R}^k)$ and for any $\varepsilon > 0$ there exists a compact $K \subset B$ such that $\mu(B - K) < \varepsilon$.*

Theorem 1.5 (Charatheodory extension)**.** *Let μ_0 be a σ-finite measure on a semialgebra $\mathcal{A}$. Then μ_0 extends uniquely to a measure μ on $\sigma(\mathcal{A})$.*

Theorem 1.6 (Kolmogorov's continuity)**.** *Let μ be a finitely additive measure on the algebra $\mathcal{A}$. It is measure if and only if $A_n \downarrow \emptyset$, $A_n \in \mathcal{A}$, implies $\lim_{n\to\infty} \mu(A_n) = 0$.*

Exercise 1.7. Prove the continuity theorem.

Proof of the consistency theorem. Since $\mathcal{F}_I = \sigma(\mathcal{C})$ we have to define Q_0 as

$$Q_0(C) = Q_J(A), \quad C = \pi_J^{-1}(A) \in \mathcal{C}\,,\ A \in \mathcal{F}_J,\ J \subset I \text{ finite}, \tag{2}$$

and hope for the best.

First we need that Q_0 is well-defined. Let J, K finite, $A \in \mathcal{F}_J$, $B \in \mathcal{F}_K$, such that

$$C = \pi_J^{-1}(A) = \pi_K^{-1}(B).$$

Let $L = J \cup K$. Then

$$\pi_L^{-1}(\pi_{L,J}^{-1}(A)) = \pi_J^{-1}(A) = C = \pi_K^{-1}(B) = \pi_L^{-1}(\pi_{L,K}^{-1}(B)).$$

This implies $\pi_{L,J}^{-1}(A) = \pi_{L,K}^{-1}(B) =: D$. By the consistency condition (1)

$$Q_J(A) = Q_L \circ \pi_{L,J}^{-1}(A) = Q_L(\pi_{L,J}^{-1}(A)) = Q_L(D) = Q_K(B),$$

that is the definition is unique.

Next we show that Q_0 is finitely additive. Let $C, D \in \mathcal{C}$ disjoint, $C \in \mathcal{C}_J$, $D \in \mathcal{C}_K$. Then $C, D \in \mathcal{C}_L$, with $L = J \cup K$, and $C = \pi_L^{-1}(A)$, $D = \pi_L^{-1}(B)$. Since $C \cap D = \emptyset$ we have $A \cap B = \emptyset$. Thus

$$\begin{aligned} Q_0(C \cup D) &= Q_0(\pi_L^{-1}(A) \cup \pi_L^{-1}(B)) = Q_0(\pi_L^{-1}(A \cup B)) \\ &= Q_L(A \cup B) = Q_L(A) + Q_L(B) \\ &= Q_0(C) + Q_0(D), \end{aligned}$$

showing the additivity.

Therefore Q_0 is a finitely additive finite measure on $\mathcal{C}$. Therefore, by the continuity theorem, to show that Q_0 is in fact a measure we need that $C_1, C_2, \ldots \in \mathcal{C}$, such that $C_n \downarrow \emptyset$, implies $Q_0(C_n) \to 0$. Equivalently, if

$$C_1, C_2, \ldots \in \mathcal{C}\,,\ C_n \downarrow \ \lim_{n\to\infty} Q_0(C_n) = \inf_{n\geq 1} Q_0(C_n) =: 2\varepsilon > 0\,, \tag{3}$$

then $\cap_{n=1}^{\infty} C_n \neq \emptyset$. This is the difficult part.

In what follows we assume $(\mathcal{X}_i, \mathcal{F}_i) = (\mathbb{R}, \mathcal{B})$. Then $(\mathcal{X}_I, \mathcal{F}_I) = (\mathbb{R}^I, \mathcal{B}^I)$.

Let $C_1, C_2, \ldots$ a sequence satisfying (3). Then there exist $B_1, B_2, \ldots$

$$C_n = \pi_{J_n}^{-1}(B_n), \quad B_n \in \mathcal{B}^{J_n}, \quad J_n \subset I, \quad J_1 \subset J_2 \subset \ldots, \quad |J_n| = k_n,$$

where $0 < k_1 < k_2 < \ldots$. Indeed, smaller B_n's can be extended by some factors of $\mathbb{R}$ if necessary. Choosing

$$\{\widetilde{C}_1, \widetilde{C}_2, \ldots\} = \{\underbrace{\mathbb{R}^I, \ldots, \mathbb{R}^I}_{k_1-1}, \underbrace{C_1, \ldots, C_1}_{k_2-k_1}, \ldots, \underbrace{C_n, \ldots, C_n}_{k_{n+1}-k_n}, \ldots\},$$

we may also assume that $k_n = n$. We suppress the $\sim$.

That is for each $n \in \mathbb{N}$

$$C_n = \pi_{J_n}^{-1}(B_n), \quad B_n \in \mathcal{B}^{J_n}, \quad J_n \subset I, \quad J_1 \subset J_2 \subset \ldots, \quad |J_n| = n.$$

By the regularity lemma there exists $K_n \in \mathbb{R}^{J_n}$ compact (bounded, closed)

$$K_n \subset B_n, \quad Q_{J_n}(B_n - K_n) < \frac{\varepsilon}{2^n}\,.$$

Since $C_n = \cap_{k=1}^n C_k$,

$$\begin{aligned} Q_0(C_n) &= Q_0(\cap_{k=1}^n C_k) = Q_0(\cap_{k=1}^n \pi_{J_k}^{-1}(B_k)) \\ &= Q_0(\cap_{k=1}^n [\pi_{J_n}^{-1} \circ \pi_{J_n,J_k}^{-1}(B_k)]) = Q_0\Big(\pi_{J_n}^{-1}\big(\cap_{k=1}^n \pi_{J_n,J_k}^{-1}(B_k)\big)\Big) \\ &= Q_{J_n}\big(\cap_{k=1}^n \ \pi_{J_n,J_k}^{-1}(B_k)\big). \end{aligned}$$

Using that $\cap A_k - \cap B_k \subset \cup(A_k - B_k)$

$$
\begin{aligned}
&Q_{J_n}(\cap_{k=1}^n \pi_{J_n,J_k}^{-1}(K_k)) \\
&= Q_0(C_n) - \left\{Q_{J_n}(\cap_{k=1}^n \pi_{J_n,J_k}^{-1}(B_k)) - Q_{J_n}(\cap_{k=1}^n \pi_{J_n,J_k}^{-1}(K_k))\right\} \\
&\geq 2\,\varepsilon - Q_{J_n}\left(\cap_{k=1}^n \pi_{J_n,J_k}^{-1}(B_k) - \cap_{k=1}^n \pi_{J_n,J_k}^{-1}(K_k)\right) \\
&\geq 2\,\varepsilon - Q_{J_n}\left(\cup_{k=1}^n \pi_{J_n,J_k}^{-1}(B_k - K_k)\right) \\
&\geq 2\,\varepsilon - \sum_{k=1}^n Q_{J_n}(\pi_{J_n,J_k}^{-1}(B_k - K_k)) = 2\,\varepsilon - \sum_{k=1}^n Q_{J_k}(B_k - K_k) \\
&\geq 2\,\varepsilon - \sum_{k=1}^n \frac{\varepsilon}{2^k} \geq 2\,\varepsilon - \sum_{k=1}^\infty \frac{\varepsilon}{2^k} = \varepsilon > 0.
\end{aligned}
$$

That is

$$\cap_{k=1}^n \pi_{J_n,J_k}^{-1}(K_k) \neq \emptyset\,, \quad n = 1, 2, \ldots\,.$$

Therefore, for each n there exists $(x_{n,1}, \ldots, x_{n,n})$ such that $(x_{n,1}, \ldots, x_{n,k}) \in K_k$, $k = 1, \ldots, n$. Since $(x_{n,1})_{n=1}^\infty \subset K_1$, and K_1 is compact, by the Bolzano–Weierstrass theorem there exists a subsequence $(n_j^{(1)})_j$ and $x_1 \in K_1 \subset \mathbb{R}^{J_1}$ such that $\lim_{j\to\infty} x_{n_j^{(1)},1} = x_1$. We can choose a further subsequence $(n_j^{(2)})_j$ of $n_j^{(1)}$ such that $\lim_{j\to\infty} x_{n_j^{(2)},2} = x_2$ for some x_2 for which $(x_1, x_2) \in K_2 \subset \mathbb{R}^{J_2}$. Continuing the process we obtain a point $(x_1, x_2, \ldots)$ such that $(x_1, \ldots, x_k) \in K_k$ for $k = 1, 2, \ldots$. Let $x \in \mathbb{R}^I$ such that $\pi_{J_k}(x) = (x_1, \ldots, x_k)$, $k = 1, 2, \ldots$, and the other coordinates are arbitrary. Then

$$x \in \cap_{k=1}^\infty \pi_{J_k}^{-1}(K_k) \subset \cap_{k=1}^\infty \pi_{J_k}^{-1}(B_k) = \cap_{k=1}^\infty C_k,$$

proving $\cap_{k=1}^\infty C_k \neq \emptyset$, thus the statement. □

As an important corollary, we obtain the coordinate representation of a stochastic process.

