Martingale theory

Lecture notes

Instructor: Arijit Chakrabarty

Last updated: 18 May 2025

1 Radon-Nikodym theorem

Definition. *If* ν *and* μ *are measures on* $(\Omega, \mathcal{A})$ *such that* $\mu(A) = 0$ *implies that* $\nu(A) = 0$ *for all* $A \in \mathcal{A}$*, then* ν *is* absolutely continuous with respect to μ*, written as* $\nu \ll \mu$*.*

Definition. *Two measures* μ *and* ν *on* $(\Omega, \mathcal{A})$ *are singular w.r.t. each other if there exists* $S \in \mathcal{A}$ *such that*

$$\mu(S) = \nu(S^c) = 0\,.$$

Suppose that $(\Omega, \mathcal{A})$ is a measurable space.

Definition. *A function* ψ *from* $\mathcal{A}$ *to* $\mathbb{R}$ *is an "additive set function" if the following is true. Whenever* $A_1, A_2, \ldots$ *is a finite or infinite sequence of disjoint sets in* $\mathcal{A}$*, it holds that*

$$\sum_n |\psi(A_n)| < \infty$$

and

$$\psi\left(\bigcup_n A_n\right) = \sum_n \psi(A_n)\,.$$

Theorem 1.1 (Hahn decomposition)**.** *For any additive set function* ψ*, there exist disjoint sets* A^+ *and* A^- *such that* $A^+ \cup A^- = \Omega$*,* $\psi(E) \geq 0$ *for all* $E \subset A^+$ *and* $\psi(E) \leq 0$ *for all* $E \subset A^-$*.*

The proof will use the following exercise.

Exercise 1.1. *Show that if* $E_n \uparrow E$ *or* $E_n \downarrow E$*, then* $\psi(E_n) \to \psi(E)$*.*

Proof of Hahn decomposition. Let $\alpha = \sup\{\psi(A) : A \in \mathcal{A}\}$. Our first claim is to show that the supremum is achieved, that is, there exists $A^+ \subset \Omega$ with

$$\psi(A^+) = \alpha \,. \tag{1.1}$$

Let $A_n \in \mathcal{A}$ be such that

$$\lim_{n\to\infty} \psi(A_n) = \alpha \,.$$

For $n \geq 1$, let $B_n \in \sigma(A_1, \ldots, A_n)$ be such that

$$\psi(B_n) = \max\{\psi(E) : E \in \sigma(A_1, \ldots, A_n)\} \,.$$

Clearly, $\psi(B_n) \geq \psi(A_n)$.

We claim that if $E \subset B_n$ and $E \in \sigma(A_1, \ldots, A_n)$, then $\psi(E) \geq 0$. If not, then

$$\psi(B_n \setminus E) = \psi(B_n) - \psi(E) > \psi(B_n) \,,$$

contradicting the choice of B_n. Thus, $\psi(E) \geq 0$ for such an E. For positive integers $m < n$,

$$(B_m \cup \ldots \cup B_n) \setminus (B_m \cup \ldots \cup B_{n-1}) \subset B_n$$

and the LHS is an element of $\sigma(A_1, \ldots, A_n)$. Therefore, ψ evaluated on the LHS is non-negative, which shows that

$$\psi(B_m \cup \ldots \cup B_{n-1}) \leq \psi(B_m \cup \ldots \cup B_n) \,.$$

In view of the preceding exercise,

$$\psi\left(\bigcup_{n=m}^{\infty} B_n\right) = \lim_{n\to\infty} \psi(B_m \cup \ldots \cup B_n) \geq \psi(B_m) \geq \psi(A_m) \,.$$

Setting

$$A^+ := \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} B_n \,,$$

it follows that

$$\psi(A^+) = \lim_{m\to\infty} \psi\left(\bigcup_{n=m}^{\infty} B_n\right) \geq \lim_{m\to\infty} \psi(A_m) = \alpha \,,$$

proving (1.1).

The above shows that, in particular, $\alpha < \infty$. If $E \subset A^+$ and $\psi(E) < 0$, then

$$\psi(A^+ \setminus E) > \psi(A^+)$$

which is impossible. If $E \subset \Omega \setminus A^+ =: A^-$, then by similar arguments it follows that $\psi(E) \leq 0$. This completes the proof. □

Lemma 1.1. *Suppose that the finite measures μ and ν on $(\Omega, \mathcal{A})$ are not mutually singular. Then, there exists A with $\mu(A) > 0$ and $\varepsilon > 0$ such that $\varepsilon\mu(E) \leq \nu(E)$ for all $E \subset A$.*

Proof. For $n \geq 1$, let $A_n^+ \cup A_n^-$ be the Hahn decomposition of the additive set function $\nu - n^{-1}\mu$. Set

$$M = \bigcup_n A_n^+ .$$

Thus, $M^c \subset A_n^-$ for every n, and hence

$$\nu(M^c) \leq n^{-1}\mu(M^c) ,$$

showing that $\nu(M^c) = 0$. The fact that ν and μ are not mutually singular implies that $\mu(M) > 0$. Therefore, $\mu(A_n^+) > 0$ for some n. Set $A := A_n^+$ and $\varepsilon := n^{-1}$. □

Theorem 1.2. *Let μ, ν be σ-finite measures on $(\Omega, \mathcal{A})$. There exists a non-negative measurable function $f : \Omega \to \mathbb{R}$ and a measure ν_s singular w.r.t. μ such that*

$$\nu(A) = \int_A f d\mu + \nu_s(A), \text{ for all } A \in \mathcal{A} .$$

Proof. Suppose first that ν and μ are finite measures. Let $\mathcal{G}$ be the class of non-negative functions g such that

$$\int_E g d\mu \leq \nu(E) \text{ for all } E .$$

If $g, g' \in \mathcal{G}$, then $g \vee g' \in \mathcal{G}$ because

$$\begin{aligned} \int_E g \vee g' d\mu &= \int_{E \cap [g \geq g']} g d\mu + \int_{E \cap [g < g']} g' d\mu \\ &\leq \nu(E) . \end{aligned}$$

MCT shows that $\mathcal{G}$ is closed under increasing limits, and therefore under countable supremum.

Set

$$\alpha := \sup_{g \in \mathcal{G}} \int g d\mu .$$

Let $g_n \in \mathcal{G}$ be such that

$$\lim_{n \to \infty} \int g_n d\mu = \alpha .$$

Set

$$f = \sup_n g_n .$$

Thus, $f \in \mathcal{G}$ and $\int f d\mu = \alpha$.

Set

$$\bar{\nu}(E) := \nu(E) - \int_E f d\mu,\ E \in \mathcal{A}\,.$$

Clearly, $\bar{\nu}$ is a measure. All that needs to be shown is $\bar{\nu}$ is singular w.r.t. μ.

Suppose that is not the case. By the preceding lemma, it follows that there exists $\varepsilon > 0$ and A with $\mu(A) > 0$ such that $\varepsilon\mu(E) \le \bar{\nu}(E)$ for all $E \subset A$. Then, for every $E \in \mathcal{A}$,

$$\begin{aligned}
\int_E (f + \varepsilon \mathbf{1}_A) d\mu &= \int_E f d\mu + \varepsilon\mu(E \cap A) \\
&\le \int_E f d\mu + \bar{\nu}(E \cap A) \\
&\le \int_E f d\mu + \bar{\nu}(E) \\
&= \nu(E)\,.
\end{aligned}$$

Therefore, $f + \varepsilon\mathbf{1}_A \in \mathcal{G}$. Notice that

$$\int (f + \varepsilon\mathbf{1}_A) d\mu = \alpha + \varepsilon\mu(A) > \alpha\,,$$

which is a contradiction. This shows that $\bar{\nu}$ and μ are mutually singular, and thus completes the proof for finite measures.

Now suppose that μ and ν are σ-finite measures on $(\Omega, \mathcal{A})$. Then there exist disjoint sets $A_1, A_2, \ldots$ such that

$$\Omega = \bigcup_{n=1}^{\infty} A_n\,,$$

and

$$\mu(A_n) + \nu(A_n) < \infty,\ n \ge 1\,.$$

For $n \ge 1$ and $A \in \mathcal{A}$, define

$$\begin{aligned}
\mu_n(A) &:= \mu(A \cap A_n)\,, \\
\nu_n(A) &:= \nu(A \cap A_n)\,.
\end{aligned}$$

Then, μ_n and ν_n are finite measures for which the result is true. Therefore, there exists a measurable $f_n : \Omega \to [0, \infty)$ and a measure ν_s^n which is singular w.r.t. μ_n such that

$$\nu_n(A) = \int_A f_n d\mu_n + \nu_s^n(A) \text{ for all } A \in \mathcal{A}\,.$$

Therefore, for all $A \in \mathcal{A}$,

$$\begin{aligned}
\nu(A) - \sum_n \nu_s^n(A) &= \sum_n \int_A f_n d\mu_n \\
&= \sum_n \int_A f_n \mathbf{1}_{A_n} d\mu = \int_A \sum_n f_n \mathbf{1}_{A_n} d\mu\,.
\end{aligned}$$

Define

$$f := \sum_n f_n \mathbf{1}_{A_n} ,$$

and the measure

$$\nu_s(A) := \sum_n \nu_s^n(A),\ A \in \mathcal{A}\,.$$

The proof follows if it can be shown that ν_s is singular w.r.t. μ. To that end, for $n \geq 1$, there exists $S_n \subset A_n$ such that

$$\nu_s^n(S_n) = 0 = \mu_n(S_n^c)\,.$$

Define

$$S := \bigcup_n S_n ,$$

and note that

$$\begin{aligned}
\nu_s(S) &= \sum_m \sum_n \nu_s^m(S_n) \\
&= \sum_n \nu_s^n(S_n) + \sum_{m \neq n} \nu_s^m(S_n) \\
&= \sum_{m \neq n} \nu_s^m(S_n) \\
&\leq \sum_{m \neq n} \nu_m(A_n) \\
&= 0\,.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mu(S^c) &= \sum_n \mu_n(S^c) \\
&\leq \sum_n \mu_n(S_n^c) \\
&= 0\,.
\end{aligned}$$

This establishes the mutual singularity of ν_s and μ, thereby completing the proof. □

Theorem 1.3 (Radon-Nikodym theorem)**.** *If μ and ν are σ-finite measures on $(\Omega, \mathcal{A})$ such that $\nu \ll \mu$, then there exists a non-negative measurable function f on Ω such that*

$$\nu(A) = \int_A f d\mu$$

for all $A \in \mathcal{A}$. If f and g both satisfy the above, then $\mu[f \neq g] = 0$.

Exercise 1.2. *If f, g are non-negative measurable functions on a σ-finite measure space $(\Omega, \mathcal{A}, \mu)$ such that*

$$\int_A f d\mu = \int_A g d\mu, \text{ for all } A \in \mathcal{A},$$

then show that $f = g$ almost everywhere.

Proof of the Radon-Nikodym theorem. By Theorem 1.2, there exists a non-negative measurable function f on Ω and a measure ν_s singular with respect to μ such that

$$\nu(A) = \int_A f d\mu + \nu_s(A),\ A \in \mathcal{S}.$$

The mutual singularity implies the existence of S such that

$$\nu_s(S) = 0 = \mu(S^c).$$

Since $\nu \ll \mu$, it follows that

$$0 = \nu(S^c) \geq \nu_s(S^c).$$

Thus, ν_s is the null measure which completes the proof. □

Corollary 1.1. *For σ-finite measures μ, ν on $(\Omega, \mathcal{A})$, there exists a measurable $f : \Omega \to [0, \infty)$ such that*

$$\nu(A) = \int_A f d\mu,\ A \in \mathcal{S}, \tag{1.2}$$

if and only if $\nu \ll \mu$.

Definition. *The function f is the "Radon-Nikodym derivative" of ν w.r.t. μ. We write*

$$f = \frac{d\nu}{d\mu},$$

$$d\nu = f\, d\mu,$$

or

$$\nu(dx) = f(x)\mu(dx),\ x \in \Omega.$$

Theorem 1.4 (Lebesgue decomposition)**.** *Let μ and ν be σ-finite measures on $(\Omega, \mathcal{A})$. Then there is a unique decomposition*

$$\nu = \nu_{ac} + \nu_s,$$

such that ν_{ac} and ν_s are respectively absolutely continuous and singular w.r.t. μ.

Proof. By Theorem (1.2), there exists a function f such that

$$\nu(A) = \int_A f d\mu + \nu_s(A), \text{ for all } A \in \mathcal{A}.$$

Defining

$$\nu_{ac}(A) := \int_A f d\mu, \; A \in \mathcal{A},$$

the desired equality holds. To show uniqueness, suppose that

$$\nu = \nu'_{ac} + \nu'_s,$$

is a different decomposition. By the Radon-Nikodym theorem, there exists a non-negative measurable g such that

$$\nu'_{ac}(A) = \int_A g d\mu.$$

The above equations put together imply that

$$\int_A f d\mu + \nu_s(A) = \int_A g d\mu + \nu'_s(A), \text{ for all } A \in \mathcal{A}. \tag{1.3}$$

Furthermore, there exist $S, S' \in \mathcal{A}$ such that

$$\nu_s(S) = \nu'_s(S') = 0 = \mu\left((S \cap S')^c\right). \tag{1.4}$$

Since μ is σ-finite, it follows that there exist $A_1, A_2, \ldots \in \mathcal{A}$ such that $\nu(A_n) < \infty$ for all n, and

$$\Omega = \bigcup_n A_n.$$

Fix $n \geq 1$ and note that $(f-g)\mathbf{1}_{A_n}$ is integrable w.r.t. μ, and hence for a fixed $A \in \mathcal{A}$,

$$\begin{aligned}
& \left|\nu_{ac}(A_n \cap A) - \nu'_{ac}(A_n \cap A)\right| \\
&= \left|\int_{A_n \cap A} (f-g) d\mu\right| \\
\text{(by (1.4))} \quad &= \left|\int_{A_n \cap A \cap S \cap S'} (f-g) d\mu\right| \\
\text{(by (1.3))} \quad &= \left|\nu'_s(A_n \cap A \cap S \cap S') - \nu_s(A_n \cap A \cap S \cap S')\right| \\
&\leq \nu_s(S) + \nu'_s(S') \\
\text{(by (1.4))} \quad &= 0.
\end{aligned}$$

Since the above holds for all n, it follows that

$$\nu_{ac}(A) = \nu'_{ac}(A) \text{ for all } A \in \mathcal{A},$$

and hence the uniqueness. This completes the proof. □

Theorem 1.5. *Suppose that μ is a σ-finite measure on $(\mathbb{R}, \mathcal{B})$. Then there exist measures μ_d, μ_{sc} and μ_{ac} such that*

$$\mu = \mu_d + \mu_{sc} + \mu_{ac} \,, \tag{1.5}$$

where μ_d is concentrated on a countable set, μ_{sc} is singular with respect to the Lebesgue measure and gives zero mass to singletons, μ_{ac} is absolutely continuous with respect to the Lebesgue measure. Furthermore, the decomposition (1.5) *is unique.*

Proof. Since μ is σ-finite, the set $D := \{x \in \mathbb{R} : \mu(\{x\}) > 0\}$ is countable. Define

$$\mu_d(A) := \mu(A \cap D),\ A \in \mathcal{A} \,,$$

and let

$$\tilde{\mu}(A) := \mu(A) - \mu_d(A) \,.$$

The proof follows by applying Lebesgue decomposition to $\tilde{\mu}$. □

Recall the question when does a random variable have a density?

Theorem 1.6. *Suppose that F is the c.d.f. of X. Then, X has a density if and only if the following holds. Given $\varepsilon > 0$, there exists $\delta > 0$ such that whenever, for some $n \geq 1$, $a_1 < b_1 < \ldots < a_n < b_n \in \mathbb{R}$ are such that*

$$\sum_{i=1}^{n} (b_i - a_i) \leq \delta \,,$$

it holds that

$$\sum_{i=1}^{n} [F(b_i) - F(a_i)] \leq \varepsilon \,.$$

Lemma 1.2. *A Borel subset A of $\mathbb{R}$ has Lebesgue measure zero if and only if the following holds. Given $\varepsilon > 0$, there exist $a_1, b_1, a_2, b_2, \ldots \in \mathbb{R}$ such that $a_n < b_n$ for all n,*

$$(a_m, b_m) \cap (a_n, b_n) = \emptyset \text{ if } m \neq n \,,$$

$$A \subset \bigcup_{n=1}^{\infty} (a_n, b_n) \,,$$

and

$$\sum_{n=1}^{\infty} (b_n - a_n) \leq \varepsilon \,.$$

Proof. The "if" part is trivial. The "only if" part follows by recalling that for any Borel set $A \subset [a, b]$,

$$\lambda(A) = \inf\{\lambda(U \cap [a, b]) : U \text{ open},\ U \supset A\} \,,$$

where λ denotes the Lebesgue measure. □

Proof of Theorem 1.6. We start with proving the "if" part. Let the probability space underlying X be $(\Omega, \mathcal{A}, P)$. Let λ denote the Lebesgue measure on $\mathbb{R}$, and define

$$\mu := P \circ X^{-1} .$$

In order to show the desired claim using the Radon-Nikodym theorem, all we need to check is that

$$\mu \ll \lambda . \tag{1.6}$$

To that end, fix a Borel set A with $\lambda(A) = 0$. We need to show that $\mu(A) = 0$. Fix an arbitrary $\varepsilon > 0$. By the assumed hypothesis, there exists $\delta > 0$ such that $a_1 < b_1 < \ldots < a_n < b_n$ and

$$\sum_{i=1}^{n} (b_i - a_i) \leq \delta ,$$

implies that

$$\sum_{i=1}^{n} [F(b_i) - F(a_i)] \leq \varepsilon .$$

Lemma 1.2 implies the existence of $a_1, b_1, \ldots$ such that $a_n < b_n$,

$$(a_m, b_m) \cap (a_n, b_n) = \emptyset \text{ if } m \neq n ,$$

$$A \subset \bigcup_{n=1}^{\infty} (a_n, b_n) ,$$

and

$$\sum_{n=1}^{\infty} (b_n - a_n) \leq \delta .$$

For a fixed n,

$$\mu\left(\bigcup_{k=1}^{n} (a_k, b_k)\right) \leq \sum_{k=1}^{n} \mu(a_k, b_k] = \sum_{k=1}^{n} (F(b_k) - F(a_k)) \leq \varepsilon ,$$

the rightmost inequality following from the choice of δ. Letting $n \to \infty$ and using the continuity of measure from below, it follows that

$$\varepsilon \geq \mu\left(\bigcup_{k=1}^{\infty} (a_k, b_k)\right) \geq \mu(A) .$$

Since ε is arbitrary, (1.6) follows, thereby establishing the "if" part.

For the "only if" part, assume that X has a density f. Fix $\varepsilon > 0$. Since

$$1 = \int_{-\infty}^{\infty} f(x)dx = \sup_{s \text{ simple}, 0 \leq s \leq f} \int_{-\infty}^{\infty} s(x)dx ,$$

there exists a simple function $0 \le s \le f$ such that

$$\int_{-\infty}^{\infty} s(x)dx \ge 1 - \varepsilon/2 .$$

Define

$$M = 1 + \max_{x \in \mathbb{R}} s(x) ,$$

and

$$\delta := \varepsilon/2M .$$

Now suppose that $a_1 < b_1 < \ldots < a_n < b_n$ are such that

$$\sum_{i=1}^{n} (b_i - a_i) \le \delta .$$

Let

$$A := \bigcup_{i=1}^{n} (a_i, b_i) ,$$

and observe that

$$\begin{aligned}
\sum_{i=1}^{n} (F(b_i) - F(a_i)) &= \mu(A) \\
&= \int_A f(x)dx \\
&= \int_A [f(x) - s(x)]\, dx + \int_A s(x)dx \\
&\le \int_{-\infty}^{\infty} [f(x) - s(x)]\, dx + M\lambda(A) \\
&\le \frac{\varepsilon}{2} + M\delta \\
&\le \varepsilon .
\end{aligned}$$

Thus, the "only if" part follows. □

2 Conditional expectation

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Exercise 2.1. *Let X be a random variable defined on Ω. Set*

$$\mathcal{F} := \sigma(X) .$$

Show that a random variable Y defined on Ω is measurable w.r.t. $\mathcal{F}$ if and only if there exists a measurable $f : \mathbb{R} \to \mathbb{R}$ such that

$$Y = f(X) .$$

Definition. *Let X be an integrable random variable on Ω, and $\mathcal{F}$ be any sub σ-field of $\mathcal{A}$. An integrable random variable Y, which is measurable w.r.t. $\mathcal{F}$, is called the* conditional expectation of X given $\mathcal{F}$ *if*

$$\int_A X\,dP = \int_A Y\,dP \text{ for all } A \in \mathcal{F}\,.$$

It is written as

$$\mathrm{E}\,(X|\mathcal{F}) = Y\,.$$

Observation:

$$\mathrm{E}(Y) = \mathrm{E}(X)\,.$$

Immediate questions:

- Why does such an Y exist?
- Even if it exists, is it unique?
- Why should such an Y be called a conditional expectation?

