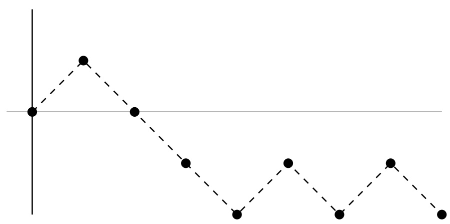
1 Probability space, events, and random variables

Examples
We begin with some common examples of probability models:

1. Coin tossing: We can encode a series of coin tosses as a random binary sequence 11010001.... Denoting it by $\omega_1\omega_2\omega_3\ldots$, we expect to see that
$$\frac{\omega_1 + \cdots + \omega_n}{n} \to \frac{1}{2}.$$

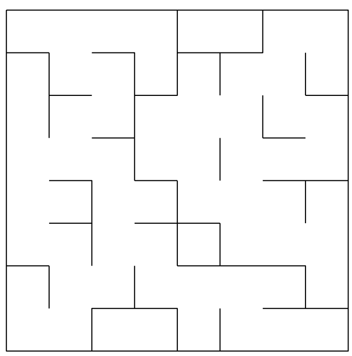
2. Random walk on $\mathbb{Z}$:



How far is S_n from the origin? $S_n \approx \sqrt{n}$.
What is the probability that the random walk will reach 20 before -10? Equals 1/3.
How long do we have to wait until we hit 20 or -10? 200 in expectation.

3. Brownian motion: "random walk in continuous time".
Its paths are nowhere differentiable with probability 1.
It is widely used in PDEs, financial maths, mathematical physics, biology, chemistry, etc.

4. Percolation:



What is the chance that there is an infinite connected cluster?
Either 0 or 1 (and never a number in between!), critical probability for bond percolation on $\mathbb{Z}^2$ is 1/2, and decreases in higher dimensions.

5. Rolling dice: Two people wait to see a particular sequence from successive rolls of a die – one waits for 456, the other for 666.

Who will wait the longest?
What is the average waiting time?

Aims of this course

1. Introduce a rigorous setup to study random objects (so far you have only used combinatorics (with its limited capacity) plus formulas for expectation without proof).
2. Strong Law of Large Numbers:
   If $X_1, X_2, \ldots$ are independent random variables with the same distribution and mean μ, then
   $$\frac{X_1 + \cdots + X_n}{n} \to \mu.$$
3. Central Limit Theorem:
   If $X_1, X_2, \ldots$ are independent random variables with the same distribution, mean μ and variance σ^2, then
   $$\frac{X_1 + \cdots + X_n}{n} \approx \mu + \frac{\sigma}{\sqrt{n}} Z,$$
   where Z is a normal random variable. So the fluctuations are μ are of order $1/\sqrt{n}$ and the randomness does not depend on the distribution of our original random variable.
4. Martingales: random processes which on average stay the same (fair games). Very useful: SLLN, examples 2 and 5 above, many more.

Reminder from measure theory

A family Σ of subsets of a set Ω is called a *σ-algebra* if

1. $\varnothing \in \Sigma$,
2. $A \in \Sigma \implies \Omega \setminus A \in \Sigma$,
3. $A_1, A_2, \ldots \in \Sigma \implies \bigcup_{n=1}^{\infty} A_n \in \Sigma$

A function $\mu : \Sigma \to [0, \infty]$ is called a *measure* if

1. $\mu(\varnothing) = 0$,
2. $A_1, A_2, \ldots \in \Sigma$ disjoint $\implies \mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$

A function $f : \Omega \to \mathbb{R}$ is called *measurable* if $f^{-1}(B) \in \Sigma$ for all Borel set B.

Definition 1.1.
A measure μ is called a *probability measure* if $\mu(\Omega) = 1$.
A *probability space* is a triple $(\Omega, \Sigma, \mathbb{P})$.
A *random variable* is a measurable function $X : \Omega \to \mathbb{R}$.

Idea
Ω = all possible outcomes of a random experiment,
Σ = all possible events,
$\mathbb{P}(A)$ = probability of an event $A \in \Sigma$.

Example 1.2.

