Lecture 3

1 $L_p$ martingale convergence theorems

Using the upcrossing inequality, we have shown that if $(X_n)_{n \in \mathbb{N}}$ is an $L_1$-bounded martingale, then it almost surely converges to a limit $X_\infty \in L_1(\Omega, \mathcal{F}, \mathbb{P})$. However, it could happen that $X_\infty \equiv 0$ while $\mathbb{E}[X_n] = 1$ for all $n \in \mathbb{N}$, such as the example of a symmetric simple random walk starting at $X_0 = 1$ and stopped when it first hits 0. We now address the question of when we would have convergence of $X_n$ to $X_\infty$ in $L_p$ for some $p \geq 1$, besides the a.s. convergence. The situation is simpler when $p > 1$.

**Theorem 1.1 [L$_p$ martingale convergence for $p > 1$]**

Let $(X_n)_{n \in \mathbb{N}}$ be a martingale adapted to the filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\sup_{n \in \mathbb{N}} \mathbb{E}|X_n|^p < \infty$ for some $p > 1$. Then there exists a random variable $X_\infty$ such that $X_n \to X_\infty$ almost surely and in $L_p(\Omega, \mathcal{F}, \mathbb{P})$. Furthermore, $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$ a.s. for all $n \in \mathbb{N}$.

**Exercise 1.2** Use Theorem 1.1 to show that if $X \in L_p(\Omega, \mathcal{F}, \mathbb{P})$ for some $p > 1$ and $X$ is measurable w.r.t. $\sigma(\bigcup_n \mathcal{F}_n)$, then $X_n := \mathbb{E}[X|\mathcal{F}_n]$ converges a.s. and in $L_p$ to $X$.

**Proof.** Since $p > 1$, we have $\sup_{n \in \mathbb{N}} \mathbb{E}|X_n| \leq \sup_{n \in \mathbb{N}}(\mathbb{E}|X_n|^p)^{\frac{1}{p}} < \infty$, and hence $\lim_{n \to \infty} X_n = X_\infty$ a.s. by the martingale convergence theorem. On the other hand, by Doob’s $L_p$ maximal inequality,

$$\mathbb{E}\left[\sup_{1 \leq i \leq n} |X_i|^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p].$$

Letting $n \to \infty$ and applying the monotone convergence theorem in the above inequality shows that $\sup_{n \in \mathbb{N}} |X_n| \in L_p(\Omega, \mathcal{F}, \mathbb{P})$. Since $|X_n - X_\infty| \leq 2\sup_{n \in \mathbb{N}} |X_n|$, it follows from the dominated convergence theorem that $\lim_{n \to \infty} \mathbb{E}[X_n - X_\infty|^p] = 0$.

For any $A \in \mathcal{F}_n$, by the martingale property, we have $\mathbb{E}[1_A X_n] = \mathbb{E}[1_A X_m]$ for any $m \geq n$. As $m \to \infty$, clearly $\mathbb{E}[1_A X_m] \to \mathbb{E}[1_A X_\infty]$ by the $L_p$ convergence of $X_m$ to $X_\infty$. Therefore $\mathbb{E}[1_A X_n] = \mathbb{E}[1_A X_\infty]$ for all $A \in \mathcal{F}_n$, and hence $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$ a.s. □

As seen in the example of the symmetric simple random walk started at 1 and stopped at 0, Theorem 1.1 cannot hold for $p = 1$. In order to have convergence in $L_1$ for a sequence of $L_1$ bounded martingales, we need to impose the extra condition of uniform integrability, which is both necessary and sufficient.

**Theorem 1.3 [L$_1$ martingale convergence]**

Let $(X_n)_{n \in \mathbb{N}}$ be a martingale adapted to the filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F} = \sigma(\bigcup_n \mathcal{F}_n)$. Then the following statements are equivalent:

(i) $(X_n)_{n \in \mathbb{N}}$ are uniformly integrable.

(ii) There exists a random variable $X_\infty$ such that $X_n \to X_\infty$ a.s. and in $L_1$.

(iii) There exists a random variable $X_\infty$ such that $X_n \to X_\infty$ in $L_1$.

(iv) There exists a random variable $X_\infty \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ with $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$ for all $n \in \mathbb{N}$. 

1
Furthermore, the random variable $X_\infty$ is the same in (ii)–(iv).

**Proof.** Clearly (ii) implies (iii), while the converse follows from the martingale convergence theorem since $X_n \to X_\infty$ in $L_1$ implies that $\sup_{n \in \mathbb{N}} \mathbb{E}|X_n| < \infty$.