From now on, X will denote a $\mathcal{A}$-measurable and integrable random variable, and $\mathcal{F}$ is a sub σ-field of $\mathcal{A}$. The following result guarantees the existence and uniqueness of conditional expectation.

Theorem 2.1. *An $\mathcal{F}$-measurable and integrable random variable Y satisfying*

$$\int_A X\,dP = \int_A Y\,dP \text{ for all } A \in \mathcal{F}\,,$$

always exists. Furthermore, if Y_1 and Y_2 are two choices satisfying the above, then

$$Y_1 = Y_2 \text{ a.s.}\,.$$

Proof. Write

$$\begin{aligned} X^+ &:= \max\{X, 0\}\,, \\ X^- &:= \max\{-X, 0\}\,. \end{aligned}$$

Then,

$$X = X^+ - X^-\,,$$

with $X^+, X^- \geq 0$. Clearly,

$$\int X^+ dP \leq \int |X| dP < \infty\,.$$

Define a measure ν^+ on $(\Omega, \mathcal{F})$ by

$$\nu^+(A) = \int_A X^+ dP,\ A \in \mathcal{F}\,.$$

Clearly, $\nu^+ \ll P$ as finite measures on $(\Omega, \mathcal{F})$. By the R.N.T., there exists $Y^+ : \Omega \to [0, \infty)$ which is $\mathcal{F}$-measurable such that

$$\nu^+(A) = \int_A Y^+ dP,\ A \in \mathcal{F}\,.$$

Once can define Y^- analogously, and set

$$Y := Y^+ - Y^-\,,$$

to achieve the desired equality. The second claim is an easy exercise. □

Theorem 2.2. *Suppose that X, Y, X_n are integrable random variables on $(\Omega, \mathcal{A}, P)$, and $\mathcal{F} \subset \mathcal{A}$ is a σ-field.*

1. *If $X = a$ a.s., then $\mathrm{E}(X|\mathcal{F}) = a$ a.s. .*
2. *For constants a and b,*
$$\mathrm{E}\left(aX + bY|\mathcal{F}\right) = a\mathrm{E}\left(X|\mathcal{F}\right) + b\mathrm{E}\left(Y|\mathcal{F}\right) \ \text{a.s.}\,.$$
3. *If $X \le Y$ a.s., then $\mathrm{E}\left(X|\mathcal{F}\right) \le \mathrm{E}\left(Y|\mathcal{F}\right)$ a.s. .*
4. *Almost surely, $|\mathrm{E}(X|\mathcal{F})| \le \mathrm{E}(|X|\,|\mathcal{F})$.*
5. *If $\lim_n X_n = X$ a.s., $|X_n| \le Y$ and Y is integrable, then*
$$\lim_n \mathrm{E}(X_n|\mathcal{F}) = \mathrm{E}(X|\mathcal{F}) \ \text{a.s.}\,.$$

Proof.

1. Easy exercise.
2. Easy exercise.
3. Follows from 2. by considering $\mathrm{E}(Y - X|\mathcal{F})$.
4. Using 3., it follows that almost surely
$$\mathrm{E}(|X|\,|\mathcal{F}) \ge \mathrm{E}(X|\mathcal{F})\,,$$
and
$$\mathrm{E}(|X|\,|\mathcal{F}) \ge \mathrm{E}(-X|\mathcal{F}) = -\mathrm{E}(X|\mathcal{F})\,,$$
the second equality following from 2. This completes the proof of 4.
5. Define
$$Z_n := \sup_{k \ge n} |X_k - X|,\ n \ge 1\,.$$
Clearly
$$Z_n \downarrow 0 \text{ a.s.}\,,$$

and 1.-3. imply that

$$|\mathrm{E}(X_n|\mathcal{F}) - \mathrm{E}(X|\mathcal{F})| \leq \mathrm{E}(Z_n|\mathcal{F}) \text{ a.s.}, \, n \geq 1\,.$$

Therefore, it suffices to show that

$$\mathrm{E}(Z_n|\mathcal{F}) \to 0 \text{ a.s.}\,.$$

By 3., $\mathrm{E}(Z_n|\mathcal{F})$ is non-increasing, and hence has a limit, say Z. Furthermore,

$$\begin{aligned}\int Z dP &\leq \int \mathrm{E}(Z_n|\mathcal{F}) dP \\ &= \int Z_n dP\,.\end{aligned}$$

Notice that $Z_n \leq 2Y$. DCT implies that

$$\lim_{n\to\infty} \int Z_n dP = 0\,.$$

This completes the proof of 5.

□

Theorem 2.3. *Suppose that X is $\mathcal{F}$-measurable, and Y, XY are integrable. Then,*

$$\mathrm{E}\left(XY|\mathcal{F}\right) = X\mathrm{E}(Y|\mathcal{F}) \;\; a.s.\,.$$

Proof. The equality is trivial if X is a simple function. When X is just an integrable function which is $\mathcal{F}$-measurable, one can get $\mathcal{F}$-measurable simple functions s_n such that $|s_n| \leq |X|$ and $s_n \to X$. The proof follows from the observation that $|s_n Y| \leq |XY|$, XY is integrable, and claim 5. of Theorem 2.2. □

The next result justifies, to an extent, the definition of conditional expectation.

Theorem 2.4. *For a square integrable random variable X,*

$$\mathrm{E}\left[\left(X - \mathrm{E}\left(X|\mathcal{F}\right)\right)^2\right] = \min_{\mathcal{F}\text{-measurable } Y} \mathrm{E}\left[(X-Y)^2\right]\,.$$

Proof. Denote

$$Z := \mathrm{E}\left(X|\mathcal{F}\right)\,.$$

Our first task is to show that

$$\mathrm{E}(Z^2) < \infty\,.$$

To that end, note that

$$\begin{aligned} 0 &\le (X-Z)^2 \\ &= X^2 - Z^2 - 2Z(X-Z)\,. \end{aligned}$$

Therefore,

$$X^2 \ge X^2\mathbf{1}(|Z| \le n) \ge Z^2\mathbf{1}(|Z| \le n) + 2Z\mathbf{1}(|Z| \le n)(X-Z)\,,$$

for all $n \ge 1$. Therefore,

$$\begin{aligned} \mathrm{E}(X^2|\mathcal{F}) &\ge \mathrm{E}\left(Z^2\mathbf{1}(|Z| \le n) + 2Z\mathbf{1}(|Z| \le n)(X-Z)\big|\,\mathcal{F}\right) \\ \text{(by Theorem 2.3)} &= Z^2\mathbf{1}(|Z| \le n) + 2Z\mathbf{1}(|Z| \le n)\mathrm{E}\left((X-Z)|\mathcal{F}\right) \\ &= Z^2\mathbf{1}(|Z| \le n)\,. \end{aligned}$$

Taking expectation on both sides, it follows that

$$\mathrm{E}\left[Z^2\mathbf{1}(|Z| \le n)\right] \le \mathrm{E}\left[\mathrm{E}(X^2|\mathcal{F})\right] = \mathrm{E}(X^2)\,.$$

Letting $n \to \infty$, it follows that

$$\mathrm{E}(Z^2) \le \mathrm{E}(X^2) < \infty\,.$$

Note that for any random variable Y with $\mathrm{E}(Y^2) = \infty$,

$$\mathrm{E}(X-Y)^2 = \infty > \mathrm{E}(X-Z)^2\,.$$

Therefore, it suffices to show that

$$\mathrm{E}(X-Z)^2 = \min_{\mathcal{F}\text{-measurable } Y : \mathrm{E}(Y^2) < \infty} \mathrm{E}\left[(X-Y)^2\right]\,.$$

To that end, fix a $\mathcal{F}$-measurable Y with $\mathrm{E}(Y^2) < \infty$, and note that $(X-Z)(Z-Y)$ is integrable by the Cauchy-Schwartz inequality. Thus,

$$\begin{aligned} \mathrm{E}(X-Y)^2 &= \mathrm{E}(X-Z)^2 + \mathrm{E}(Z-Y)^2 + 2\mathrm{E}\left[(X-Z)(Z-Y)\right] \\ &\ge \mathrm{E}(X-Z)^2 + 2\mathrm{E}\left[(X-Z)(Z-Y)\right] \\ &= \mathrm{E}(X-Z)^2 + 2\mathrm{E}\mathrm{E}\left[(X-Z)(Z-Y)|\mathcal{F}\right] \\ &= \mathrm{E}(X-Z)^2 + 2\mathrm{E}\left[(Z-Y)\mathrm{E}\left[X-Z|\mathcal{F}\right]\right] \\ &= \mathrm{E}(X-Z)^2\,. \end{aligned}$$

This completes the proof. □

Theorem 2.5. *If X is integrable and the σ-fields $\mathcal{F}_1, \mathcal{F}_2$ satisfy*

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{A}\,,$$

then

$$\mathrm{E}\left(X|\mathcal{F}_1\right) = \mathrm{E}\left(\mathrm{E}\left(X|\mathcal{F}_2\right)|\mathcal{F}_1\right) \;\; a.s.\,.$$

Proof. Denoting

$$Y := \mathrm{E}\left(X|\mathcal{F}_2\right) ,$$

and

$$Z := \mathrm{E}\left(X|\mathcal{F}_1\right) ,$$

all that needs to be checked is

$$\int_A Z dP = \int_A Y dP,\ A \in \mathcal{F}_1 . \tag{2.1}$$

To that end, fix $A \in \mathcal{F}_1$ and note that

$$\begin{aligned}\int_A Z dP &= \int_A X dP \\ &= \int_A Y dP ,\end{aligned}$$

the equality in the second line using the fact that $A \in \mathcal{F}_2$. This proves (2.1) and thereby completes the proof. ☐

Theorem 2.6 (Jensen)**.** *Suppose that ϕ is a convex function, and $X, \phi(X)$ are integrable random variables. For a sub σ-field $\mathcal{F}$,*

$$\phi\left(\mathrm{E}(X|\mathcal{F})\right) \le \mathrm{E}\left(\phi(X)|\mathcal{F}\right) \ \ a.s. .$$

Lemma 2.1. *If $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function, the right derivative $\phi'_+(x_0)$ exists at all $x_0 \in \mathbb{R}$, and*

$$\phi(x_0) + (x - x_0)\phi'_+(x_0) \le \phi(x) \text{ for all } x, x_0 \in \mathbb{R} .$$

Proof. Fix $x_0 \in \mathbb{R}$ and notice that if $x_0 < y < z$, then

$$\begin{aligned}\phi(y) &= \phi\left(\frac{z-y}{z-x_0}x_0 + \frac{y-x_0}{z-x_0}z\right) \\ &\le \frac{z-y}{z-x_0}\phi(x_0) + \frac{y-x_0}{z-x_0}\phi(z) \\ &= \left[1 - \frac{y-x_0}{z-x_0}\right]\phi(x_0) + \frac{y-x_0}{z-x_0}\phi(z) \\ &= \phi(x_0) + (y-x_0)\frac{\phi(z)-\phi(x_0)}{z-x_0} .\end{aligned} \tag{2.2}$$

Therefore,

$$\frac{\phi(y)-\phi(x_0)}{y-x_0} \le \frac{\phi(z)-\phi(x_0)}{z-x_0} ,$$

that is,

$$x \mapsto \frac{\phi(x)-\phi(x_0)}{x-x_0} , x > x_0 , \tag{2.3}$$

is a non-decreasing function.

A restatement of (2.2) is that

$$\frac{z-y}{z-x_0}\left[\phi(y)-\phi(x_0)\right] \le \frac{y-x_0}{z-x_0}\left[\phi(z)-\phi(y)\right] ,$$

which is equivalent to

$$\frac{\phi(y)-\phi(x_0)}{y-x_0} \le \frac{\phi(z)-\phi(y)}{z-y} .$$

For any $v < x_0 < w$, by replacing x_0, y, z by v, x_0, w, respectively, it follows that

$$\frac{\phi(v)-\phi(x_0)}{v-x_0} \le \frac{\phi(w)-\phi(x_0)}{w-x_0} ,$$

that is, the function in (2.3) is bounded below.

Thus, $\phi'_+(x_0)$ exists, and in fact, for all $v < x_0$,

$$\frac{\phi(v)-\phi(x_0)}{v-x_0} \le \phi'_+(x_0) = \inf_{x>x_0} \frac{\phi(x)-\phi(x_0)}{x-x_0} .$$

Therefore, for all $x \ne x_0$,

$$\phi(x)-\phi(x_0) \ge (x-x_0)\phi'_+(x_0) ,$$

which can be checked separately for the cases $x > x_0$ and $x < x_0$. This completes the proof. This completes the proof. □

Proof of Jensen's inequality. Denote $Y = \mathrm{E}(X|\mathcal{F})$. We first prove the claim when Y is a bounded random variable, that is, $a \le Y \le b$ for some finite a, b. The preceding lemma implies that

$$\phi(Y) + (X-Y)\phi'_+(Y) \le \phi(X) .$$

Continuity of ϕ implies that $\phi(Y)$ takes values in the compact set $\phi([a,b])$, and convexity implies that $\phi'_+(a) \le \phi'_+(Y) \le \phi'_+(b)$. Thus, the left hand is integrable. Hence,

$$\begin{aligned} \mathrm{E}\left[\phi(X)|\mathcal{F}\right] &\ge \mathrm{E}\left[\phi(Y) + (X-Y)\phi'_+(Y)|\mathcal{F}\right] \\ &= \phi(Y) + \phi'_+(Y)\mathrm{E}(X-Y|\mathcal{F}) \\ &= \phi(Y) , \end{aligned}$$

which is the desired inequality.

To prove the result in general, note that for $n \ge 1$,

$$\phi(Y)\mathbf{1}(|Y| \le n) + (X-Y)\phi'_+(Y)\mathbf{1}(|Y| \le n) \le \phi(X)\mathbf{1}(|Y| \le n) .$$

Taking conditional expectation of both sides w.r.t. $\mathcal{F}$ and going through the same steps, it follows that

$$\phi(Y)\mathbf{1}(|Y| \le n) \le \mathbf{1}(|Y| \le n)\mathrm{E}\left(\phi(X)|\mathcal{F}\right) .$$

Letting $n \to \infty$, the proof follows. □

Example 2.1. *Let X be an integrable random variable on $(\Omega, \mathcal{A}, P)$. Suppose that $A_1, A_2, \ldots \in \mathcal{A}$ are* ***disjoint*** *such that*

$$\Omega = \bigcup_n A_n \,.$$

Define

$$\mathcal{F} := \sigma\left(A_1, A_2, \ldots\right) \,.$$

Let

$$Y := \sum_{n:P(A_n)>0} \mathbf{1}_{A_n} \frac{1}{P(A_n)} \int_{A_n} X \, dP \,.$$

Our claim is that

$$\mathrm{E}(X|\mathcal{F}) = Y \,.$$

Y is clearly $\mathcal{F}$ measurable, and

$$\begin{aligned} \int |Y| dP &= \sum_{n:P(A_n)>0} \left| \int_{A_n} X dP \right| \\ &\le \sum_{n:P(A_n)>0} \int_{A_n} |X| dP \\ &= \int |X| dP < \infty \,. \end{aligned}$$

To see that, fix $B \in \mathcal{F}$. Then, there exists $I \subset \mathbb{N}$ such that

$$B = \bigcup_{i \in I} A_i \,,$$

and hence,

$$\begin{aligned} \int_B Y dP &= \sum_{i \in I:P(A_i)>0} \int_{A_i} Y dP \\ &= \sum_{i \in I:P(A_i)>0} \int_{A_i} X dP \\ &= \int_B X dP \,. \end{aligned}$$

Definition. *For any random variable Y and integrable X, $\mathrm{E}(X|\sigma(Y))$ is also denoted by $\mathrm{E}(X|Y)$.*

Example 2.2. *Let X be an integrable random variable, and Y a discrete random variable taking values $y_1, y_2, \ldots$. Then,*

$$\mathrm{E}(X|Y) = \sum_i \mathbf{1}(Y = y_i) \frac{1}{P(Y = y_i)} \int_{[Y=y_i]} X dP \,.$$

Example 2.3. *Let (X, Y) have a joint density $f(\cdot, \cdot)$ with respect to the Lebesgue measure on $\mathbb{R}^2$. Assume X to be integrable. Let*

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx, \; y \in \mathbb{R}\,.$$

Define

$$g(y) := \begin{cases} \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} x f(x, y)dx\,, & \textit{if } \int_{-\infty}^{\infty} |x| f(x, y)dx < \infty \textit{ and } f_Y(y) > 0\,, \\ 0\,, & \textit{otherwise}\,. \end{cases}$$

Our claim is that

$$\mathrm{E}(X|Y) = g(Y)\,.$$

First note that for a fixed n,

$$\begin{aligned} \infty > \mathrm{E}(|X|) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x| f(x, y)\,dx\,dy \\ &\geq \int_{\{y: f_Y(y) \geq 1/n\}} \left\{ \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} |x| f(x, y)\,dx \right\} f_Y(y)\,dy \\ &\geq \frac{1}{n} \int_{\{y: f_Y(y) \geq 1/n\}} \left\{ \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} |x| f(x, y)\,dx \right\} dy\,. \end{aligned}$$

Therefore,

$$\lambda\left(\left\{ y \in \mathbb{R} : f_Y(y) \geq \frac{1}{n} \textit{ and } \int_{-\infty}^{\infty} |x| f(x, y)dx = \infty \right\}\right) = 0\,.$$

Letting $n \to \infty$ implies that

$$\lambda\left(\left\{ y \in \mathbb{R} : f_Y(y) > 0 \textit{ and } \int_{-\infty}^{\infty} |x| f(x, y)dx = \infty \right\}\right) = 0\,.$$

Therefore,

$$\begin{aligned} \mathrm{E}(|g(Y)|) &= \int_{-\infty}^{\infty} |g(y)| f_Y(y)\,dy \\ &= \int_{\{y: f_Y(y) > 0\}} |g(y)| f_Y(y)\,dy \\ &= \int_{\{y: f_Y(y) > 0\}} \left| \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} x f(x, y)dx \right| f_Y(y)\,dy \\ &\leq \int_{\{y: f_Y(y) > 0\}} \int_{-\infty}^{\infty} |x| f(x, y)\,dx\,dy \\ &= \mathrm{E}(|X|) < \infty\,. \end{aligned}$$

Fix $A \in \sigma(Y)$. Then, there exists $B \in \mathcal{B}$ such that $A = Y^{-1}B$. Furthermore,

$$\begin{aligned}
\mathrm{E}\left(g(Y)\mathbf{1}_A\right) &= \int_{B\cap\{y:f_Y(y)>0\}} g(y) f_Y(y)\,dy \\
&= \int_{B\cap\{y:f_Y(y)>0\}} \int_{-\infty}^{\infty} x f(x,y)\,dx\,dy \\
&= \mathrm{E}\left(X\mathbf{1}(Y\in B)\right) \\
&= \mathrm{E}(X\mathbf{1}_A)\,.
\end{aligned}$$

This proves the claim that $g(Y)$ is the conditional expectation of X given Y.

Example 2.4. *Let X be integrable, and $\mathcal{F}$ a sub σ-field independent of $\mathcal{F}$. Then,*

$$\mathrm{E}(X|\mathcal{F}) = \mathrm{E}(X)\,.$$

To see this, fix any $A \in \mathcal{F}$. Then X and $\mathbf{1}_A$ will be independent, implying the claim.

Definition. *Let $A \in \mathcal{A}$, and $\mathcal{F}$ be a sub σ-field of $\mathcal{A}$. The conditional probability of A given $\mathcal{F}$ is same as $\mathrm{E}(\mathbf{1}_A|\mathcal{F})$.*

Exercise 2.2. *Let $(N_t : t \ge 0)$ be a Poisson process with intensity λ. For $0 \le t \le u$, calculate*

$$\mathrm{E}\left(N_u|\sigma(N_s : 0 \le s \le t)\right)\,.$$

3 Uniform integrability

Definition. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. A family $\{f_\alpha : \alpha \in I\}$ of measurable functions from Ω to $\mathbb{R}$ is **uniformly integrable** if*

$$\lim_{T\to\infty} \sup_{\alpha\in I} \int_{[|f_\alpha|>T]} |f_\alpha|\,d\mu = 0\,.$$

Theorem 3.1. *A family $\{f_\alpha : \alpha \in I\}$ of measurable functions is uniformly integrable if and only if the following hold:*

$$\sup_{\alpha\in I} \int |f_\alpha|\,d\mu < \infty\,,$$

and given $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\sup_{\alpha\in I} \int_A |f_\alpha|\,d\mu \le \varepsilon \text{ whenever } \mu(A) \le \delta\,.$$

Proof. Exercise. □

Theorem 3.2. *1. If $(f_\alpha : \alpha \in I)$ and $(g_{\alpha'} : \alpha' \in I')$ are uniformly integrable, then so is $(f_\alpha + g_{\alpha'} : \alpha \in I, \alpha' \in I')$.*

2. *If $(f_\alpha : \alpha \in I)$ is uniformly integrable, and g_α is a measurable function with*
$$|g_\alpha| \le |f_\alpha|\,, \alpha \in I\,,$$
then $(g_\alpha : \alpha \in I)$ is uniformly integrable.