Theorem 1.8 (Kolmogorov's existence theorem)**.** *Under the conditions of the consistency theorem there exists a probability space* $(\Omega, \mathcal{A}, \mathbf{P})$ *and a stochastic process* $X : \Omega \to \mathcal{X}_I$, $X(\omega)(i) = X_i(\omega)$, $\omega \in \Omega$, $i \in I$, *such that the finite dimensional distributions of* $Q = \mathbf{P} \circ X^{-1}$ *is the given collection of consistent measures*

$$\{Q \circ \pi_J^{-1} : J \subset I, J \text{ finite}\}.$$

Proof. Choose $(\Omega, \mathcal{A}, \mathbf{P}) = (\mathcal{X}_I, \mathcal{F}_I, Q)$, and $X : \mathcal{X}_I \to \mathcal{X}_I : x \mapsto x$. $\square$

Example 1.9. Markov chains. A process (X_n) is a Markov chain, if

$$\mathbf{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n),$$

for any n, $i_0, \ldots, i_{n+1}$. It is homogeneous if the latter equals $= p_{i_n, i_{n+1}}$. For the k-step transition probabilities write

$$p_{i,j}^{(k)} = \mathbf{P}(X_k = j | X_0 = i).$$

Then $P^k = (p_{i,j}^{(k)})$, the k-step transition probabilities can be calculated by matrix multiplication. These are the Chapman–Kolmogorov equations.

Let $I = \{0, 1, \ldots\}$, and $\mathcal{X} = \mathbb{Z}$. A homogeneous Markov chain $(X_n)_{n \geq 0}$ is determined by an initial distribution μ, and a transition matrix $P = (p_{ij})$. Then the finite dimensional distributions are determined as

$$Q_{\{0,\ldots,n\}}(\{(i_0, \ldots, i_n)\}) = \mathbf{P}(X_0 = i_0, \ldots, X_n = i_n) = \mu_{i_0} p_{i_0, i_1} \ldots p_{i_{n-1}, i_n}.$$

Formally, $J = \{0, 1, \ldots, n\}$.

Let's check that these indeed satisfy the consistency condition. For simplicity $J = \{0, 1, 2\}$, $K = \{0, 1\}$. Then

$$\pi_{J,K}^{-1}(\{(i,j)\}) = \{(i,j,k) : k \in \mathbb{Z}\}.$$

Thus

$$Q_J(\pi_{J,K}^{-1}(\{(i,j)\})) = \sum_{k \in \mathbb{Z}} \mu_i p_{i,j} p_{j,k} = \mu_i p_{i,j} = Q_K(\{(i,j)\}),$$

as claimed. Similarly, if $J = \{0, 1, 2\}$, $K = \{0, 2\}$, then

$$\pi_{J,K}^{-1}(\{(i,j)\}) = \{(i,k,j) : k \in \mathbb{Z}\},$$

and

$$Q_J(\pi_{J,K}^{-1}(\{(i,j)\})) = \sum_{k \in \mathbb{Z}} \mu_i p_{i,k} p_{k,j} = \mu_i p_{i,j}^{(2)} = Q_K(\{(i,j)\}),$$

as claimed.

2 Conditional expectation

This is a from Csörgő [2] and Durrett [3].

2.1 Existence and uniqueness

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space, and $\mathcal{G} \subset \mathcal{A}$ a sub-σ-algebra. Let X be an integrable random variable, $\mathbf{E}|X| < \infty$. The *conditional expectation of X given the σ-algebra $\mathcal{G}$* is a random variable $\mathbf{E}[X|\mathcal{G}]$ which is

(i) $\mathcal{G}$ measurable and integrable,

(ii) for any $G \in \mathcal{G}$

$$\int_G \mathbf{E}[X|\mathcal{G}] \mathrm{d}\mathbf{P} = \int_G X \mathrm{d}\mathbf{P}.$$

Note that conditional expectation is a random variable. Therefore, all equalities, inequalities involving conditional expectation are meant almost surely.

The *conditional probability* of A given $\mathcal{G}$ is defined by

$$\mathbf{P}(A|\mathcal{G}) = \mathbf{E}[\mathbb{I}(A)|\mathcal{G}].$$

Example 2.1. If $\mathcal{G} = \{\emptyset, \Omega\}$ then $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}X$, while if $\mathcal{G} = \mathcal{A}$ (or more generally, X is $\mathcal{G}$-measurable) then $\mathbf{E}[X|\mathcal{G}] = X$.

If X is independent of the σ-algebra $\mathcal{G}$, then $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}X$.

Theorem 2.2. *Conditional expectation exists and it is unique.*

Proof. **Existence.** First recall the Radon–Nikodym theorem. Let μ, ν be measure on the measurable space $(\mathcal{X}, \mathcal{F})$. Then ν is absolute continuous with respect to μ, $\nu \ll \mu$, if $\mu(A) = 0$ implies $\nu(A) = 0$. The Radon–Nikodym theorem states that whenever $\nu \ll \mu$ then there exists a measurable function f such that

$$\nu(A) = \int_A f \mathrm{d}\mu.$$

Then $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ is called the Radon–Nikodym derivative of ν with respect to μ.

Let's go back to our setup. First assume that $X \geq 0$, and define a new measure ν on $(\Omega, \mathcal{G})$ as

$$\nu(G) = \int_G X \mathrm{d}\mathbf{P}, \quad G \in \mathcal{G}.$$

Clearly, $\nu \ll \mathbf{P}$. Therefore, there exists f $\mathcal{G}$-measurable, for which

$$\nu(G) = \int_G f \mathrm{d}\mathbf{P}.$$

Then $f = \mathbf{E}[X|\mathcal{G}]$ satisfies (i) and (ii) in the definition above.

In the general case write $X = X^+ - X^-$, and apply the result above.

Uniqueness. Let Y, Z be random variables satisfying (i) and (ii). Choose $G = \{Y > Z\} \in \mathcal{G}$. Then

$$\int_G Y \mathrm{d}\mathbf{P} = \int_G Z \mathrm{d}\mathbf{P},$$

implying $\mathbf{P}(G) = 0$. Symmetry gives the statement. □

We need some results from measure theory. A π-system is a collection of sets closed under intersection. A class of sets $\mathcal{L}$ on Ω is λ-system if

(i) $\emptyset, \Omega \in \mathcal{L}$;

(ii) if $A, B \in \mathcal{L}$ and $A \subset B$ then $B - A \in \mathcal{L}$; (i.e. closed under proper difference)

(iii) $A_n \uparrow$, $A_n \in \mathcal{L}$ implies $\cup_n A_n \in \mathcal{L}$ (closed under monotone union).

Exercise 2.3. Let μ and ν be finite measure on $(\Omega, \mathcal{A})$ such that $\mu(\Omega) = \nu(\Omega)$. Show that

$$\mathcal{L} = \{A : \mu(A) = \nu(A)\}$$

is λ-system.

The following result and its proof is similar to the monotone class theorem, but it more useful.

Theorem 2.4 (π–λ theorem)**.** *Let $\mathcal{C}$ be a π-system. Then $\sigma(\mathcal{C}) = \lambda(\mathcal{C})$.*

It is enough to check the defining integral condition (ii) on a generating π-system.

