4.2 Martingales

\((\Omega, \mathcal{F}, P)\)

A filtration is an increasing sequence of \(\sigma\)-algebras:
\[ \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \subseteq \mathcal{F}. \]

Think of these as being made up of the \(r\) variables whose values we can observe or compute at times \(0, 1, 2, \ldots\).

Read §4 in my SDE notes.

We will often write \(\mathcal{F}\) for the whole filtration, \(\mathcal{F} = (\mathcal{F}_0, \mathcal{F}_1, \ldots)\).

**Definition** Martingale, submartingale, supermartingale

Let \(X_0, X_1, X_2, \ldots\) be real-valued random variables on \((\Omega, \mathcal{F})\).

Suppose

(i) \(\mathbb{E}|X_n| < \infty\) for all \(n\) (integrability, so we can do conditional expectations)

(ii) \(X_n \in \mathcal{F}_n\) for all \(n\) (we can observe \(X_n\) at time \(n\))

(iii) \(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n\) for all \(n\), then \(X = (X_n)\) is a martingale w.r.t. \(\mathcal{F}\)

(iii)' \(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n\) for all \(n\), \ldots submartingale\ldots

(iii)'' \(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n\) for all \(n\), \ldots supermartingale\ldots

**Theorem 4.2.1**, p. 232

Given (iii), in fact, \(\mathbb{E}[X_{nm} \mid \mathcal{F}_n] = X_n\), all \(n, m \geq 0\).

Similarly with (iii)' and (iii)''.

**Proof**

\[
\mathbb{E}[X_{nm} \mid \mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X_{nm} \mid \mathcal{F}_{nm-1}] \mid \mathcal{F}_n] \quad \text{property 9}
\]

\[= \mathbb{E}[X_{nm-1} \mid \mathcal{F}_n].\]

Continue until \(\mathbb{E}[X_{n1} \mid \mathcal{F}_n] = X_n\).

Similarly for the others.

**Show** Suppose \(X\) is a martingale w.r.t. \(\mathcal{F}\) and set \(Y_n = (X_0, \ldots, X_n)\).

Then \(\mathcal{Y} = (\mathcal{Y}_n)\) is a filtration and \(X\) is a martingale w.r.t. \(\mathcal{Y}\).

\(\mathcal{Y}\) is called the filtration generated by \(X\).
Example

1. Let $A_1, A_2, \ldots$ be i.i.d. real valued integrable r.v. with mean 0.
   
   Set $X_0 = 0$
   
   $X_n = A_1 + \cdots + A_n$.
   
   Let $\mathcal{F}$ be the filtration generated by $X$ $(\mathcal{F}_n = \sigma(X_0, \ldots, X_n) = \sigma(A_1, \ldots, A_n))$.
   
   Then $X$ is a martingale w.r.t. $\mathcal{F}$ because $(X_0, \ldots, X_n) \in \mathcal{F}(A_1, \ldots, A_n)$ and $(A_1, \ldots, A_n) \in \mathcal{F}(X_0, X_1, \ldots, X_n)$.

   (i) $\mathbb{E}|X_n| < \infty$
   
   (ii) $X_n \in \mathcal{F}_n$
   
   (iii) $\mathbb{E}[X_{n+1} | \mathcal{F}_n] = \mathbb{E}[X_n + A_{n+1} | \mathcal{F}_n]$
       
       $= \mathbb{E}[X_n | \mathcal{F}_n] + \mathbb{E}[A_{n+1} | \mathcal{F}_n]$
       
       $= X_n + \mathbb{E}[A_{n+1}]$, properties 2, 3.

   The key here is that $\mathbb{E}[A_{n+1} | \mathcal{F}_n] = 0$.

   Similarly, if $\mathbb{E}[A_n] > 0$, $X$ is a submartingale,
   
   if $\mathbb{E}[A_n] < 0$, $X$ is a supermartingale.

2. Let $h_n$ be a sequence of real numbers. Deterministic.
   
   Set $X_0 = 0$
   
   $X_n = h_1 A_1 + \cdots + h_n A_n$.
   
   In the same way, $X$ is a martingale.
   
   We can scale up or down the sizes of the additive increments without changing the fact that we have a martingale.

3. Set $X_0 = 0$
   
   $X_1 = A_1$
   
   $X_n = A_1 + A_2 + A_3 + A_4 + \cdots + A_n + A_n$
   
   Provided $\mathbb{E}[|A_1, A_2|] < \infty$, $X$ is again a martingale:

   $\mathbb{E}[X_{n+1} | \mathcal{F}_n] = \mathbb{E}[X_n + A_{n+1} | \mathcal{F}_n]$
   
   $= X_n + \mathbb{E}[A_{n+1} | \mathcal{F}_n]$
   
   $= X_n$.

   We can scale up or down the size of $A_n$ randomly.
(4) Let $H_n \in \mathbb{A}_{n-1}$ be bounded for each $n$.

Set $X_0 = 0$

$X_n = H_1 A_1 + \cdots + H_n A_n$

Then $X$ is a martingale.

Show Check all three parts of the definition.
**Martingale intuition**

**General martingale now.**

It is helpful to make a sort of tree diagram to show how the values of $X_n$ can evolve over time.

$$X_0, X_1, X_2, \ldots$$

no matter what has happened so far, the new value has to be centered on the old. There is balance.

$$E[X_n] = E\left[E[X_1 | X_0]\right] = E[X_0]$$

Similarly, $E[X_n] = E[X_0]$ for all $n$.

1. $X_n = A_1 + \cdots + A_n$

   At every "leaf" the options are the same.

2. Imagine $h_n \to 0$ as $n \to \infty$.

   the new additive increments are smaller, but still centered

   May converge as $n \to \infty$, $h_n = \frac{1}{2^n}$, say.

3. $X_n = A_1 + A_1A_2 + A_1A_2A_3 + \cdots$

   When $A_n = 0$, $X$ does not branch out between time $n$ and $n+1$.


Turn these pictures $90^\circ$ clockwise and imagine them as a hanging mobile with weights on the nodes. There must always be balance or the mobile will tilt.

*Note:* $E[X_{n+1} | Y_n] = X_n$ means that no matter what path was followed to get to the current value of $X_n$, the (future) value of $X_{n+1}$ is still centered on $X_n$. 
General random scaling of martingale increments

\((\Omega, \mathcal{F}, \mathbb{P})\)

Filtration \( \mathcal{F} = (\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots) \)

Real-valued r.v. \( X = (X_0, X_1, X_2, \ldots) \)

Suppose \( X \) is a martingale \( \text{w.r.t.} \ \mathcal{F} \)

(i) the \( X_n \) are integrable
(ii) \( X \) is adapted to \( \mathcal{F} \)
(iii) \( \mathbb{E} \left[ X_{n+1} \mid \mathcal{F}_n \right] = X_n \) or \( \mathbb{E} \left[ (X_{n+1} - X_n) \mid \mathcal{F}_n \right] = 0 \).

Note that \( \mathcal{F} \) may be larger than the filtration generated by \( X \).

Let \( H = (H_0, H_1, H_2, \ldots) \) be a predictable sequence:

\[ H_n \in \mathcal{F}_n, \quad \text{for all } n. \]

Intuitively, \( H_n \) can be computed from the r.v.'s observed up to time \( n-1 \).

The next value of \( H \) is decided in advance.

Earlier, we used \( H \) to scale up and down the additive increments of \( X \).

Here is how we do this in general:

Write \( X_n = X_0 + (X_1 - X_0) + (X_2 - X_1) + \cdots + (X_n - X_{n-1}) \)

Set \( (H \cdot X)_0 = 0 \)

\( (H \cdot X)_1 = H_1 (X_1 - X_0) \)

\[ \vdots \]

\( (H \cdot X)_n = H_n (X_n - X_{n-1}) \)

Then \( H \cdot X \) is a new stochastic process.

It scales up (or down) the increments of \( X \), making a process that spreads out more (or less) than \( X \).