Proof. Follows from Theorem 3.1. □

Since the theory of uniform integrability is useful and interesting only when the measure is finite, we shall henceforth work on a probability space $(\Omega, \mathcal{A}, P)$.

Exercise 3.1. *Show that $\{X_\alpha : \alpha \in I\}$ is uniformly integrable if for some $p > 1$,*
$$\sup_{\alpha \in I} \mathrm{E}\left(|X_\alpha|^p\right) < \infty\,.$$

Exercise 3.2. *Show that a family of finitely many integrable random variables is uniformly integrable.*

Theorem 3.3. *Let X_n, X be integrable random variables. Then, as $n \to \infty$,*
$$X_n \to X \text{ in } L^1\,,$$
if and only if
$$X_n \xrightarrow{P} X\,,$$
as $n \to \infty$, and $(X_n : n \ge 1)$ is uniformly integrable.

Proof. We first prove the "if" part, that is, we assume that $X_n \xrightarrow{P} X$ and $(X_n : n \ge 1)$ is U.I. There is a subsequence (X_{n_k}) of (X_n) such that
$$X_{n_k} \to X \text{ a.s.}$$

Thus,
$$\begin{aligned}
\mathrm{E}(|X|) &= \mathrm{E}\left(\liminf_{k\to\infty} |X_{n_k}|\right) \\
\text{(Fatou)} &\le \liminf_{k\to\infty} \mathrm{E}\left(|X_{n_k}|\right) \\
&\le \sup_{n\ge1} \mathrm{E}|X_n| \\
&< \infty\,.
\end{aligned}$$

Define
$$Y_n = |X_n - X|\,.$$

The assumptions, in view of the previous exercise, mean that $Y_n \xrightarrow{P} 0$ and $(Y_n : n \ge 1)$ is U.I. What needs to be shown is
$$\lim_{n\to\infty} \mathrm{E}(Y_n) = 0\,.$$

Fix $\varepsilon > 0$. Use (Y_n) is U.I. to get T s.t.

$$\sup_n \mathrm{E}\left(Y_n \mathbf{1}(Y_n \geq T)\right) \leq \varepsilon\,.$$

Write

$$\begin{aligned}\limsup_{n\to\infty} \mathrm{E}(Y_n) &\leq \limsup_{n\to\infty} \mathrm{E}(Y_n \mathbf{1}(Y_n < T)) + \limsup_{n\to\infty} \mathrm{E}\left(Y_n \mathbf{1}(Y_n \geq T)\right) \\ &\leq \varepsilon + \limsup_{n\to\infty} \mathrm{E}(Y_n \mathbf{1}(Y_n < T)) \\ &= \varepsilon\,,\end{aligned}$$

by DCT because $Y_n \mathbf{1}(Y_n < T) \xrightarrow{P} 0$. Since ε is arbitrary, the "if" part follows.

For the "only if" part, assume that $X_n \to X$ in L^1. Then, $X_n \xrightarrow{P} X$, trivially. Letting $Y_n = |X_n - X|$, it suffices to show that (Y_n) is U.I. Fix $\varepsilon > 0$. The hypothesis implies that for some N,

$$\sup_{n\geq N} \mathrm{E}(Y_n) \leq \varepsilon\,.$$

For $i = 1, \ldots, N-1$, there exists T_i with

$$\mathrm{E}(Y_i \mathbf{1}(Y_i > T_i)) \leq \varepsilon\,.$$

Letting $T = T_1 \vee \ldots \vee T_{N-1}$, it follows that

$$\mathrm{E}(Y_n \mathbf{1}(Y_n > T)) \leq \varepsilon\,, n \geq 1\,.$$

Thus, $(Y_n : n \geq 1)$ is U.I., which completes the proof. □

Exercise 3.3. *If* $(X_\alpha : \alpha \in I)$ *is dominated by an integrable random variable* Y*, that is,*

$$|X_\alpha| \leq Y\,, \alpha \in I\,,$$

prove that $(X_\alpha : \alpha \in I)$ *is U.I. Hence, show that the dominated convergence theorem is weaker than Theorem 3.3.*

Theorem 3.4. *Let* $\mathrm{E}|X| < \infty$ *and* $(\mathcal{F}_\alpha : \alpha \in I)$ *be a family of sub-*σ*-fields. Then* $(\mathrm{E}(X|\mathcal{F}_\alpha) : \alpha \in I)$ *is uniformly integrable.*

Proof. For a fixed $\alpha \in I$ and $T > 0$, $[X_\alpha > T] \in \mathcal{F}_\alpha$. Hence,

$$\int_{[X_\alpha > T]} X_\alpha\, dP = \int_{[X_\alpha > T]} X\, dP\,.$$

Thus,

$$\begin{aligned}\int_{[|X_\alpha| > T]} |X_\alpha|\, dP &= \int_{[X_\alpha > T]} X_\alpha\, dP - \int_{[X_\alpha < -T]} X_\alpha\, dP \\ &= \int_{[X_\alpha > T]} X\, dP + \int_{[X_\alpha < -T]} (-X)\, dP \\ &\leq \int_{[|X_\alpha| > T]} |X|\, dP\,.\end{aligned}$$

Fix $\varepsilon > 0$. By integrability of X, get $\delta > 0$ s.t.

$$\int_A |X|\, dP \leq \varepsilon\,, \text{ if } P(A) \leq \delta\,.$$

Let $T = \frac{\mathrm{E}|X|}{\delta}$. Markov inequality implies that for any $\alpha \in I$,

$$\begin{aligned} P(|X_\alpha| > T) &\leq T^{-1}\mathrm{E}|X_\alpha| \\ &\leq T^{-1}\mathrm{E}|X| \\ &= \delta\,. \end{aligned}$$

Thus,

$$\int_{[|X_\alpha|>T]} |X_\alpha|\, dP \leq \int_{[|X_\alpha|>T]} |X|\, dP \leq \varepsilon\,.$$

This completes the proof. □

Exercise 3.4. *If $X_1, X_2, \ldots$ are i.i.d. with finite mean, and*

$$S_n = \sum_{i=1}^n X_i\,, n \geq 1\,,$$

show that $(S_n/n : n \geq 1)$ is U.I.

Exercise 3.5. *If $X_1, X_2, \ldots$ are i.i.d. with mean μ, and finite variance, and*

$$S_n = \sum_{i=1}^n X_i\,, n \geq 1\,,$$

show that

$$n^{-1} S_n \to \mu \text{ in } L^1\,, n \to \infty\,.$$

In the following theorem, the equivalence of **1** and **3** for $p = 1$ is known as Scheffé's lemma.

Theorem 3.5. *If X_n, X are such that $X_n \xrightarrow{P} X$ and*

$$\mathrm{E}\left(|X|^p\right) < \infty\,,$$

for some $p \in [1, \infty)$, then the following are equivalent.

1. *As $n \to \infty$,*
$$X_n \to X \text{ in } L^p\,.$$
2. *As $n \to \infty$,*
$$|X_n| \to |X| \text{ in } L^p\,.$$
3. *It holds that*
$$\limsup_{n\to\infty} \mathrm{E}\left(|X_n|^p\right) \leq \mathrm{E}\left(|X|^p\right)\,.$$

4. As $n \to \infty$,

$$|X_n|^p \to |X|^p \text{ in } L^1 .$$

5. The family $\{|X_n|^p : n \geq 1\}$ is uniformly integrable.

Proof. It will be proved that **1**⇒**2**⇒ … ⇒**5**⇒**1**.

Proof of **1**⇒**2**. Follows trivially from the observation that

$$\big||X_n| - |X|\big| \leq |X_n - X| .$$

□

Proof of **2**⇒**3**. The hypothesis **2** implies that

$$\lim_{n\to\infty} \mathrm{E}(|X_n|^p) = \mathrm{E}(|X|^p) ,$$

from which **3** follows. □

Proof of **3**⇒**4**. Write

$$\mathrm{E}\left(\big||X_n|^p - |X|^p\big|\right) = \mathrm{E}\left(|X_n|^p\right) + \mathrm{E}\left(|X|^p\right) - 2\mathrm{E}\left(|X_n|^p \wedge |X|^p\right) . \tag{3.1}$$

The assumption $X_n \xrightarrow{P} X$ implies that

$$|X_n|^p \wedge |X|^p \xrightarrow{P} |X|^p , n \to \infty .$$

The assumption that the p-th moment of $|X|$ is finite along with DCT implies

$$\lim_{n\to\infty} \mathrm{E}\left(|X_n|^p \wedge |X|^p\right) = \mathrm{E}\left(|X|^p\right) .$$

Use the above with (3.1) and **3** to get

$$\limsup_{n\to\infty} \mathrm{E}\left(\big||X_n|^p - |X|^p\big|\right) \leq 0 ,$$

which implies **4**. □

Proof of **4**⇒**5**. Follows from the 'only if' part of Theorem 3.3. □

Proof of **5**⇒**1**. Let

$$Y_n = |X_n - X|^p , n \geq 1 .$$

The assumption $X_n \xrightarrow{P} X$ is equivalent to $Y_n \xrightarrow{P} 0$. Furthermore,

$$0 \leq Y_n \leq 2^p \left(|X_n|^p + |X|^p\right) , n \geq 1 .$$

The hypothesis **5** and 1. of Theorem 3.2 imply that $\{|X_n|^p + |X|^p : n \geq 1\}$ is uniformly integrable. Furthermore, 2. of Theorem 3.2 shows uniform integrability of $(Y_n : n \geq 1)$. The 'if part' of Theorem 3.3 establishes

$$Y_n \to 0 \text{ in } L^1 ,$$

which is equivalent to **1**. □

This completes the proof. □

4 Martingales: definition and convergence results

Definition. *Let $(\Omega, \mathcal{A}, P)$ be a probability space. A family of σ-fields $(\mathcal{F}_n : n \in \mathbb{N})$ is a* ***filtration*** *if $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{A}$. A sequence of random variables $(X_n : n \in \mathbb{N})$ is a* ***martingale w.r.t. a filtration*** *$(\mathcal{F}_n)$ if for all $n \in \mathbb{N}$,*

- *X_n is $\mathcal{F}_n$-measurable,*
- *$\mathrm{E}|X_n| < \infty$,*
- *and $\mathrm{E}(X_{n+1}|\mathcal{F}_n) = X_n$.*

Dictionary meaning: "A gambling system of continually doubling the stakes in the hope of an eventual win that must yield a net profit."

Example 4.1. *Imagine a gambler who starts with a certain capital, say X_0. At each stage, he puts on stake his entire asset at that point of time. With probability half he doubles his asset, and with probability $1/2$, he looses everything. Let X_n denote his asset after n iterations. Then,given $X_1, \ldots, X_n$, X_{n+1} equals $2X_n$ or 0, both with probability $1/2$. Then, $(X_n : n \geq 1)$ is a martingale w.r.t. $(\sigma(X_0, \ldots, X_n) : n \geq 1)$.*

Example 4.2. *Let $X_1, X_2, \ldots$ are independent random variables with mean $\mu_1, \mu_2, \ldots$ respectively. Define*

$$\begin{aligned} S_n &:= \sum_{i=1}^{n} (X_i - \mu_i)\,,\ n \geq 1\,, \\ \mathcal{F}_n &:= \sigma(X_1, \ldots, X_n),\ n \geq 1\,. \end{aligned}$$

Then, $(S_n, \mathcal{F}_n)_{n\geq 1}$ is a martingale.

Example 4.3. *Suppose that $(N_t : t \geq 0)$ is a Poisson process with mean λ. Define*

$$\begin{aligned} X_n &:= N_n - n\lambda,\ n \geq 1\,, \\ \mathcal{F}_n &:= \sigma(N_t : t \leq n),\ n \geq 1\,. \end{aligned}$$

Then, $(X_n, \mathcal{F}_n)$ is a martingale.

Example 4.4. *Consider an urn with w white balls and b black balls. At each stage, a ball is selected at random from the urn, and a ball of the same color is added. Let X_n denote the proportion of white balls after n draws. Then X_n is a martingale.*

Example 4.5. *Let $(X_n : n \geq 0)$ be a Markov chain with a countable state space $\mathcal{S}$ and transition matrix P. Let $\pi : \mathcal{S} \to \mathbb{R}$ be a function such that for all $i \in \mathcal{S}$,*

$$\sum_{j \in \mathcal{S}} P(i,j)|\pi(j)| < \infty\,,$$

and

$$\sum_{j \in \mathcal{S}} P(i,j)\pi(j) = \pi(i)\,.$$

Then, $\pi(X_n)$ is a martingale w.r.t. $(\sigma(X_0, \ldots, X_n) : n \geq 1)$.

Definition. *A function $\tau : \Omega \to \mathbb{N} \cup \{0, \infty\}$ is a **stopping time** w.r.t. a filtration $(\mathcal{F}_n)$ if*

$$[\tau = n] \in \mathcal{F}_n \text{ for all } n \geq 0\,.$$

Exercise 4.1. *Given a filtration $(\mathcal{F}_n : n \geq 0)$ and a $\mathbb{N} \cup \{0, \infty\}$-valued random variable τ, show that τ is a stopping time if and only if*

$$[\tau \leq k] \in \mathcal{F}_k\,, k \geq 0\,.$$

Definition. *For a stopping time τ w.r.t. $(\mathcal{F}_n)$,*

$$\mathcal{F}_\tau := \{A \in \mathcal{A} : A \cap [\tau \leq k] \in \mathcal{F}_k \text{ for all } k \geq 0\}\,.$$

Definition. *The family $(X_n : n \geq 1)$ is adapted to a filtration $(\mathcal{F}_n : n \geq 1)$ if X_n is $\mathcal{F}_n$-measurable for all $n \in \mathbb{N}$.*

Exercise 4.2. *If (X_n) is adapted to $(\mathcal{F}_n)$, and τ is a stopping time, then show that X_τ is $\mathcal{F}_\tau$-measurable.*

Definition. *A family $(X_n : n \geq 1)$ of random variables is a **submartingale w.r.t. a filtration** $(\mathcal{F}_n)$ if for all $n \in \mathbb{N}$,*

- *X_n is $\mathcal{F}_n$-measurable,*
- *$\mathrm{E}|X_n| < \infty$,*
- *and $\mathrm{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$ a.s. .*

*If the inequality in the last line is reversed and everything else is unchanged, then it defines a **supermartingale**.*

Theorem 4.1 (Doob's optional stopping theorem)**.** *Let $(X_n, \mathcal{F}_n)_{n \geq 1}$ be a submartingale. Let τ_1, τ_2 be stopping times such that $1 \leq \tau_1 \leq \tau_2 \leq n$ for some fixed $n \in \mathbb{N}$. Then, X_{τ_1}, X_{τ_2} are integrable and*

$$\mathrm{E}\left(X_{\tau_2} | \mathcal{F}_{\tau_1}\right) \geq X_{\tau_1} \text{ a.s.}\,.$$

Proof. For $i = 1, 2$,

$$|X_{\tau_i}| = \left| \sum_{j=1}^{n} X_j \mathbf{1}(\tau_i = j) \right| \le \sum_{j=1}^{n} |X_j| ,$$

and hence X_{τ_i} is integrable. In view of Exc. 4.2, it suffices to show that

$$\int_A (X_{\tau_2} - X_{\tau_1}) dP \ge 0 \text{ for all } A \in \mathcal{F}_{\tau_1} .$$

Fix $A \in \mathcal{F}_{\tau_1}$ and note that for all $k \in \mathbb{N}$,

$$A \cap [\tau_1 < k \le \tau_2] = A \cap [\tau_1 \le k-1] \cap [\tau_2 \le k-1]^c \in \mathcal{F}_{k-1} .$$

Therefore,

$$\begin{aligned} \int_A (X_{\tau_2} - X_{\tau_1})\, dP &= \int_A \sum_{k=\tau_1+1}^{\tau_2} (X_k - X_{k-1}) dP \\ &= \sum_{k=2}^{n} \int_{A \cap [\tau_1 < k \le \tau_2]} (X_k - X_{k-1}) dP \\ &= \sum_{k=2}^{n} \int_{A \cap [\tau_1 < k \le \tau_2]} (\mathrm{E}(X_k | \mathcal{F}_{k-1}) - X_{k-1})\, dP \\ &\ge 0 , \end{aligned}$$

the inequality in the last line following from the submartingale property. This completes the proof. □

An immediate corollary of Doob's optional sampling theorem is the following.

Theorem 4.2. *Let $(X_n, \mathcal{F}_n)_{n \ge 1}$ be a martingale. Suppose that τ is a finite stopping time with respect to $(\mathcal{F}_n : n \ge 1)$ such that $(X_{\tau \wedge n} : n \ge 1)$ is U.I. Then, X_τ is integrable and*

$$\mathrm{E}(X_\tau) = \mathrm{E}(X_1) .$$

The proof uses the following trivial exercises.

Exercise 4.3. *1. If τ_1 and τ_2 are stopping times, show that so is $\tau_1 \wedge \tau_2$.*

2. Show that any deterministic $k \in \{0, 1, \ldots, \infty\}$, $\tau = k$ is a stopping time with respect to any filtration $(\mathcal{F}_n : n \ge 0)$, and that

$$\mathcal{F}_\tau = \mathcal{F}_k .$$

Proof of Theorem 4.2. Doob's optional sampling theorem implies that

$$X_1 = \mathrm{E}\left(X_{\tau\wedge n}|\mathcal{F}_1\right) , n \geq 1 ,$$

because 1 is a stopping time, as is $\tau \wedge n$, by the above exercise, and $1 \leq (\tau \wedge n) \leq n$. Thus,

$$\mathrm{E}(X_1) = \mathrm{E}\left(X_{\tau\wedge n}\right) , n \geq 1 . \tag{4.1}$$

Since τ is finite,

$$X_{\tau\wedge n} \to X_\tau , n \to \infty .$$

Uniform integrability of $(X_{\tau\wedge n} : n \geq 1)$ along with Theorem 3.3 implies that

$$\lim_{n\to\infty} \mathrm{E}\left(X_{\tau\wedge n}\right) = \mathrm{E}\left(X_\tau\right) ,$$

which, in conjunction with (4.1), completes the proof. □

Theorem 4.3. *Let $(X_n, \mathcal{F}_n)_{n\geq 1}$ be a martingale. Suppose that τ is a stopping time with respect to $(\mathcal{F}_n : n \geq 1)$ such that either of the following holds:*

1. *τ is bounded,*

2. *or there is a finite constant c satisfying*

$$|X_{n+1} - X_n| \leq c \text{ a.s. for all } n \geq 1 ,$$

and $\mathrm{E}(\tau) < \infty$.

Then,

$$\mathrm{E}(X_\tau) = \mathrm{E}(X_1) .$$

Proof. Under 1., the claim follows trivially from the Optional Sampling Theorem. If 2. holds, then

$$|X_{\tau\wedge n}| = \left| X_1 + \sum_{i=2}^{\tau\wedge n}(X_i - X_{i-1}) \right| \leq |X_1| + \sum_{i=2}^{\tau} |X_i - X_{i-1}| \leq |X_1| + c\tau .$$

Exc 3.3 implies that $(X_{\tau\wedge n} : n \geq 1)$ is U.I. from which the proof follows. □

Example 4.6. *Consider independent successive tosses of a fair coin. Let the outcome of the n-th toss be denoted by Z_n, that is, $Z_n = h$ or $Z_n = t$. Define*

$$\tau = \inf\{n \geq 2 : Z_{n-1} = Z_n = h\} .$$

We are interested in calculating $\mathrm{E}(\tau)$.