To show (ii) implies (i), note that for any fixed $M > 0$, we have

$$
\mathbb{E}[1_{\{|X_n| > M\}}|X_n|] \leq \mathbb{E}[1_{\{|X_n| > M\}}|X_n - X_\infty|] + \mathbb{E}[1_{\{|X_n| > M\}}|X_\infty|]
$$

$$
\leq \mathbb{E}|X_n - X_\infty| + \mathbb{E}[1_{\{|X_n| > M\}}|X_\infty|],
$$

where for any $\epsilon > 0$, the first term can be made smaller than $\epsilon/2$ by choosing $n$ sufficiently large, say $n \geq n_0$. For the second term, note that

$$
\mathbb{P}(|X_n| > M) \leq \frac{\mathbb{E}|X_n|}{M},
$$

which can be made arbitrarily small uniformly in $n$ by choosing $M$ sufficiently large. Since $|X_\infty| \in L_1(\Omega, \mathcal{F}, \mathbb{P})$, this implies that if $M$ is sufficiently large, say $M \geq M_0$, then the second term in (1.1) is bounded by $\epsilon/2$ uniformly in $n$ (exercise: prove this). Therefore $(X_n)_{n \geq n_0}$ are uniformly integrable, and so are $(X_n)_{n \in \mathbb{N}}$.

To verify (i) implies (ii), note that the uniform integrability of $(X_n)_{n \in \mathbb{N}}$ implies that $\sup_{n \in \mathbb{N}} \mathbb{E}|X_n| < \infty$, and hence by the martingale convergence theorem, $X_n \to X_\infty$ a.s. Let $X_n^{(M)} := X_n \wedge M \vee (-M)$. Then

$$
\mathbb{E}|X_n - X_\infty| \leq \mathbb{E}|X_n^{(M)} - X_\infty^{(M)}| + \mathbb{E}|X_n - X_n^{(M)}| + \mathbb{E}|X_\infty - X_\infty^{(M)}|,
$$

where the second and third terms can be made arbitrarily small uniformly in $n$ by choosing $M$ large, thanks to uniform integrability, then the first term tends to 0 as $n \to \infty$ by the bounded convergence theorem. Therefore $X_n \to X_\infty$ in $L_1$.

The implication (ii)$\Rightarrow$(iv) follows by the same argument as in the proof of Theorem 1.1. To show that (iv)$\Rightarrow$(i), note that $X_n := \mathbb{E}[X_\infty|\mathcal{F}_n]$ is an $L_1$ bounded martingale, which converges a.s. to a limit $\tilde{X}_\infty$. For $M > 0$,

$$
\mathbb{E}|\mathbb{E}[X_n|\mathcal{F}_n]|1_{\{|X_n| \geq M\}}| = \mathbb{E}\left[\mathbb{E}[X_\infty|\mathcal{F}_n]|1_{\{|X_n| \geq M\}}\right] \\
\leq \mathbb{E}\left[\mathbb{E}[X_\infty|\mathcal{F}_n]|1_{\{|X_n| \geq M\}}\right] \\
= \mathbb{E}|X_\infty||1_{\{|X_n| \geq M\}}|,
$$

where we used Jensen’s inequality and the fact that $\{\{|X_n| \geq M\}\} \in \mathcal{F}_n$. Note further that

$$
\mathbb{P}(\{|X_n| \geq M\}) \leq \frac{\mathbb{E}|X_n|}{M} \leq \frac{\mathbb{E}|X_\infty|}{M},
$$

which tends to 0 uniformly in $n \in \mathbb{N}$ as $M \to \infty$. Therefore by the integrability of $|X_\infty|$, the RHS of (1.2) tends to 0 uniformly in $n \in \mathbb{N}$ as $M \to \infty$, which proves the uniform integrability of $(X_n)_{n \in \mathbb{N}}$. In particular, $X_n \to \tilde{X}_\infty$ a.s. and in $L_1$. The implication (ii)$\Rightarrow$(iv) implies that $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n] = \mathbb{E}[\tilde{X}_\infty|\mathcal{F}_n]$ for all $n \in \mathbb{N}$. Since $X_\infty$ is measurable w.r.t. $\sigma(\bigcup_n \mathcal{F}_n)$, we must have $\tilde{X}_\infty = X_\infty$ a.s., which concludes the proof of the theorem.