You can think of \( H_n \) as the number of shares of stock you own over time interval \( n-1 \) to \( n \), and \( X \) as the stock price, so \( H_n (X_n - X_{n-1}) \) is your profit or loss over that time interval.

Remark

Stieltjes integrals \( \int \delta_t \, dF(x) \) are approximated by sums \( \sum \delta_t (x_i) (F(x_{i+1}) - F(x_i)) \).

Then the maximum width \( |x_{i+1} - x_i| \) goes to 0.

Our process \( H \cdot X \) is very much like \( \int H \, dX \), except that, in discrete time, it is not possible to let the width of the interval go to 0.
Theorem 4.2.7

(a) Let $X$ be a martingale w.r.t. $\mathcal{F}$ and $H$ be predictable and $|H_n| \leq h_n$ for some sequence $(h_n)$. Then $H \cdot X$ is a martingale w.r.t. $\mathcal{F}$.

(b) Let $X$ be a submartingale w.r.t. $\mathcal{F}$, $H$ predictable, positive, bounded. Then $H \cdot X$ is a submartingale w.r.t. $\mathcal{F}$.

(c) $X$ a supermartingale ... $H \cdot X$ a supermartingale.

Lesson

No matter how you bet on a martingale (who seeing the future), your earnings are a martingale, expected value $0$.

Bet on a submartingale instead, and never on a supermartingale.

"the martingale" double $H_n$ if you lose $\#1$, make $H_{n+1}$ if you win.

$A_n = \pm 1$

Proof

Case (a)

(i) $E|H_n \cdot X_n| \leq \sum_{i=1}^{n} E|H_n (X_n - X_{n-1})| \leq \sum_{i=1}^{n} h_n (E[X_n | \mathcal{F}_n]) < \infty$

(ii) $(H \cdot X)_n \in \mathcal{F}_n$ is clear

(iii) $E[(H \cdot X)_n | \mathcal{F}_n] = E[(H \cdot X)_n + H_{n+1} (X_{n+1} - X_n) | \mathcal{F}_n]$

$= (H \cdot X)_n + H_{n+1} E[X_{n+1} - X_n | \mathcal{F}_n]$

$= (H \cdot X)_n$

Cases (b) and (c) are similar.
General comments about martingales

The term 'martingale' apparently comes from horse racing.
A martingale is a set of straps that keep the horse's head steady or check its upward movement.

The 'martingale' is a betting strategy in which, every time you lose, you bet twice the amount the next time.
If you bet $5 and lose, next time you bet $10.
If you win, then your balance is $5, but if you lose, your balance is $15. In that case, you bet $20, so that, if you win, your balance is $5, but if you lose...

Martingales have direct relevance to gambling.
They are also reasonably good models for stock prices.
If we could anticipate today that tomorrow's price would be higher than today's (that is, if $E[X_n, 1/\gamma_t] \geq X_n$), then people would raise today's price by buying the stock, until $E[X_n, 1/\gamma_t] = X_n$.
Similarly, if we could anticipate the price going down, people would sell, lowering $X_n$, to restore $E[X_n, 1/\gamma_t] = X_n$.
This is called the martingale hypothesis.
It is pretty good, but not perfect.
Theorem 4.2.3

If $X$ is a martingale w.r.t. $\mathcal{F}$ and $\varphi$ is convex, $\mathbb{E}|\varphi(X_n)| < \infty$ for all $n$, then $\varphi(X_n), n=0,1,2,\ldots$ is a submartingale w.r.t. $\mathcal{F}$.

Proof

$$\mathbb{E} \left[ \varphi \left( X_{n+1} \right) \mid \mathcal{F}_n \right] \geq \varphi \left( \mathbb{E} \left[ X_{n+1} \mid \mathcal{F}_n \right] \right) \quad \text{Jensen}$$

$$= \varphi \left( X_n \right) \quad \text{mart.}$$

$\mathbb{E} \left[ |X_{n+1}|^{p} \right] < \infty$ for $p \geq 1$.

---

Theorem 4.2.5

If $X$ is a submartingale and $\varphi$ is increasing, convex, $\mathbb{E}|\varphi(X_n)| < \infty$, then $\varphi(X_n)$ is a submartingale.
Theorem 4.2.10  Martingale convergence

If \( X \) is a submartingale with \( \sup_n E X_n^+ < \infty \), then \( X_n \)
converges a.s. to a limit \( X_\infty \) with \( E |X_\infty| < \infty \).

Proof

1. Claim: \( \sup_n E |X_n| < \infty \).

\[
E X_0 \leq E X_n \quad \text{submartingale}
\]
\[
= E X_n^+ - E X_n^-
\]
Solving, \( E X_n^- \leq E X_n^+ - E X_0 \).

So \( E |X_n| = E X_n^+ + E X_n^- \leq 2 E X_n^+ - E X_0 \).

Thus, \( \sup_n E |X_n| \leq 2 \sup_n E X_n^+ - E X_0 < \infty \).

Thus, we could just as well have required \( \sup_n E |X_n| < \infty \) in the statement of the theorem.

2. By Fatou's Lemma,

\[
E \liminf_{n \to \infty} |X_n| \leq \lim_{n \to \infty} E |X_n| = \sup_n E |X_n| < \infty.
\]

Thus, \( \liminf_{n \to \infty} |X_n| < \infty \) a.s.

Then \( P(\lim_{n \to \infty} X_n = -\infty) = 0 \) and \( P(\lim_{n \to \infty} X_n = +\infty) = 0 \),

since in each case, \( \liminf_{n \to \infty} |X_n| \) would be \( \infty \), but it never can be.

3. The only way that \( X_n \) can fail to converge now is by oscillation.

We will show that, essentially, the number of oscillations is controlled by \( E |X_n| \), which is bounded.

First, \( P(\lim_{n \to \infty} X_n \text{ does not exist}) \)
\[
= P(\liminf_{n \to \infty} X_n < \limsup_{n \to \infty} X_n)
\]
\[
= P(\bigcup_{a,b} \{ \liminf_{n \to \infty} X_n < a < b < \limsup_{n \to \infty} X_n \})
\]
\[
\leq \sum_{a,b} P(\liminf_{n \to \infty} X_n < a < b < \limsup_{n \to \infty} X_n)
\]
\[
\leq \sum_{a,b} P( X \text{ crosses the interval } [a,b] \text{ infinitely often})
\]
Counting crossings of \([a, b] \)

Let \( A_i = \min \{ m \geq 0 : X_n < a \} \)
\( B_k = \min \{ m > A_i : X_n > b \} \)
\( A_k = \min \{ m > B_k : X_n < a \} \)

Let \( H_n = \begin{cases} 1 & \text{if } A_i < n \leq B_i \\ 0 & \text{otherwise} \end{cases} \)
\[ = \sum_{i=1}^{\infty} 1(A_i, B_i)(n) \]

Then \( H_{n+1} \in \mathbb{N} \) because \( H_n = \mathbb{I}(X_0, \ldots, X_n) \), and moreover, we can write \( f \) to keep track of going up and down:
If \( H_n = 0 \) (going down), and \( X_n \geq a \) (still going down), then \( H_{n+1} = 0 \).
If \( H_n = 0 \) and \( X_n < a \) (starting going down), then \( H_{n+1} = 1 \).
If \( H_n = 1 \) and \( X_n > b \), then \( H_{n+1} = 0 \); no longer going up.
If \( H_n = 1 \) and \( X_n \leq b \), then \( H_{n+1} \) will equal 1 as well, even if \( X_{n+1} > b \).

Summing increments

Let \( U_n \) be the number of "upcrossings" of the interval \([a, b]\) completed by time \( n \).

Then \( (H - X)_n \) adds up increments of \( X \) only during upcrossings, each of which contributes \( b - a \) or a little more to the sum.

so \( (H - X)_n \geq (b-a)U_n + \text{partial upcrossing} \).

But the partial upcrossing may be negative, and in this case the inequality doesn't mean much, nor does it help us bound \( U_n \) above.