Observe that

$$\mathrm{E}\left[\mathbf{1}(Z_n = Z_{n+1} = h)|\mathcal{F}_n\right] = \frac{1}{2}\mathbf{1}(Z_n = h) , \tag{4.2}$$

where $\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n)$. Thus, the first task is to compensate

$$\sum_{i=2}^{n} \mathbf{1}\left(Z_{i-1} = Z_i = h\right) + \frac{1}{2}\mathbf{1}(Z_n = h)$$

in order to turn it into a martingale. Some calculations using (4.2) *show that letting*

$$X_n = \frac{1}{2}\mathbf{1}(Z_n = h) - \frac{n}{4} + \sum_{i=2}^{n} \mathbf{1}\left(Z_{i-1} = Z_i = h\right), n = 1, 2, \ldots,$$

$(X_n, \mathcal{F}_n)$ *is a martingale.*

Note that

$$\frac{\tau}{2} \le \inf\{n \ge 1 : Z_{2n-1} = Z_{2n} = h\},$$

The random variable on the RHS is geometric and has mean 4*. Thus*

$$\mathrm{E}(\tau) \le 8 .$$

Further, $|X_{n+1} - X_n| \le 2$*. Theorem 4.3 shows*

$$\mathrm{E}(X_\tau) = \mathrm{E}(X_1) = 0 .$$

Since

$$X_\tau = \frac{1}{2} - \frac{\tau}{4} + 1 ,$$

it follows that

$$\mathrm{E}(\tau) = 6 .$$

Example 4.7. *Let (S_n) be a simple symmetric random walk. Suppose that $-a < 0 < b$ be integers. Recall that for a gambler who starts with a capital of a rupees and wishes to quit once he earns b rupees, his probability of ruin is defined as*

$$p_r := P\left(\text{for some } n \ge 1,\, S_n = -a,\, S_m \ne b \text{ for all } 1 \le m \le n-1\right) .$$

Define

$$\tau := \inf\big\{n \in \mathbb{N} : S_n \in \{-a, b\}\big\} .$$

It is immediate that τ is a stopping time w.r.t. $(\mathcal{F}_n : n \ge 1)$*, where*

$$\mathcal{F}_n := \sigma(S_1, \ldots, S_n) .$$

Since (S_n) is a martingale and $|S_{n\wedge\tau}| \le a + b$, it follows by Theorem 4.2 that

$$\mathrm{E}(S_\tau) = \mathrm{E}(S_1) = 0 .$$

It is immediate that S_τ takes values $-a$ and b with probabilities p_r and $1-p_r$ respectively. Therefore,

$$0 = \mathrm{E}(S_\tau) = -ap_r + b(1-p_r)\,.$$

Therefore,

$$p_r := \frac{b}{a+b}\,.$$

Exercise 4.4. *If $(S_n : n \geq 0)$ is once again the simple symmetric random walk, and*

$$\tau = \inf\{n \geq 0 : S_n = 1\}\,,$$

show that $\tau < \infty$ a.s. and

$$\mathrm{E}(\tau) = \infty\,.$$

Hint: *Use Theorem 4.3.*

Theorem 4.4. *1. If $(X_n, \mathcal{F}_n)$ is a martingale and $\phi : \mathbb{R} \to \mathbb{R}$ is convex such that $\phi(X_n)$ is integrable, then $(\phi(X_n))$ is a submartingale.*
2. If $(X_n, \mathcal{F}_n)$ is a submartingale and $\phi : \mathbb{R} \to \mathbb{R}$ is convex and non-decreasing such that $\phi(X_n)$ is integrable, then $(\phi(X_n))$ is a submartingale.

Proof. Exercise. □

Theorem 4.5 (Doob's maximal inequality)**.** *If (X_n) is a submartingale w.r.t. a filtration $(\mathcal{F}_n)$, then for $\alpha > 0$,*

$$P\left(\max_{1\leq i\leq n} X_i \geq \alpha\right) \leq \frac{1}{\alpha}\mathrm{E}|X_n|,\ n \geq 1\,.$$

Exercise 4.5. *Recall Kolmogorov's maximal inequality, which says the following. If $X_1, \ldots, X_n$ are independent random variables each having zero mean and finite variance, then*

$$P\left(\max_{1\leq k\leq n} |S_k| \geq \alpha\right) \leq \frac{1}{\alpha^2}\mathrm{E}(S_n^2)\,,$$

where $S_k := X_1 + \ldots + X_k$. Show that this inequality is a corollary of Doob's maximal inequality.

Proof of Doob's maximal inequality. Let τ be the smallest k such that $X_k \geq 1$, if there is such a k, and n otherwise. Let

$$M_k := \max_{1\leq i\leq k} X_i,\ 1 \leq k \leq n\,.$$

Then, for all $1 \le k \le n-1$, $[M_n \ge \alpha] \cap [\tau \le k] = [M_k \ge \alpha] \in \mathcal{F}_k$, and for $k \ge n$, $[M_n \ge \alpha] \cap [\tau \le k] = [M_n \ge \alpha] \in \mathcal{F}_n \subset \mathcal{F}_k$. Thus, $[M_n \ge \alpha] \in \mathcal{F}_\tau$. Note that

$$\begin{aligned} \alpha P(M_n \ge \alpha) &\le \int_{[M_n \ge \alpha]} X_\tau \, dP \\ \text{(By the Optional Sampling Theorem)} &\le \int_{[M_n \ge \alpha]} X_n \, dP \\ &\le \mathrm{E}|X_n| \,, \end{aligned}$$

which completes the proof. □

Exercise 4.6. *If $(X_n, \mathcal{F}_n)$ is a martingale, then show that $|X_n|, \mathcal{F}_n)$ is a submartingale.*

The following is an immediate corollary of Doob's maximal inequality and the above exercise.

Theorem 4.6. *If $(X_n, \mathcal{F}_n)$ is a martingale, then for $\alpha > 0$,*

$$P\left(\max_{1 \le i \le n} |X_i| \ge \alpha\right) \le \frac{1}{\alpha} \mathrm{E}|X_n|, \; n \ge 1 \,.$$

Definition. *Let $(X_n : n \ge 1)$ be a process and $-\infty < \alpha < \beta < \infty$. Define*

$$\begin{aligned} \tau_0 &= 0 \,, \\ \tau_k &= \begin{cases} \inf\{n > \tau_{k-1} : X_n \ge \beta\}, \; k \text{ even} \,, \\ \inf\{n > \tau_{k-1} : X_n \le \alpha\}, \; k \text{ odd} \,. \end{cases} \end{aligned}$$

The ***number of upcrossings of*** (X_n) ***till time*** n *is defined as*

$$U_n := \#\{2 \le k \le n : \; k \text{ even and } \tau_k \le n\} \,.$$

Theorem 4.7. *For a submartingale (X_n), the number of upcrossings U_n, till time n, of $[\alpha, \beta]$ for $\alpha < \beta$, satisfies*

$$\mathrm{E}(U_n) \le \frac{|\alpha| + \mathrm{E}|X_n|}{\beta - \alpha} \,.$$

Proof. Define

$$\begin{aligned} Y_n &:= \max\{0, X_k - \alpha\}, \; n \ge 1 \,, \\ \theta &:= \beta - \alpha \,. \end{aligned}$$

By Exc 4.4.2, (Y_n) is a submartingale. Define

$$\begin{aligned} \tau_0 &= 0 \,, \\ \tau_k &= \begin{cases} \inf\{n > \tau_{k-1} : Y_n \ge \theta\}, \; k \text{ even} \,, \\ \inf\{n > \tau_{k-1} : Y_n = 0\}, \; k \text{ odd} \,. \end{cases} \end{aligned}$$

Fix $n \geq 1$, and observe that $\tau_k \wedge n$ is a stopping time for every $k \geq 1$. Notice that $n \wedge \tau_{k-1} < n \wedge \tau_k$ if $\tau_{k-1} < n$. Therefore, $\tau_n \geq n$, and hence,

$$\begin{aligned} Y_n &= Y_{\tau_n \wedge n} \\ \text{(because } Y_0 \geq 0) &\geq Y_{\tau_n \wedge n} - Y_{\tau_1 \wedge n} \\ &= \sum_{k=2}^{n} (Y_{\tau_k \wedge n} - Y_{\tau_{k-1} \wedge n}) \\ &= \sum_e + \sum_o , \end{aligned}$$

where $\sum_e$ and $\sum_o$ denote the sums over even k and odd k respectively in $2 \leq k \leq n$. By the O.S.T., for all k,

$$\mathrm{E}(Y_{\tau_k \wedge n}) \geq \mathrm{E}(Y_{\tau_{k-1} \wedge n}) , \tag{4.3}$$

and hence

$$\begin{aligned} \mathrm{E}(Y_n) &\geq \mathrm{E} \sum_e \\ &= \mathrm{E} \sum_{2 \leq k \leq n : k \text{ even}} (Y_{\tau_k \wedge n} - Y_{\tau_{k-1} \wedge n}) . \end{aligned}$$

For k even,

$$(Y_{\tau_k \wedge n} - Y_{\tau_{k-1} \wedge n}) \mathbf{1} (\tau_{k-1} \geq n) = 0 ,$$

and

$$(Y_{\tau_k \wedge n} - Y_{\tau_{k-1} \wedge n}) \mathbf{1} (\tau_{k-1} < n < \tau_k) = Y_n \geq 0 .$$

Therefore

$$\begin{aligned} \mathrm{E}(Y_n) &\geq \mathrm{E} \sum_{2 \leq k \leq n : k \text{ even}} (Y_{\tau_k \wedge n} - Y_{\tau_{k-1} \wedge n}) \mathbf{1}(\tau_k \leq n) \\ &= \mathrm{E} \sum_{2 \leq k \leq n : k \text{ even}} (Y_{\tau_k} - Y_{\tau_{k-1}}) \mathbf{1}(\tau_k \leq n) \\ &\geq \mathrm{E} \left(\theta \# \{ k : 2 \leq k \leq n, \, k \text{ even and } \tau_k \leq n \} \right) \\ &= \theta \mathrm{E}(U_n) . \end{aligned}$$

Thus,

$$(\beta - \alpha) \mathrm{E}(U_n) \leq \mathrm{E}(Y_n) \leq |\alpha| + \mathrm{E}|X_n| .$$

Thus the proof follows. □

Theorem 4.8 (Convergence of L^1-bounded submartingales)**.** *If $(X_n : n \geq 1)$ is a submartingale satisfying*

$$\sup_{n \geq 1} \mathrm{E}(|X_n|) < \infty ,$$

then there exists an integrable random variable X such that

$$X_n \to X \ \ a.s.$$

Proof. For fixed $\alpha < \beta$, let $U_n(\alpha, \beta)$ denote the number of upcrossings of $[\alpha, \beta]$ by (X_n) till time n, for $1 \le n \le \infty$. The upcrossings inequality implies that

$$\mathrm{E}\left(U_n(\alpha, \beta)\right) \le (\beta - \alpha)^{-1}\left(|\alpha| + \mathrm{E}|X_n|\right),\, 1 \le n < \infty\,.$$

Letting $n \to \infty$ and using the hypothesis, we get

$$\mathrm{E}\left(U_\infty(\alpha, \beta)\right) < \infty\,,$$

with the aid of the M.C.T.

Thus,

$$U_\infty(\alpha, \beta) < \infty \text{ a.s.}$$

Therefore,

$$P\left(U_\infty(\alpha, \beta) < \infty \text{ for all } \alpha, \beta \in \mathbb{Q},\, \alpha < \beta\right) = 1\,.$$

Hence,

$$\lim_{n\to\infty} X_n$$

exists, a.s. That is, if

$$X = \limsup_{n\to\infty} X_n\,,$$

then $X_n \to X$ a.s. Fatou's lemma implies that

$$\mathrm{E}|X| \le \liminf_{n\to\infty} \mathrm{E}|X_n| < \infty\,,$$

and hence X is integrable. This completes the proof. □

Example 4.8. *Let $(S_n : n \ge 0)$ be a simple symmetric random walk. It is trivial that it is a martingale w.r.t. $(\sigma(S_1, \ldots, S_n) : n \ge 1)$. It is well known that a.s.,*

$$\liminf_{n\to\infty} S_n = -\infty,\, \limsup_{n\to\infty} S_n = +\infty\,.$$

Therefore, the almost sure convergence of a martingale may fail in the absence of L^1-boundedness.

Example 4.9. *Recall the example of a gambler who enters a Casino with a positive capital, and each stage he bets his entire asset. At the next stage, his asset either gets doubled or becomes zero, each with probability $1/2$, independently of what has happened so far. If X_n denotes his asset at time n, then clearly*

$$\sup_{n\ge1} \mathrm{E}(X_n) = X_0\,.$$

However, $X_n \to 0$ a.s., but not in L^1. This shows that (X_n) is not uniformly integrable, even though it is L^1-bounded.

Theorem 4.9 (Convergence of uniformly integrable submartingales)**.** *Let $(X_n, \mathcal{F}_n : n \geq 1)$ be an uniformly integrable submartingale. Then there exists an integrable random variable X_∞ such that*

$$X_n \to X_\infty,\, n \to \infty \text{ a.s. and in } L^1 .$$

Furthermore,

$$X_n \leq \mathrm{E}(X_\infty | \mathcal{F}_n) \text{ a.s. for all } n \geq 1 . \tag{4.4}$$

Proof. Since (X_n) is uniformly integrable, it follows that

$$\sup_{n \geq 1} \mathrm{E}|X_n| < \infty .$$

By Theorem 4.8, there exists X_∞, which is integrable, such that

$$X_n \to X_\infty \text{ a.s.}$$

Using uniform integrability once again, it follows that

$$X_n \to X_\infty \text{ in } L^1 .$$

Fix $n \in \mathbb{N}$. It is immediate that

$$\mathrm{E}(X_m | \mathcal{F}_n) \to \mathrm{E}(X_\infty | \mathcal{F}_n) \text{ in } L^1 \text{ as } m \to \infty .$$

Since for all $m \geq n$,

$$X_n \leq \mathrm{E}(X_m | \mathcal{F}_n) \text{ a.s.},$$

(4.4) follows, which completes the proof. □

An immediate corollary of Theorem 4.9 is the following which follows by working with $-X_n$ as well, and hence the proof is left as an exercise.

Theorem 4.10. *Let $(X_n, \mathcal{F}_n : n \geq 1)$ be an uniformly integrable martingale. Then there exists an integrable random variable X_∞ such that*

$$X_n \to X_\infty,\, n \to \infty \text{ a.s. and in } L^1 .$$

Furthermore,

$$X_n = \mathrm{E}(X_\infty | \mathcal{F}_n) \text{ a.s. for all } n \geq 1 .$$

The following result, known as Lévy's upward theorem, is a consequence of the above.

Theorem 4.11. *If $(\mathcal{F}_n)$ is a filtration and X an integrable random variable, then*

$$\mathrm{E}\left(X | \mathcal{F}_n\right) \to \mathrm{E}(X | \mathcal{F}_\infty)\,, n \to \infty\,,$$

a.s. and in L^1, where

$$\mathcal{F}_\infty = \sigma\left(\bigcup_{n=1}^\infty \mathcal{F}_n\right) .$$

Proof. Let

$$X_n = \mathrm{E}(X|\mathcal{F}_n)\,, n \geq 1\,.$$

Then, for all $n \geq 1$,

$$\begin{aligned}\mathrm{E}(X_{n+1}|\mathcal{F}_n) &= \mathrm{E}\left(\mathrm{E}(X|\mathcal{F}_{n+1})|\mathcal{F}_n\right) \\ &= \mathrm{E}(X|\mathcal{F}_n) = X_n\,,\end{aligned}$$

because $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. Thus (X_n) is a martingale. Furthermore, (X_n) is uniformly integrable by Theorem 3.4. Thus, there exists X_∞ such that

$$X_n \to X_\infty\,, n \to \infty\,,$$

a.s. and in L^1. Since each X_n is $\mathcal{F}_\infty$-measurable, redefining

$$X_\infty = \limsup_{n\to\infty} X_n\,,$$

it follows that X_∞ is $\mathcal{F}_\infty$-measurable. Fix

$$A \in \bigcup_{n=1}^{\infty} \mathcal{F}_n\,,$$

that is, $A \in \mathcal{F}_{n_0}$ for some n_0. For all $n \geq n_0$, since $A \in F_n$, it holds that

$$\int_A X_n\,dP = \int_A X\,dP\,.$$

Since $X_n \to X_\infty$ in L^1, what follows is

$$\int_A X_\infty\,dP = \int_A X\,dP\,.$$

Thus, for all

$$A \in \bigcup_{n=1}^{\infty} \mathcal{F}_n\,,$$

it holds that

$$\int_A X_\infty\,dP = \int_A X\,dP\,.$$

Since the class of sets A for which the above hold is a monotone class, and $\bigcup_{n=1}^{\infty} \mathcal{F}_n$ is a field, the monotone class theorem implies that

$$\int_A X_\infty\,dP = \int_A X\,dP\,, A \in \mathcal{F}_\infty\,.$$

Thus,

$$X_\infty = \mathrm{E}(X|\mathcal{F}_\infty)\,.$$

Hence the proof. □

Henceforth, we shall use the notation

$$\bigvee_{n=1}^{\infty} \mathcal{F}_n = \sigma\left(\bigcup_{n=1}^{\infty} \mathcal{F}_n\right).$$

Theorem 4.12 (Convergence of L^p-bounded martingales)**.** *If* $(X_n : \mathcal{F}_n : n \geq 1)$ *is a martingale such that*

$$\sup_{n\geq 1} \mathrm{E}(|X_n|^p) < \infty\,,$$

for some $1 < p < \infty$*, then there exists* $X_\infty \in L^p(\Omega)$ *such that*

$$X_n \to X_\infty \text{ a.s. and in } L^p\,,$$

and

$$X_n = \mathrm{E}(X_\infty|\mathcal{F}_n) \text{ a.s. for all } n \geq 1\,.$$

Proof. In view of Theorem 4.9, it suffices to show that $(|X_n|^p : n \geq 1)$ is uniformly integrable. Since X_n, is a martingale, $(|X_n|^p)$ is a submartingale with respect to the same filtration. Note that

$$\begin{aligned}
\mathrm{E}\left(\max_{1\leq i\leq n}|X_i|^p\right) &= \int_0^\infty P\left(\max_{1\leq i\leq n}|X_i|^p \geq x\right)dx \\
&= \int_0^\infty P\left(\max_{1\leq i\leq n}|X_i| \geq x^{1/p}\right)dx \\
&= p\int_0^\infty z^{p-1}P\left(\max_{1\leq i\leq n}|X_i| \geq z\right)dz \\
&\leq p\int_0^\infty z^{p-2}\int_{\left[\max_{1\leq i\leq n}|X_i|\geq z\right]}|X_n|\,dP\,dz \\
&= p\int_\Omega |X_n|\int_0^{\max_{1\leq i\leq n}|X_i|} z^{p-2}\,dz\,dP \\
&= \frac{p}{p-1}\int_\Omega |X_n|\left(\max_{1\leq i\leq n}|X_i|\right)^{p-1}dP \\
\left(q := \frac{p}{p-1}\right) \quad &= q\mathrm{E}\left(|X_n|\max_{1\leq i\leq n}|X_i|^{p-1}\right) \\
&\leq q\left(\mathrm{E}(|X_n|^p)\right)^{1/p}\left(\mathrm{E}\left(\max_{1\leq i\leq n}|X_i|^{(p-1)q}\right)\right)^{1/q},
\end{aligned}$$

the last line invoking the Holder's inequality, along with the observation that $p^{-1} + q^{-1} = 1$. Therefore,

$$\mathrm{E}\left(\max_{1\leq i\leq n}|X_i|^p\right) \leq q^p\mathrm{E}(|X_n|^p)\,.$$

Letting $n \to \infty$,

$$\mathrm{E}\left(\sup_{n\geq 1}|X_n|^p\right) \leq q^p \sup_{n\geq 1}\mathrm{E}(|X_n|^p) < \infty\,.$$

This shows that $(|X_n|^p : n \geq 1)$ is uniformly integrable, and thereby completes the proof. □

Example 4.10. *Consider an urn having black and white balls, the initial proportion of white balls being X_0. A ball is drawn at random and replaced. A ball of the same color is added to the urn. If X_n denotes the proportion of white balls after n draws, then Theorem 4.12 shows that there exists a random variables X_∞ such that*

$$X_n \to X_\infty \text{ a.s. and in } L^p\,,$$

for all $1 \leq p < \infty$, $0 \leq X_\infty \leq 1$ and $\mathrm{E}(X_\infty) = X_0$. Since X_n^2 is a submartingale, it follows that

$$\mathrm{Var}(X_n) = \mathrm{E}(X_n^2) - X_0^2 \uparrow \mathrm{E}(X_\infty^2) - X_0^2 = \mathrm{Var}(X_\infty)\,.$$

Therefore, if $0 < X_0 < 1$, then X_1 is non-degenerate and hence so is X_∞.

Definition. *In a probability space $(\Omega, \mathcal{A}, P)$, a family of σ-fields $(\mathcal{F}_n : n \geq 1)$ is a* ***reverse filtration*** *if $\mathcal{A} \supset \mathcal{F}_1 \supset \mathcal{F}_2 \supset \ldots$. A family $(X_n : n \geq 1)$ of integrable random variables is a reverse martingale w.r.t. a reverse filtration $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_n$ measurable for all $n \in \mathbb{N}$ and*

$$\mathrm{E}(X_{n-1}|\mathcal{F}_n) = X_n,\, n \geq 2\,.$$

Theorem 4.13 (Reverse martingale convergence theorem)**.** *If $(X_n, \mathcal{F}_n)$ is a reverse martingale, then as $n \to \infty$,*

$$X_n \to \mathrm{E}\left(X_1 \middle| \bigcap_{n=1}^{\infty} \mathcal{F}_n\right)\,,$$

a.s. and in L^1.