Proposition 2.5. *Let Z be $\mathcal{G}$-measurable, $\mathcal{C}$ be a π-system such that $\sigma(\mathcal{C}) = \mathcal{G}$. If*

$$\int_C Z \mathrm{d}\mathbf{P} = \int_C X \mathrm{d}\mathbf{P} \tag{4}$$

for each $C \in \mathcal{C}$, then $Z = \mathbf{E}[X|\mathcal{G}]$.

Proof. Let $\mathcal{L}$ be the set of all sets C for which (4) holds. Then, by assumption $\mathcal{C} \subset \mathcal{L}$. It is easy to check that $\mathcal{L}$ is a λ-system. The π–λ-theorem gives the result. □

We also frequently use the following simple lemma.

Lemma 2.6. *Let $\mathcal{G}$ be a sub-σ-algebra, X, Y $\mathcal{G}$-measurable. Then $X \leq Y$ a.s. iff*

$$\int_G X \mathrm{d}\mathbf{P} \leq \int_G Y \mathrm{d}\mathbf{P} \quad \forall G \in \mathcal{G}.$$

Next we relate the new notion to the familiar notion of conditional probability.

Example 2.7. Recall the definition of conditional probability from the elementary probability course: $\mathbf{P}(A|B) = \frac{\mathbf{P}(A\cap B)}{\mathbf{P}(B)}$. Assume that $B_1, B_2, \ldots$ are a countable partition of Ω, i.e. they are disjoint, and the union is Ω. Let $\mathcal{G} = \sigma(B_1, B_2, \ldots)$. Then $\mathbf{P}(A|\mathcal{G}) = \sum_i I_{B_i} \mathbf{P}(A|B_i)$.

Using the definition of conditional probability we have that $\mathbf{P}(A|\mathcal{G})$ is a random varible, which is (i) $\mathcal{G}$-measurable; (ii) for each $G \in \mathcal{G}$ we have

$$\int_G \mathbf{P}(A|\mathcal{G}) \mathrm{d}\mathbf{P} = \mathbf{P}(A \cap G).$$

The measurability implies that $\mathbf{P}(A|\mathcal{G})$ has to be a constant on each B_i, and (ii) with $G = B_i$ implies that the constant on B_i has to be $\mathbf{P}(A|B_i)$, as stated.

Exercise 2.8. In the setup above, determine $\mathbf{E}[X|\mathcal{G}]$.

2.2 Conditional expectation with respect to random variable

Let Y be a random variable. Then

$$\mathbf{E}[X|Y] = \mathbf{E}[X|\sigma(Y)].$$

The next lemma implies that $\mathbf{E}[X|Y]$ is a function of Y.

Lemma 2.9. *Let Y be a random element on $(\mathcal{Y}, \mathcal{F})$, and let Z be a $\sigma(Y)$-measurable random variable. Then there exists an $\mathcal{F}$-measurable function $g : \mathcal{Y} \to \mathbb{R}$, such that $Z = g(Y)$ a.s.*

Proof. Writing $Z = Z_+ - Z_-$, we may assume that $Z \geq 0$. Consider a sequence of $\sigma(Y)$-simple functions

$$Z_m = \sum_{i=1}^{k_m} c_{m,i} \mathbb{I}(A_{m,i}), \quad A_{m,i} \in \sigma(Y),$$

such that $Z_m \uparrow Z$. Since $A_{m,i} = Y^{-1}(F_{m,i})$, for some $F_{m,i} \in \mathcal{F}$, we see that

$$Z_m = \sum_{i=1}^{m} c_{m,i} \mathbb{I}(Y \in F_{m,i}) =: g_m(Y).$$

Clearly, g_m is $\mathcal{F}$-measurable, and $g_m(Y(\omega)) \uparrow Z(\omega)$ a.s. Let

$$g(y) = \limsup_{m \to \infty} g_m(y).$$

Then g is measurable, and $g(Y) = \limsup_{m\to\infty} g_m(Y) = Z$, as claimed. $\square$

Since $\mathbf{E}[X|Y]$ is $\sigma(Y)$-measurable (by definition), there exists a g_X measurable such that $\mathbf{E}[X|Y] = g_X(Y)$ a.s. The *conditional expectation of X given $Y = y$* is

$$\mathbf{E}[X|Y = y] = g_X(y), \quad y \in \mathcal{Y}.$$

Note that the event $\{Y = y\}$ has probability 0 for continuous Y, therefore the definition above does not make sense in the old (elementary) definition of conditional probability.

By the definition, for any $F \in \mathcal{F}$

$$\begin{aligned}\int_{\{Y \in F\}} X \mathrm{d}\mathbf{P} &= \int_{\{Y \in F\}} \mathbf{E}[X|Y] \, \mathrm{d}\mathbf{P} = \int_{\{Y \in F\}} g_X(Y) \, \mathrm{d}\mathbf{P} \\ &= \int_F g_X(y) \, \mu_Y(\mathrm{d}y) = \int_F \mathbf{E}[X|Y = y] \, \mu_Y(\mathrm{d}y).\end{aligned}$$

For a real random variable, with $F = \mathbb{R}$

$$\mathbf{E}X = \int_{\mathbb{R}} \mathbf{E}[X|Y = y] \, \mu_Y(\mathrm{d}y).$$

In particular, if Y is discrete

$$\mathbf{E}X = \sum_j \mathbf{E}[X|Y = y_j] \, \mathbf{P}(Y = y_j),$$

while if it is continuous with density f_Y

$$\mathbf{E}X = \int_{\mathbb{R}} \mathbf{E}[X|Y = y] f_Y(y) \, \mathrm{d}y.$$

2.3 Elementary setup

2.3.1 Discrete case

Let X, Y be discrete random variables with possible values $x_1, x_2, \ldots$, and $y_1, y_2, \ldots$. According to previous definitions from elementary probability,

$$\mathbf{P}(X = x_k | Y = y_\ell) = \frac{\mathbf{P}(X = x_k, Y = y_\ell)}{\mathbf{P}(Y = y_\ell)},$$

and

$$\mathbf{E}[X | Y = y_\ell] = \sum_k \mathbf{P}(X = x_k | Y = y_\ell) x_k.$$

Let $\mathbf{E}[X|Y]$ denote the random variable defined on the probability space $(\Omega, \mathcal{A}, \mathbf{P})$ whose value on the event $Y = y_\ell$ is given by $\mathbf{E}[X|Y = y_\ell]$. Formally,

$$\mathbf{E}[X|Y](\omega) = \sum_i \mathbf{E}[X|Y = y_i] I(Y = y_i),$$

where $I(\cdot)$ denotes the indicator function. Notice that $\mathbf{E}[X|Y]$ is a random variable that is a function of Y.

Theorem 2.10 (Law of Total Probability and Expectation for the Discrete Case)**.** *Let X, Y be discrete random variables with possible values $x_1, x_2, \ldots$, and $y_1, y_2, \ldots$. Then,*

$$\mathbf{P}(X = x_k) = \sum_i \mathbf{P}(X = x_k | Y = y_i) \mathbf{P}(Y = y_i)$$

$$\mathbf{E}(X) = \sum_i \mathbf{E}[X | Y = y_i] \mathbf{P}(Y = y_i).$$

Proof. The first equality follows from the law of total probability with the partition $Y = y_i$. The second equality follows from the first and the definition:

$$\begin{aligned}
\mathbf{E}(X) &= \sum_k \mathbf{P}(X = x_k) x_k \\
&= \sum_k \sum_i \mathbf{P}(X = x_k | Y = y_i) \mathbf{P}(Y = y_i) x_k \\
&= \sum_i \sum_k \mathbf{P}(X = x_k | Y = y_i) \mathbf{P}(Y = y_i) x_k \\
&= \sum_i \mathbf{E}[X | Y = y_i] \mathbf{P}(Y = y_i).
\end{aligned}$$

□

2.3.2 Continuous case

Let X, Y be jointly continuous random variables with density function h. Let

$$f_X(x) = \int_{-\infty}^{\infty} h(x,y)\mathrm{d}y, \quad f_Y(y) = \int_{-\infty}^{\infty} h(x,y)\mathrm{d}x$$

denote the density functions of X and Y, respectively. Then, the conditional density function of X given Y is

$$f_{X|Y}(x|y) = \begin{cases} \frac{h(x,y)}{f_Y(y)}, & \text{if } f_Y(y) \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Observe that if $f_Y(y) > 0$, then $f_{X|Y}(\cdot|y)$ is indeed a density function. The conditional expectation of X given Y is

$$\mathbf{E}[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\mathrm{d}x,$$

provided that the integral is well-defined.