1. Coin toss, or X equals 0 or 1 with probabilities 1/2:
   $\Omega = \{H, T\}$,
   $\Sigma = \{\varnothing, \{H\}, \{T\}, \Omega\}$
   $\mathbb{P}(\varnothing) = 0$, $\mathbb{P}(\{H\}) = 1/2$, $\mathbb{P}(\{T\}) = 1/2$, $\mathbb{P}(\Omega) = 1$,
   $X : \{H, T\} \to \mathbb{R}$, $X(H) = 1$, $X(T) = 0$

2. Roll a die and spell the number you get – let X be the number of letters:
   $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\Sigma = 2^\Omega$ (set of all subsets of Ω),
   $\mathbb{P}(\{1\}) = \cdots = \mathbb{P}(\{6\}) = 1/6$ and extend to Σ by additivity,
   $X : \{1, 2, 3, 4, 5, 6\} \to \mathbb{R}$,
   $X(1) = 3$, $X(2) = 3$, $X(3) = 5$, $X(4) = 4$, $X(5) = 4$, $X(6) = 3$.
   Then $\mathbb{P}(X = 3) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = 3)\}) = \mathbb{P}(\{1, 2, 6\}) = 1/2$.

3. We can represent the die experiment in more than one way (i.e. on different probability spaces), e.g.:
   $\Omega = \{0, 1, 2, 3, 4, 5, 6\}$, $\Sigma = 2^\Omega$,
   $\mathbb{P}(\{0\}) = 0$, $\mathbb{P}(\{1\}) = \cdots = \mathbb{P}(\{6\}) = 1/6$ and extend to Σ by additivity,
   $X : \{0, 1, 2, 3, 4, 5, 6\} \to \mathbb{R}$,
   $X(0) = 4$, $X(1) = 3$, $X(2) = 3$, $X(3) = 5$, $X(4) = 4$, $X(5) = 4$, $X(6) = 3$.

4. Toss a coin infinitely many times:
   Ω = set of all sequences consisting of 0s and 1s. We can code such a sequence as a number in $[0, 1]$ by considering each sequence as a binary number. So $\Omega = [0, 1]$ but what is Σ? What kind of events do we want to measure? We certainly want to be able to measure events that a sequence begins with a certain pattern, e.g. 110:

   $$\{\omega = 0.\omega_1\omega_2\omega_3 \cdots \in [0, 1] : \omega_1 = 1, \omega_2 = 1, \omega_3 = 0\} = [3/4, 7/8].$$

   So Σ should contain all binary intervals. The smallest σ-algebra with this property is $\mathcal{B}$, the Borel σ-algebra, so $\Sigma = \mathcal{B}$.
   How should we define $\mathbb{P}$? We want $\mathbb{P}(110 * * \cdots) = (1/2)^3 = 1/8$, and observing that Leb([3/4,7/8])=1/8 (and similarly for all patterns), we set $\mathbb{P} = \text{Leb}$. So our probability space is $([0, 1], \mathcal{B}, \text{Leb})$. We can consider various random variables on this space, for example, $X(\omega) = \omega_n$ (representing the nth toss).

We often care about the probabilities

$$\mathbb{P}(X \in B) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}),$$

for Borel sets B, rather than which ω occur. We can record these probabilities with a probability measure μ on $(\mathbb{R}, \mathcal{B})$ defined by

$$\mu(B) = \mathbb{P}(X \in B), \quad B \in \mathcal{B}.$$

This measure is entirely determined by its values on the sets $(-\infty, t]$, $t \in \mathbb{R}$. Denote

$$F(t) = \mu((-\infty, t]) = \mathbb{P}(X \leq t), \quad t \in \mathbb{R}.$$

Definition 1.3.
The measure μ is called the *distribution* or *law* of X.
The function F is called the *distribution function* of X.

Conversely, we can ask: what properties should a function F have to be a distribution function for some random variable? We begin with some properties of a distribution function and then show that if a function F has these properties, there exists a random variable with F its distribution function.

Theorem 1.4 (Properties of a distribution function)**.**
Let F be a distribution function. Then

1. *F is increasing,*
2. $\lim_{t\to\infty} F(t) = 1$*,* $\lim_{t\to-\infty} F(t) = 0$*,*
3. *F is right continuous.*

Proof.

1. Let $s \leq t$. Then
$$F(t) = \mathbb{P}(\{\omega : X(\omega) \leq s\} \cup \{\omega : X(\omega) \in (s, t]\}) \geq \mathbb{P}(\{\omega : X(\omega) \leq s\}) = F(s).$$
2. Let $t_n \to \infty$ (and set $t_0 = -\infty$). Then
$$\begin{aligned} F(t_n) &= \mathbb{P}\left(\bigcup_{i=1}^{n} \{\omega : X(\omega) \in (t_{i-1}, t_i]\}\right) = \sum_{i=1}^{n} \mathbb{P}(\{\omega : X(\omega) \in (t_{i-1}, t_i]\}) \\ &\to \sum_{i=1}^{\infty} \mathbb{P}(\{\omega : X(\omega) \in (t_{i-1}, t_i]\}) = \mathbb{P}(\Omega) = 1. \end{aligned}$$
Similarly, letting $t_n \to -\infty$,
$$F(t_n) = \mathbb{P}(X \leq t_n) \to \mathbb{P}(\varnothing) = 0.$$
3. Let $t_n \downarrow t$. Then
$$F(t_n) = \mathbb{P}\left(\{X \leq t\} \cup \{X \in (t, t_n]\}\right) = \mathbb{P}(X \leq t) + \mathbb{P}(X \in (t, t_n]) \to F(t).$$

$\square$

Theorem 1.5 (Skorokhod Representation)**.**
Let $F : \mathbb{R} \to [0, 1]$ have the three properties of a distribution function above. Then there is a random variable X on $([0, 1], \mathcal{B}, \mathrm{Leb})$ with distribution function F.