Proof. Fix $n_0 \geq 1$ and define

$$k(n) := (n_0 + 1 - n) \vee 1\,.$$

Clearly $(X_{k(n)}, \mathcal{F}_{k(n)} : n \geq 1)$ is a martingale. Furthermore, for fixed $-\infty < \alpha < \beta < \infty$

$$\begin{aligned}
& \mathrm{E}\left[\#\text{ times } X_{n_0}, \ldots, X_1 \text{ upcrosses } [\alpha,\beta]\right] \\
= \; & \mathrm{E}\left[\#\text{ times } X_{k(1)}, \ldots, X_{k(n_0)} \text{ upcrosses } [\alpha,\beta]\right] \\
\leq \; & \frac{|\alpha| + \mathrm{E}|X_{k(n_0)}|}{\beta - \alpha} \\
= \; & \frac{|\alpha| + \mathrm{E}|X_1|}{\beta - \alpha}\,.
\end{aligned}$$

Since the right hand side is independent of n_0, it follows that X_n converges almost surely to a random variable X_∞. without loss of generality, one can define

$$X_\infty = \limsup_{n\to\infty} X_n \,,$$

and it is clear that X_∞ is measurable w.r.t. $\mathcal{F}_\infty = \bigcap_{n=1}^\infty \mathcal{F}_n$. Furthermore, for any $A \in \mathcal{F}_\infty$,

$$\int_A X_\infty dP = \lim_{n\to\infty} \int_A X_n dP = \int_A X_1 dP \,.$$

Thus,

$$\mathrm{E}(X_1 | \mathcal{F}_\infty) = X_\infty \,,$$

and hence the proof follows. □

A restatement of the above result is Lévy's downward theorem which says the following.

Theorem 4.14. *If Z is an integrable random variable and $(\mathcal{F}_n : n \ge 1)$ a reverse filtration, then*

$$\mathrm{E}(Z|\mathcal{F}_n) \to \mathrm{E}\left(Z \Big| \bigcap_{n=1}^\infty \mathcal{F}_n \right) \text{ a.s. and in } L^1 \,, n \to \infty \,.$$

Combining Theorems 4.11 and 4.14 yields the following.

Theorem 4.15. *Let Z be an integrable random variable. If $(\mathcal{F}_n)$ is a filtration or a reverse filtration, then*

$$\mathrm{E}(Z|\mathcal{F}_n) \to \mathrm{E}\left(Z|\mathcal{F}_\infty\right) \,,$$

as $n \to \infty$, a.s. and in L^1, where

$$\mathcal{F}_\infty := \begin{cases} \bigvee_{n=1}^\infty \mathcal{F}_n \,, & \textit{filtration} \,, \\ \bigcap_{n=1}^\infty \mathcal{F}_n \,, & \textit{reverse filtration} \,. \end{cases}$$

5 Applications of martingale and reverse martingale convergence theorems

In this chapter, we'll see several applications of the results in the last one. The first one is Kolmogorov's 0-1 law.

Theorem 5.1 (Kolmogorov's 0-1 law)**.** *Let $\mathcal{A}_1, \mathcal{A}_2, \ldots$ be independent sub σ-fields of the master σ-field. The tail σ-field is defined as*

$$\mathcal{T} = \bigcap_{n=1}^\infty \bigvee_{i=n}^\infty \mathcal{A}_i \,.$$

Then, for all $A \in \mathcal{T}$, $P(A)$ is either 0 or 1, that is, $\mathcal{T}$ is a trivial σ-field.

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

$$\mathcal{F}_n = \bigvee_{i=1}^{n} \mathcal{A}_i \,,$$
$$\mathcal{T}_n = \bigvee_{i=n+1}^{\infty} \mathcal{A}_i \,.$$

Observe that for every n, $\mathcal{F}_n$ and $\mathcal{T}_n$ are independent. Fix $A \in \mathcal{T}$. Therefore, for every $n \in \mathbb{N}$, $A \in \mathcal{T}_n$, and hence

$$\mathrm{E}\left(\mathbf{1}_A | \mathcal{F}_n\right) = P(A) \,.$$

The observations that $(\mathcal{F}_n)$ is a filtration and $\bigvee_{n=1}^{\infty} \mathcal{F}_n = \bigvee_{i=1}^{\infty} \mathcal{A}_i$ imply that

$$\mathrm{E}\left(\mathbf{1}_A | \mathcal{F}_n\right) \to \mathrm{E}\left(\mathbf{1}_A \middle| \bigvee_{i=1}^{\infty} \mathcal{A}_i\right) \text{ a.s. as } n \to \infty \,.$$

Since $\bigvee_{i=1}^{\infty} \mathcal{A}_i \supset \mathcal{T}$, it follows that

$$\mathrm{E}\left(\mathbf{1}_A \middle| \bigvee_{i=1}^{\infty} \mathcal{A}_i\right) = \mathbf{1}_A, \text{ almost surely.}$$

Thus,

$$P(A) = \mathbf{1}_A \text{ almost surely.}$$

Since the right hand side takes only two values which are 0 and 1, the proof follows. □

An immediate application of the above result and the reverse martingale convergence theorem is the strong law of large numbers.

Theorem 5.2 (SLLN)**.** *If $X_1, X_2, \ldots$ are i.i.d. with finite mean μ, and*

$$S_n = \sum_{i=1}^{n} X_i \,, n \ge 1 \,,$$

then, as $n \to \infty$,

$$n^{-1} S_n \to \mu \textit{ a.s. and in } L^1 \,.$$

For the proof, the following exercise is needed.

Exercise 5.1. *Let X be an integrable random variable and $\mathcal{F}, \mathcal{G}$ be σ-fields such that $\mathcal{G} \vee \sigma(X)$ is independent of $\mathcal{F}$. Show that*

$$\mathrm{E}\left(X | \mathcal{F} \vee \mathcal{G}\right) = \mathrm{E}\left(X | \mathcal{G}\right) \textit{ a.s.}$$

Solution. Let
$$Y = \mathrm{E}\left(X|\mathcal{G}\right) .$$
Clearly, Y is $\mathcal{F}\vee\mathcal{G}$-measurable. To complete the solution, it suffices to show that
$$\mathrm{E}\left(Y\mathbf{1}_E\right) = \mathrm{E}(X\mathbf{1}_E)\,, E \in \mathcal{F}\vee\mathcal{G}\,. \tag{5.1}$$
Let
$$\mathcal{A} = \{A\cap B : A \in \mathcal{F},\, B \in \mathcal{G}\}\,.$$
Clearly, if $E, F \in \mathcal{A}$ then $E \cap F \in \mathcal{A}$ and if $E = A \cap B$ for $A \in \mathcal{F}$, $B \in \mathcal{G}$, then
$$E^c = A^c \cup (A \cap B^c)\,,$$
and $A^c, A \cap B^c \in \mathcal{A}$ are disjoint. Thus, $\mathcal{A}$ is a semi-field. Furthermore, for E as above, $X\mathbf{1}_B$ and $\mathbf{1}_A$ are measurable with respect to $\sigma(X)\vee\mathcal{G}$ and $\mathcal{F}$, respectively, and hence are independent. Therefore,
$$\begin{aligned}
\mathrm{E}\left(X\mathbf{1}_E\right) &= \mathrm{E}\big((X\mathbf{1}_B)\,\mathbf{1}_A\big)\\
&= \mathrm{E}(X\mathbf{1}_B)\mathrm{E}(\mathbf{1}_A)\\
\big(\text{because } B \in \mathcal{G} \text{ and } Y = \mathrm{E}(X|\mathcal{G})\big) &= \mathrm{E}(Y\mathbf{1}_B)\mathrm{E}(\mathbf{1}_A)\\
&= \mathrm{E}\left(Y\mathbf{1}_E\right) ,
\end{aligned}$$
the last line following from the observation that $Y\mathbf{1}_B$ and $\mathbf{1}_A$ are measurable with respect to $\mathcal{G}$ and $\mathcal{F}$, respectively, which are independent.

Thus, (5.1) holds for all $E \in \mathcal{A}$. Since $\mathcal{A}$ is a semi-field, sets of the form $E_1 \cup \ldots \cup E_n$ for disjoint $E_1, \ldots, E_n \in \mathcal{A}$ and $n \geq 1$ constitute the field generated by $\mathcal{A}$. Thus, (5.1) holds for all E in the field generated by $\mathcal{A}$. It is easy to see that the collection of sets E for which (5.1) holds is a monotone class. The monotone class theorem proves that it holds for all E in the σ-field generated by $\mathcal{A}$ which is $\mathcal{F}\vee\mathcal{G}$. This completes the solution.

Proof of Theorem 5.2. Let
$$\mathcal{F}_n = \sigma(S_n, S_{n+1}, S_{n+2}, \ldots)\,, n \geq 1\,.$$
Clearly, $(\mathcal{F}_n)$ is a reverse filtration, and
$$\mathcal{F}_n = \sigma(S_n, X_{n+1}, X_{n+2}, \ldots) = \sigma(S_n) \vee \sigma(X_{n+1}, X_{n+2}, \ldots)\,.$$
Fix $n \geq 1$ and denote $X = X_1$, $\mathcal{F} = \sigma(X_{n+1}, X_{n+2}, \ldots)$ and $\mathcal{G} = \sigma(S_n)$. Clearly, $\sigma(X)\vee\mathcal{G}$ and $\mathcal{F}$ are independent, which in view of the above exercise implies that
$$\mathrm{E}(X|\mathcal{F}\vee\mathcal{G}) = \mathrm{E}(X|\mathcal{G})\,,$$
that is,
$$\mathrm{E}\left(X_1|\mathcal{F}_n\right) = \mathrm{E}(X_1|S_n) = \frac{1}{n}S_n\,,$$

the second equality being a trivial consequence of the fact that $X_1, \ldots, X_n$ are i.i.d.

Since $(\mathcal{F}_n)$ is a reverse filtration, it follows that

$$n^{-1}S_n = \mathrm{E}\left(X_1 | \mathcal{F}_n\right) \to \mathrm{E}\left(X_1 \middle| \bigcap_{n=1}^{\infty} \mathcal{F}_n\right) ,$$

in L^1 and a.s.

Let

$$Z = \limsup_{n\to\infty} n^{-1}S_n .$$

It is immediate that for any $k \geq 1$,

$$Z = \limsup_{n\to\infty} n^{-1} \sum_{i=k}^{n} X_i ,$$

and hence Z is $\sigma(X_k, X_{k+1}, \ldots)$-measurable for all k. Therefore, Z is measurable with respect to

$$\bigcap_{k=1}^{\infty} \sigma(X_k, X_{k+1}, \ldots) .$$

Kolmogorov's 0-1 law implies that the above σ-field is trivial, that is, every event therein has probability 0 or 1, and hence Z is degenerate. Since

$$Z = \mathrm{E}\left(X_1 \middle| \bigcap_{n=1}^{\infty} \mathcal{F}_n\right) \text{ a.s.} ,$$

it follows that $Z = \mathrm{E}(X_1) = \mu$ a.s. Thus, we have shown that

$$n^{-1}S_n \to \mu ,$$

a.s. and in L^1, as desired. □

If $\mathcal{A}_1, \mathcal{A}_2, \ldots$ are σ-fields generated by i.i.d. random variables, then Kolmogorov's 0-1 law can be strengthened to a result due to Hewitt and Savage. For stating the same, a few notions have to be introduced. Recall that

$$\mathbb{R}^{\mathbb{N}} = \{(x_1, x_2, \ldots) : x_n \in \mathbb{R} \text{ for all } n \in \mathbb{N}\} ,$$

and $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ is the smallest σ-field on $\mathbb{R}^{\mathbb{N}}$ with respect to which the coordinate function $\theta_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, defined by $\theta_n(x) = x_n$, is measurable for each $n \in \mathbb{N}$. Fix $n \in \mathbb{N}$ and a permutation π of $\{1, \ldots, n\}$. Define the function

$$f_\pi : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}} ,$$

by

$$f_\pi(x) = (x_{\pi(1)}, \ldots, x_{\pi(n)}, x_{n+1}, \ldots)\,.$$

Let

$$\mathcal{S}_\pi = \Big\{ B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}}) : f_\pi(B) = B \Big\}\,. \tag{5.2}$$

Clearly, $\mathcal{S}_\pi$ is a σ-field. Define

$$\mathcal{S} = \bigcap_{n=1}^{\infty} \bigcap_{\pi \text{ permutation of } \{1,\ldots,n\}} \mathcal{S}_\pi\,.$$

Theorem 5.3 (Hewitt-Savage 0-1 law)**.** *If $X_1, X_2, , \ldots$ are i.i.d. random variables, then for all $B \in \mathcal{S}$,*

$$P\left((X_1, X_2, \ldots) \in B\right) = 0 \ \textit{or}\ 1\,.$$

First note that the Hewitt-Savage law implies the Kolmogorov law in the i.i.d. case. That is, for random variables $Z_1, Z_2, \ldots$ defined on $(\Omega, \mathcal{A}, P)$,

$$\bigcap_{n=1}^{\infty} \sigma(Z_n, Z_{n+1}, \ldots) \subset$$

$$\big\{ A \in \mathcal{A} :\ A = \{\omega \in \Omega : (Z_1(\omega), Z_2(\omega), \ldots) \in B\} \text{ for some } B \in \mathcal{S} \big\}\,.$$

This follows from the observation that for every $n \in \mathbb{N}$ and every permutation π of $\{1, \ldots, n\}$,

$$\sigma(Z_{n+1}, Z_{n+2}, \ldots) \subset \big\{ A \in \mathcal{A} :\ A = [(Z_1, Z_2, \ldots) \in B] \text{ for some } B \in \mathcal{S}_\pi \big\}\,.$$

Proof of the Hewitt-Savage law. Fix $B \in \mathcal{S}$, and define

$$W = \mathbf{1}\left[(X_1, X_2, \ldots) \in B\right]\,.$$

In order to complete the proof, it suffices to show that

$$\mathrm{E}(W^2) = \mathrm{E}^2(W)\,. \tag{5.3}$$

Setting

$$\mathcal{F}_n := \sigma(X_1, \ldots, X_n),\ n \geq 1\,,$$

it follows from the observation that W is $\bigvee_{n=1}^{\infty} \mathcal{F}_n$-measurable, that

$$\lim_{n\to\infty} \mathrm{E}\left|\mathrm{E}(W|\mathcal{F}_n) - W\right| = 0\,.$$

Fix $\varepsilon > 0$. There exists $n \geq 1$ such that

$$\mathrm{E}\left|\mathrm{E}(W|\mathcal{F}_n) - W\right| \leq \varepsilon\,. \tag{5.4}$$

Since $0 \leq W \leq 1$, there exists a measurable function $g : \mathbb{R}^n \to [0,1]$ such that

$$Y := \mathrm{E}(W|\mathcal{F}_n) = g(X_1, \ldots, X_n)\,.$$

A restatement of (5.4) is that

$$\mathrm{E}\left|\mathbf{1}\left[(X_1, X_2, \ldots) \in B\right] - g(X_1, \ldots, X_n)\right| \leq \varepsilon\,.$$

Since $(X_1, X_2, \ldots) \stackrel{d}{=} (X_{n+1}, \ldots, X_{2n}, X_1, \ldots, X_n, X_{2n+1}, X_{2n+2}, \ldots)$, it follows that

$$\begin{aligned}
&\mathrm{E}\Big|\mathbf{1}\left[(X_{n+1}, \ldots, X_{2n}, X_1, \ldots, X_n, X_{2n+1}, X_{2n+2}, \ldots) \in B\right] \\
&\quad - g(X_{n+1}, \ldots, X_{2n})\Big| \\
&\leq \varepsilon\,.
\end{aligned}$$

Since $B \in \mathcal{S}$, it is immediate that

$$W = \mathbf{1}\left[(X_{n+1}, \ldots, X_{2n}, X_1, \ldots, X_n, X_{2n+1}, X_{2n+2}, \ldots) \in B\right]\,.$$

Denoting

$$Z := g(X_{n+1}, \ldots, X_{2n})\,,$$

it is immediate that

$$\mathrm{E}\left|W - Z\right| \leq \varepsilon\,. \tag{5.5}$$

The fact that Y and Z are i.i.d. implies that

$$\begin{aligned}
\left|\mathrm{E}^2(W) - \mathrm{E}(YZ)\right| &= \left|\mathrm{E}^2(W) - \mathrm{E}^2(Z)\right| \\
&\leq 2\mathrm{E}\left|W - Z\right| \\
&\leq 2\varepsilon\,,
\end{aligned}$$

the last line following from (5.5). Similarly,

$$\begin{aligned}
\left|\mathrm{E}(W^2) - \mathrm{E}(YZ)\right| &\leq \mathrm{E}|W^2 - WY| + \mathrm{E}|WY - YZ| \\
&\leq \mathrm{E}|W - Y| + \mathrm{E}|W - Z| \\
&\leq 2\varepsilon\,,
\end{aligned}$$

the last line following from (5.4) and (5.5). The two inequalities put together imply that

$$\left|\mathrm{E}(W^2) - \mathrm{E}^2(W)\right| \leq 4\varepsilon\,.$$

Since ε is arbitrary, this shows (5.3) from which the proof follows. □

The next application is the strong law of large numbers for U-statistics, of which Theorem 5.2 is a special case.

Definition. *Let $f : \mathbb{R}^r \to \mathbb{R}$ be a measurable function. The U-statistic, associated with f, based on a sample $x_1, \ldots, x_n$ of size $n \geq r$ is defined as*

$$U_n(x_1, \ldots, x_n) = \frac{1}{r!\binom{n}{r}} \sum_{1 \leq i_1, \ldots, i_r \leq n : i_j \neq i_k \text{ for all } j \neq k} f(x_{i_1}, \ldots, x_{i_r}) .$$

Theorem 5.4 (SLLN for U-statistics)**.** *Suppose that $X_1, X_2, \ldots$ are i.i.d. random variables. Let U_n be the U-statistic based on $X_1, \ldots, X_n$ associated with an $f : \mathbb{R}^r \to \mathbb{R}$ for $n \geq r$. If*

$$\mathrm{E}|f(X_1, \ldots, X_r)| < \infty ,$$

then, as $n \to \infty$,

$$U_n \to \mathrm{E}\left(f(X_1, \ldots, X_r)\right) \text{ a.s. and in } L^1 .$$

Proof. Let the underlying probability space be denoted by $(\Omega, \mathcal{A}, P)$. Let $\mathcal{S}_\pi$ be as in (5.2). Define for all $n \in \mathbb{N}$,

$$\mathcal{S}_n = \bigcap_{\pi \text{ permutation of } \{1,\ldots,n\}} \mathcal{S}_\pi ,$$

and

$$\mathcal{F}_n = \{A \in \mathcal{A} : A = [(X_1, X_2, \ldots) \in B] \text{ for some } B \in \mathcal{S}_n\} .$$

It is immediate that $\mathcal{S}_n \supset \mathcal{S}_{n+1}$, and hence $\mathcal{F}_n \supset \mathcal{F}_{n+1}$.

Fix $n \geq r$. Clearly if a measurable function $g : \mathbb{R}^n \to \mathbb{R}$ is such that

$$g(x_1, \ldots, x_n) = g(x_{\pi(1)}, \ldots, x_{\pi(n)})$$

for every permutation π of $\{1, \ldots, n\}$ and $(x_1, \ldots, x_n) \in \mathbb{R}^n$, then g is $\mathcal{F}_n$-measurable. Hence U_n is $\mathcal{F}_n$-measurable. Since $X_1, X_2, \ldots$ are i.i.d., it follows that for distinct $i_1, \ldots, i_r \in \{1, \ldots, n\}$,

$$\mathrm{E}\big(f(X_{i_1}, \ldots, X_{i_r})|\mathcal{F}_n\big) = \mathrm{E}\big(f(X_1, \ldots, X_r)|\mathcal{F}_n\big) .$$

Adding over all distinct $i_1, \ldots, i_r \in \{1, \ldots, n\}$, we get

$$\mathrm{E}\left(\sum_{i_1, \ldots, i_r \in \{1, \ldots, n\} \text{ distinct}} f(X_{i_1}, \ldots, X_{i_r}) \middle| \mathcal{F}_n \right) = r!\binom{n}{r} \mathrm{E}\big(f(X_1, \ldots, X_r)|\mathcal{F}_n\big) .$$

Since the left hand side is the conditional expectation of $r!\binom{n}{r}U_n$ which is $\mathcal{F}_n$-measurable, it follows that

$$\mathrm{E}\big(f(X_1, \ldots, X_r)|\mathcal{F}_n\big) = U_n .$$

Therefore, there exists a random variable U_∞ such that

$$U_n \to U_\infty \,, \tag{5.6}$$

as $n \to \infty$, a.s. and in L^1.

In order to complete the proof, it suffices to show that U_∞ is degenerate. Denote

$$\mathcal{N} = \{A \in \mathcal{A} : P(A) = 0\} \,,$$

and

$$\mathcal{A}_i = \sigma(X_i) \vee \mathcal{N} \,, i \ge 1 \,.$$

Then, $\mathcal{A}_1, \mathcal{A}_2, \ldots$ are independent σ-fields.