Theorem 2.11 (Law of Total Probability and Expectation for the Continuous Case)**.** *Let X, Y be jointly continuous random variables with density function h. Then,*

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\mathrm{d}y \\ \mathbf{E}(X) &= \int_{-\infty}^{\infty} \mathbf{E}[X|Y=y] f_Y(y)\mathrm{d}y. \end{aligned}$$

Proof. By definition,

$$\begin{aligned} \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\mathrm{d}y &= \int_{y: f_Y(y)>0} f_{X|Y}(x|y) f_Y(y)\mathrm{d}y \\ &= \int_{y: f_Y(y)>0} \frac{h(x,y)}{f_Y(y)} f_Y(y)\mathrm{d}y \\ &= \int_{y: f_Y(y)>0} h(x,y)\mathrm{d}y \\ &= \int_{-\infty}^{\infty} h(x,y)\mathrm{d}y = f_X(x). \end{aligned}$$

The law of total expectation follows as

$$\begin{aligned}
\mathbf{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\mathrm{d}y \right) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f_{X|Y}(x|y) x \mathrm{d}x \right) f_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \mathbf{E}[X|Y=y] f_Y(y)\mathrm{d}y.
\end{aligned}$$

□

From this, the continuous version of Bayes' theorem follows.

Theorem 2.12 (Continuous Bayes' Theorem)**.** *Let x, y be such that $f_X(x) > 0$ and $f_Y(y) > 0$. Then,*

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|u) f_X(u)\mathrm{d}u}.$$

Proof. This follows directly from the previous results:

$$\frac{h(x,y)}{f_Y(y)} = f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|u) f_X(u)\mathrm{d}u} = \frac{\frac{h(x,y)}{f_X(x)} f_X(x)}{f_Y(y)}.$$

□

2.4 Properties

Theorem 2.13. *Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space, $\mathcal{G} \subset \mathcal{A}$ a sub-σ-algebra, X, Y, X_n are integrable random variables, and $a, b \in \mathbb{R}$.*

(i) *(linearity)* $\mathbf{E}[aX + bY|\mathcal{G}] = a\mathbf{E}[X|\mathcal{G}] + b\mathbf{E}[Y|\mathcal{G}]$.

(ii) *(monotonicity) If $X \leq Y$ then* $\mathbf{E}[X|\mathcal{G}] \leq \mathbf{E}[Y|\mathcal{G}]$.

(iii) $|\mathbf{E}[X|\mathcal{G}]| \leq \mathbf{E}[|X|\,|\mathcal{G}]$.

(iv) *(monotone convergence)* $0 \leq X_n \uparrow X$ *a.s., then* $\mathbf{E}[X_n|\mathcal{G}] \uparrow \mathbf{E}[X|\mathcal{G}]$.

(v) *If* $0 \geq X_n \downarrow X$ *a.s., then* $\mathbf{E}[X_n|\mathcal{G}] \downarrow \mathbf{E}[X|\mathcal{G}]$.

(vi) *(conditional Fatou) If* $X_n \geq 0$ *and* $\liminf_{n\to\infty} X_n$ *is integrable, then*

$$\mathbf{E}[\liminf_{n\to\infty} X_n|\mathcal{G}] \leq \liminf_{n\to\infty} \mathbf{E}[X_n|\mathcal{G}].$$

(vii) *(conditional Lebesgue's dominated convergence) If* $\lim_{n\to\infty} X_n = X$ *a.s., and* $|X_n| \leq Y$ *a.s., for all* $n \in \mathbb{N}$*, with an integrable* Y *then* $\lim_{n\to\infty} \mathbf{E}[X_n|\mathcal{G}] = \mathbf{E}[X|\mathcal{G}]$*.*

(viii) *If* $\lim_{n\to\infty} X_n = X$ *a.s., and* $\mathbf{E}[\sup_n |X_n|] < \infty$*, then* $\mathbf{E}[X_n|\mathcal{G}] \to \mathbf{E}[X|\mathcal{G}]$*.*

(ix) *If* Z *is* $\mathcal{G}$*-measurable,* X *and* ZX *are integrable, then*

$$\mathbf{E}[ZX|\mathcal{G}] = Z\mathbf{E}[X|\mathcal{G}].$$

(x) *(tower rule) If* $\mathcal{G}_1 \subset \mathcal{G}_2$ *sub-*σ*-algebras then*

$$\mathbf{E}\Big[\mathbf{E}[X|\mathcal{G}_2]|\mathcal{G}_1\Big] = \mathbf{E}[X|\mathcal{G}_1] = \mathbf{E}\Big[\mathbf{E}[X|\mathcal{G}_1]|\mathcal{G}_2\Big].$$

(xi) *(conditional Jensen inequality) If* φ *is convex and* $\mathbf{E}|X| < \infty$*,* $\mathbf{E}|\varphi(X)| < \infty$*, then*

$$\varphi\left(\mathbf{E}[X|\mathcal{G}]\right) \leq \mathbf{E}\left[\varphi(X)|\mathcal{G}\right].$$

Proof. (i) The RHS is $\mathcal{G}$-measurable, and satisfies the integral equation defining the LHS.

(ii) Both sides are $\mathcal{G}$-measurable, thus it is enough to prove that

$$\int_G \mathbf{E}[X|\mathcal{G}]\mathrm{d}\mathbf{P} \leq \int_G \mathbf{E}[Y|\mathcal{G}]\mathrm{d}\mathbf{P}$$

for any $G \in \mathcal{G}$. This follows from the defining integral equation of the conditional expectation.

(iii) follows from (ii).

(iv) From (ii) we see that $\mathbf{E}[X_n|\mathcal{G}]$ is increasing, therefore it has a $\mathcal{G}$-measurable limit, say Z. Monotone convergence theorem implies that Z satisfies the defining integral equation of $\mathbf{E}[X|\mathcal{G}]$.

(v) follows from (iv).

(vi) Put $Y_n = \inf_{m\geq n} X_m$. Then $Y_n \uparrow \liminf_{n\to\infty} X_n$, thus by (iv)

$$\mathbf{E}[Y_n|\mathcal{G}] \uparrow \mathbf{E}[\liminf_{n\to\infty} X_n|\mathcal{G}].$$

Simply $Y_n \le X_n$, thus by (ii) $\mathbf{E}[Y_n|\mathcal{G}] \le \mathbf{E}[X_n|\mathcal{G}]$, so

$$\liminf \mathbf{E}[Y_n|\mathcal{G}] \le \liminf \mathbf{E}[X_n|\mathcal{G}].$$

(vii) By assumption $Z_n := \sup_{m\ge n} |X_m - X| \downarrow 0$ a.s., and by Lebesgue's dominated convergence theorem, $\mathbf{E}Z_n \to 0$. Then

$$|\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}]| \le \mathbf{E}[|X_n - X|\,|\mathcal{G}] \le \mathbf{E}[Z_n|\mathcal{G}].$$

Since $Z_n \downarrow 0$, we have that $\mathbf{E}[Z_n|\mathcal{G}] \downarrow V$, for some V nonnegative, $\mathcal{G}$-measurable. Then

$$\mathbf{E}V \le \mathbf{E}Z_n \to 0,$$

implying that $V = 0$ a.s.

(viii) follows from (vii) with $Y = \sup |X_n|$.