Let \( Y_n = \max (a, X_n), \quad n=0, 1, 2, \ldots \)

The function \( x \rightarrow \max(a, x) \) is increasing and convex, so \( Y_i \) is also a submartingale w.r.t. \( f \). Thus, as above,

\[ (H - Y)_n \geq (b-a)U_n + \text{partial upcrossing} \],
and this partial one is positive, so
\[ (b-a)U_n \leq (H - Y)_n \]

Now we want to take the expectation and bound the RHS.

Very similar to the \( X^+ \) \( X^- \) part above.
Note that \((H \cdot Y)_n + ((1-H) \cdot Y)_n = Y_n - Y_0\). So,

\[(b-a) \mathbb{E} U_n \leq \mathbb{E} (H \cdot Y)_n \]

\[= \mathbb{E} \left[ Y_n - Y_0 - ((1-H) \cdot Y)_n \right] \]

\[\leq \mathbb{E} \left[ Y_n - Y_0 \right] - \mathbb{E} ((1-H) \cdot Y)_n \]

\[= \mathbb{E} \left[ Y_n - Y_0 \right] \]

\[= \mathbb{E} \max(a, x_n) - \max(a, x_0) \]

\[\leq \mathbb{E} (|a| + |x_n|) - \max(a, x_0) \]

This is called the **Uppercrossing Inequality**.

6. Let \(n \to \infty\) to get

\[(b-a) \lim_{n \to \infty} \mathbb{E} U_n \leq \sup_n \mathbb{E} \left[ |a| + |x_n| - \max(a, x_0) \right] \]

\[= |a| + \sup_n \mathbb{E} |x_n| - \mathbb{E} \max(a, x_0) \]

\[< \infty \]

by assumption. But, by the Monotone Convergence Theorem,

\[\lim_{n \to \infty} \mathbb{E} U_n = \mathbb{E} \lim_{n \to \infty} U_n, \quad \text{and} \quad U = \lim_{n \to \infty} U_n \]

is the number of upcrossings of \([a, b]\) that \(X\) makes over all time. Because \(\mathbb{E} U < \infty\), \(U < \infty\) almost surely, so \(\mathbb{P}(U = \infty) = 0\).

Thus, going back a few pages,

\[\mathbb{P}(\lim \inf X_n < \lim \sup X_n) \]

\[\leq \sum_{a, b, \in \mathbb{Q}} \mathbb{P}(X \text{ upcrosses } [a, b], i.o.) \]

\[= \sum_{a, b} \mathbb{P}(U = \infty) \]

\[= 0, \]

and thus \(\lim \inf X_n = \lim \sup X_n\) almost surely.

Let \(X_\infty = \lim_{n \to \infty} X_n\). Then, also,

\[\mathbb{E} |X_\infty| = \mathbb{E} \lim \inf |x_n| \leq \lim \inf \mathbb{E} |x_n| \leq \sup_n \mathbb{E} |x_n| < \infty, \]

so \(X_\infty\) is integrable.

**Remark** Perhaps the most amazing thing about this proof is that \((1-H) \cdot Y\), which adds up increments of \(Y\) during downcrossings, is still a submartingale, meaning that it tends to go up. Why? Even on a "downcrossing," \(Y\) can go up.
Recap - martingale convergence

If $X$ is a sub/super/mart w.r.t. $Y$ and $\sup_n E|X_n| < \infty$, then $X_n$ converges a.s. to a r.v. $X_\infty$ with $E|X_\infty| < \infty$.

Sufficient conditions:

- $X$ is a submart with $\sup_n E X_n^+ < \infty$
- $X$ is a supermart with $\sup_n E X_n^- < \infty$
- $X$ is a supermart with $X_n \geq 0$ a.s.

In this case, $E X_\infty = E \lim_n X_n$

= $E \liminf_n X_n$

$\leq \liminf_n E X_n$

$\leq E X_\infty$.

- $X$ is a martingale with $X_n \geq 0$ for all $n$.

Then $-X$ is a martingale, so also a submartingale, and $\ell E(X_n)^+ = 0 < \infty$.

Wiggly martingales have large moments.

Draw a tree diagram, try to have lots of paths of $X$ that wiggle a particular way.

Want a certain fraction of the paths to be wiggling in the given interval.

In order for a positive fraction of the paths to still be wiggling in the given interval, some paths from outside that interval must re-enter it. When they do, other paths must split off in the other direction, and so spread out the picture as a whole.

Thus, oscillation causes the martingale paths to spread out.
Stopping times

Please read §4 of the SDE notes, especially p. 34 ff.

Let $\mathcal{F}$ be a filtration

$\mathcal{F}_n$ consists of all random variables whose value can be observed or computed at time $n$.

If $A \in \mathcal{F}_n$, then $1_A \in \mathcal{F}_n$; we can compute its value, either 0 or 1.

So at time $n$, we may not know everything about the realization $\omega$, but we will know whether or not $\omega \in A$.

Let $N: \Omega \to \{0,1,2,\ldots\}$ be a random variable.

Think of it as the time that something happens.

Example #1

$N$ is the day that you meet your future spouse.

Is today the day? You can't tell yet, even at the end of the day.

Example #2

$N$ is the day you get married (for the first time).

Is today the day? You'll know, probably ahead of time!

Definition

$N$ is a stopping time if $\{N=n\} \in \mathcal{F}_n$ for all $n$.

That is, on day $n$, you'll know "is today the day?"

Show $\{N=n\} \in \mathcal{F}_n$

Example

Let $X$ be adapted to $\mathcal{F}$.

Let $A \in \mathcal{B}^R$.

Let $N = \min \{ n \geq 0 : X_n \in A \}$.

Then $\{N=n\} = \{ X_0 \notin A, X_1 \notin A, \ldots, X_n \in A \} \in \mathcal{F}_n$,

so $N$ is a stopping time.

Example

Let $N$ be a stopping time w.r.t. $\mathcal{F}$, let $A \in \mathcal{B}^R$,

and set $T = \min \{ n > N : X_n \in A \}$.

Then $\{T=n\} = \{ n > N, X_{n+1} \notin A, X_{n+2} \notin A, \ldots, X_n \in A \} = \bigcup_{m=0}^{n-1} \{N=m, X_{m+1} \notin A, X_{m+2} \notin A, \ldots, X_n \in A \} \in \mathcal{F}_n$,

so $T$ is also a stopping time.

Show If $M, N$ are stopping times, then $M \wedge N = \min(M, N)$

$M \vee N = \max(M, N)$ are as well.
Stopping times as predictable processes

Let $N$ be a stopping time w.r.t. $\mathcal{F}$.
Let $H_n = 1\{n \leq N\}$, so that $H$ is 1 until $N$, then 0.

Then $H$ is predictable:

$$\{n \leq N\} = \{N < n\}^c = \{N \leq n-1\}^c \in \mathcal{F}_{n-1}.$$  

(Notice that $H$ is continuous from the left; that is a typical feature of predictable processes.)

*Example*  When deriving the upcrossing inequality, we had random times $A_1 < B_1 < A_2 < B_2 < \ldots$.

There are stopping times, from above.

We had

$$H_n = \sum_{i=1}^{B_i} 1(A_i, B_i](n)$$

$$= \sum_{n \leq \tau \leq N} (1(0, A_{\tau}](n) - 1(0, A_{\tau}](n)), $$

so we see (again) that $H$ is predictable.

**Stopping martingales**

Let $X$ be a martingale w.r.t. $\mathcal{F}$.

Consider $X_{N+n}$, $n = 0, 1, 2, \ldots$

$$X_{N+n} = \begin{cases} X_n & \text{if } n \leq N \\ X_N & \text{if } n \geq N \end{cases}$$

**Corollary 4.2.8** $X_{N+n}$, $n = 0, 1, 2, \ldots$ is a martingale w.r.t. $\mathcal{F}$.

**Proof**

1. $X_{N+n} = X_0 + (X_1 - X_0) + \ldots$

   $= X_0 + (H_n X)_n$

   But $H$ is positive, bounded, predictable, so $H_n X$ is a mart.