As a corollary of the $L_1$ martingale convergence Theorem 1.3, we have

**Corollary 1.4 [Lévy’s 0-1 law]**

Let $A \in \mathcal{F}$ with $\mathcal{F} = \sigma(\bigcup_n \mathcal{F}_n)$ for a filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$. Then $\lim_{n \to \infty} \mathbb{P}(A|\mathcal{F}_n) = 1_A$ almost surely.
The statement may seem trivial, but the 0-1 dichotomy on the possible limits of conditional probabilities can be useful in determining some a.s. properties of a stochastic process, provided we have some information on the conditional probabilities of events. Here is an example.

**Example 1.5 [a.s. extinction vs unbounded growth for branching processes]**

Let \((\xi_{i,j})_{i,j \in \mathbb{N}}\) be a two-parameter family of i.i.d. random variables with common distribution \(\mu\) on \(\mathbb{N}_0 := \{0\} \cup \mathbb{N}\). A discrete time branching process with offspring distribution \(\mu\) is defined recursively by \(S_n = \sum_{i=1}^{\infty} \xi_{n-1,i}\), which represents the population size in the \(n\)-th generation, and \(\xi_{i,j}\) represents the number of children of the \(j\)-th individual in the \(i\)-th generation. Assume that \(\mu(0) > 0\), i.e., each individual has a positive probability of having no offspring. Then \(\mathbb{P}(S_n = 0|S_{n-1} = k) = \mu(0)^k > 0\). Let \(\mathcal{F}_n := \sigma((\xi_{i,j})_{1 \leq i \leq n, j \in \mathbb{N}})\), and denote the event of extinction by \(D := \{S_n = 0\}\). By Lévy’s 0-1 law, on the event \(D^c\), we have \(\mathbb{P}(D|\mathcal{F}_n) \to 0\) a.s. However, \(\mathbb{P}(D|S_n = k) \geq \mu(0)^k > 0\) for all \(k, n \in \mathbb{N}\). Therefore on \(D^c\), necessarily \(S_n \to \infty\) a.s.

**Exercise 1.6** Use the second Borel-Cantelli Lemma to prove the above dichotomy between almost sure extinction and unbounded growth for branching processes.

## 2 Doob’s decomposition and quadratic variation

Here is another way to find martingales. Every sub/super-martingale can be decomposed as the sum of a martingale and a **predictable** increasing/decreasing sequence. This applies in particular to the square of a martingale.

**Definition 2.1 [Predictable sequence]**

A sequence of random variables \((V_n)_{n \in \mathbb{N}}\) is called predictable w.r.t. a filtration \((\mathcal{F}_n)_{n \geq 0}\) if \(V_n\) is measurable w.r.t. \(\mathcal{F}_{n-1}\) for all \(n \in \mathbb{N}\).

**Remark 2.2** Recall that a martingale transform of a martingale \((X_n)_{n \geq 0}\) is defined as \(\tilde{X}_n = \sum_{i=1}^n (X_i - X_{i-1})V_i\) for a predictable sequence \((V_i)_{i \in \mathbb{N}}\).

**Theorem 2.3 [Doob’s decomposition]**

Let \((X_n)_{n \in \mathbb{N}}\) be a sub-martingale adapted to \((\mathcal{F}_n)_{n \in \mathbb{N}}\). Then we can write \(X_n = Y_n + A_n\), where \(Y_n\) is a martingale adapted to \(\mathcal{F}_n\), \(A_n\) is an increasing predictable sequence with \(A_1 \equiv 0\). Furthermore, \(Y_n\) and \(A_n\) are uniquely determined by the above properties.

**Proof.** Let \(D_n := \mathbb{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}] = \mathbb{E}[X_n|\mathcal{F}_{n-1}] - X_{n-1}\). Then \(D_n\) is a predictable sequence, and is furthermore non-negative since \(X_n\) is a sub-martingale. Note that \(D_n\) encodes the bias of the increment of the sub-martingale \(X_n\) at time \(n\), conditional on \(\mathcal{F}_{n-1}\). Therefore, if we let \(A_1 \equiv 0\) and \(A_n = \sum_{i=2}^{n} D_i\) for \(n \geq 2\), then \(A_n\) satisfies the desired properties, and \(Y_n := X_n - A_n\) is a martingale adapted to \(\mathcal{F}_n\).

For any decomposition \(X_n = Y_n + A_n\) satisfying the stated properties, we have

\[ \mathbb{E}[X_n|\mathcal{F}_{n-1}] = Y_{n-1} + A_n. \]

These two equalities together uniquely determine \(A_n\) and \(Y_n\) inductively.
Remark 2.4 Note that

\[ Y_n := X_n - \sum_{i=2}^{n} D_n = X_1 + \sum_{i=2}^{n} (X_i - X_{i-1} - \mathbb{E}[X_i - X_{i-1}|\mathcal{F}_{i-1}]) \]

is a martingal regardless of whether or not \( X_n \) is a sub-martingale. Therefore we can always extract a martingale \( Y_n \) from \( X_n \) in this way, except that the predictable compensator \( A_n := X_n - Y_n \) is not increasing in general.