(ix) As usual we may assume that X, Z are nonnegative. We show that $Z\mathbf{E}[X|\mathcal{G}]$ satisfies the defining properties of $\mathbf{E}[ZX|\mathcal{G}]$. Measurability is clear. If $Z = \mathbb{I}_A$, $A \in \mathcal{G}$, then

$$\int_G \mathbb{I}_A \mathbf{E}[X|\mathcal{G}]\mathrm{d}\mathbf{P} = \int_{A\cap G} \mathbf{E}[X|\mathcal{G}]\mathrm{d}\mathbf{P} = \int_{A\cap G} X\mathrm{d}\mathbf{P} = \int_G \mathbb{I}_A X\mathrm{d}\mathbf{P}.$$

Thus the statement holds for indicators. Using linearity and monotone convergence the statement follows.

(x) Since $\mathbf{E}[X|\mathcal{G}_1]$ is $\mathcal{G}_2$-measurable, the second equality follows. For the first equality, for $G \in \mathcal{G}_1$

$$\int_G \mathbf{E}\left[\mathbf{E}[X|\mathcal{G}_2]|\mathcal{G}_1\right]\mathrm{d}\mathbf{P} = \int_G \mathbf{E}[X|\mathcal{G}_2]\mathrm{d}\mathbf{P} = \int_G X\mathrm{d}\mathbf{P},$$

as claimed, where the second equality holds since $G \in \mathcal{G}_2$.

(xi) By convexity, for any x, a

$$\varphi(x) \ge \varphi(a) + \varphi'(a)(x-a).$$

Thus

$$\varphi(X) \ge \varphi\left(\mathbf{E}[X|\mathcal{G}]\right) + \varphi'\left(\mathbf{E}[X|\mathcal{G}]\right)\left(X - \mathbf{E}[X|\mathcal{G}]\right).$$

Take conditional expectation $\mathbf{E}[\cdot|\mathcal{G}]$ and note that on the RHS $\varphi'\left(\mathbf{E}[X|\mathcal{G}]\right)$ is $\mathcal{G}$-measurable. If φ is not differentiable, take left derivative, which exists. $\square$

Exercise 2.14 (Chebyshev's inequality)**.** Prove that

$$\mathbf{P}(|X| \geq x|\mathcal{G}) \leq \frac{\mathbf{E}[X^2|\mathcal{G}]}{x^2}.$$

Theorem 2.15. *Assume that* $\mathbf{E}X^2 < \infty$*. Then*

$$\inf_{Y\mathcal{G}-measurable} \mathbf{E}(X-Y)^2 = \mathbf{E}\left(X - \mathbf{E}[X|\mathcal{G}]\right)^2;$$

that is $\mathbf{E}[X|\mathcal{G}]$ *is that* $\mathcal{G}$*-measurable random variable which minimizes the error* $\mathbf{E}(X-Y)^2$*.*

This means that the conditional expectation is a projection to

$$L^2(\mathcal{G}) = \{Z : \Omega \to \mathbb{R}, \mathbf{E}Z^2 < \infty,\ Z\ \mathcal{G} - \text{measurable}\}.$$

Proof. For $Z \in L^2(\mathcal{G})$

$$\mathbf{E}\left(Z\mathbf{E}[X|\mathcal{G}]\right) = \mathbf{E}(ZX),$$

that is

$$\mathbf{E}\left(Z(X - \mathbf{E}[X|\mathcal{G}])\right) = 0.$$

Thus for $Y \in L^2(\mathcal{G})$

$$\begin{aligned}\mathbf{E}[(X-Y)^2] &= \mathbf{E}[(X - \mathbf{E}[X|\mathcal{G}] + \mathbf{E}[X|\mathcal{G}] - Y)^2] \\ &= \mathbf{E}[(X - \mathbf{E}[X|\mathcal{G}])^2] + \mathbf{E}[(\mathbf{E}[X|\mathcal{G}] - Y)^2].\end{aligned}$$

Thus we see that the minimum is attained at $Y = \mathbf{E}[X|\mathcal{G}]$. □

2.5 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 $\mathbf{P}(A|\mathcal{G}) = \mathbf{E}[\mathbb{I}_A|\mathcal{G}]$. Since $0 \leq \mathbb{I}_A \leq 1$,

$$\mathbf{P}(A|\mathcal{G}) \in [0,1] \quad \text{a.s.} \tag{5}$$

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

$$\mathbf{P}(\cup_{i=1}^{\infty} A_i|\mathcal{G}) = \sum_{i=1}^{\infty} \mathbf{P}(A_i|\mathcal{G}) \quad \text{a.s.} \tag{6}$$

Therefore, $\mathbf{P}(\cdot|\mathcal{G})$ behaves as a probability measure. However, both (5) and (6) 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 (5) 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 (6). 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 $A \in \mathcal{S}$ 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 2.16. *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 $\mathbf{P}(X \leq q|\mathcal{G})$, $q \in \mathbb{Q}$. For each $q \in \mathbb{Q}$ 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

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

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

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 2.17. *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 $G(q, \omega) = \mathbf{P}(X \le q|Y)(\omega)$, a.s., $q \in \mathbb{Q}$. 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 $F(x, Y(\omega)) = \mathbf{P}(X \le x|Y)(\omega)$. 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 2.18. *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*

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

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

3 Discrete time martingales

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

3.1 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 3.1. 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 3.2. *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 3.3. *(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 submartingale.*

(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 3.4. *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.*

3.2 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

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

Note that this is a discrete stochastic integral.

Theorem 3.5. *Let (X_n) be a submartingale and (H_n) be predictable, non-negative, and bounded. Then $(H \cdot X)_n$ is a submartingale.*

Proof. Follows from the submartingale property. $\square$

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

Corollary 3.6. *Let τ be a stopping time, and (X_n) a submartingale. Then $(X_{\tau \wedge n})$ is a submartingale.*

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

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

$$H_m = \begin{cases} 1, & \text{if } \tau_{2k-1} < m \leq \tau_{2k} \text{ for some } k, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

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 3.7 (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 (7) 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 3.8 (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

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

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

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

□

3.3 Examples

3.3.1 Bounded increments

Theorem 3.9. *Let (X_n) be a martingale with $|X_{n+1} - X_n| \leq M < \infty$ for some M. Let*

$$C = \{\lim_{n\to\infty} X_n \text{ exists and finite}\},$$
$$D = \{\limsup_{n\to\infty} X_n = \infty, \quad \liminf_{n\to\infty} X_n = -\infty\}.$$

Then $\mathbf{P}(C \cup D) = 1$.

Proof. Since $(X_n - X_0)$ is martingale, we may and do assume that $X_0 = 0$. Fix $K > 0$, and put $\tau = \inf\{n : X_n \leq -K\}$. Then τ is stopping time, and $X_{\tau\wedge n} \geq -K-M$. Thus $X_{\tau\wedge n}+K+M$ is a nonnegative martingale. Therefore, by the martingale convergence theorem $\lim X_n$ exists and finite on the set $\{\tau = \infty\}$. Letting $K \to \infty$ we see that $\lim X_n$ exists on $\{\liminf X_n > -\infty\}$. Applying the same argument for $-X_n$, the result follows. $\square$

Corollary 3.10 (Conditional Borel–Cantelli lemma)**.** *Let $\mathcal{F}_n$ be a filtration with $\mathcal{F}_0 = \{\emptyset, \Omega\}$, and let $A_n \in \mathcal{F}_n$. Then*

$$\{A_n \text{ infinitely often}\} = \left\{ \sum_{n=1}^{\infty} \mathbf{P}(A_n|\mathcal{F}_{n-1}) = \infty \right\}.$$

Proof. Put

$$X_n = \sum_{k=1}^{n} [I_{A_k} - \mathbf{P}(A_k|\mathcal{F}_{k-1})].$$

Then (X_n) is a martingale, with $|X_n - X_{n-1}| \leq 1$. Using the notation of the theorem above, clearly both on C and on D $\sum_n I_{A_n} = \infty$ iff $\sum_n \mathbf{P}(A_n|\mathcal{F}_{n-1}) = \infty$. $\square$

3.3.2 Pólya urns

An urn contains red and green balls. Each time we draw a ball randomly with replacement, and additionally add c balls of the same color. Let X_n denote the fraction of green balls after the nth draw. Show that X_n is a martingale.

Since X_n is nonnegative, by the martingale convergence theorem $\lim_n X_n = X_\infty$ exists a.s.