2. Tree diagram for $X^n$, $X_{N+n}$ just stops the splitting, based on what has occurred already. Some branches stop splitting earlier than others.
Martingale convergence in $L^1$? Don't count on it

Even when $\sup_n E|X_n| < \infty$, $X_n \to X_\infty$ a.s. and $E|X_\infty| < \infty$, it can be the case that $\|X_n - X_\infty\|_{L^1} = E|X_n - X_\infty|$ does not go to 0.

Example 4.2.2

Let $S$ be the simple random walk $S_n = A_1 + \ldots + A_n$, $A_i$ iid Bernoulli($\frac{1}{2}$).

Then $S$ is a martingale.

Let $N = \min\{n > 0 : S_n = -1\}$.

Then $(S_{N+n})$ is a martingale with $E S_{N+n} < 1$.

From the proof of the martingale convergence theorem, $\sup_n E|S_{N+n}| < \infty$ and $S_{N+n}$ converges a.s. to a finite limit as $n \to \infty$.

But it must converge to $-1$; that's the only place it can stop moving.

So $S_{\infty} = -1$.

But $E|S_{\infty} - (-1)| = E S_{\infty} + 1 = 1$; doesn't go to 0!

The problem is this: although $S_n \to -1$ a.s., it does not converge (anywhere close to 0) uniformly.

* Do the exercises at the end of §4.2.
Martingales with bounded increments

Theorem 4.3.1

Let \( X \) be a martingale w.r.t. \( \mathcal{F} \) and suppose there exists \( b < \infty \) for which \( \{X_{n+1} - X_n\} \leq b \) for all \( n \).

With probability one, either

- \( \lim_{n \to \infty} X_n \) exists and is finite
- \( \sup_n X_n = \infty \) and \( \inf_n X_n = -\infty \).

Proof: Note that \( Y_n = X_n - X_0, \ n \geq 0, 1, 2, \ldots \) is a martingale w.r.t. \( \mathcal{F} \). Thus, we may assume \( X_0 = 0 \).

Let \( K > 0 \).

Let \( N_K = \min \{ n \geq 0 : X_n > K \} \).

Note: \( N_K < \infty \) if \( X \) never goes above \( K \).

Then \( X_{N_K} \) is a martingale.

But by bounded increments, \( X_{N_K} \) is bounded below by \( K + b \) almost surely.

So \( X_{N_K} \) converges almost surely as \( n \to \infty \).

This is true even for those \( w \) for which \( N_K(w) = \infty \); never even hits \( K \)!

In that case, \( N_K \) is null, so \( X_n \) converges on \( \{N_K = \infty\} \).

That is, \( \{N_K = \infty\} \subseteq \{\lim_{n \to \infty} X_n \text{ exists}\} \).

This is true for all \( K > 0 \), so

\[
\bigcup_{K=1}^{\infty} \{N_K = \infty\} \subseteq \{\lim_{n \to \infty} X_n \text{ exists}\}.
\]

But

\[
\bigcup_{K=1}^{\infty} \{N_K = \infty\} = \{w : X_n(w) \text{ doesn't get above } K, \text{ for some } K\} = \{\sup_n X_n < \infty\}.
\]

So \( \{\sup_n X_n < \infty\} \subseteq \{\lim X_n \text{ exists}\} \).

By considering \( -X_n \) which is also a martingale,

\( \{\inf_n X_n > -\infty\} \subseteq \{\lim X_n \text{ exists}\} \).

Thus,

\[
\{\sup_n X_n < \infty\} \cup \{\inf_n X_n < \infty\} \subseteq \{\lim X_n \text{ exists}\}.
\]

Contrapositively, if \( \lim X_n \) does not exist, then both \( \sup X_n = \infty \) and \( \inf X_n = -\infty \).
Sums with a random number of terms

We saw an example similar to this when we discussed the compound Poisson distribution, during the infinite divisibility section.

The technique here is more sophisticated, and uses Exercise 33.

Let $A_1, A_2, \ldots$ be iid and integrable.
Let $N$ take values in $\{0, 1, 2, \ldots\}$ be integrable and independent of $(A_1, A_2, \ldots)$.
Let $Z = \begin{cases} 0, & \text{if } N = 0 \\ A_1 + A_2 + \cdots + A_N, & \text{if } N > 0. \end{cases}$

Then $\mathbb{E}Z = \mathbb{E}A_1 \cdot \mathbb{E}N$

Proof

\[
\mathbb{E}Z = \mathbb{E}[A_1 + A_2 + \cdots + A_N] = \mathbb{E} \left[ \mathbb{E} \left[ f(A_1, A_2, \ldots, N) \mid N \right] \right] = \mathbb{E} h(N),
\]

where $h(n) = \mathbb{E} f(A_1, A_2, \ldots, n)$, by Exercise 33.

But $h(n) = \mathbb{E} [A_1 + A_2 + \cdots + A_n] = n \cdot \mathbb{E}A_1$, by identically distributed property and integrability.

so

\[
\mathbb{E}Z = \mathbb{E} h(N) = \mathbb{E} \left[ N \cdot \mathbb{E}A_1 \right] = \mathbb{E}A_1 \cdot \mathbb{E}N.
\]

Sticky point

In order to use Exercise 33, we need to have $\mathbb{E} \left| f(A_1, A_2, \ldots, N) \right| < \infty$.

Now \[
\left| f(A_1, A_2, \ldots, N) \right| \leq \left| f(|A_1|, |A_2|, \ldots, N) \right|,
\]
and this function of $\omega$ is positive. Exercise 33 also applies to positive integrands (although the integral may be infinite), and so the method above can be used to show that $\mathbb{E} |A_1| + \cdots + |A_N| = \mathbb{E}N \cdot \mathbb{E}|A_1| < \infty$ before going on to the case of signed $A_i$. 

Let $A_1, A_2, \ldots$ be iid and integrable, and let $S_n = A_1 + \cdots + A_n$.
Let $\mathcal{F}$ be the filtration generated by the $A_i$, and let $N$ be a stopping time with respect to $\mathcal{F}_n$ for which $\mathbb{E}N < \infty$.

Then $\mathbb{E}S_N = \mathbb{E}A_i \cdot \mathbb{E}N$. [Note, $S_N = 0$ when $N = 0$.]

Proof

1. Suppose that $A_i \geq 0$, a.s.

   Then
   
   $$\mathbb{E}S_N = \mathbb{E} \sum_{n=1}^{\infty} A_n \mathbf{1}_{\{n \leq N\}}$$
   
   $$= \sum_{n=1}^{\infty} \mathbb{E} A_n \mathbf{1}_{\{n \leq N\}}$$
   
   $$= \sum_{n=1}^{\infty} \mathbb{E} A_n \mathbf{1}_{\{n \leq N, n-1\}}$$
   
   $$= \sum_{n=1}^{\infty} \mathbb{E} \left[ \mathbb{E} \mathbf{1}_{\{n \leq N, n-1\}} \mathbf{1}_{\mathcal{F}_{n-1}} \right]$$
   
   $$= \sum_{n=1}^{\infty} \mathbb{E} \left[ \mathbf{1}_{\{n \leq N, n-1\}} \mathbf{1}_{\mathcal{F}_{n-1}} \mathbb{E} A_n \mid \mathcal{F}_{n-1} \right]$$
   
   $$= \sum_{n=1}^{\infty} \mathbb{E} \mathbf{1}_{\{n \leq N\}} \mathbb{E} A_n$$

   $$= \mathbb{E}A_i \cdot \mathbb{E} \sum_{n=1}^{\infty} \mathbf{1}_{\{n \leq N\}}$$

   $$= \mathbb{E}A_i \cdot \mathbb{E}N$$

2. Suppose that $\mathbb{E}|A_i| < \infty$ and $\mathbb{E}N < \infty$.

   We may write $S_n = (A_1^+ + \cdots + A_n^+) - (A_1^- + \cdots + A_n^-)$, and so

   $$S_N = S_N^{(+) - S_N^{(-)}},$$

   where $S_N^{(+)}$ is the sum of the first $N$ values of $A_n^+$, which is different from $S_N^+ = \max(0, S_N)$.