Fix $k \ge 2$ and $n \ge k + r$. Write

$$\begin{aligned} U_n &= \frac{1}{r!\binom{n}{r}} \sum_{\text{distinct } i_1,\ldots,i_r \in \{k,\ldots,n\}} f(X_{i_1}, \ldots, X_{i_r}) \\ &\quad + \frac{1}{r!\binom{n}{r}} \sum_{\text{distinct } i_1,\ldots,i_r \in \{1,\ldots,n\} : \{i_1,\ldots,i_r\} \cap \{1,\ldots,k-1\} \neq \emptyset} f(X_{i_1}, \ldots, X_{i_r}) \\ &=: V_n + W_n \,. \end{aligned}$$

It is immediate that V_n is $\bigvee_{i=k}^\infty \mathcal{A}_k$-measurable. For the second term, let

$$A_n = \big\{\{i_1, \ldots, i_r\} : i_1, \ldots, i_r \in \{1, \ldots, n\} \text{ are distinct and } \\ \{i_1, \ldots, i_r\} \cap \{1, \ldots, k-1\} \neq \emptyset\}\big\} \,,$$

and note that

$$\begin{aligned} \mathrm{E}|W_n| &\le \frac{1}{r!\binom{n}{r}} \# A_n \, \mathrm{E} \, |f(X_1, \ldots, X_r)| \\ &= \frac{1}{\binom{n}{r}} \left(\binom{n}{r} - \binom{n-k+1}{r} \right) \mathrm{E} \, |f(X_1, \ldots, X_r)| \\ &\to 0 \,, \, n \to \infty \,. \end{aligned}$$

Thus,

$$U_n - V_n \xrightarrow{P} 0 \,, n \to \infty \,.$$

Recalling (5.6), it follows that

$$V_n \xrightarrow{P} U_\infty \,, n \to \infty \,,$$

and hence there exists a subsequence of (V_n) which converges to U_∞ a.s. Since $\mathcal{A}_i \supset \mathcal{N}$, it follows that U_∞ is measurable with respect to

$$\bigvee_{i=k}^\infty \mathcal{A}_i \,.$$

The above being true for all k, U_∞ is measurable with respect to

$$\bigcap_{k=1}^{\infty} \bigvee_{i=k}^{\infty} \mathcal{A}_i \,.$$

By Kolmogorov's 0-1 law, it follows that U_∞ is degenerate, and hence

$$U_\infty = \mathrm{E}\left(f(X_1, \ldots, X_r)\right) \,,$$

which in view of (5.6) completes the proof. ☐

The next result of interest is the SLLN for exchangeable random variables, where the limit is not necessarily degenerate any more.

Definition. *A sequence of random variables $X_1, X_2, \ldots$ is exchangeable if for all $n \ge 1$ and every permutation π of $\{1, \ldots, n\}$,*

$$(X_1, \ldots, X_n) \stackrel{d}{=} (X_{\pi(1)}, \ldots, X_{\pi(n)}) \,.$$

Example 5.1. *Let $Z_0, Z_1, \ldots$ be i.i.d. standard Normal random variables, and $0 \le \rho < 1$. Define*

$$X_n := \sqrt{\rho} Z_0 + \sqrt{1-\rho}\, Z_n,\ n \ge 1 \,.$$

Then, for every n, $(X_1, X_2, \ldots, X_n)$ follows multivariate normal with

$$\begin{aligned} \mathrm{E}(X_i) &= 0\,, \\ \mathrm{Cov}(X_i, X_j) &= \begin{cases} \rho, & i \ne j\,, \\ 1, & i = j\,, \end{cases} \end{aligned}$$

for all $1 \le i, j \le n$. It is easy to see that the sequence $(X_1, X_2, \ldots)$ is exchangeable.

Example 5.2. *Let Θ follow standard Uniform, and conditional on Θ, let $X_1, X_2, \ldots$ be i.i.d. from Bernoulli(Θ). Then $X_1, X_2, \ldots$ are exchangeable.*

Theorem 5.5 (SLLN for exchangeable random variables)**.** *Let $X_1, X_2, \ldots$ be an exchangeable sequence of integrable random variables. Then there exists an integrable random variable Y such that*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to Y \text{ in } L^1 \text{ and a.s., as } n \to \infty \,.$$

Proof. Let

$$S_n := \sum_{i=1}^{n} X_i,\ n \ge 1 \,.$$

Define the σ-fields

$$\begin{aligned}\mathcal{F}_{mn} &:= \sigma(S_m, S_{m+1}, \ldots, S_n),\ 1 \le m \le n < \infty\,,\\ \mathcal{F}_{m\infty} &:= \sigma(S_m, S_{m+1}, \ldots),\ m \in \mathbb{N}\,,\\ \mathcal{F}_{\infty\infty} &:= \bigcap_{m=1}^{\infty} \mathcal{F}_{m\infty}\,.\end{aligned}$$

Fix $2 \le m \le n$. There exists a measurable function $f : \mathbb{R}^{n-m+1} \to \mathbb{R}$ such that

$$\mathrm{E}(X_1|\mathcal{F}_{mn}) = f(S_m, \ldots, S_n)\,.$$

For any $i \in \{1, \ldots, m\}$,

$$(X_1, X_2, \ldots, X_n) \stackrel{d}{=} (X_i, X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)\,,$$

and hence

$$(X_1, S_m, \ldots, S_n) \stackrel{d}{=} (X_i, S_m, \ldots, S_n)\,.$$

Therefore,

$$\mathrm{E}(X_i|\mathcal{F}_{mn}) = f(S_m, \ldots, S_n),\ 1 \le i \le m\,.$$

Thus,

$$mf(S_m, \ldots, S_n) = \sum_{i=1}^{m} \mathrm{E}(X_i|\mathcal{F}_{mn}) = \mathrm{E}(S_m|\mathcal{F}_{mn}) = S_m\,.$$

That is,

$$\mathrm{E}(X_1|\mathcal{F}_{mn}) = \frac{1}{m} S_m\,.$$

For a fixed m,

$$\mathcal{F}_{m\infty} = \bigvee_{n=1}^{\infty} \mathcal{F}_{mn}\,,$$

and hence

$$\mathrm{E}(X_1|\mathcal{F}_{m\infty}) = \lim_{n\to\infty} \mathrm{E}(X_1|\mathcal{F}_{mn}) = \frac{1}{m} S_m\,.$$

Since $(\mathcal{F}_{m\infty} : m \in \mathbb{N})$ is a reverse filtration, by the reverse martingale convergence theorem,

$$\frac{1}{m} S_m \to \mathrm{E}(X_1|\mathcal{F}_{\infty\infty})\,,$$

a.s. and in L^1, as $m \to \infty$. This completes the proof. □

De Finetti's theorem is a non-trivial application of the above result.

Theorem 5.6 (de Finetti's theorem)**.** *If $X_1, X_2, \ldots$ are exchangeable and each X_n takes value 0 or 1, and Θ denotes the a.s. limit of $n^{-1}(X_1+\ldots+X_n)$, then*

$$P(X_1 = u_1, \ldots, X_n = u_n | \sigma(\Theta)) = \Theta^{\sum_{i=1}^n u_i}(1-\Theta)^{n-\sum_{i=1}^n u_i},$$

for all $u_1, \ldots, u_n \in \{0,1\}$ with the convention that $0^0 = 1$.

Proof. Let S_n for $n \in \mathbb{N}$ and $\mathcal{F}_{mn}$ for $1 \le m \le n \le \infty$ be defined as in the above proof. Fix $1 \le m \le n < \infty$. Fix $v_m, \ldots, v_n$ such that

$$P(S_m = v_m, \ldots, S_n = v_n) > 0.$$

Fix $u_1, \ldots, u_m \in \{0,1\}$ such that $u_1 + \ldots + u_m = v_m$ and observe that

$$\begin{aligned} & P\left(X_1 = u_1, \ldots, X_m = u_m | S_m = v_m, \ldots, S_n = v_n\right) \\ = \; & \frac{1}{P(S_m = v_m, \ldots, S_n = v_n)} P(X_1 = u_1, \ldots, X_m = u_m, \\ & \quad X_{m+1} = v_{m+1} - v_m, \ldots, X_n = v_n - v_{n-1}). \end{aligned}$$

Clearly, there exist $1 \le i_1 < \ldots < i_{v_m} \le m$ such that $u_{i_1} = \ldots = u_{i_{v_m}} = 1$ and $u_j = 0$ for $j \in \{1, \ldots, m\} \setminus \{i_1, \ldots, i_m\}$. Therefore,

$$\begin{aligned} & P(X_1 = u_1, \ldots, X_m = u_m, X_{m+1} = v_{m+1} - v_m, \ldots, X_n = v_n - v_{n-1}) \\ = \; & P(X_j = 1 \text{ for } j \in \{i_1, \ldots, i_{v_m}\}, X_j = 0 \text{ for} \\ & \; j \in \{1, \ldots, m\} \setminus \{i_1, \ldots, i_m\}, X_j = v_j - v_{j-1} \text{ for } j = m+1, \ldots, n) \\ = \; & P(X_j = 1 \text{ for } j \in \{1, \ldots, v_m\}, X_j = 0 \text{ for} \\ & \; j \in \{v_m + 1, \ldots, m\}, X_j = v_j - v_{j-1} \text{ for } j = m+1, \ldots, n), \end{aligned}$$

the last inequality following by exchangeability. That is, for all $u_1, \ldots, u_m \in \{0,1\}$ such that $u_1 + \ldots + u_m = v_m$,

$$P\left(X_1 = u_1, \ldots, X_m = u_m | S_m = v_m, \ldots, S_n = v_n\right)$$

depends only on $v_m, \ldots, v_n$. Since there are $\binom{m}{v_m}$ many possible choices for $u_1, \ldots, u_m$, it follows that

$$P\left(X_1 = u_1, \ldots, X_m = u_m | S_m = v_m, \ldots, S_n = v_n\right) = \frac{1}{\binom{m}{v_m}},$$

for all $u_1, \ldots, u_m \in \{0,1\}$ such that $u_1 + \ldots + u_m = v_m$. In other words,

$$P\left(X_1 = u_1, \ldots, X_m = u_m | \mathcal{F}_{mn}\right) = \binom{m}{S_m}^{-1} \mathbf{1}\left(S_m = \sum_{i=1}^m u_i\right),$$

for all $1 \le m \le n < \infty$. Letting $n \to \infty$, it follows that for all $1 \le m < \infty$,

$$P\left(X_1 = u_1, \ldots, X_m = u_m | \mathcal{F}_{m\infty}\right) = \binom{m}{S_m}^{-1} \mathbf{1}\left(S_m = \sum_{i=1}^m u_i\right).$$

Fix once again $1 \le m \le n < \infty$, and $u_1, \ldots, u_m \in \{0,1\}$. Define

$$j := \sum_{i=1}^m u_i,\ k := m - j\,,$$

and note that

$$\begin{aligned}
& P\left(X_1 = u_1, \ldots, X_m = u_m | \mathcal{F}_{n\infty}\right) \\
= \quad & \frac{\binom{n-m}{S_n - j}}{\binom{n}{S_n}} \mathbf{1}(j \le S_n \le n-k) \\
= \quad & \mathbf{1}(j \le S_n \le n-k) \\
& \frac{S_n(S_n-1)\ldots(S_n-j+1)\cdot(n-S_n)\ldots(n-S_n-k+1)}{n(n-1)\ldots(n-m+1)} \\
= \quad & \mathbf{1}(j \le S_n \le n-k)\frac{S_n}{n}\ldots\frac{S_n-j+1}{n-j+1}\frac{n-S_n}{n-j}\ldots\frac{n-S_n-k+1}{n-m+1} \\
=: \quad & \mathbf{1}(j \le S_n \le n-k) Y_{mn}\,.
\end{aligned}$$

Fixing m and letting $n \to \infty$, it is easy to see by the SLLN for exchangeable sequence that

$$Y_{mn} \to \Theta^j (1-\Theta)^k\,,$$

where Θ is the a.s. limit of S_n/n. Thus, it is immediate that on $[0 < \Theta < 1]$,

$$P\left(X_1 = u_1, \ldots, X_m = u_m | \mathcal{F}_{n\infty}\right) \to \Theta^j(1-\Theta)^k, \text{ a.s., as } n \to \infty\,. \tag{5.7}$$

Now fix ω such that

$$\frac{1}{n} S_n(\omega) \to \Theta(\omega) = 0\,.$$

If $j \ge 1$, then it is easy to see that

$$P\left(X_1 = u_1, \ldots, X_m = u_m | \mathcal{F}_{n\infty}\right)(\omega) \le \frac{1}{n} S_n(\omega) \to 0\,,$$

and if $j = 0$, then

$$\begin{aligned}
& P\left(X_1 = u_1, \ldots, X_m = u_m | \mathcal{F}_{n\infty}\right)(\omega) \\
= \quad & \mathbf{1}(S_n(\omega) \le n-m)\frac{n - S_n(\omega)}{n}\ldots\frac{n-S_n(\omega)-m+1}{n-m+1} \\
\to \quad & 1, \text{ as } n \to \infty\,.
\end{aligned}$$

Thus, (5.7) holds on $[\Theta = 0]$ also, recalling the convention that $0^0 = 1$. The convergence can be established similarly on $[\Theta = 1]$.

Since,

$$P\left(X_1 = u_1, \ldots, X_m = u_m | \mathcal{F}_{n\infty}\right) \to P\left(X_1 = u_1, \ldots, X_m = u_m | \mathcal{F}_{\infty\infty}\right),$$

a.s., as $n \to \infty$, it follows from (5.7) that

$$P\left(X_1 = u_1, \ldots, X_m = u_m | \mathcal{F}_{\infty\infty}\right) = \Theta^{\sum_{i=1}^m u_i}(1-\Theta)^{m-\sum_{i=1}^m u_i}.$$

The proof follows by conditioning with respect to $\sigma(\Theta)$ which is a subset of $\mathcal{F}_{\infty\infty}$. □

The next result, which is another application of the martingale convergence theorems, is due to Kakutani.

Theorem 5.7. *Let $X_1, X_2, \ldots$ be independent non-negative random variables with mean 1. Let*

$$M_n = X_1 \ldots X_n, n \ge 1,$$

$$a_n = \mathrm{E}\left(\sqrt{X_n}\right), n \ge 1,$$

and

$$A = \prod_{n=1}^{\infty} a_n.$$

If $A > 0$, then

$$M_n \to M_\infty,$$

in L^1 and a.s. for some random variable M_∞. On the contrary, if $A = 0$, then

$$M_n \to 0 \text{ a.s.}$$

Proof. For proving the first claim, assume that $A > 0$. Define

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n), n \ge 1,$$

and

$$W_n = \prod_{i=1}^{n} \frac{\sqrt{X_i}}{a_i}, n \ge 1.$$

Then, $(W_n, \mathcal{F}_n)$ is a martingale. Further,

$$\mathrm{E}(W_n^2) = (a_1 \ldots a_n)^{-1}, n \ge 1.$$

Since $a_n \le 1$ by Jensen, it follows that

$$\sup_{n\ge1} \mathrm{E}(W_n^2) \le A^{-1}.$$

Theorem 4.12 with $p = 2$ implies that there exists a random variable W_∞ such that $W_n \to W_\infty$ in L^2 and a.s. Thus,

$$W_n^2 \to W_\infty^2 \text{ a.s. and in } L^1 .$$

Since

$$W_n^2 = (a_1 \dots a_n)^{-2} M_n ,$$

it follows that

$$M_n \to A^2 W_\infty^2 \text{ a.s. and in } L^1 .$$

Letting

$$M_\infty = A^2 W_\infty^2 ,$$

the desired claim for the case $A > 0$ follows.

Suppose that $A = 0$. If W_n is as above, then it is still a non-negative martingale and hence converges a.s. to some finite random variable W_∞. Thus,

$$M_n = (a_1 \dots a_n W_n)^2 \to A^2 W_\infty^2 = 0 \text{ a.s.}$$

This completes the proof. □

The final application we'll study is the so-called discrete Dirichlet problem. The Dirichlet problem, which arises in physics, is the following. Let D be the open unit disc in $\mathbb{R}^2$, $\bar{D}$ its closure, and

$$\partial D = \bar{D} \setminus D .$$

Given a continuous function $f : \partial D \to \mathbb{R}$, the Dirichlet problem is to find a continuous function $h : \bar{D} \to \mathbb{R}$ which is harmonic on D, that is, for all $(x, y) \in D$ and $r > 0$ such that the ball of radius r around (x, y) is in D, it holds that

$$h(x, y) = \frac{1}{2\pi} \int_0^{2\pi} h\left(x + r\cos\theta, y + r\sin\theta\right) \, d\theta ,$$

and satisfies

$$h = f \text{ on } \partial D .$$

If such an h exists, the next question is whether it is unique.

Solving the Dirichlet problem is beyond the scope of this course. We shall consider a discrete version of the above problem.

Theorem 5.8. *Let $D \subset \mathbb{Z}^2$ be a finite set. Define*

$$\bar{D} = \{y \in \mathbb{Z}^2 : \|y - x\| \leq 1 \text{ for some } x \in D\} ,$$

where $\|\cdot\|$ denotes the L^1 norm, and

$$\partial D = \bar{D} \setminus D .$$

Given any $f : \partial D \to \mathbb{R}$, there exists a unique

$$h : \bar{D} \to \mathbb{R}\,,$$

which agrees with f on ∂D, and satisfies for all $(x, y) \in D$

$$h(x,y) = \frac{1}{4}\left[h(x-1,y) + h(x+1,y) + h(x,y-1) + h(x,y+1)\right]\,. \quad (5.8)$$

Proof. We start with showing existence. Let $(S_n : n \geq 0)$ be the simple symmetric random walk on $\mathbb{Z}^2$ with $S_0 = 0$, that is, it takes a step in one of the four directions, each with probability $1/4$. Define

$$\tau_{x,y} = \min\{n \geq 0 : (x,y) + S_n \notin D\}\,, (x,y) \in \bar{D}\,,$$

the right hand side being finite a.s. because D is a finite set, and

$$h(x,y) = \mathrm{E}\left[f\left((x,y) + S_{\tau_{x,y}}\right)\right]\,, (x,y) \in \bar{D}\,. \quad (5.9)$$

The right hand side makes sense because $(x,y) + S_{\tau_{x,y}} \in \partial D$. It is easy to see that for $(x,y) \in \partial D$, $\tau_{x,y} = 0$ and hence h agrees with f on ∂D.

To show (5.8), fix $(x,y) \in D$ and define a Markov chain $(X_n : n \geq 0)$ on $\bar{D}$ as follows. Declare the states in ∂D to be absorbing, and if the chain is in $(x,y) \in D$, then it moves to one of $(x-1,y)$, $(x+1,y)$, $(x,y-1)$ and $(x,y+1)$, each with probability $1/4$. Since this is a chain on a finite state space with every non-absorbing state leading to at least one absorbing state, the chain eventually gets absorbed, that is, there exists X_∞ such that

$$X_n \to X_\infty \text{ a.s.}$$

A moment's thought reveals that (5.9) is equivalent to

$$h(x,y) = \mathrm{E}\left(f(X_\infty) | X_0 = (x,y)\right)\,, (x,y) \in \bar{D}\,. \quad (5.10)$$

Denoting $e_1 = (-1,0)$, $e_2 = (1,0)$, $e_3 = (0,-1)$ and $e_4 = (0,1)$,

$$\begin{aligned} h(x,y) &= \frac{1}{4}\sum_{i=1}^{4} \mathrm{E}\left(f(X_\infty) | X_0 = (x,y), X_1 = (x,y) + e_i\right) \\ &= \frac{1}{4}\sum_{i=1}^{4} \mathrm{E}\left(f(X_\infty) | X_1 = (x,y) + e_i\right) \\ &= \frac{1}{4}\sum_{i=1}^{4} \mathrm{E}\left(f(X_\infty) | X_0 = (x,y) + e_i\right)\,, \end{aligned}$$

the second and third lines following from the Markov properties. Using (5.10), the above line is same as the right hand side of (5.8). This proves the existence.

For uniqueness, let $h : \bar{D} \to \mathbb{R}$ agree with f on ∂D and satisfy (5.8). Fix $(x, y) \in D$ and set

$$Z_n = h\left((x, y) + S_n\right) \,, n \ge 1\,.$$

Setting $\mathcal{F}_n = \sigma(S_0, \ldots, S_n)$, (5.8) implies that $(Z_n, \mathcal{F}_n)_{n\ge 0}$ is a martingale. Since $(Z_{\tau_{x,y}\wedge n} : n \ge 0)$ is uniformly bounded, as D is a finite set, the optional sampling theorem implies that

$$\mathrm{E}\left(Z_{\tau_{x,y}}\right) = \mathrm{E}(Z_0) = h(x, y)\,.$$

Further, $\tau_{x,y} + S_{\tau_{x,y}} \in \partial D$ and hence

$$Z_{\tau_{x,y}} = f\left((x, y) + S_{\tau_{x,y}}\right)\,.$$

Thus, h is given by (5.9) which establishes uniqueness. Hence the proof follows. □

6 Martingale central limit theorem

Suppose that for each fixed $n \in \mathbb{N}$, $(S_{nk}, \mathcal{F}_{nk})_{k\ge 0}$ is a martingale, with $S_{n0} = 0$. Let

$$Y_{nk} = S_{nk} - S_{n\,k-1}\,, k \ge 1\,.$$

Assume that

$$\sum_{k=1}^{\infty} \mathrm{E}(Y_{nk}^2) < \infty\,. \tag{6.1}$$

Then, for every n,

$$S_{nk} \to S_n\,, k \to \infty\,,$$

a.s. for some random variable S_n. Define

$$\sigma_{nk} = \sqrt{\mathrm{E}\left(Y_{nk}^2|\mathcal{F}_{n\,k-1}\right)}\,, k \ge 1\,.$$

The assumption (6.1) also ensures that

$$\sum_{k=1}^{\infty} \sigma_{nk}^2 < \infty \text{ a.s.} \tag{6.2}$$

Theorem 6.1 (Martingale CLT)**.** *Assume that for all $\varepsilon > 0$,*

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} \mathrm{E}\left(Y_{nk}^2 \mathbf{1}(|Y_{nk}| \ge \varepsilon)\right) = 0\,, \tag{6.3}$$

and that

$$\sum_{k=1}^{\infty} \sigma_{nk}^2 \xrightarrow{P} \sigma^2\,, n \to \infty\,, \tag{6.4}$$

where $\sigma > 0$ is a constant. Then,

$$S_n \Rightarrow Z\,, n \to \infty\,,$$

where Z follows $N(0, \sigma^2)$.