To determine the limit, let G_n and R_n denote the number of green and red balls after the nth draw. Starting with r red, g green balls,

$$\begin{aligned}\mathbf{P}(G_n = g+mc) = &\binom{n}{m}\frac{g}{g+r}\frac{g+c}{g+r+c}\cdots\frac{g+(m-1)c}{g+r+(m-1)c}\\ &\times\frac{r}{g+r+mc}\frac{r+c}{g+r+(m+1)c}\cdots\frac{r+(n-m-1)c}{g+r+(n-1)c}.\end{aligned}$$

Assume $c=1$ and initially the urn contains 1 red and 1 green ball. Then

$$\mathbf{P}(G_n = m+1) = \binom{n}{m}\frac{m!(n-m)!}{(n+1)!} = \frac{1}{n+1}.$$

Thus $X_n = G_n/(n+1)$ converges to the uniform distribution on $[0,1]$.

Still with $c=1$, but r and g general,

$$\begin{aligned}\mathbf{P}(G_n = g+m) = &\frac{(g+r-1)!}{(g-1)!(r-1)!}\\ &\times\frac{(m+1)\dots(m+g-1)(n-m+1)\dots(n-m+r-1)}{(n+1)\dots(n+g+r-1)}\\ \sim &\frac{1}{n}\frac{(g+r-1)!}{(g-1)!(r-1)!}x^{g-1}(1-x)^{r-1},\end{aligned}$$

as $n\to\infty$ and $m/n\to x$. This shows that the limiting distribution has density function

$$\frac{(g+r-1)!}{(g-1)!(r-1)!}x^{g-1}(1-x)^{r-1},\quad x\in(0,1),$$

which is the beta distribution with parameters g and r.

In the genreal case using the Stirling formula $\Gamma(x)\sim\left(\frac{x}{e}\right)^x\sqrt{2\pi/x}$

$$\begin{aligned}\mathbf{P}(G_n = g+mc) &= \binom{n}{m}\frac{c^m\prod_{j=0}^{m-1}(g/c+j)c^{n-m}\prod_{j=0}^{n-m-1}(r/c+j)}{c^n\prod_{j=0}^{n-1}((g+r)/c+j)}\\ &= \binom{n}{m}\frac{\Gamma(g/c+m)\Gamma(r/c+n-m)\Gamma((g+r)/c)}{\Gamma(g/c)\Gamma(r/c)\Gamma((g+r)/c+n)}\\ &\sim \frac{1}{n}x^{g/c-1}(1-x)^{r/c-1}\frac{\Gamma((g+r)/c)}{\Gamma(g/c)\Gamma(r/c)},\end{aligned}$$

where the last limit holds as $m/n \to x$. Therefore, the limit distribution in the general case is beta with parameters g/c and r/c.

To see this, fix $0 < y < x < 1$. Then $X_n = G_n/(r+g+nc)$, thus

$$\begin{aligned}\mathbf{P}(y < X_n \le x) &= \sum_{m=((r+g)y-g)/c+ny}^{((r+g)x-g)/c+nx} \mathbf{P}(G_n = g+mc)\\ &\sim \sum_{m=((r+g)y-g)/c+ny}^{((r+g)x-g)/c+nx} \frac{1}{n}\left(\frac{m}{n}\right)^{g/c-1}\left(1-\frac{m}{n}\right)^{r/c-1}\frac{\Gamma((g+r)/c)}{\Gamma(g/c)\Gamma(r/c)}\\ &\to \int_y^x u^{g/c-1}(1-u)^{r/c-1}\frac{\Gamma((g+r)/c)}{\Gamma(g/c)\Gamma(r/c)}\mathrm{d}u.\end{aligned}$$

3.3.3 Galton–Watson process

Fix an offspring distribution on the nonnegative integers (p_k). Let $Z_0 = 1$ and

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{i,n},$$

where $\{\xi_{i,n} : i = 1, 2, \ldots, n = 0, 1, \ldots\}$ are independent and identically distributed, with common distribution (p_k). Let

$$m = \sum_{k=1}^{\infty} k p_k$$

denote the offspring mean. Then the process $X_n = Z_n/m^n$ is a martingale. Indeed, if $(\mathcal{F}_n)$ stands for the natural filtration, then

$$\mathbf{E}[Z_n|\mathcal{F}_{n-1}] = \mathbf{E}\left[\sum_{i=1}^{Z_{n-1}} \xi_{i,n-1}|\mathcal{F}_{n-1}\right] = mZ_{n-1}.$$

Since X_n is nonnegative, the limit $X_\infty = \lim_{n\to\infty} X_n$ exists a.s. It is difficult in general to determine its distribution. If $m < 1$ then it is easy to show that $X_\infty \equiv 0$. This is an example for which $\mathbf{E}(X_\infty) \neq \mathbf{E}(X_0)$.

Lemma 3.11. *If $m < 1$ then $Z_n = 0$ for n large enough.*

Proof. By Markov

$$\mathbf{P}(Z_n > 0) = \mathbf{P}(Z_n \geq 1) \leq \mathbf{E}(Z_n) = m^n.$$

Therefore $\mathbf{P}(Z_n > 0) \to 0$ exponentially fast. Since $\{Z_n > 0\}$ is decreasing sequence of events, the result follows. □

Let $f(s) = \mathbf{E}s^\xi = \sum_{k=0}^\infty p_k s^k$ denote the offspring generating function. Since $X_n \to X_\infty$

$$\mathbf{E}e^{-\lambda X_n} \to \mathbf{E}e^{-\lambda X_\infty} = \varphi(\lambda),$$

where φ is the Laplace transform of X_∞. On the other hand, by the branching structure

$$\begin{aligned}
\mathbf{E}e^{-\lambda \frac{Z_n}{m^n}} &= \mathbf{E}\exp\left\{-\frac{\lambda}{m}\sum_{i=1}^{\xi}\frac{Z_{n-1}^i}{m^{n-1}}\right\} \\
&= \sum_{k=0}^\infty p_k \left(\mathbf{E}e^{-\frac{\lambda}{m}X_{n-1}}\right)^k \\
&= f\left(\mathbf{E}e^{-\frac{\lambda}{m}X_{n-1}}\right).
\end{aligned}$$

Letting $n \to \infty$ we obtain that φ is a solution to the functional equation

$$\varphi(\lambda) = f(\varphi(\lambda/m)).$$

It can be shown that this is the only solution (up to multiplicative constant).

3.4 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 3.12 (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. $\square$

3.5 Doob's maximal inequality

This part is mainly from Csörgő [2].

Our first optional stopping theorem is the following.

Theorem 3.13. *Let $(X_n)_n$ be a submartingale and let τ be a bounded stopping time, i.e. $\tau \le k$ a.s. for some $k \in \mathbb{N}$. Then*

$$\mathbf{E}X_0 \le \mathbf{E}X_\tau \le \mathbf{E}X_k.$$

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

$$\mathbf{E}X_0 = \mathbf{E}X_{\tau\wedge 0} \le \mathbf{E}X_{\tau\wedge k} = \mathbf{E}X_\tau.$$

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

$$(K \cdot X)_n = \sum_{i=1}^{n} \mathbb{I}(\tau \le i-1)(X_i - X_{i-1}) = X_n - X_{\tau\wedge n}.$$

That is

$$\mathbf{E}X_k - \mathbf{E}X_\tau = \mathbf{E}(K \cdot X)_k \ge \mathbf{E}(K \cdot X)_0 = 0.$$

$\square$

An easy consequence is Doob's maximal inequality.

Theorem 3.14 (Doob's maximal inequality)**.** *Let $(X_k, \mathcal{F}_k)_k$ be a submartingale, and put*

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

Then for any $x > 0$

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

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

Proof. The second inequality is obvious.

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

$$x\mathbf{P}(M_n \geq x) \leq \int_{\{M_n \geq x\}} X_\tau \mathrm{d}\mathbf{P}.$$

By Theorem 3.13 $\mathbf{E}X_\tau \leq \mathbf{E}X_n$, and $X_\tau = X_n$ on the event $\{M_n < x\}$, thus

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

proving the statement. $\square$

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

Example 3.15 (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 3.16. *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. $\square$

Exercise 3.17. Prove the corollary.

For the L^p version we need a lemma.