   Part 1 shows that $\mathbb{E}S_N^{(+) = \mathbb{E}A_i^+ \cdot \mathbb{E}N < \infty}$ and

   $\mathbb{E}S_N^{(-)} = \mathbb{E}A_i^- \cdot \mathbb{E}N < \infty$.

   Combining these,

   $$\mathbb{E}S_N = \mathbb{E}A_i^+ \cdot \mathbb{E}N - \mathbb{E}A_i^- \cdot \mathbb{E}N$$

Remark

If the $A_i$ are iid and $N$ is simply independent of $(A_1, A_2, \ldots)$, then we may go directly from Equation (1) to Equation (2) and conclude that $\mathbb{E}S_N = \mathbb{E}A_i \cdot \mathbb{E}N$; similarly (2) follows if the $A_i$ and $N$ are integrable. \(\square\)
Martingale convergence overview

§4.2 If \( X \) is a sub/super/mart and \( \sup_n \mathbb{E}|X_n| < \infty \), then \( X_n \) converges a.s. to a limit \( X_\infty \).

However, \( \mathbb{E}|X_n - X_\infty| \) might not converge to 0, so \( |X_n - X_\infty| \) does not go to 0 uniformly enough over \( \omega \).

One could use the dominated convergence theorem to get \( \mathbb{E}|X_n - X_\infty| \to 0 \)
if one could find a single r.v. dominating \( |X_n - X_\infty| \).

§4.4 We will see that if \( X \) is a martingale for which \( \sup_n \mathbb{E}|X_n|^p < \infty \) for some \( p > 1 \), then \( (\sup_n \mathbb{E}|X_n| < \infty \) so \( X_n \to X_\infty \) a.s. and \( \mathbb{E}|X_n - X_\infty|^p \to 0 \) as \( n \to \infty \). This is convergence in \( L^p \), \( p > 1 \), which is stronger than convergence in \( L^1 \).

However, the condition \( \sup_n \mathbb{E}|X_n|^p < \infty \) is stronger than \( \sup_n \mathbb{E}|X_n| < \infty \).

§4.5 For convergence in \( L^1 \), we need a condition in between these two, called \textbf{uniform integrability}:

\[
\lim_{M \to \infty} \sup_{n} \mathbb{E}|X_n| 1_{\{X_n > M\}} = 0.
\]

This resembles tightness.
4.4 Doob's inequality and convergence in $L^p$

Probability space $(\Omega, \mathcal{F}, P)$, filtration $\mathcal{F}$.

Let $X$ be a submartingale with respect to $\mathcal{F}$.

Then $E X_0 \leq E \left[ E(X_n \mid \mathcal{F}_k) \right] = E X_n$ for all $n$.

Let $N$ be a stopping time with respect to $\mathcal{F}$.

Then $X_{N\wedge n}$, $n = 0, 1, 2, \ldots$ is a submartingale as well.

Now $X_N$ is the value of $X$ when it is stopped, and it is also the limit as $n \to \infty$ of $X_{N\wedge n}$.

Must we have $E X_0 \leq E X_N$?

In general, no

Let $S$ be the symmetric simple random walk with $S_0 = 0$.

Let $N = \min \{ n > 0 : S_n = -1 \}$.

Then $S$ is a submartingale and $N$ is a stopping time with respect to the filtration generated by $S$.

Then $E S_0 = 0$, but $S_N = -1$, and so $E S_N = -1$.

The trouble is that $N$ can be very very large.

When $N$ is bounded, yes

Theorem 4.4.1, p. 246

Let $X$ be a submartingale and let $N$ be a stopping time for which $N \leq k$ a.s. Then $E X_0 \leq E X_N \leq E X_k$.

Proof

1. $E X_0 = E X_{N \wedge 0} \leq E X_{N \wedge k} = E X_k$ since $0 \leq N$ and $X_{N \wedge n}$, $n = 0, 1, 2, \ldots$ is a submartingale.

2. For the next inequality, look at the increments $X_k - X_{N\wedge n}$, essentially:

Now $X_n - X_{N\wedge n} = \text{sum of increments from } N \text{ to } n$, or zero if $n < N$

$= (H \cdot X)_n$, where

$H_n = 1 \{ n > N \}$

$= 1 \{ N \leq n - 1 \} \in \mathcal{F}_{n-1}$,

so $H$ is predictable. Thus, $H \cdot X$ is a submartingale, so,

$E X_k - E X_{N \wedge k}$ since $N \leq k$

$= E (H \cdot X)_k$

$\geq E (H \cdot X)_0$ $H \cdot X$ is a submartingale.

\[ \square \]
Running maximum

Let $X$ be a stochastic process.

Set $\bar{X}_n = \max_{0 \leq m \leq n} X^+_m$ 

$= \max (0, x_0, x_1, ..., x_n)$

For dominated convergence, it would help to have a bound on the size of $\bar{X}$.

One can get a bound on $\bar{X}_n$ in terms of $X_n$, which reduces the maximum over many r.v. to the value of one r.v., which is very helpful.

4.4.2 Doob's inequality

Let $X$ be a submartingale, $\lambda > 0$, $n > 0$.

Then $P(\bar{X}_n \geq \lambda) \leq \frac{1}{\lambda} E[X_n 1_{\{\bar{X}_n \geq \lambda\}}] \leq \frac{1}{\lambda} E[X^+_n]$.

Thus, the probability that the running maximum $\bar{X}_n$ is greater than $\lambda$ can be bounded in terms of $\lambda$ and the size of $X^+_n$.

Proof

Let $N = \min (m : X_m \geq \lambda) \wedge n$ (to make $N$ bounded).

If $\bar{X}_n \geq \lambda$, then $X_N \geq \lambda$; $X$ has already hit height $\lambda$, $N$ is just the first time that occurred.

$P(\bar{X}_n \geq \lambda) = E 1_{\{\bar{X}_n \geq \lambda\}}$ 

$\leq E \frac{X_N}{\lambda} 1_{\{\bar{X}_n \geq \lambda\}}$

$= \frac{1}{\lambda} \left[ E X_N - E X_N 1_{\{\bar{X}_n < \lambda\}} \right]$ 

$\leq \frac{1}{\lambda} \left[ E X_n - E X_n 1_{\{\bar{X}_n < \lambda\}} \right]$ 

$= \frac{1}{\lambda} E X_n 1_{\{\bar{X}_n \geq \lambda\}}$ 

$\leq \frac{1}{\lambda} E X^+_n 1_{\{\bar{X}_n \geq \lambda\}}$ 

$\leq \frac{1}{\lambda} E X^+_n$

As in Chebychev's inequality,

$\bar{X}_n \geq \lambda \Rightarrow X_N \geq \lambda$.

N.B., Thm 4.4.1.

If $\bar{X}_n < \lambda$, then $N = n$.

(the first inequality)

trivial.
Example 4.4.1  Kolmogorov's inequality

If $A_1, A_2, \ldots$ are independent, mean 0, finite variance,
and we set $S_n = A_1 + \cdots + A_n$, then $S$ is a martingale, so
$S_n^2$, $n = 0, 1, 2, \ldots$ is a submartingale.

Then
\[
P\left( \max_{0 \leq m \leq n} |S_m| \geq x \right) = P\left( \max_{0 \leq m \leq n} S_m^2 \geq x^2 \right)
\leq \frac{1}{x^2} E S_n^2
= \frac{\text{Var}(S_n)}{x^2}.
\]

Remarks
Always remember that there are techniques to bound the running maximum of
sums of iid random variables.