Proof. If F is continuous and strictly increasing, then we can just take $X(\omega) = F^{-1}(\omega)$. Then

$$\text{Leb}(X \leq u) = \text{Leb}(\{\omega : F^{-1}(\omega) \leq u\}) = \text{Leb}(\{\omega : \omega \leq F(u)\}) = F(u).$$

For more general F, we proceed as follows: Define the "inverse" $G : [0,1] \to \mathbb{R}$ by

$$G(\omega) = \inf\{t : F(t) > \omega\}.$$

We set $X(\omega) = G(\omega)$ and prove that $\text{Leb}(X \leq u) = F(u)$. It suffices to show that

$$[0, F(u)) \subseteq \{\omega \in [0,1] : \inf\{t : F(t) > \omega\} \leq u\} \subseteq [0, F(u)].$$

First suppose $\omega \in [0, F(u))$. Then $F(u) > \omega$ so $u \in \{t : F(t) > \omega\}$ and so

$$\inf\{t : F(t) > \omega\} \leq u.$$

On the other hand, set $\omega \in [0,1]$ be such that $\inf\{t : F(t) > \omega\} \leq u$. Then by monotonicity of F,

$$F(\inf\{t : F(t) > \omega\}) \leq F(u),$$

and so by right-continuity,

$$\inf\{F(t) : F(t) > \omega\} \leq F(u).$$

But now since $\omega \leq \inf\{F(t) : F(t) > \omega\}$ we have that $\omega \leq F(u)$. $\square$

Definition 1.6.
If F (the distribution function of a random variable X) can be written as an integral:

$$F(t) = \int_{-\infty}^{t} f(u)\, du,$$

for some measurable function f, then X is called a *continuous random variable* with *density* f.

Remark 1.7.

1. If F is differentiable then X has density $f = F'$.
2. If f is a density then $\int_{-\infty}^{\infty} f(t)\, dt = \mathbb{P}(\Omega) = 1$.
3. If X has density f and law μ, then $\mu \ll \text{Leb}$ ($\text{Leb}(A) = 0 \implies \mu(A) = 0$) and f is the Radon-Nikodym derivative $\frac{d\mu}{d\text{Leb}}$.

Example 1.8 (Examples of distributions)**.**

1. Uniform random variable on $[0,1]$ ("random number between 0 and 1").

$$F(t) = \begin{cases} 1, & \text{if } t > 1, \\ t, & \text{if } 0 \leq t \leq 1, \\ 0, & \text{if } t < 0. \end{cases} \quad \text{and} \quad f(t) = \begin{cases} 1, & \text{if } 0 \leq t \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

2. Exponential random variable with mean μ.

$$F(t) = \begin{cases} 1 - e^{-t/\mu}, & \text{if } t > 0, \\ 0, & \text{if } t \leq 0. \end{cases} \quad \text{and} \quad f(t) = \begin{cases} \frac{1}{\mu} e^{-t/\mu}, & \text{if } t > 0, \\ 0, & \text{if } t \leq 0. \end{cases}$$

3. Normal (or Gaussian) random variable with mean μ and variance σ^2 (denoted $N(\mu, \sigma^2)$).

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

4. Bernoulli (discrete) random variable with parameter $p \in [0, 1]$.

$$\mathbb{P}(X = 0) = 1 - p, \qquad \mathbb{P}(X = 1) = p.$$

   Can compute F:

$$F(t) = \begin{cases} 1, & \text{if } t \geq 1, \\ 1 - p, & \text{if } 0 \leq t < 1 \\ 0, & \text{if } t < 0. \end{cases}$$

5. Poisson (discrete) random variable with mean λ.

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

Definition 1.9.
Let X be a random variable on $(\Omega, \Sigma, \mathbb{P})$. If X is integrable, then

$$\mathbb{E}(X) = \int X \, d\mathbb{P}$$

is called the *expectation* of X. (If X is not integrable, but $X \geq 0$ we write $\mathbb{E}(X) = \infty$).