Lemma 3.18. *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,

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

The result follows using Hölder's inequality as

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

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

Theorem 3.19 (L^p maximal inequality)**.** *(i) Let $(X_k)_{k=1}^n$ be a nonnegative submartingale and $p \in (1, \infty)$. Then*

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

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

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

Proof. Statement (i) follows from Doob's maximal inequality and Lemma 3.18.

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

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

□

3.6 Optional stopping theorem

Let $(\Omega, \mathcal{A}, \mathbf{P})$ 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 3.20. *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 3.21. 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

$$\mathcal{F}_\tau = \{A \in \mathcal{F} :\ A \cap \{\tau \le n\} \in \mathcal{F}_n,\ n = 1, 2, \dots\}. \tag{8}$$

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

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

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

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

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

Exercise 3.22. 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 3.23. *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 $\sigma \le \tau$, 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 3.24 (Optional stopping theorem, Doob)**.** *Let $(X_n)_n$ be a submartingale, and $\sigma \le \tau$ stopping times such that*

$$\mathbf{E}(|X_\sigma|) < \infty\,, \qquad \mathbf{E}(|X_\tau|) < \infty \tag{9}$$

and

$$\liminf_{n \to \infty} \int_{\{\tau > n\}} |X_n| \,\mathrm{d}\mathbf{P} = 0. \tag{10}$$

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

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

Clearly, conditions (9) and (10) 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

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

This is the same as

$$\int_A (X_\tau - X_\sigma)\,\mathrm{d}\mathbf{P} \geq 0 \text{ for all } A \in \mathcal{F}_\sigma. \tag{11}$$

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

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

thus

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

proving (11).

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

$$\int_A (X_{\tau\wedge n} - X_{\sigma\wedge n})\mathrm{d}\mathbf{P} = \int_{A\cap\{\sigma<n\}} (X_{\tau\wedge n} - X_{\sigma\wedge n})\mathrm{d}\mathbf{P} \geq 0\,, \quad n \in \mathbb{N}, \tag{12}$$

where the inequality follows from the previous case.

By condition (10) 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 (12) tends to 0. For the second term

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

Similarly, for the third term

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

Using (9) both upper bounds tend to 0. □

Corollary 3.25. *Assume that (X_n) is (super-, sub-) martingale, τ is a stopping time, $\mathbf{E}(|X_\tau|) < \infty$ and (10) 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 3.26 (Simple symmetric random walk)**.** Let $\xi, \xi_1, \xi_2, \ldots$ 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 (10) does not hold.

Theorem 3.27 (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. Consider the decomposition $S_\tau = S_\tau^{(+)} - S_\tau^{(-)}$ where

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

and

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

Then $\mathbf{E}S_\tau = \mathbf{E}S_\tau^{(+)} - \mathbf{E}S_\tau^{(-)}$. We have

$$\begin{aligned}
\mathbf{E}(S_\tau^{(+)}) &= \mathbf{E}\left(\sum_{k=1}^{\infty} \mathbb{I}(\tau \geq k) X_k^+ \right) \\
&= \sum_{k=1}^{\infty} \mathbf{E}(\mathbb{I}(\tau \geq k) X_k^+) \\
&= \sum_{k=1}^{\infty} \mathbf{E}\mathbb{I}(\tau \geq k) \mathbf{E}(X_k^+) \\
&= \mathbf{E}(X^+) \sum_{k=1}^{\infty} \mathbf{P}(\tau \geq k) \\
&= \mathbf{E}(X^+)\, \mathbf{E}(\tau),
\end{aligned}$$

where at the third equality we used that $\mathbb{I}(\tau \geq k) \in \mathcal{F}_{k-1}$, thus it is independent of X_k^+.

Similarly, $\mathbf{E}(S_\tau^{(+)}) = \mathbf{E}(X^-)\, \mathbf{E}(\tau)$, and the statement follows. □

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 3.28 (Gambler's ruin)**.** Let $X, X_1, X_2, \dots$ 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

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

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

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

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 3.29. Show that $\tau < \infty$ a.s.

4 Continuous time martingales

4.1 Definitions and simple properties

This part is from Karatzas and Shreve [4].

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space and $(\mathcal{F}_t)_{t\geq 0}$ 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 4.1. Let U be uniform$(0,1)$, and $X_t \equiv 0$, $t \in [0,1]$, and $Y_t = \mathbb{I}(U = t)$. 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

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

Exercise 4.2. Show that $\mathcal{F}_\tau$ is indeed a σ-algebra.

The next result is obvious, but very useful.

Proposition 4.3. *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 4.4. *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 4.5. 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 4.6. *If $(X_t)_t$ is right continuous and adapted, then it is progressively measurable.*

Example 4.7 (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 4.8. Show that $(N_t - \lambda t)$ is martingale.

Proposition 4.9. *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 4.10 (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 4.11. Show that (W_t) and $(W_t^2 - t)$ are martingales.

4.2 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 4.12 (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 $\mathbb{Q} \cap [S, T]$ as

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

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 4.13 (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. □

Exercise 4.14. 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.

4.3 Inequalities

Theorem 4.15 (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

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

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 4.16. 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 4.17. *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

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

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

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

Applying Borel–Cantelli again, we are done. □

4.4 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 $\mathcal{F}_\infty = \sigma\left(\cup_{t\ge 0}\mathcal{F}_t\right)$, $\mathbf{E}|X_\infty| < \infty$ and for all $t \ge 0$ $\mathbf{E}[X_\infty|\mathcal{F}_t] \ge X_t$ 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 4.18 (Optional stopping)**.** *Let $(X_t, \mathcal{F}_t)_{t\ge 0}$ be a right-continuous submartingale with last element X_∞. Let $\sigma \le \tau$ be stopping times. Then*

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

Proof. Assume that τ is bounded, i.e. $\tau \le K$. 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 $\sigma_n \le \tau_n$. 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

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

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

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

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 4.19. Prove that σ_n, τ_n are indeed stopping times.

4.5 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 4.20 (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 4.21. 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

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

5 Wiener process

This part is from Karatzas and Shreve [4].

5.1 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 5.1. Show that (W2) and (W3) with $s = 0$ (i.e. $W_t \sim \mathrm{N}(0, t)$) implies (W3).

Proposition 5.2. *(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

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

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

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

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.

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

Definition 5.3. 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 $t_1, \ldots, t_k$ the random vector $(X_{t_1}, \ldots, X_{t_k})$ has multivariate normal distribution with mean $(m(t_1), \ldots, m(t_k))$ 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 5.4. *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 5.5. Prove the statement.

Exercise 5.6. 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 existence of Gaussian processes.

Theorem 5.7. *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 5.8 (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 $\delta > 0$

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

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

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

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

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

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

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

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 $\omega \in \Omega_0$

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

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

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

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 $n \ge n_0$ such that $2^{-n-1} \le t - s < 2^{-n}$, thus

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

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 5.9 (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

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

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 $\delta > 0$

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

Exercise 5.10. 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 $C_n = \mathbf{E}(Z^{2n})$, $Z \sim N(0,1)$.

Corollary 5.11. *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 5.12. 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 5.13. *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. □

Exercise 5.14 (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 5.15. 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 5.16. *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 5.17. *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 5.16 we obtain a better criteria for continuity.

Corollary 5.18. *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.

5.2 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 5.19. *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 $t \geq n$, are rational at k/n for $k = 0, 1, \dots, n^2 - 1$, and linear between. $\square$

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 measures 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 bounded 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 $\pi_{t_1,\dots,t_k} : C[0,\infty) \to \mathbb{R}^k$

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

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

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

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

The converse is not true in general.