We have seen that the fluctuation of a martingale \( (X_i)_{1 \leq i \leq n} \) can be controlled by the \( L_p \) norm of \( X_n \). Alternatively, if \( X \) is square-integrable, its fluctuation can also be controlled via its quadratic variation.

Definition 2.5 [Quadratic variation process of a square-integrable martingale]
Let \( (X_n)_{n \in \mathbb{N}} \) be a square-integrable martingale adapted to the filtration \( (\mathcal{F}_n)_{n \in \mathbb{N}} \). Then \( (X^2_n)_{n \in \mathbb{N}} \) is a sub-martingale. By Doob’s decomposition, we can write \( X^2_n = M_n + \langle X \rangle_n \), where \( M_n \) is a martingale adapted to \( \mathcal{F}_n \), and \( \langle X \rangle_n \) is an increasing predictable sequence. More explicitly,

\[ \langle X \rangle_n = \sum_{i=2}^{n} (\mathbb{E}[X^2_i|\mathcal{F}_{i-1}] - X^2_{i-1}) = \sum_{i=2}^{n} \mathbb{E}[(X_i - X_{i-1})^2|\mathcal{F}_{i-1}] . \tag{2.3} \]

Note that \( \mathbb{E}\langle X \rangle_n = \mathbb{E}[X^2_n] - \mathbb{E}[X^2] \). The increasing process \( \langle X \rangle_n \) is called the quadratic variation process for the square-integrable martingale \( X_n \). The proper definition of the quadratic variation process for continuous time martingales is much more subtle, although the basic idea is the same.

Since \( \langle X \rangle_n \) is an increasing process for a square-integrable martingale \( X_n \), almost surely \( \langle X \rangle_n \) increases to a limit \( \langle X \rangle_\infty \) which might be \( +\infty \). We next show that on the event \( \langle X \rangle_\infty < \infty \), which may have probability between 0 and 1, \( X_n \) converges a.s. to a finite limit. This complements the martingale convergence theorems, which require the martingale to have uniformly bounded \( L_p \) norm for some \( p \geq 1 \) and establish convergence with probability 1.

Theorem 2.6 Let \( (X_n)_{n \in \mathbb{N}} \) be a square-integrable martingale, and \( \langle X \rangle_n \) its quadratic variation process. Then \( \lim_{n \to \infty} X_n \) exists and is finite a.s. on the event \( \{\langle X \rangle_\infty < \infty\} \).

Proof. Let \( a > 0 \), and define \( \tau_a = \inf\{n \geq 1 : \langle X \rangle_{n+1} > a^2\} \). Because \( \langle X \rangle_n \) is a predictable sequence, \( \tau_a \) is a stopping time. Now let us consider the stopped martingales \( X_{n \wedge \tau_a} \), and \( X^2_{n \wedge \tau_a} - \langle X \rangle_{n \wedge \tau_a} \). The latter by the martingale property gives

\[ \mathbb{E}[X^2_{n \wedge \tau_a}] = \mathbb{E}[\langle X \rangle_{n \wedge \tau_a}] \leq a^2. \]

Therefore \( X_{n \wedge \tau_a} \) is a \( L_2 \) bounded martingale, which converges a.s. to a finite limit by the martingale convergence theorem. In particular, \( X_n \) converges to a finite limit on the event that \( \langle X \rangle_\infty \leq a^2 \). Sending \( a \to \infty \) then yields the theorem.

Remark 2.7 Note that Theorem 2.6 is still valid if we replace \( \langle X \rangle_n \) by the compensator for \( |X_n|^p \) for any \( p \geq 1 \). However, the quadratic variation process has more structure and is easier to work with.

We can also say something about \( X_n \) on the event \( \langle X \rangle_\infty = \infty \).
Theorem 2.8 Let $X_n$ be a square-integrable martingale. On the event \{$(X)_\infty = \infty$\}, almost surely $\lim_{n \to \infty} \frac{X_n}{(X)_n} = 0$.

**Proof.** Note that $\lim_{n \to \infty} \frac{X_n}{(X)_n} = 0$ is equivalent to $\lim_{n \to \infty} \frac{X_n}{1 + (X)_n} = 0$. We need the following lemma, which replaces the convergence to 0 for a sequence by the convergence of a series.