Does this definition agree with the calculation rules used in elementary probability? Such as

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} t f(t) \, dt, \qquad \text{for continuous random variables,} \\ \mathbb{E}(X) &= \sum_i a_i \mathbb{P}(X = a_i), \qquad \text{for discrete random variables.} \end{aligned}$$

Theorem 1.10.
Let X be a random variable on $(\Omega, \Sigma, \mathbb{P})$ with law μ and h an integrable measurable function on $(\mathbb{R}, \mathcal{B}, \mu)$. Then

$$\int h(X) \, d\mathbb{P} =: \mathbb{E}(h(X)) = \mu(h) := \int h(t) \, d\mu(t).$$

In particular, if X has density f, then

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(t) f(t) \, dt,$$

and if X is discrete (taking countably many values) then

$$\mathbb{E}(h(X)) = \sum_i h(a_i) \mathbb{P}(X = a_i).$$

In order to prove this, we recall some results from measure theory.

Reminder from measure theory
Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and $f : \Omega \to \mathbb{R}$ a measurable function. The integral $\int f \, d\mathbb{P}$ is defined in the following three steps:

1. If $f = \sum_{i=1}^n a_i \mathbf{1}_{\{A_i\}}$ (a simple function), then $\int f \, d\mathbb{P} = \sum_{i=1}^n a_i \mathbb{P}(A_i)$.
2. If $f \geq 0$ then $\int f \, d\mathbb{P} = \lim_{n\to\infty} \int f_n \, d\mathbb{P}$, where f_n are simple, nonnegative and with $f_n \uparrow f$.
3. If f is arbitrary define $f_+ = \max(f, 0)$ and $f_- = \max(-f, 0)$, write $f = f_+ - f_-$ and define $\int f \, d\mathbb{P} = \int f_+ \, d\mathbb{P} - \int f_- \, d\mathbb{P}$ (if both of these parts are finite; otherwise we say that f is not integrable).

Theorem 1.11 (MON)**.**
If f and (f_n) are measurable functions from Ω to $[0, \infty]$ such that $f_n \uparrow f$ then $\int f_n \, d\mathbb{P} \uparrow \int f \, d\mathbb{P}$.

Theorem 1.12 (DOM)**.**
If f and (f_n) are measurable functions from Ω to $\mathbb{R}$ such that $f_n \to f$ and $|f_n| \leq g$ for some measurable integrable function g then $\int f_n \, d\mathbb{P} \to \int f \, d\mathbb{P}$.

Proof of Theorem 1.10.
We will check the statement for more and more general h. The essence of the proof is contained in the first step and the standard machinery is used to do the next steps. To begin with we suppose h is an indicator function.

1. $h = \mathbf{1}_{\{B\}}$, $B \in \mathcal{B}$:
$$\mathbb{E}(h(X)) = \int h(X) \, d\mathbb{P} = \mathbb{P}(X \in B),$$
$$\int_{\mathbb{R}} h(t) \, d\mu(t) = \int_{\mathbb{R}} \mathbf{1}_{\{B\}}(t) \, d\mu(t) = \mu(B) = \mathbb{P}(X \in B).$$
2. $h = \sum_{i=1}^n a_i \mathbf{1}_{\{B_i\}}$, all $a_i \in \mathbb{R}$, $B_i \in \mathcal{B}$:
Use linearity of the expectation and the integral.
3. $h \geq 0$ integrable:
Choose simple $h_n \uparrow h$. Then $h_n(X) \uparrow h(X)$ and we apply MON to both sides.
4. h integrable:
Write $h = h_+ - h_-$ and use additivity of both sides.

If X has density f, then it is the Radon-Nikodym derivative of μ with respect to Leb and
$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(t) \, d\mu(t) = \int_{\mathbb{R}} h(t) f(t) \, dt.$$
If X is discrete then $\mu(\{a_i\}) = \mathbb{P}(X = a_i)$ and so
$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(t) \, d\mu(t) = \sum h(a_i) \mathbb{P}(X = a_i).$$
□

Definition 1.13.
Let X be a square-integrable (i.e. $\mathbb{E}(X^2) < \infty$) random variable. Then

$$\operatorname{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2,$$

is called the *variance* of X.

Lemma 1.14 (Square-integrability implies integrability)**.**
If $\mathbb{E}(X^2) < \infty$, then $\mathbb{E}|X| < \infty$ and so $\mathbb{E}(X) < \infty$.

Proof. By the Cauchy-Schwarz inequality,

$$\mathbb{E}|X| \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(1^2)} < \infty.$$

□

Theorem 1.15 (Markov's/Chebyshev's Inequality)**.**
Let X be a non-negative random variable and $c > 0$. Then $c\mathbb{P}(X \geq c) \leq \mathbb{E}(X)$.