Proof. We first prove it under the additional assumption, which will later be removed, that

$$\sum_{k=1}^{\infty} \sigma_{nk}^2 \le C\,, n \ge 1\,, \tag{6.5}$$

for some constant C.

Denote

$$\Sigma_{nk} = \sum_{j=1}^{k} \sigma_{nj}^2\,, n \in \mathbb{N}\,, 1 \le k \le \infty\,.$$

In order to prove the claim, it suffices to show that

$$\lim_{n\to\infty} \mathrm{E}\left(e^{itS_n}\right) = e^{-t^2\sigma^2/2}\,, t \in \mathbb{R}\,, \tag{6.6}$$

where $i = \sqrt{-1}$. Fix t and observe that

$$\begin{aligned}
&\left|\mathrm{E}\left(e^{itS_n}\right) - e^{-t^2\sigma^2/2}\right| \\
&= \left|\mathrm{E}\left(e^{itS_n} - e^{itS_n + (\Sigma_{n\infty} - \sigma^2)t^2/2}\right) + \mathrm{E}\left(e^{itS_n + (\Sigma_{n\infty} - \sigma^2)t^2/2} - e^{-t^2\sigma^2/2}\right)\right| \\
&\le \mathrm{E}\left|e^{itS_n}\left(1 - e^{(\Sigma_{n\infty} - \sigma^2)t^2/2}\right)\right| + e^{-t^2\sigma^2/2}\left|\mathrm{E}\left(e^{itS_n + \Sigma_{n\infty}t^2/2}\right) - 1\right| \\
&\le \mathrm{E}\left|1 - e^{(\Sigma_{n\infty} - \sigma^2)t^2/2}\right| + \left|\mathrm{E}\left(e^{itS_n + \Sigma_{n\infty}t^2/2}\right) - 1\right|\,.
\end{aligned}$$

The hypothesis (6.4) implies that

$$\left|1 - e^{(\Sigma_{n\infty} - \sigma^2)t^2/2}\right| \xrightarrow{P} 0\,,$$

and (6.5) implies that the left hand side is at most $1 + e^{Ct^2}$. Therefore,

$$\lim_{n\to\infty} \mathrm{E}\left|1 - e^{(\Sigma_{n\infty} - \sigma^2)t^2/2}\right| = 0\,,$$

and (6.6) would follows once it is shown that

$$\lim_{n\to\infty} \mathrm{E}\left(e^{itS_n + \Sigma_{n\infty}t^2/2}\right) = 1\,. \tag{6.7}$$

Proceeding towards the above, fix n and note that

$$\lim_{k\to\infty} \mathrm{E}\left(e^{itS_{nk} + \Sigma_{nk}t^2/2}\right) = \mathrm{E}\left(e^{itS_n + \Sigma_{n\infty}t^2/2}\right)\,,$$

(6.5) being used once again, and hence

$$
\begin{aligned}
&\mathrm{E}\left(e^{itS_n+\Sigma_{n\infty}t^2/2}-1\right)\\
&=\sum_{k=1}^{\infty}\mathrm{E}\left(e^{itS_{nk}+\Sigma_{nk}t^2/2}-e^{itS_{n\,k-1}+\Sigma_{n\,k-1}t^2/2}\right)\\
&=\sum_{k=1}^{\infty}\mathrm{E}\left[e^{itS_{n\,k-1}}e^{\Sigma_{nk}t^2/2}\left(e^{itY_{nk}}-e^{-\sigma_{nk}^2t^2/2}\right)\right]\\
&=\sum_{k=1}^{\infty}\mathrm{E}\left[e^{itS_{n\,k-1}}e^{\Sigma_{nk}t^2/2}\mathrm{E}\left(e^{itY_{nk}}-e^{-\sigma_{nk}^2t^2/2}\big|\mathcal{F}_{n\,k-1}\right)\right]\,.
\end{aligned}
$$

Hence,

$$
\left|\mathrm{E}\left(e^{itS_n+\Sigma_{n\infty}t^2/2}-1\right)\right|\le e^{Ct^2}\sum_{k=1}^{\infty}\mathrm{E}\left|\mathrm{E}\left(e^{itY_{nk}}-e^{-\sigma_{nk}^2t^2/2}\big|\mathcal{F}_{n\,k-1}\right)\right|\,. \tag{6.8}
$$

Fix $k\in\mathbb{N}$. Taylor's theorem implies that

$$
e^{itY_{nk}}=1+itY_{nk}-\frac{1}{2}t^2Y_{nk}^2+\theta\,,
$$

where

$$
|\theta|\le|tY_{nk}|^3\wedge(tY_{nk})^2\,.
$$

Fix $\varepsilon>0$ and denote

$$
\mathbf{1}_{nk}=\mathbf{1}(|Y_{nk}|\ge\varepsilon)\,.
$$

Thus, there exists a constant K_t, depending only on t, such that

$$
\begin{aligned}
|\theta|&\le K_t\left(|Y_{nk}|^3\wedge Y_{nk}^2\right)\\
&\le K_t\left(Y_{nk}^2\mathbf{1}_{nk}+|Y_{nk}|^3\mathbf{1}(|Y_{nk}|<\varepsilon)\right)\\
&\le K_t\left(Y_{nk}^2\mathbf{1}_{nk}+\varepsilon Y_{nk}^2\right)\,.
\end{aligned}
$$

Taylor's theorem also implies that

$$
e^{-\sigma_{nk}^2t^2/2}=1-\frac{1}{2}t^2\sigma_{nk}^2+\theta'\,,
$$

where

$$
|\theta'|\le\bar{K}_t\sigma_{nk}^4\,,
$$

where $\bar{K}_t\ge K_t$.

Recalling $\mathrm{E}(Y_{nk}|\mathcal{F}_{n\,k-1}) = 0$ and $\mathrm{E}(Y_{nk}^2|\mathcal{F}_{n\,k-1}) = \sigma_{nk}^2$,

$$\begin{aligned}
&\left|\mathrm{E}\left(e^{itY_{nk}} - e^{-\sigma_{nk}^2 t^2/2}\middle|\mathcal{F}_{n\,k-1}\right)\right| \\
&= \left|\mathrm{E}\left(e^{itY_{nk}} - 1 - itY_{nk} + \frac{1}{2}t^2 Y_{nk}^2\middle|\mathcal{F}_{n\,k-1}\right)\right. \\
&\quad \left. - \mathrm{E}\left(e^{-\sigma_{nk}^2 t^2/2} - 1 + \frac{1}{2}t^2\sigma_{nk}^2\middle|\mathcal{F}_{n\,k-1}\right)\right| \\
&\leq \mathrm{E}\left(|\theta| + |\theta'|\middle|\mathcal{F}_{n\,k-1}\right) \\
&\leq \bar{K}_t \mathrm{E}\left(Y_{nk}^2 \mathbf{1}_{nk} + \varepsilon Y_{nk}^2 + \sigma_{nk}^4\middle|\mathcal{F}_{n\,k-1}\right) \\
&= \bar{K}_t\left[\mathrm{E}(Y_{nk}^2\mathbf{1}_{nk}|\mathcal{F}_{n\,k-1}) + \varepsilon\sigma_{nk}^2 + \sigma_{nk}^4\right] .
\end{aligned}$$

Going back to (6.8), we get

$$\begin{aligned}
&\left|\mathrm{E}\left(e^{itS_n + \Sigma_{n\infty}t^2/2} - 1\right)\right| \\
&\leq \bar{K}_t e^{Ct^2}\sum_{k=1}^{\infty}\mathrm{E}\left(Y_{nk}^2\mathbf{1}_{nk} + \varepsilon\sigma_{nk}^2 + \sigma_{nk}^4\right) \\
&\leq \bar{K}_t e^{Ct^2}\mathrm{E}\left(\left(\varepsilon + \sup_{k\geq 1}\sigma_{nk}^2\right)\Sigma_{n\infty} + \sum_{k=1}^{\infty} Y_{nk}^2\mathbf{1}_{nk}\right) \\
&\leq \bar{K}_t e^{Ct^2}\left[C\left(\varepsilon + \mathrm{E}\left(\sup_{k\geq 1}\sigma_{nk}^2\right)\right) + \sum_{k=1}^{\infty}\mathrm{E}(Y_{nk}^2\mathbf{1}_{nk})\right] \\
&\leq \bar{K}_t e^{Ct^2}\left[C(\varepsilon + \varepsilon^2) + (1+C)\sum_{k=1}^{\infty}\mathrm{E}(Y_{nk}^2\mathbf{1}_{nk})\right] , \qquad (6.9)
\end{aligned}$$

the last line following from the observation

$$\sigma_{nk}^2 = \mathrm{E}(Y_{nk}^2|\mathcal{F}_{n\,k-1}) \leq \varepsilon^2 + \mathrm{E}(Y_{nk}^2\mathbf{1}_{nk}|\mathcal{F}_{n\,k-1}) \leq \varepsilon^2 + \sum_{j=1}^{\infty}\mathrm{E}(Y_{nj}^2\mathbf{1}_{nj}|\mathcal{F}_{n\,j-1}) ,$$

and hence

$$\mathrm{E}\left(\sup_{k\geq 1}\sigma_{nk}^2\right) \leq \varepsilon^2 + \sum_{k=1}^{\infty}\mathrm{E}(Y_{nk}^2\mathbf{1}_{nk}) .$$

Letting $n \to \infty$ in (6.9) and using (6.3), it follows that

$$\limsup_{n\to\infty}\left|\mathrm{E}\left(e^{itS_n + \Sigma_{n\infty}t^2/2} - 1\right)\right| \leq \bar{K}_t e^{Ct^2}(\varepsilon + \varepsilon^2)C .$$

Arbitrariness of ε shows that the left hand side of (6.8) goes to zero as $n \to \infty$, that is, (6.7) holds. In turn, (6.7) implies (6.6) which completes the proof under the assumption (6.5).

To complete the proof when (6.5) possibly fails, fix $C > \sigma^2$ and let for all $1 \le n < \infty$ and $1 \le k \le \infty$,

$$A_{nk} = \left[\sum_{j=1}^{k} \sigma_{nj}^2 \le C \right] ,$$

and

$$Z_{nk} = Y_{nk} \mathbf{1}_{A_{nk}} .$$

Since $\mathbf{1}_{A_{nk}} \in \mathcal{F}_{n\,k-1}$, it follows that

$$\mathrm{E}\left(Z_{nk} | \mathcal{F}_{n\,k-1}\right) = 0 \, , k, n \in \mathbb{N} .$$

Let

$$\theta_{nk}^2 = \mathrm{E}\left(Z_{nk}^2 | \mathcal{F}_{n\,k-1}\right) = \sigma_{nk}^2 \mathbf{1}_{A_{nk}} .$$

Observe that for n fixed, $A_{nk} \downarrow A_{n\infty}$ as $k \to \infty$, and hence

$$\Omega = A_{n\infty} \cup \left(\bigcup_{j=1}^{\infty} (A_{n\,j-1} \setminus A_{nj}) \right) ,$$

where $A_{n0} = \Omega$. Thus, for an $\omega \in \Omega$, either $\omega \in A_{n\infty}$ or $\omega \in A_{n\,j-1} \setminus A_{nj}$ for some $j \in \mathbb{N}$. Further,

$$\sum_{j=1}^{\infty} \theta_{nj}^2(\omega) = \begin{cases} \sum_{j=1}^{k} \sigma_{nj}^2(\omega) \, , & \omega \in A_{nk} \setminus A_{n\,k+1} \text{ for some } k \ge 0 \, , \\ \sum_{j=1}^{\infty} \sigma_{nj}^2(\omega) \, , & \omega \in A_{n\infty} \, . \end{cases}$$

In all the cases, it holds that $\sum_{j=1}^{\infty} \theta_{nj}^2(\omega) \le C$. Besides,

$$\mathbf{1}_{A_{n\infty}} \sum_{j=1}^{\infty} \theta_{nj}^2 = \mathbf{1}_{A_{n\infty}} \sum_{j=1}^{\infty} \sigma_{nj}^2 ,$$

and that $C > \sigma^2$ along with (6.4) implies that $\mathbf{1}_{A_{n\infty}} \xrightarrow{P} 1$ as $n \to \infty$. Therefore,

$$\sum_{j=1}^{\infty} \theta_{nj}^2 \xrightarrow{P} \sigma^2 \, , n \to \infty .$$

By the case that has already been proved,

$$\sum_{k=1}^{\infty} Z_{nk} \Rightarrow Z ,$$

where $Z \sim N(0, \sigma^2)$. Finally,

$$S_n = S_n \mathbf{1}_{A_{n\infty}} + S_n \mathbf{1}_{A_{n\infty}^c} = \mathbf{1}_{A_{n\infty}} \sum_{k=1}^{\infty} Z_{nk} + S_n \mathbf{1}_{A_{n\infty}^c} \Rightarrow Z .$$

This completes the proof. $\square$

It is immediate that the Lévy-Lindeberg central limit theorem, which is the following, is a corollary of the martingale CLT.

Theorem 6.2. *Suppose that for each n, $k_n \in \mathbb{N} \cup \{\infty\}$, and the collection $(Y_{ni} : i \in \mathbb{N} \cap [1, k_n])$ comprises independent zero mean random variables. If it holds that*

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} \mathrm{E}\left(Y_{ni}^2 \mathbf{1}(|Y_{ni}| > \varepsilon\right) = 0\,,\ \textit{for all}\ \varepsilon > 0\,,$$

and

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} \mathrm{E}(Y_{ni}^2) = \sigma^2 > 0\,,$$

then

$$\sum_{i=1}^{k_n} Y_{ni} \Rightarrow Z\,, n \to \infty\,,$$

where $Z \sim N(0, \sigma^2)$.

The following is a special case of above and gives an easy way to check the Lindberg condition based on the inequality

$$\mathrm{E}(|Y|^{2+\delta}) \geq \varepsilon^\delta \mathrm{E}\left(Y^2 \mathbf{1}(|Y| > \varepsilon)\right)\,, \varepsilon, \delta > 0\,.$$

Theorem 6.3. *Suppose that for each n, $k_n \in \mathbb{N} \cup \{\infty\}$, and the collection $(Y_{ni} : i \in \mathbb{N} \cap [1, k_n])$ comprises independent zero mean random variables. If for some $\delta > 0$ it holds that*

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} \mathrm{E}\left(|Y_{ni}|^{2+\delta}\right) = 0\,,$$

and

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} \mathrm{E}(Y_{ni}^2) = \sigma^2 > 0\,,$$

then

$$\sum_{i=1}^{k_n} Y_{ni} \Rightarrow Z\,, n \to \infty\,,$$

where $Z \sim N(0, \sigma^2)$.

Theorem 6.1 has several applications. The first one is a CLT for Markov chains. However, before, that let us quickly state and prove the SLLN for Markov chains, which doesn't need martingales, though.

Theorem 6.4. *If $(X_n : n \geq 0)$ is an irreducible Markov chain on a finite state space $\mathcal{S}$ with stationary distribution π, then for any $f : \mathcal{S} \to \mathbb{R}$ it holds that*

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \sum_{j\in\mathcal{S}} \pi_j f(j) \text{ a.s., as } n \to \infty\,.$$

Proof. Since cardinality of $\mathcal{S}$ is finite,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} f(X_i) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\mathcal{S}} f(j)\mathbf{1}(X_i = j)\\ &= \sum_{j\in\mathcal{S}} f(j)\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(X_i = j)\,.\end{aligned}$$

Recalling that for every $j \in \mathcal{S}$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(X_i = j) \to \pi_j \text{ a.s.}\,,$$

as $n \to \infty$, because the chain is irreducible and positive recurrent, the proof follows. □

For the general CLT, one should center by $\sum_j \pi_j f(j)$. Instead, in the following we consider the special case where the quantity vanishes.

Theorem 6.5. *Suppose that $(X_n : n \geq 0)$ is an irreducible Markov chain with state space $\mathcal{S}$ and stationary distribution π. If $f : \mathcal{S} \to \mathbb{R}$ is such that*

$$\sum_{i\in\mathcal{S}} \pi_i f(i) = 0\,,$$

then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f(X_i) \Rightarrow Z\,, n \to \infty\,,$$

where $Z \sim N(0, \sigma^2)$ for some $\sigma^2 \geq 0$.

Proof. Without loss of generality, let $\mathcal{S} = \{1, \ldots, k\}$ for some $k \geq 2$. Writing $f_i = f(i)$, and defining $1 \times k$ vectors,

$$f = [f_1, \ldots, f_k]\;,$$

and

$$\pi = [\pi_1, \ldots, \pi_k]\,,$$

the hypothesis implies that

$$\pi' f = 0\,.$$

Recall that $\{c\pi : c \in \mathbb{R}\}$ is the eigenspace of 1 as a left eigenvalue of P, that is, the null space of $(I - P')$. Therefore, f belongs to the orthogonal complement of $(I - P')$, which is same as the row space of $(I - P')$, that is, there exists a $k \times 1$ vector u with

$$u'(I - P') = f\,.$$

In other words,

$$f' = u - Pu\,.$$

Thinking of row and column vectors in $\mathbb{R}^k$ as functions from $\{1, \ldots, k\}$ to $\mathbb{R}$, set

$$M_n = \sum_{i=1}^n \left[u(X_i) - (Pu)(X_{i-1})\right], n \ge 1\,.$$

Thus,

$$\begin{aligned} M_n - \sum_{i=1}^n f(X_i) &= \sum_{i=1}^n \left[u(X_i) - (Pu)(X_{i-1})\right] - \sum_{i=1}^n \left[u(X_i) - (Pu)(X_i)\right] \\ &= (Pu)(X_n) - (Pu)(X_0)\,. \end{aligned}$$

Therefore,

$$n^{-1/2}\left[M_n - \sum_{i=1}^n f(X_i)\right] \xrightarrow{P} 0\,, n \to \infty\,,$$

and hence it suffices to show that

$$n^{-1/2} M_n \Rightarrow Z\,. \tag{6.10}$$

Proceeding towards proving the above, let $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ for $n \ge 0$, and observe that

$$\begin{aligned} \mathrm{E}\left(M_{n+1} | \mathcal{F}_n\right) &= M_n - (Pu)(X_n) + \mathrm{E}\left(u(X_{n+1}) | \mathcal{F}_n\right) \\ \text{(Markov property)} &= M_n - (Pu)(X_n) + \mathrm{E}\left(u(X_{n+1}) | X_n\right) \\ &= M_n - (Pu)(X_n) + \sum_{j=1}^k P_{X_n j} u(j) \\ &= M_n\,. \end{aligned}$$

Thus, $(M_n, \mathcal{F}_n)$ is a martingale. Let

$$Y_{ni} = \begin{cases} \frac{1}{\sqrt{n}}\left(M_i - M_{i-1}\right), & 1 \le i \le n\,, \\ 0\,, & i > n\,. \end{cases}$$

Clearly, for every n, $(Y_{ni}, \mathcal{F}_i)_{i \ge 1}$ is a martingale difference. Furthermore,

$$Y_{ni} = n^{-1/2}\left[u(X_i) - (Pu)(X_{i-1})\right], 1 \le i \le n\,,$$

and hence for all $\varepsilon > 0$ and n large, $\mathbf{1}(|Y_{ni}| \geq \varepsilon) = 0$ for all i. Finally,

$$\begin{aligned}\sum_{i=1}^{\infty} \mathrm{E}\left(Y_{ni}^2 | \mathcal{F}_{i-1}\right) &= \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left[(u(X_i) - u(X_{i-1}))^2 \big| \mathcal{F}_{i-1}\right] \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} P_{X_{i-1}j} (u(j) - u(X_{i-1}))^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} g(X_{i-1}) \,,\end{aligned}$$

where

$$g(i) = \sum_{j=1}^{k} P_{ij} \left(u(j) - u(i)\right)^2 \,, 1 \leq i \leq k \,.$$

Theorem 6.4 implies that as $n \to \infty$, a.s.

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g(X_{i-1}) = \sum_{i=1}^{k} \pi_i g(i) = \sum_{i=1}^{k} \sum_{j=1}^{k} \pi_i P_{ij} \left(u(j) - u(i)\right)^2 \,.$$

Calling the extreme right hand side σ^2, (6.10) follows, which completes the proof. □

7 Miscellaneous topics

As the name suggests, this chapter covers a few short topics which are not directly related to each other.