This bound looks much like Chebyshev's inequality, which would read:
\[
P\left( |S_n| \geq x \right) \leq \frac{\text{Var}(S_n)}{x^2}.
\]

What we have just shown is better, though, because we have shown
something stronger, since
\[
P\left( |S_n| \geq x \right) \leq P\left( \max_{0 \leq m \leq n} |S_m| \geq x \right) \leq \frac{\text{Var}(S_n)}{x^2}.
\]
Theorem 4.4.3  $L^p$ maximum inequality

(i) If $X$ is a submartingale and $p > 1$, then $E X_n^p \leq \left( \frac{p}{p-1} \right)^b E (X_n^+)^p$.

(ii) If $Y$ is a martingale then

$$E \max_{0 \leq n \leq m} |Y_n|^b \leq \left( \frac{p}{p-1} \right)^b E |Y_m|^p.$$

Proof  Similar to the book, but simpler.

(i) $E X_n^p = \int_0^\infty \lambda^{p-1} P(X_n - \lambda > 0) d\lambda$

$\leq \int_0^\infty d\lambda \lambda^{p-1} \frac{1}{\lambda} E X_n 1 \{ X_n > \lambda \} $

$$\leq \int_0^\infty d\lambda \lambda^{p-2} E X_n^+ 1 \{ X_n > \lambda \}$$

$$= \int_0^\infty d\lambda \lambda^{p-1} 1 \{ X_n > \lambda \}$$

$$= \int_0^\infty d\lambda \lambda^{p-2} X_n^+$$

$$= \frac{p}{p-1} E X_n^+ \left[ X_n^{p-1} - O^{p-1} \right]$$

[We have our foot in the door: $E X_n^p$ is replaced by $E X_n^+ X_n^{p-1}$; involves $X_n$ directly]

Hölder: If $\frac{1}{a} + \frac{1}{b} = 1$, then $E |XY| \leq \|X\|_a \|Y\|_b = (E |X|^a)^{\frac{1}{a}} (E |Y|^b)^{\frac{1}{b}}$

Use this to define $b$; then $b = \frac{1}{1 - \frac{1}{p}} = \frac{p}{p-1}$, so $(p-1) b = p$:

$E X_n^p \leq \frac{p}{p-1} \left( E (X_n^+)^p \right)^{\frac{1}{p}} \left( E (X_n^{p-1})^b \right)^{\frac{1}{p}}$

$$= \frac{p}{p-1} \left( E (X_n^+)^p \right)^{\frac{1}{p}} \left( E X_n^p \right)^{1 - \frac{1}{p}}$$

Divide through by $\left( E X_n^p \right)^{1 - \frac{1}{p}}$ to get

$$\left( E X_n^p \right)^{\frac{1}{p}} \leq \frac{p}{p-1} \left( E (X_n^+)^p \right)^{\frac{1}{p}}$$

Now raise to the $p$th power:

$$E X_n^p \leq \left( \frac{p}{p-1} \right)^b E (X_n^+)^b.$$

(ii) Let $X_n = |Y_n|$. Then $X$ is a submartingale and

$X_n = \max_{0 \leq n \leq m} |Y_n|$

$X_n^+ = |Y_n|$. Easy!
Theorem 4.4.5 \(L^p\) convergence theorem.

If \(X\) is a martingale with \(\sup_n E |X_n|^p < \infty\) where \(p > 1\),
then \(X_\infty = \lim_{n \to \infty} X_n\) exists a.s. and \(X_n \to X_\infty\) in \(L^p\).

Proof

\((E |X_n|)^p \leq E |X_n|^p\) by Jensen's inequality, since \(p > 1\).

So \(\sup_n E |X_n| < \infty\), so \(X_n\) converges by the martingale conv. thm.

Let \(X_\infty = \lim_{n \to \infty} X_n\).

For the dominated convergence part, note that

\[|X_n - X| \leq |X_n| + |X|\]

\[\leq \sup_n |X_n| + \sup_n |X_n|\]

so \(|X_n - X|^p \leq 2^p \sup_n |X_n|^p\).

But

\[E \sup_n |X_n|^p = E \lim_{n \to \infty} \max_{0 \leq \omega \leq n} |X_n|^p\]

\[= \lim_{n \to \infty} E \max_{0 \leq \omega \leq n} |X_n|^p\quad \text{MCT}\]

\[\leq \lim_{n \to \infty} \left( \frac{p}{p-1} \right)^p E |X_n|^p\quad \text{L^p max ineq.}\]

\[\leq \left( \frac{p}{p-1} \right)^p \sup_n E |X_n|^p\]

\[< \infty,\]

by assumption.

By dominated convergence, \(E |X_n - X_\infty|^p \to 0\) as \(n \to \infty\). \(\square\)
4.5 Uniform integrability and $L^1$ convergence of martingales

$(\Omega, \mathcal{F}, (\mathcal{F}_n), P)$ filtration $\mathcal{F} = (\mathcal{F}_1, \mathcal{F}_2, \ldots)$, stochastic process $X = (X_0, X_1, \ldots)$

If $X$ is a sub/super/martingale with respect to $\mathcal{F}$ and $\sup_n E |X_n| < \infty$, then the Martingale Convergence Theorem says that there exists $X_\infty \in L^1$ for which $X_n \to X_\infty$ almost surely.

However, it is possible that $X_n$ does not converge to $X_\infty$ in $L^1$.

We need a slightly stronger condition than $\sup_n E |X_n| < \infty$.

We will give a condition that is equivalent to $E |X_n - X_\infty| \to 0$, but is easier to check.

Uniform integrability

Let $I$ be a set, called the index set. It may be countable or uncountable.

A collection $\{X_i, i \in I\}$ of real-valued random variables is uniformly integrable if $\lim_{M \to \infty} \sup_{i \in I} E |X_i| 1\{|X_i| > M\} = 0$.

Relation to integrability

1. If $E |X_i| < \infty$, then $1_M < \infty$ almost surely, so $|X_i| 1\{|X_i| > M\} \to 0$ a.s. as $M \to \infty$. These random variables are dominated by $|X_i|$, which is integrable, so by Dominated Convergence, $\lim_{M \to \infty} E |X_i| 1\{|X_i| > M\} = 0$.

Uniform integrability simply makes this limit uniform over the set $I$.

2. Conversely, if $\lim_{M \to \infty} E |X_i| 1\{|X_i| > M\} = 0$, then for some value of $M$,

$E |X_i| 1\{|X_i| > M\} < 1$, and so

$E |X_i| = E |X_i| 1\{|X_i| > M\} + E |X_i| 1\{|X_i| \leq M\}$

< 1 + M < \infty,

and so $X_i$ is integrable.

3. If $\{X_i, i \in I\}$ is uniformly integrable, then $\sup_{i \in I} E |X_i| < \infty$:

For some $M$, $\sup_{i \in I} E |X_i| 1\{|X_i| > M\} < 1$, so

$\sup_{i \in I} E |X_i| 1 \leq \sup_{i \in I} E |X_i| 1\{|X_i| > M\} + \sup_{i \in I} E |X_i| 1\{|X_i| \leq M\}$

< 1 + M < \infty.

I do not think the converse is true. Can you find a counterexample to it? Note: uniform integrability is thus stronger than the condition in the mart. conv. thm.
Relation to tightness

For each $i \in I$, define a measure $\nu_i$ on $\mathbb{R}$ by

$$\nu_i(A) = \mathbb{E} \left[ 1_{\mathbb{R}} \mathbb{1}_A(X_i) \right]$$

This is not necessarily a probability measure on $\mathbb{R}$, for $\nu_i(\mathbb{R}) = \mathbb{E}|X_i|$, and all we know is that this is finite.

The measure $\nu_i$ is not the distribution of $X_i$. The distribution $\mu_i$ is defined by $\mu_i(A) = \mathbb{P} (X_i \in A) = \mathbb{E} \mathbb{1}_A(X_i) = \nu_i(A)$.

By contrast, $\nu_i$ puts relatively more mass on sets $A$ where $X_i$ is large.

Still, $\nu_i(A) = \int_A |x| \mu_i(dx)$.

Tightness is defined for probability measures, but if we simply apply the definition to $\{\nu_i, i \in I\}$, it would read:

For all $\epsilon > 0$, there exists $M$ such that

for all $i \in I$, $\nu_i([-M,M]^c) < \epsilon$.