Proof. Let

$$Y(\omega) = \begin{cases} c, & \text{if } X(\omega) \geq c, \\ 0, & \text{otherwise.} \end{cases}$$

Then $X(\omega) \geq Y(\omega)$ for all ω and so taking expectations

$$\mathbb{E}(X) \geq \mathbb{E}(Y) = c\,\mathbb{P}(X \geq c).$$

□

2 Independence

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space.

Definition 2.1.

- Two events $A, B \in \Sigma$ are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
- Two σ-algebras $\Sigma_1, \Sigma_2 \subset \Sigma$ are *independent* if

$$P(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2) \quad \text{for all } A_1 \in \Sigma_1,\ A_2 \in \Sigma_2.$$

- Finitely many σ-algebras $\Sigma_1, \ldots, \Sigma_n$ are *independent* if

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n) \quad \text{for all } A_1 \in \Sigma_1, \ldots, A_n \in \Sigma_n.$$

- Countably many σ-algebras $\Sigma_1, \Sigma_2, \ldots$ are *independent* if $\Sigma_1, \ldots, \Sigma_n$ are independent for each n.

Definition 2.2.
Let X be a random variable on $(\Omega, \Sigma, \mathbb{P})$. Then

$$\sigma(X) := \{\{\omega : X(\omega) \in B\} : B \in \mathcal{B}\}$$

is called the *σ-algebra generated by X*.
Random variables (X_n) are *independent* if $(\sigma(X_n))$ are independent.

Example 2.3.

1. Toss a coin and roll a die: gives two random numbers X and Y. We can model this experiment on a probability space

$$\begin{aligned}\Omega &= \{0,1\}\times\{1,2,3,4,5,6\},\ \Sigma = 2^{\Omega},\ \mathbb{P}(\{\omega\}) = 1/12,\\ \omega &= (\omega_1,\omega_2),\ X(\omega) = \omega_1,\ Y(\omega) = \omega_2,\\ \sigma(X) &= \{\varnothing, \Omega, \{X=0\}, \{X=1\}\},\\ \sigma(Y) &= \{\varnothing, \Omega, \{Y=1\}, \dots, \{Y=6\}, \{Y=1, Y=2\}, \dots\}.\end{aligned}$$

   We can check if X and Y are independent, for example if

$$\frac{1}{12} = \mathbb{P}(X=0,\, Y=1) = \mathbb{P}(X=0)\mathbb{P}(Y=1) = \frac{1}{2}\cdot\frac{1}{6}.$$

   X and Y are indeed independent.

2. We first toss a coin to get X. If the coin comes up tails, we set $Y = 6$, otherwise we roll a die to obtain Y. We can model this experiment on the space:

$$\Omega = \{0,1\}\times\{1,2,3,4,5,6\},\ \Sigma = 2^{\Omega},\ \mathbb{P}(\{1,\omega_2\}) = 1/12,\ \mathbb{P}(\{0,6\}) = 1/2,\ \mathbb{P}(\{0,\omega_2\}) = 0,\ \omega_2 \neq 6,$$

   and with $X, \sigma(X)$ and $Y, \sigma(Y)$ as before. X and Y are not independent since

$$\mathbb{P}(X=0, Y=1) = 0 \neq \mathbb{P}(X=0)\mathbb{P}(Y=1) = 1/24.$$

Definition 2.4.
A subcollection $\mathcal{I} \subset \Sigma$ is a π-*system* if $A \cap B \in \mathcal{I}$ whenever $A, B \in \mathcal{I}$ (closed under finite intersections).

Example 2.5.

1. $\{(-\infty, t],\, t \in \mathbb{R}\}$ and $\{\varnothing, (a,b),\, a < b\}$ are two π-systems which generate $\mathcal{B}$.
2. $\{\{X < t\},\, t \in \mathbb{R}\}$ and $\{\{X \leq t\},\, t \in \mathbb{R}\}$ are two π-systems which generate $\sigma(X)$ for a random variable X.
3. If X takes countably many values $a_1, a_2, \dots$ then $\{\varnothing, \{X = a_1\}, \{X = a_2\}, \dots\}$ is a π-system generating $\sigma(X)$.

Theorem 2.6.
Let $\mathcal{I}$ be a π-system and μ_1 and μ_2 be two measures on $(\Omega, \sigma(\mathcal{I}))$ such that $\mu_1 = \mu_2$ on $\mathcal{I}$ and $\mu_1(\Omega) = \mu_2(\Omega) < \infty$. Then $\mu_1 = \mu_2$ on $\sigma(\mathcal{I})$.