**Kronecker’s Lemma:** If $a_n \uparrow \infty$ and $\sum_{n=1}^{\infty} \frac{b_n}{a_n}$ is convergent, then $\lim_{n \to \infty} \frac{1}{a_n} \sum_{i=1}^{n} y_i = 0$. (2.4)

When $(y_n)_{n \geq 1}$ are non-negative, (2.4) is a simple consequence of the Dominated Convergence Theorem. For general $(y_n)_{n \geq 1}$, the proof is based on summation by parts, which we leave as an exercise.

Applying Kronecker’s lemma with $a_n = 1 + (X)_n$ and $y_n = X_n - X_{n-1}$, it suffices to show that $\sum_{n=1}^{\infty} \frac{X_n - X_{n-1}}{1 + (X)_n}$ is convergent on the event $\{(X)_\infty = \infty\}$. Note that because $(X)_n$ is a predictable sequence, $W_n := \sum_{i=1}^{n} \frac{X_i - X_{i-1}}{1 + (X)_i}$ is a martingale transform of $X_n$. In particular, its quadratic variation process is given by

$$\langle W \rangle_n = \sum_{i=1}^{n} \frac{(X)_i - (X)_{i-1}}{(1 + (X)_i)^2} \leq \sum_{i=1}^{n} \int_{(X)_{i-1}}^{(X)_i} \frac{dx}{(1 + x)^2} = \int_0^{(X)_n} \frac{dx}{(1 + x)^2} \leq 1.$$  

Therefore $W_n$ is an $L_2$ bounded martingale and hence converges almost surely to a finite limit. The theorem then follows from Kronecker’s lemma.

**Exercise 2.9** Show that in Theorem 2.8, the result still holds if we replace $\frac{X_n}{(X)_n}$ by $\frac{X_n}{f((X)_n)}$ for any increasing function $f \geq 1$ with $\int_0^\infty \frac{dx}{f(x)^2} < \infty$.

**Corollary 2.10** Let $X_n = \sum_{i=1}^{n} \xi_i$, where $\xi_i$ are i.i.d. with $\mathbb{E}[\xi_i] = 0$ and $\mathbb{E}[\xi_i^2] = 1$. Then $X$ is a martingale with $(X)_n = n$. Therefore by Theorem 2.8, $\lim_{n \to \infty} \frac{X_n}{n} = 0$ almost surely, which is the strong law of large numbers. By Exercise 2.9, we in fact have $\lim_{n \to \infty} \frac{X_n}{\sqrt{n \log n}} = 0$ almost surely.

Another corollary of Theorems 2.6 and 2.8 is the following refined version of the second Borel-Cantelli lemma.

**Corollary 2.11 [Second Borel-Cantelli Lemma Refined]** Let $(F_n)_{n \geq 0}$ be a filtration on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $F_0 = \\{\emptyset, \Omega\}$. Let $A_n \in \mathcal{F}_n$ for $n \geq 1$, and let $W_n = \mathbb{P}(A_n | \mathcal{F}_{n-1})$. Then

$$\lim_{n \to \infty} \sum_{i=1}^{n} \frac{1_{A_i}(\omega)}{W_i} = 1 \quad \text{almost surely on} \quad \{\omega : \sum_{i=1}^{\infty} W_i = \infty\}. \quad (2.5)$$

**Proof.** The martingale $X_n := \sum_{i=1}^{n} (1_{A_i} - W_i)$ has quadratic variation process $(X)_n = \sum_{i=1}^{n} W_i(1 - W_i)$. Therefore by Theorems 2.6 and 2.8, either $(X)_n = \sum_{i=1}^{n} W_i(1 - W_i)$ is finite, in which case $X_n \rightarrow X_\infty \in \mathbb{R}$, and on the event $\sum_{i=1}^{n} W_i = \infty$ this leads to $\frac{X_n}{\sum_{i=1}^{n} W_i} \rightarrow 1$; or $\sum_{i=1}^{n} W_i(1 - W_i) = \infty$, in which case $\sum_{i=1}^{n} W_i \geq \sum_{i=1}^{n} W_i(1 - W_i) = \infty$, and by Theorem 2.8 we have $\sum_{i=1}^{n} \frac{X_n}{\sum_{i=1}^{n} W_i(1 - W_i)} \rightarrow 0$, together they imply $\sum_{i=1}^{n} \frac{X_n}{W_i} \rightarrow 0$ and hence $\frac{X_n}{\sum_{i=1}^{n} W_i} = 1 + \sum_{i=1}^{n} \frac{X_n}{W_i(1 - W_i)} \rightarrow 1$. □