This is equivalent to:

for all $\epsilon > 0$, there exists $M$ such that $\sup_{i \in I} \nu_i([-M,M]^c) < \epsilon$.

Or,

$$\lim_{M \to \infty} \sup_{i \in I} \nu_i([-M,M]^c) = 0$$

Or finally,

$$\lim_{M \to \infty} \sup_{i \in I} \mathbb{E} |X_i| \mathbb{1}_{\{|X_i| > M\}} = 0.$$

Thus, uniform integrability is somehow equivalent to the size-weighted measures $\{\nu_i, i \in I\}$ being tight. They must put essentially all their mass in a compact set.

Main example

Let $\mathcal{F}$ be a filtration and let $X_n \in \mathcal{F}$ be integrable.

For each $n = 0, 1, 2, \ldots$, set $X_n = \mathbb{E}[X_n | \mathcal{F}_n]$.

Show $X$ is a martingale:

(i) $X_n \in \mathcal{F}_n$

(ii) $X_n$ is integrable

(iii) $\mathbb{E}[X_n + \mathcal{F}_n] = X_n$

We get $X_n$ by smoothing/averaging $X_n$ given the information available at time $n$.

We will see that this martingale is uniformly integrable, and that every uniformly integrable martingale is of this form.
Theorem 4.5.1 \((\Omega, \mathcal{F}, P)\). Let \(X : \Omega \to \mathbb{R}\) be measurable and integrable.

Let \(\mathcal{U} = \{ E[X|\mathcal{F}] : \mathcal{F} \subseteq \mathcal{F}\\text{ is a }\sigma\text{-algebra} \}\).

Then \(\mathcal{U}\) is uniformly integrable. \(\text{Intuition: } E[X|\mathcal{F}] \leq X\) for the most part.

Proof

1. Claim If \(P(A) \leq P(B)\), then \(P(A \cap B^c) \leq P(A^c \cap B)\).

\[
P(A) = P(A \cap B) + P(A \cap B^c) \\
\leq P(B) = P(A \cap B) + P(A^c \cap B) .
\]

2. Let \(\epsilon > 0\)

Let \(M\) be large enough that \(E_{|X| > M} < \epsilon\).

Let \(B = \{ |x| > M \}\). Then \(E_{|X| > B} < \epsilon\).

3. Claim If \(A \in \mathcal{F}\) has \(P(A) \leq P(B)\), then \(E_{|X| > B} < E_{|X| > A}\).

\[
E_{|X| > B} = E_{|X| > B} + E_{|X| > A} \\
\leq E_{|X| > B} + E_{|X| > M} \text{ for } A \subseteq B.
\]

4. Let \(N\) be such that \(\frac{E_{|X| > N}}{N} < P(B)\). \(N\) depends on \(M\).

Let \(\mathcal{F} \subseteq \mathcal{F}\) be a \(\sigma\)-algebra.

Now \(E(E[X|\mathcal{F}]) \leq E(|X|)\) mononotonicity of conditional expectation.

So \(E_{E|X| > N} \leq E_{E|X| > N} \leq E_{E|X| > N} \leq E_{E|X| > N}\) (easy)

Let \(A = \{ E|X| > N \}\). Then \(A \in \mathcal{F}\).

Then \(P(A) = P(E|X| > N) \leq \frac{E_{E|X| > N}}{N} \text{ Chebyshev, \(c_p(a) = a\)}

\[
= \frac{E_{E|X| > N}}{N} \leq P(B).
\]

By 3 and 2, \(E_{|X| > B} < \epsilon\), and thus,

\(E_{E|X| > N} \leq \epsilon\)

for all \(\mathcal{F} \subseteq \mathcal{F}\) a \(\sigma\)-algebra, so that \(\mathcal{U}\) is uniformly integrable. \(\square\)
The next result shows that, in the presence of convergence in probability, uniform integrability is equivalent to convergence in $L^1$. This result is not restricted to martingales. \underline{Example} $X_n = n \mathbb{1}_{[0, \frac{1}{n}]}$ converges in prob. to 0, but not in $L^1$. Not uniformly integrable.

\textbf{Theorem 4.5.2}

Suppose $X_n \to X$ in probability as $n \to \infty$. Then the following are equivalent:

1. \{X_n, n \geq 0\} is uniformly integrable
2. $X_n \to X$ in $L^1$
3. $\mathbb{E} |X_n|, \mathbb{E} |X| < \infty$, and $\mathbb{E} |X_n| \to \mathbb{E} |X|$

\textbf{Proof}

(i) $\Rightarrow$ (ii)

Idea: When $X_n$ and $X$ are big, we can control them using uniform integrability. When $X_n$ and $X$ are small, they are close by convergence in probability.

Let $\epsilon > 0$.

Fix $M > 0$ and set $\phi(x) = \min(M, \max(-M, x))$

Then $\phi$ is bounded and continuous, and $|x - \phi(x)| \leq |x| 1_{\{|x| > M\}}$.

We break the difference $|X_n - X|$ into three parts using the triangle inequality:

$$|X_n - X| \leq |X_n - \phi(X_n)| + |\phi(X_n) - \phi(x)| + |\phi(x) - X|$$

$$\leq |X_n| 1_{\{|X_n| > M\}} + |\phi(X_n) - \phi(x)| + |x| 1_{\{|x| > M\}}$$

Take expected values and $\limsup$. Then, for all $M > 0$,

$$\limsup_{n \to \infty} \mathbb{E} |X_n - x| \leq \sup_n \mathbb{E} |X_n| 1_{\{|X_n| > M\}} + \limsup_{n \to \infty} \mathbb{E} |\phi(x) - \phi(x)| + \mathbb{E} |X| 1_{\{|x| > M\}}$$

- By uniform integrability, by making $M$ large enough the first term is less than $\frac{\epsilon}{2}$.
- We will show that $\lim_{n \to \infty} \mathbb{E} |\phi(x) - \phi(x)| = 0$.

Let $n_k$ be a subsequence. Then $X_{n_k} \to X$ in probability, and so there exists a further subsequence $n_{k_2}$, for which $X_{n_{k_2}} \to X$ a.s.

Because $\phi$ is continuous, $|\phi(X_{n_{k_2}}) - \phi(x)| \to 0$ a.s. as $k \to \infty$, and because $\phi$ is bounded, $\mathbb{E} |\phi(X_{n_{k_2}}) - \phi(x)| \to 0$. Thus, $\lim_{n \to \infty} \mathbb{E} |\phi(x) - \phi(x)| = 0$.

Let $n_k$ be a subsequence for which $X_{n_k} \to X$ a.s. Then,

$$\mathbb{E} |x| = \mathbb{E} \liminf_{n \to \infty} |X_n|$$

$$\leq \liminf_{n \to \infty} \mathbb{E} |X_n|$$

$$\leq \sup_{n \to \infty} \mathbb{E} |X_n|$$

by (3) on the first page of this section. Thus, by (1) of the first page,

$$\limsup_{n \to \infty} \mathbb{E} |X_n - X| \to 0$$

Thus, $\limsup_{n \to \infty} \mathbb{E} |X_n - X| < \epsilon$, for all $\epsilon > 0$, and thus $\limsup_{n \to \infty} \mathbb{E} |X_n - X| = 0$. 

(ii) \implies (iii)

This is simple. \( |E[X_n| - E[X]| \leq E |X_n - X| \leq E |X_n - X|, \)
by the "reverse triangle inequality." By the Squeeze Law, \( E|X_n| \to E|X|, \)
and the fact that \( X_n \to X \) in \( L^1 \) requires \( E|X_n|, E|X| < \infty \) in the first place.

(iii) \implies (i)

Here again we need to use convergence in probability, so we need a bounded, continuous function.
Let \( \varepsilon > 0 \). Fix \( n > 0 \).
Fix \( M > 1 \).