Theorem 2.7.
If two π-systems $\mathcal{I}$ and $\mathcal{J}$ are independent, that is,

$$\mathbb{P}(I \cap J) = \mathbb{P}(I)\mathbb{P}(J)$$

whenever $I \in \mathcal{I}$, $J \in \mathcal{J}$, then $\sigma(\mathcal{I})$ and $\sigma(\mathcal{J})$ are independent.

Proof. Fix $I \in \mathcal{I}$ and set

$$\mu_1(B) = \mathbb{P}(I \cap B),$$
$$\mu_2(B) = \mathbb{P}(I)\mathbb{P}(B)$$

for $B \in \sigma(\mathcal{J})$. They agree for $B \in \mathcal{J}$ and satisfy $\mu_1(\Omega) = \mu_2(\Omega) = \mathbb{P}(I) < \infty$. Hence by Theorem 2.6 they agree on $\sigma(\mathcal{J})$, i.e.

$$\mathbb{P}(I \cap B) = \mathbb{P}(I)\mathbb{P}(B)$$

whenever $I \in \mathcal{I}$, $B \in \sigma(\mathcal{J})$. Now fix $B \in \sigma(\mathcal{J})$ and consider

$$\mu_1(A) = \mathbb{P}(A \cap B),$$
$$\mu_2(A) = \mathbb{P}(A)\mathbb{P}(B),$$

for $A \in \sigma(\mathcal{I})$. They agree on $\mathcal{I}$ and satisfy $\mu_1(\Omega) = \mu_2(\Omega) = \mathbb{P}(B) < \infty$. Hence by Theorem 2.6 they agree on $\sigma(\mathcal{I})$. This implies

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

whenever $A \in \sigma(\mathcal{I})$, $B \in \sigma(\mathcal{J})$. □

Corollary 2.8.
Continuous random variables X and Y are independent if and only if

$$\mathbb{P}(X \le x, \, Y \le y) = \mathbb{P}(X \le x)\, \mathbb{P}(Y \le y),$$

for all x, y.
Discrete random variables X and Y taking values $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ respectively, are independent if and only if

$$\mathbb{P}(X = a_i, \, Y = b_j) = \mathbb{P}(X = a_i)\, \mathbb{P}(Y = b_j)$$

for all i, j.

Infinite and Finite Occurrence of Events

Example 2.9.
Let (X_n) be independent Bernoulli random variables. How do we formalise the intuitively clear statement "1 occurs infinitely often with probability 1"?
Let $E_n = \{X_n = 1\}$ and denote

$$\begin{aligned} E &= \{1 \text{ occurs infinitely often}\} \\ &= \{\forall N \, \exists n \ge N : \, X_n = 1\} \\ &= \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} E_n. \end{aligned}$$

The statement "1 occurs infinitely often with probability 1" means $\mathbb{P}(E) = 1$.

Notation:
Let (E_n) be a sequence of events.

- The event which contains all ω belonging to infinitely many of (E_n) is denoted by

$$\{E_n \text{ i.o.}\} = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} E_n.$$

- The complement of this events contains all ω belonging to only finitely many of (E_n) and is

$$\{E_n \text{ i.o}\}^c = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} E_n^c.$$

We have to deal with infinite/finite occurence of events when we discuss limits, eg. in the Strong Law of Large Numbers:

Theorem 2.10 (SLLN).
Let (X_n) be independent identically distributed random variables such that $\mathbb{E}|X_1| < \infty$. Then

$$\frac{X_1 + \ldots + X_n}{n} \to \mathbb{E}(X_1) \qquad \textit{almost surely.}$$

Notation:
We shall use the abbreviation *i.i.d.* for "independent identically distributed" and *a.s.* for "almost surely".

The Strong Law of Large Numbers is a very important and hard-to-prove theorem. We will prove it towards the end of the course in this form, but along the way we shall prove it for specific cases:

- Bernoulli r.v.s via a direct calculation.
- Bounded r.v.s (exercise – similar to Bernoulli)
- Square-integrable r.v.s – an easy proof once we have developed the theory of martingales.
- General case – hard proof using martingale theory.