Example 5.21. Let

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

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 5.22 (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 5.23. 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 5.24 (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 5.25. *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 5.26. *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. $\square$

5.3 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 5.27 (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 5.26 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 5.25 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

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

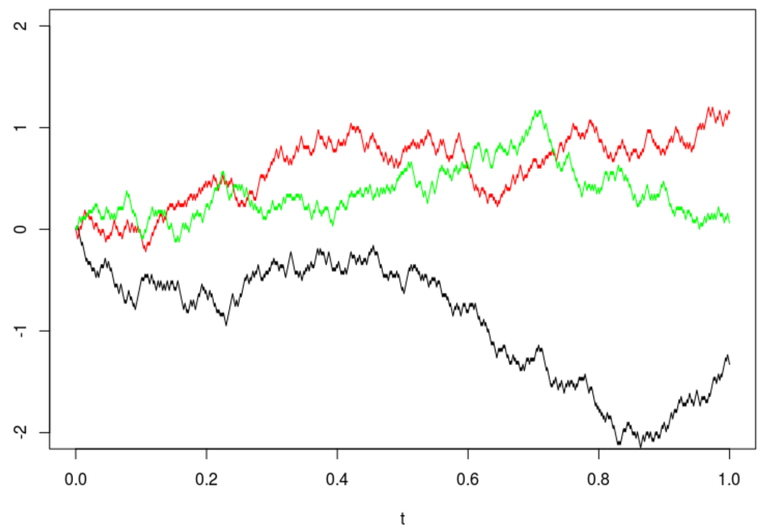

Figure 1: Simulation of 3 independent SBM

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. $\square$

In the proof above we used the following simple statements.

Exercise 5.28. 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 5.29. 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

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

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

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

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

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

5.4 Path properties

Theorem 5.30. *Almost surely the sample path of a SBM is not monotone in any interval.*

Proof. Let

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

Clearly

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

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. □

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 5.31. *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

$$\begin{aligned} &\mathbf{E} \left(\sum_{i=1}^{n} (W_{t_i} - W_{t_{i-1}})^2 - 1 \right)^2 = \\ &\sum_{i,j=1}^{n} \mathbf{E} \left(\left[(W_{t_i} - W_{t_{i-1}})^2 - (t_i - t_{i-1}) \right] \left[(W_{t_j} - W_{t_{j-1}})^2 - (t_j - t_{j-1}) \right] \right). \end{aligned} \tag{13}$$

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 (13) 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 5.32. 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 5.33. *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 5.34 (Paley, Wiener, Zygmund (1933))**.** *Almost surely the path $W(\cdot,\omega)$ is nowhere differentiable.*

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

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

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

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

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

Theorem 5.35 (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

$$\limsup_{t\downarrow 0} \frac{W_t(\omega)}{h(t)} \le 1. \tag{14}$$

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

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

where we use Lemma 5.17. The lower bound is a divergent series in n, therefore the event A_n occur infinitely often. On the other hand by (14) (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 5.36. 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.

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 $t, u > 0$

$$\mathbf{P}(X_{t+u} \in B | X_s, s \leq t) = \mathbf{P}(X_{t+u} \in B | X_t).$$

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 $u < s < t$, $B \in \mathcal{B}$. By the tower rule, the Markov property, and the properties of regular conditional distribution

$$\begin{aligned}
\mathbf{P}(X_t \in B|X_u) &= \mathbf{E}\left[\mathbf{P}(X_t \in B|X_u, X_s)|X_u\right] \\
&= \mathbf{E}\left[\mathbf{P}(X_t \in B|X_s)|X_u\right] \\
&= \mathbf{E}\left[h(X_s)|X_u\right] \\
&= \int h(y)\mathrm{d}\mathbf{P}(X_s \in \mathrm{d}y|X_u) \\
&= \int \mathbf{P}(X_t \in B|X_s = y)\mathbf{P}\left(X_s \in \mathrm{d}y|X_u\right) \\
&= \int_{\mathbb{R}} p_{t,s}(B|y)p_{s,u}(\mathrm{d}y|X_u).
\end{aligned}$$

That is

$$p_{t,u}(B|x) = \int p_{t,s}(B|y)p_{s,u}(\mathrm{d}y|x).$$

We proved the following.

Theorem 6.1 (Chapman–Kolmogorov equations)**.** *The transition probabilities of a Markov process satisfies the equations*

$$p_{t,u}(B|x) = \int p_{t,s}(B|y)p_{s,u}(\mathrm{d}y|x), \quad u < s < t, B \in \mathcal{B}. \tag{15}$$

The expression $p_{t,u}(B|x)$ is the probability that starting from x in time u we end up in B at time t. Consider any s between u and t. The distribution of X_s given $X_u = x$ is $p_{s,u}(\cdot|x)$, that is the probability being in y is $p_{s,u}(\mathrm{d}y|x)$. 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 (15) holds for fixed $u < s < t$ *almost surely* with respect to the probability $\mathbf{P}(X_u \in \cdot)$. Indeed, in the proof we calculated conditional probabilities, where each equality is only an almost sure equality. In what follows *we assume that* (15) *holds for every* x.

The Markov process (X_t) is *stationary* if the transition probabilities depend only on the time increment, i.e. $p_{t,u}(B|x) = p_{t-u}(B|x)$. Then $p_t(B|x) =$

$p_{t,0}(B|x)$, and the Chapman–Kolmogorov equations simplify to

$$p_{t+s}(B|x) = \int p_t(B|y) p_s(\mathrm{d}y|x). \tag{16}$$

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 6.2 (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 (16) 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 6.3 (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 (16) 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

$$f \mapsto Sf : Sf(x) = \lim_{t \to 0+} \frac{1}{t}\mathbf{E}_x\left[f(X_t) - f(x)\right], \tag{17}$$

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 6.4 (Poisson process)**.** Let (N_t) be a Poisson process with intensity 1, and let f be a bounded measurable function. By definition $N_t - N_0 \sim \text{Poisson}(t)$, 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 6.5 (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

$$\frac{\partial}{\partial t} \varphi_t(x) = \left(S \varphi_t \right)(x).$$

Substituting back the definition of φ, we obtain *Kolmogorov's backward equation*

$$\frac{\partial}{\partial t} p_t(B|x) = (S p_t(B|\cdot))(x). \tag{18}$$

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 $\tau \downarrow 0$

$$\int f(y) \frac{\partial}{\partial t} p_t(\mathrm{d}y|x) = \int (Sf)(y) p_t(\mathrm{d}y|x). \tag{19}$$

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 (19)

$$\int f(y) \frac{\partial}{\partial t} p_t(\mathrm{d}y|x) = \int f(y) \left(S^* p_t(\cdot|x) \right)(\mathrm{d}y),$$

from which we get *Kolmogorov's forward equation*

$$\frac{\partial}{\partial t} p_t(B|x) = \left(S^* p_t(\cdot|x) \right)(B). \tag{20}$$

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 (19)) 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 (18) and (20) 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 6.6 (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

$$\frac{\partial}{\partial t}p_t(B|x) = p_t(B|x+1) - p_t(B|x). \tag{21}$$

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 (21). 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 6.7 (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 stochastic differential equations. 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 5.8). 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

$$\begin{aligned}\mathbf{E}\left[\Delta Y_t | Y_t = y\right] &= \mu(y)h + o(h), \\ \mathbf{E}\left[(\Delta Y_t)^2 | Y_t = y\right] &= \sigma^2(y)h + o(h).\end{aligned}$$

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

$$\Delta_\varepsilon Y_t = \begin{cases} \Delta Y_t, & \text{if } |\Delta Y_t| \leq \varepsilon, \\ 0, & \text{otherwise.} \end{cases}$$

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

$$\left(S^*p_t(\cdot|x)\right)(\mathrm{d}y) = \left[-\frac{\mathrm{d}}{\mathrm{d}y}\left(\mu(y)\rho_t(y|x)\right) + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}y^2}\left(\sigma^2(y)\rho_t(y|x)\right)\right]\mathrm{d}y,$$

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 6.8 (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 $\sigma(W_s : s \geq 0)$.

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

$$\frac{\partial}{\partial t}\rho_t(y|x) = -\frac{\partial}{\partial y}\left(-\mu y\rho_t(y|x)\right) + \frac{\sigma^2}{2}\frac{\partial^2}{\partial y^2}\rho_t(y|x).$$

The solution to the Kolmogorov equations is given by the transition density

$$\rho_t(y|x) = \sqrt{\frac{\mu}{\pi\sigma^2(1-e^{-2\mu t})}}\,\exp\left\{-\frac{\mu(y-e^{-\mu t}x)^2}{\sigma^2(1-e^{-2\mu t})}\right\}.$$

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.

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] 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.