Define \( \eta_M \) by
\[
\eta_M(x) = \begin{cases} |x|, & |x| \leq M - 1 \\ \text{linear}, & M - 1 < |x| < M \\ 0, & |x| \geq M 
\end{cases}
\]
so that \( \eta_M \) is continuous.

Then, although bounded, \( \eta_M \) is not continuous; see dotted lines above.

the sum of three useful terms. Start with the last one.

\[
E |X_n| 1 \{ |X_n| > M \} = E |X_n| - E |X_n| \cdot 1 \{ |X_n| \leq M \} \\
\leq E |X_n| - E \eta_M(X_n) \\
= E |X_n| - E |X| \\
- E \{ \eta_M(X_n) - \eta_M(x) \} - E \eta_M(x) + E |X|,
\]

By choosing \( M \) large enough, we can make this less than \( \frac{\varepsilon}{3} \), since \( E |X| < \infty \).

Now that \( M \) is decided, we know that \( E |\eta_M(X_n) - \eta_M(x)| \to 0 \)
as \( n \to \infty \) by Bounded Convergence, so for \( n > N \), this is less than \( \frac{\varepsilon}{3} \).

Similarly, for \( n > N_2 \), \( |E |X_n| - E |X| | < \frac{\varepsilon}{3} \), by assumption (iii).

Thus, \( \sup_{n \geq N} E |X_n| 1 \{ |X_n| > M \} < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \).

Moreover, by increasing \( M \), the left-hand side will decrease even further.

Choose \( M_2 \) large enough that \( M_2 > M \) and also,
\[
\max_{0 \leq n \leq N} E |X_n| 1 \{ |X_n| > M_2 \} < \varepsilon.
\]

This is the maximum of a finite collection of numbers.
All of them are finite, by the assumption that \( E|X_n| < \infty \),
and we know that for each \( 0 \leq n \leq N \), \( E |X_n| 1 \{ |X_n| > M \} \to 0 \) as \( M \to \infty \).

Thus, we may choose \( M_2 \) to make all of them less than \( \varepsilon \).

Thus, for all \( \varepsilon > 0 \), there exists \( M_3 > 0 \) such that
\[
\sup_{n} E |X_n| 1 \{ |X_n| > M_3 \} < \varepsilon,
\]
and so the collection \( \{ X_n, n=0,1,2, \ldots \} \) is uniformly integrable.
Theorem 4.5.6, p. 262

Let $X$ be a martingale with respect to $\mathcal{F}$. The following are equivalent:

(i) $X$ is uniformly integrable
(ii) $X_n$ converges almost surely and in $L^1$
(iii) $X_n$ converges in $L^1$
(iv) There exists an integrable random variable $X_\infty$ for which $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$, $n=0,1,2,...$

Proof

(i) $\Rightarrow$ (ii)

From (i) on the first page of this section, $\sup_n \mathbb{E} |X_n| < \infty$.

By the Martingale Convergence Theorem, $X_n$ converges a.s. to a random variable $X_\infty$.

Almost sure convergence implies convergence in probability, so by Theorem 4.5.2, $X_n$ converges to $X_\infty$ in $L^1$.

(ii) $\Rightarrow$ (iii)

This is obvious.

(iii) $\Rightarrow$ (iv)

Suppose $X_n$ converges in $L^1$. Let $X_\infty$ denote its limit.

Claim: $\mathbb{E}[X_\infty | \mathcal{F}_n] = X_n$ for each $n=0,1,2,...$

Fix a value of $n$. We must check the definition of conditional expectation:

(a) $X_n \in \mathcal{F}_n$, by the fact that $X$ is a martingale

(b) Let $A \in \mathcal{F}_n$. We need to show that $\mathbb{E} 1_A X_n = \mathbb{E} 1_A X_\infty$. Now,

$$\mathbb{E} 1_A X_n = \mathbb{E} [1_A \mathbb{E}[X_n | \mathcal{F}_n]]$$

for all $n \geq n$; $X$ is a martingale

$$= \mathbb{E} \left[ \mathbb{E} [1_A X_\infty | \mathcal{F}_n] \right]$$

since $A \in \mathcal{F}_n$

$$= \mathbb{E} 1_A X_\infty$$

Property #9

This is true for all $n \geq n$.

But as $m \to \infty$, $\mathbb{E} 1_A X_m \to \mathbb{E} 1_A X_\infty$, for

$$| \mathbb{E} 1_A X_m - \mathbb{E} 1_A X_\infty | \leq \mathbb{E} |X_m - X_\infty|$$

$$\leq \mathbb{E} |X_m - X_n|,$$

which goes to 0 as $m \to \infty$, since $X_n \to X_\infty$ in $L^1$.

Thus, $\mathbb{E} 1_A X_n = \mathbb{E} 1_A X_\infty$ and the claim is proven.

(iv) $\Rightarrow$ (i)

This follows from Theorem 4.5.1.
Theorem 4.5.7 Suppose $\mathcal{F}$ is a filtration and set $\mathcal{F}_\infty = \sigma \left( \bigcup_{n=1}^{\infty} \mathcal{F}_n \right)$.

Let $X \in L^1$. Then $\mathbb{E}[X | \mathcal{F}_n] \to \mathbb{E}[X | \mathcal{F}_\infty]$ a.s.

Remark
The meaning is that these successively finer views of $X$ converge to the right limit. However, they need not converge to $X$ unless $X \in \mathcal{F}_\infty$.

Proof
We have seen that the sequence $X_n = \mathbb{E}[X | \mathcal{F}_n]$ is a martingale. Theorem 4.5.1 says it is uniformly integrable, and Theorem 4.5.2 says it converges in $L^1$ to some $X_\infty$, and in addition,

$$X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$$

Claim $\mathbb{E}[X | \mathcal{F}_\infty] = X_\infty$.

(i) $X_n \in \mathcal{F}_n \subseteq \mathcal{F}_\infty$ for all $n$,
so $X_\infty = \lim_{n \to \infty} X_n$ is in $\mathcal{F}_\infty$ by the usual measurability of limits.

(ii) We need to show that, for all $A \in \mathcal{F}_\infty$, $\mathbb{E}X_\infty 1_A = \mathbb{E}X 1_A$.
Let $\mathcal{L} = \{ A \in \mathcal{F}_\infty : \mathbb{E}X_\infty 1_A = \mathbb{E}X 1_A \}$.
It is easy to check that $\mathcal{L}$ is a $\lambda$-system.
The collection $\mathcal{P} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$ is a $\pi$-system, and for all $A \in \mathcal{P}$, $A \in \mathcal{F}_n$ for some $n$, and

$$\mathbb{E}X_\infty 1_A = \mathbb{E}X_n 1_A$$

since $X_n = \mathbb{E}[X_n | \mathcal{F}_n]$

$$= \mathbb{E}X 1_A$$

since $X_n = \mathbb{E}[X | \mathcal{F}_n]$.
By the $\pi$-$\lambda$ theorem, $\mathcal{L}$ contains $\sigma(\mathcal{P}) = \mathcal{F}_\infty$.

Corollary: Lévy's 0-1 Law
If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and $A \in \mathcal{F}_\infty$, then $\mathbb{E}[1_A | \mathcal{F}_n] \to 1_A$ a.s.

Note: If $X$ is in the tail of $\mathcal{F}$, it could be that $X \not\in \mathcal{F}_n$, so that $\mathbb{E}[X | \mathcal{F}_n] = \mathbb{E}X$ for all $n$, and so $X \in \mathcal{F}_\infty$ has to be constant. This is similar to a step in the proof of Kolmogorov's 0-1 law.
Running a martingale backwards - intuition

Write $E[X_n, |F_n] = X_n$ or $X_n = E[X_{n+1} | F_n]$, so that $X_n$ is an averaged or smoothed version of $X_{n+1}$.

$X_n \xrightarrow{\text{smoothing}} X_{n+1}$

$X_n$ is a simpler version of $X_{n+1}$, in a sense.

Example with opposite intuition

It can happen that there is some "consolidation" in the values of $X$. 

![Diagram](image-url)