What will we actually need to prove?
Denote $S_n = X_1 + \cdots + X_n$ and $\mu = \mathbb{E}(X_1)$. We want to show that $\frac{S_n}{n} \to \mu$ as $n \to \infty$ with probability 1. Consider the event

$$E = \left\{\omega \in \Omega : \frac{S_n(\omega)}{n} \to \mu\right\}.$$

We want to show that $\mathbb{P}(E) = 1$. We have

$$\begin{aligned} E &= \left\{ \forall k \, \exists N : \forall n \geq N, \, \left|\frac{S_n}{n} - \mu\right| < 1/k \right\} \\ &= \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ \left|\frac{S_n}{n} - \mu\right| < 1/k \right\} \\ &= \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ \left|\frac{S_n}{n} - \mu\right| \geq 1/k \right\}^c \\ &= \bigcap_{k=1}^{\infty} \left\{ \left|\frac{S_n}{n} - \mu\right| \geq 1/k \text{ i.o.} \right\}^c . \end{aligned}$$

Proving $\mathbb{P}(E) = 1$ is the same as proving that

$$\begin{aligned} \mathbb{P}\left(\left\{ \left|\frac{S_n}{n} - \mu\right| \geq 1/k \text{ i.o.} \right\}^c \right) &= 1 \quad \text{for all } k, \text{ i.e.} \\ \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq 1/k \text{ i.o.} \right) &= 0 \quad \text{for all } k. \end{aligned}$$

Theorem 2.11 (Borel-Cantelli Lemma 1: BC1)**.**
Let (E_n) be events such that $\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$. Then $\mathbb{P}(E_n \text{ i.o.}) = 0$.

Theorem 2.12 (Borel-Cantelli Lemma 2: BC2)**.**
Let (E_n) be independent events such that $\sum \mathbb{P}(E_n) = \infty$. Then $\mathbb{P}(E_n \text{ i.o.}) = 1$.

The requirement of independence cannot be removed. For example suppose E is an event with $0 < \mathbb{P}(E) < 1$ and set $E_n = E$ for all n. Then $\sum \mathbb{P}(E_n) = \infty$ but $\mathbb{P}(E_n \text{ i.o.}) = \mathbb{P}(E) \neq 1$.

Proof of BC1. For each N,

$$\mathbb{P}\left(\bigcup_{n=N}^{\infty} E_n \right) \leq \sum_{n=N}^{\infty} \mathbb{P}(E_n) \to 0,$$

as $N \to \infty$ by the assumption that the series converges. Now,

$$\mathbb{P}(E_n \text{ i.o.}) = \mathbb{P}\left(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} E_n \right) \leq \lim_{N \to \infty} \mathbb{P}\left(\bigcup_{n=N}^{\infty} E_n \right) = 0. \qquad \square$$

Proof of BC2. We want to prove

$$\mathbb{P}\left(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} E_n \right) = 1, \text{ i.e. } \mathbb{P}\left(\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} E_n^c \right) = 0, \text{ i.e. } \mathbb{P}\left(\bigcap_{n=N}^{\infty} E_n^c \right) = 0, \forall N.$$

Using $1 - x \le e^{-x}$ for all $x \in \mathbb{R}$, we have

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{n=N}^{\infty} E_n^c\right) &\le \lim_{k\to\infty} \mathbb{P}\left(\bigcap_{n=N}^{k} E_n^c\right) = \lim_{k\to\infty} \prod_{n=N}^{k} \mathbb{P}(E_n^c) \\
&= \lim_{k\to\infty} \prod_{n=N}^{k} (1 - \mathbb{P}(E_n)) \le \lim_{k\to\infty} \prod_{n=N}^{k} e^{-\mathbb{P}(E_n)} \\
&= \lim_{k\to\infty} e^{-\sum_{n=N}^{k} \mathbb{P}(E_n)} \\
&= e^{-\lim_{k\to\infty} \sum_{n=N}^{k} \mathbb{P}(E_n)} \\
&= 0,
\end{aligned}$$

where the first equality uses the independence assumption, and the last uses $\sum \mathbb{P}(E_n) = \infty$. $\square$

Example 2.13.
Let (X_n) be an iid sequence of exponential random variables. We shall show that

$$\limsup_{n\to\infty} \frac{X_n}{\log n} = 1. \tag{1}$$

Recall: $\limsup a_n =$ largest number a such that there is a subsequence (a_{n_k}) converging to a. To prove $\limsup a_n = a$ it suffices to show that

- $\forall\, b > a$, $a_n > b$ only finitely often,
- $\forall\, b < a$, $a_n > b$ infinitely often.

Let us prove (1):

- Let $b > 1$. We want to show that

  $$\mathbb{P}(X_n > b \log n \text{ i.o.}) = 0.$$

  We shall use BC1:

  $$\sum_{n=1}^{\infty} \mathbb{P}(X_n > b\log n) = \sum_{n=1}^{\infty} e^{-b\log n} = \sum_{n=1}^{\infty} \frac{1}{n^b} < \infty,$$

  since $b > 1$.

- Let $b < 1$. We want to show that

  $$\mathbb{P}(X_n > b \log n \text{ i.o.}) = 1.$$

  We shall use BC2:

  $$\sum_{n=1}^{\infty} \mathbb{P}(X_n > b\log n) = \sum_{n=1}^{\infty} \frac{1}{n^b} = \infty,$$

  as $b < 1$.