7.1 Martingale decomposition theorems

Theorem 7.1 (Doob's decomposition theorem)**.** *Let $((X_n, \mathcal{F}_n) : n \geq 1)$ be a submartingale. Then, X_n can be decomposed as*

$$X_n = Y_n + A_n \,,$$

where

1. *$(Y_n, \mathcal{F}_n)$ is a martingale,*
2. *$A_{n+1} \geq A_n$ a.s. for all $n \geq 1$,*
3. *$A_1 = 0$ a.s.*
4. *and for all $n \geq 2$, A_n is $\mathcal{F}_{n-1}$-measurable.*

The components Y_n and A_n are uniquely determined.

Proof. Let $A_1 = 0$, and for $n \geq 2$, define

$$A_n = A_{n-1} + \mathrm{E}\left(X_n - X_{n-1} | \mathcal{F}_{n-1}\right) ,$$

and

$$Y_n = X_n - A_n .$$

Clearly, (3) and (4) hold, and hence, for $n \geq 2$,

$$\begin{aligned}
\mathrm{E}(Y_n|\mathcal{F}_{n-1}) &= \mathrm{E}(X_n|\mathcal{F}_{n-1}) - A_n \\
&= \mathrm{E}(X_n|\mathcal{F}_{n-1}) - A_{n-1} - \mathrm{E}\left(X_n - X_{n-1}|\mathcal{F}_{n-1}\right) \\
&= X_{n-1} - A_{n-1} \\
&= Y_{n-1} ,
\end{aligned}$$

showing that (1) holds. Next observe that

$$A_n - A_{n-1} = \mathrm{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} \geq 0 \text{ a.s.} ,$$

showing that (2) holds.

Now suppose that $X_n = Z_n + B_n$ is a decomposition such that (1) – (4) hold with A_n and Y_n replaced by B_n and Z_n respectively. Therefore,

$$Y_n - Z_n = B_n - A_n,\ n \geq 1 .$$

From (4), it follows that for $n \geq 2$,

$$\begin{aligned}
B_n - A_n &= \mathrm{E}(B_n - A_n|\mathcal{F}_{n-1}) \\
&= \mathrm{E}(Y_n - Z_n|\mathcal{F}_{n-1}) \\
\text{(by (1))} \quad &= Y_{n-1} - Z_{n-1} \\
&= B_{n-1} - A_{n-1} .
\end{aligned}$$

Using (3), it follows inductively that $Y_n = Z_n$ a.s. for all n, from which the uniqueness follows. □

7.2 Regular conditional distribution

This topic should have been covered right after conditional expectation. However, it was decided to get to this later only if time permitted, which is why it's being done now.

Let X be a random variable defined on $(\Omega, \mathcal{A}, P)$, and $\mathcal{F}$ be a sub σ-field of $\mathcal{A}$.

Definition. *A function* $\mu : \mathcal{B} \times \Omega \to [0, 1]$ *satisfying*

- *for all fixed* $\omega \in \Omega$*,* $\mu(\cdot, \omega)$ *is a probability measure on* $(\mathbb{R}, \mathcal{B})$*,*
- *and for all* $A \in \mathcal{B}$*,* $\mu(A, \cdot)$ *is a version of* $P(X \in A|\mathcal{F})$*,*

is a ***regular conditional distribution*** *(RCD) of X given $\mathcal{F}$.*

Theorem 7.2. *An RCD of X given $\mathcal{F}$ exists.*

The following elementary exercise will be used in the proof.

Exercise 7.1. *Suppose that $F : \mathbb{Q} \to [0,1]$ satisfies*

$$F(r) \le F(s)\,, r, s \in \mathbb{Q}\,, r \le s\,,$$

$$\lim_{n\to-\infty} F(n) = 0\,, \text{ and } \lim_{n\to\infty} F(n) = 1\,.$$

Define

$$G(x) = \inf\left\{F(r) : r > x, r \in \mathbb{Q}\right\}\,, x \in \mathbb{R}\,.$$

1. *Show that G is a C.D.F.*

2. *If in addition, F satisfies*

$$\lim_{n\to\infty} F\left(r + n^{-1}\right) = F(r)\,, r \in \mathbb{Q}\,,$$

show that $F(r) = G(r)$ for all $r \in \mathbb{Q}$, that is, G is the unique C.D.F. whose restriction to $\mathbb{Q}$ is F.

Proof of Theorem 7.2. Let $G : \mathbb{Q} \times \Omega \to \mathbb{R}$ be a function such that for all $r \in \mathbb{Q}$, $G(r, \cdot)$ is a version of $P(X \le r|\mathcal{F})$. It holds that

$$\begin{aligned} G(r) &\le G(s) \text{ for all } r \le s\,, && (7.1)\\ 0 \le G(r) &\le 1\,, && (7.2)\\ \lim_{n\to\infty} G(r + n^{-1}) &= G(r)\,, && (7.3)\\ \lim_{n\to-\infty} G(n) &= 0\,, && (7.4)\\ \lim_{n\to\infty} G(n) &= 1\,, && (7.5) \end{aligned}$$

for all $r, s \in \mathbb{Q}$ a.s.

Let Ω_0 be the set on which all the above hold. Define for all $r \in \mathbb{Q}$ and $\omega \in \Omega$,

$$F(r,\omega) = \begin{cases} G(r,\omega)\,, & \omega \in \Omega_0\,,\\ \mathbf{1}(r \ge 0)\,, & \omega \in \Omega_0^c\,. \end{cases}$$

Thus, (7.1)–(7.5) hold for every $\omega \in \Omega$ with G replaced by F.

Furthermore, $\Omega_0 \in \mathcal{F}$ and hence,

$$F(r, \cdot) = P(X \le r|\mathcal{F})\,, r \in \mathbb{Q}\,.$$

Define

$$F(x,\omega) := \inf\{F(r,\omega) : r > x, r \in \mathbb{Q}\}, x \in \mathbb{Q}^c,\, \omega \in \Omega\,.$$

The previous exercise along with (7.1)–(7.5) shows that for all $\omega \in \Omega$, $F(\cdot, \omega)$ is a valid C.D.F.

If $x \in \mathbb{Q}^c$, and $r_n \in \mathbb{Q}$ is such that $r_n \downarrow x$, then the above shows

$$F(r_n, \omega) \to F(x, \omega)\,, \omega \in \Omega\,.$$

Since $F(r_n, \cdot)$ is a version of $P(X \le r_n | \mathcal{F})$, it follows that

$$F(x, \cdot) = P(X \le x | \mathcal{F})\,. \tag{7.6}$$

For fixed $\omega \in \Omega$, let $\mu(\cdot, \omega)$ be the probability measure on $(\mathbb{R}, \mathcal{B})$ such that

$$\mu((a, b], \omega) = F(b, \omega) - F(a, \omega),\ -\infty < a < b < \infty\,.$$

It is immediate from (7.6) that

$$\mu((a, b], \cdot) = P(a < X \le b | \mathcal{F})\,.$$

Standard good set argument along with the monotone class theorem shows that

$$\mu(B, \cdot) = P(X \in B | \mathcal{F})\,, B \in \mathcal{B}\,,$$

which completes the proof. □

Theorem 7.3. *Let μ be an RCD of X given $\mathcal{F}$. For any measurable function $f : \mathbb{R} \to \mathbb{R}$ such that $f(X)$ is integrable,*

$$\left(\mathrm{E}\left(f(X)|\mathcal{F}\right)\right)(\omega) = \int_{-\infty}^{\infty} f(x)\mu(dx, \omega) \text{ for all } P\text{-a.s. } \omega\,.$$

Proof. Follows by first doing it for simple functions f and then taking limits. □

7.3 Azuma-Hoeffding inequality

The following inequality due to Azuma and Hoeffding gives a concentration bound for a martingale with bounded differences around its mean.

Theorem 7.4 (Azuma-Hoeffding)**.** *Suppose that $(X_n, \mathcal{F}_n)_{n\ge 0}$ is a martingale with $X_0 = 0$. Assume that*

$$|X_n - X_{n-1}| \le \sigma_n,\ n \ge 1\,,$$

for some $\mathcal{F}_0$-measurable strictly positive random variable σ_n. Then for all $t \ge 0$ and $n \ge 1$,

$$P(X_n \ge t | \mathcal{F}_0) \le \exp\left(\frac{-t^2}{2\sum_{k=1}^n \sigma_k^2}\right)\,.$$

Let $Y_1, Y_2, \ldots$ be i.i.d. taking values ± 1 each with probability $1/2$. Let

$$X_n = Y_1 + \ldots + Y_n .$$

For $t > 0$, Chebyshev implies

$$P(X_n > t\sqrt{n}) \le \frac{1}{t^2} .$$

Azuma-Hoeffding gives

$$P(X_n > t\sqrt{n}) \le e^{-t^2/2} .$$

CLT tells us

$$\lim_{n\to\infty} P(X_n > t\sqrt{n}) = 1 - \Phi(t) .$$

Note that

$$1 - \Phi(t) < e^{-t^2/2} < \frac{1}{t^2} \text{ for all } t > 0 .$$

Furthermore, as $t \to \infty$, t^{-2} is much larger than $e^{-t^2/2}$, even on the logarithmic scale, that is,

$$\lim_{t\to\infty} \left| \frac{\log e^{-t^2/2}}{\log t^{-2}} \right| = \infty .$$

However,

$$\lim_{t\to\infty} \frac{\log(1 - \Phi(t))}{-t^2/2} = 1 ,$$

see page 175 of Feller Vol. 1. Furthermore, for a fixed $t > 0$ and $\alpha > 1/2$, the Chebyshev and Azuma-Hoeffding upper bounds on $P(S_n > tn^\alpha)$ are

$$t^{-2}n^{1-2\alpha}, \text{ and } \exp\left(-\frac{1}{2}t^2 n^{2\alpha-1}\right) ,$$

respectively. Clearly, the Azuma-Hoeffding bound decays much faster than the Chebyshev one as $n \to \infty$.

The following elementary lemma will be used in the proof of Theorem 7.4.

Lemma 7.1. *For any probability measure μ on $[0,1]$ such that*

$$\int_0^1 x\mu(dx) = \frac{1}{2} ,$$

and a convex $\phi : [0,1] \to \mathbb{R}$, it holds that

$$\int_0^1 \phi(x)\mu(dx) \le \frac{\phi(0) + \phi(1)}{2} .$$

Proof. Let Y and Z be independent random variables with distribution μ and Uniform$(0,1)$ respectively. Note that

$$\phi(Y) = \phi\left(\mathrm{E}(\mathbf{1}(Z \le Y)|Y)\right) \le \mathrm{E}\left(\phi\left(\mathbf{1}(Z \le Y)\right)|Y\right) ,$$

the last inequality following from Jensen. Taking unconditional expectation,

$$\begin{aligned} \int_0^1 \phi(x)\mu(dx) &= \mathrm{E}\phi(Y) \\ &\le \mathrm{E}\phi\left(\mathbf{1}(Z \le Y)\right) \\ &= \frac{\phi(0) + \phi(1)}{2} , \end{aligned}$$

the equality following from the observation that $\mathbf{1}(Z \le Y)$ follows Bernoulli$(1/2)$, because

$$P(Z \le Y) = \mathrm{E}P(Z \le Y|Y) = \mathrm{E}(Y) = \frac{1}{2} .$$

Hence the proof follows. □

The following exercise, which is a generalization of Theorem 7.3, will also be used for proving Theorem 7.4.

Exercise 7.2. *Suppose that X is a random variable defined on $(\Omega, \mathcal{A}, P)$ and $\mathcal{F}$ is a sub-σ-field of $\mathcal{A}$. If μ is the RCD of X given $\mathcal{F}$, Y is an $\mathcal{F}$-measurable random variable and $\phi : \mathbb{R}^2 \to \mathbb{R}$ is bounded measurable, show that*

$$\left(\mathrm{E}\left[\phi(X, Y)|\mathcal{F}\right]\right)(\omega) = \int_{\mathbb{R}} \phi(x, Y(\omega))\mu(dx, \omega) \text{ for almost all } \omega .$$

Proof of Theorem 7.4. Note that for all $t, \theta > 0$,

$$P(X_n \ge t|\mathcal{F}_0) = P(e^{\theta X_n} \ge e^{\theta t}|\mathcal{F}_0) \le e^{-\theta t}\mathrm{E}\left[e^{\theta X_n}|\mathcal{F}_0\right] . \tag{7.7}$$

The following two steps will complete the proof.
Step 1. For all $n \ge 1$ and $\theta > 0$,

$$\mathrm{E}\left[e^{\theta X_n}|\mathcal{F}_0\right] \le \exp\left(\frac{1}{2}\theta^2 \sum_{j=1}^n \sigma_j^2\right) .$$

Step 2. For all $n \ge 1$ and $t > 0$,

$$\min_{\theta > 0} \exp\left(-\theta t + \frac{\theta^2}{2}\sum_{j=1}^n \sigma_j^2\right) = \exp\left(-\frac{t^2}{2\sum_{j=1}^n \sigma_j^2}\right) .$$

Proof of Step 1. The claim will be proved by induction on n. For $n = 1$, need to upper bound $\mathrm{E}(e^{\theta X_1})$. This is achieved by the following claim.

Recalling that $|X_1| \le \sigma_1$, define

$$Y = \frac{X_1 + \sigma_1}{2\sigma_1}\,,$$

and $\phi : \mathbb{R}^2 \to \mathbb{R}$ by

$$\phi(y,z) = e^{\theta(2yz - z)}\mathbf{1}(y \in [0,1])\,, (y,z) \in \mathbb{R}^2\,.$$

Denoting by μ the RCD of Y given $\mathcal{F}_0$, for all $\omega \in \Omega$ it holds that

$$\begin{aligned}\left(\mathrm{E}(e^{\theta X_1}\big|\mathcal{F}_0)\right)(\omega) &= \left(\mathrm{E}\left(\phi(Y,\sigma_1)\big|\mathcal{F}_0\right)\right)(\omega) \\ &= \int_0^1 \phi(y, \sigma_1(\omega))\mu(dy,\omega)\,,\end{aligned}$$

the preceding exercise and the fact that σ_1 is $\mathcal{F}_0$-measurable implying the second line.

Lemma 7.1 and the observations that ϕ is convex in the first variable on $[0,1]$ and

$$\int_0^1 y\mu(dy,\omega) = \mathrm{E}(Y|\mathcal{F}) = \frac{1}{2}\,,$$

which is a consequence of σ_1 being $\mathcal{F}_0$-measurable and $\mathrm{E}(X_1|\mathcal{F}_0) = 0$, imply that for every $\omega \in \Omega$,

$$\begin{aligned}\int_0^1 \phi(y,\sigma_1(\omega))\mu(dy,\omega) &\le \frac{1}{2}\left[\phi(0,\sigma_1(\omega)) + \phi(1,\sigma_1(\omega))\right] \\ &= \frac{1}{2}\left(e^{-\theta\sigma_1(\omega)} + e^{\theta\sigma_1(\omega)}\right)\,.\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{E}(e^{\theta X_1}\big|\mathcal{F}_0) &\le \frac{1}{2}\left(e^{-\theta\sigma_1} + e^{\theta\sigma_1}\right) && (7.8)\\ &= \sum_{n=0}^{\infty}\frac{(\theta\sigma_1)^{2n}}{(2n)!} && (7.9)\\ &\le \sum_{n=0}^{\infty}\frac{(\theta\sigma_1)^{2n}}{2^n n!} && (7.10)\\ &= e^{\theta^2\sigma_1^2/2}\,, && (7.11)\end{aligned}$$

the inequality in (7.10) following from the observation that

$$\frac{(2n)!}{2^n n!} = \#\big\{\text{Pairings of } \{1,\ldots,2n\}\big\} \ge 1\,.$$

Thus, the claim of Step 1 for $n = 1$ follows. Assume the claim to hold for $n-1$ for some $n \geq 2$. Write

$$\mathrm{E}\left[e^{\theta X_n}|\mathcal{F}_0\right] = \mathrm{E}\left[e^{\theta X_{n-1}}\mathrm{E}\left[e^{\theta(X_n - X_{n-1})}|\mathcal{F}_{n-1}\right]|\mathcal{F}_0\right].$$

Define

$$Z = \frac{1}{2\sigma_n}\left(X_n - X_{n-1} + \sigma_n\right),$$

and $\psi : [0,1] \times \mathbb{R} \to \mathbb{R}$ by

$$\psi(z, v) = e^{\theta(2vz - v)}.$$

The hypothesis $|X_n - X_{n-1}| \leq \sigma_n$ implies that $0 \leq Z \leq 1$. Using the previous exercise and the fact that σ_n is measurable with respect to $\mathcal{F}_0$ and hence $\mathcal{F}_n$, we get

$$\mathrm{E}\left[e^{\theta(X_n - X_{n-1})}|\mathcal{F}_{n-1}\right] = \int_0^1 \psi(z, \sigma_n) P(Z \in dz|\mathcal{F}_{n-1}),$$

$P(Z \in \cdot|\mathcal{F}_{n-1})$ denoting the RCD of Z given $\mathcal{F}_{n-1}$. Using Lemma 7.1 and the fact that a.s.

$$\int_0^1 z P(Z \in dz|\mathcal{F}_{n-1}) = \mathrm{E}(Z|\mathcal{F}_{n-1}) = \frac{1}{2},$$

it follows that

$$\int_0^1 \psi(z, \sigma_n) P(Z \in dz|\mathcal{F}_{n-1}) \leq \frac{1}{2}\left[\psi(0, \sigma_n) + \psi(1, \sigma_n)\right] \text{ a.s.}$$

Arguments similar to (7.8) – (7.11) show that

$$\mathrm{E}\left[e^{\theta(X_n - X_{n-1})}|\mathcal{F}_{n-1}\right] \leq e^{\theta^2\sigma_n^2/2}.$$

Therefore,

$$\begin{aligned}\mathrm{E}\left[e^{\theta X_n}|\mathcal{F}_0\right] &\leq e^{\theta^2\sigma_n^2/2}\mathrm{E}\left(e^{\theta X_{n-1}}|\mathcal{F}_0\right) \\ &\leq \exp\left(\frac{1}{2}\theta^2\sum_{j=1}^n \sigma_j^2\right),\end{aligned}$$

using the induction hypothesis in the last inequality, from which Step 1 follows. □

Proof of Step 2. Since exponential is a strictly increasing function, for any $\alpha, t > 0$,

$$\begin{aligned}\min_{\theta>0} \exp(-\theta t + \theta^2\alpha) &= \exp\left(\min_{\theta>0}(-\theta t + \theta^2\alpha)\right) \\ &= e^{-t^2/4\alpha},\end{aligned}$$

because the minimum is attained at $\theta = t/2\alpha$. Putting $\alpha = \sum_{j=1}^n \sigma_j^2/2$, Step 2 follows. □

Steps 1 and 2 complete the proof of Theorem 7.4 in view of (7.7). □

The following corollary of Theorem 7.4 is the usually stated version of Azuma-Hoeffding.

Theorem 7.5. *Suppose that* $(X_n, \mathcal{F}_n)_{n\geq 0}$ *is a martingale with* $X_0 = 0$. *Assume that*

$$|X_n - X_{n-1}| \leq \sigma_n,\ n \geq 1\,,$$

for some strictly positive constant σ_n. *Then for all* $t \geq 0$ *and* $n \geq 1$,

$$P(X_n \geq t) \leq \exp\left(\frac{-t^2}{2\sum_{k=1}^n \sigma_k^2}\right)\,.$$

Let us complete this chapter with an example. Let $Z_1, Z_2, \ldots$ be non-zero random variables with finite mean and $Y_1, Y_2, \ldots$ be independent zero mean random variables, independent of $(Z_1, Z_2, \ldots)$ with

$$|Y_n| \leq 1\,.$$

Denote

$$X_n = \sum_{j=1}^n Y_j Z_j\,, n \geq 1\,.$$

The goal is to obtain a bound on $P(X_n > t)$ for a fixed $t > 0$.

The obvious impediment in applying Azuma-Hoeffding is that Z_n's are not necessarily bounded. To overcome this, set

$$\mathcal{F}_0 = \sigma\left(Z_1, Z_2, \ldots\right)\,,$$

and

$$\mathcal{F}_n = \sigma\left(Y_1, \ldots, Y_n, Z_1, Z_2, \ldots\right)\,, n \geq 1\,.$$

Zero-mean property of Y_n's imply that $(X_n, \mathcal{F}_n)_{n\geq 0}$ is a martingale, where $X_0 = 0$. Further,

$$|X_n - X_{n-1}| = |Y_n Z_n| \leq |Z_n|\,, n \geq 1\,,$$

$|Z_n|$ being $\mathcal{F}_0$-measurable for all n. Theorem 7.4 implies that

$$P(X_n \geq t | \mathcal{F}_0) \leq \exp\left(\frac{-t^2}{\sum_{i=1}^n Z_i^2}\right)\,.$$

Thus,

$$P(X_n \geq t) \leq \mathrm{E}\left(\exp\left(\frac{-t^2}{\sum_{i=1}^n Z_i^2}\right)\right)\,.$$