Example 2.14. let (X_n) be an iid sequence.

- Suppose $\mathbb{E}|X_1| = \mu < \infty$. Then by SLLN

$$\frac{X_n}{n} = \frac{X_1 + \cdots + X_n}{n} - \frac{X_1 + \cdots + X_{n-1}}{n-1} \cdot \frac{n-1}{n} \to \mu - \mu \cdot 1 = 0,$$

  almost surely.

- Suppose $\mathbb{E}|X_1| = \infty$ (so SLLN is not applicable). Then

$$\limsup_{n\to\infty} \frac{|X_n|}{n} = \infty.$$

  We can prove this using BC2. We want

$$\begin{aligned} \mathbb{P}\left(\limsup_{n\to\infty} \frac{|X_n|}{n} = \infty\right) &= 1 \\ \text{i.e.} \quad \mathbb{P}\left(\forall m \, \forall N \, \exists n \geq N : \frac{|X_n|}{n} > m\right) &= 1 \\ \text{i.e.} \quad \mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \left\{\frac{|X_n|}{n} > m\right\}\right) &= 1 \\ \text{i.e.} \quad \mathbb{P}\left(\bigcap_{m=1}^{\infty} \left\{\frac{|X_n|}{n} > m \text{ i.o.}\right\}\right) &= 1 \\ \text{i.e.} \quad \mathbb{P}\left(\frac{|X_n|}{n} > m \text{ i.o.}\right) &= 1 \quad \forall m. \end{aligned}$$

  The events $\{|X_n|/n > m\}$, $n \in \mathbb{N}$ are independent and

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}\left(\frac{|X_n|}{n} > m\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(\frac{|X_1|}{n} > m\right) &= \sum_{n=1}^{\infty} \mathbb{E}\mathbf{1}_{\{\frac{|X_1|}{n} > m\}} \\ &\overset{\text{MON}}{=} \mathbb{E}\sum_{n=1}^{\infty} \mathbf{1}_{\{\frac{|X_1|}{n} > m\}} = \mathbb{E}\sum_{n=1}^{\infty} \mathbf{1}_{\{n < |X_1|/m\}} = \mathbb{E}\left(\frac{|X_1|}{m} - 1\right) = \infty. \end{aligned}$$

  The result now follows from BC2.

Natural model for independent random variables

Let μ_1 and μ_2 be two laws. We want to construct random variables X_1 and X_2 such that

- X_1 has law μ_1 and X_2 has law μ_2,
- X_1 and X_2 are independent.

Method:

- Construct a random variable $\tilde{X}_1$ on $(\Omega_1, \Sigma_1, \mathbb{P}_1)$ with law μ_1 (exists by Skorokhod Representation Theorem).

- Construct a random variable $\tilde{X}_2$ on $(\Omega_2, \Sigma_2, \mathbb{P}_2)$ with law μ_2.
- Let $(\Omega, \Sigma, \mathbb{P}) = (\Omega_1 \times \Omega_2, \Sigma_1 \times \Sigma_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$.
- Define $X_1(\omega_1, \omega_2) = \tilde{X}_1(\omega_1)$, $X_2(\omega_1, \omega_2) = \tilde{X}_2(\omega_2)$. Exercise: show that X_1 and X_2 satisfy the two desired properties above.

Theorem 2.15 (Independence and expectation/variance)**.**
Let X and Y be two independent random variables.

1. *If $\mathbb{E}|X| < \infty$ and $\mathbb{E}|Y| < \infty$ then $\mathbb{E}|XY| < \infty$ and $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.*
2. *If $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$ then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.*

Proof. 1. It suffices to prove the result for simple random variables (and the result will then hold generally using the standard machine). So let

$$X = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{\{A_i\}}, \qquad Y = \sum_{j=1}^{m} \beta_j \mathbf{1}_{\{B_j\}},$$

with $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_m \in \mathbb{R}$ and $A_1, \ldots, A_n, B_1, \ldots, B_m \in \Sigma$. Then we have

$$\begin{aligned}
\mathbb{E}(XY) = \mathbb{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i\beta_j \mathbf{1}_{\{A_i\}}\mathbf{1}_{\{B_j\}}\right) &= \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i\beta_j \mathbb{E}\left(\mathbf{1}_{\{A_i \cap B_j\}}\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i\beta_j \mathbb{P}(A_i \cap B_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i\beta_j \mathbb{P}(X_i = \alpha_i,\, Y = \beta_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i\beta_j \mathbb{P}(X = \alpha_i)\mathbb{P}(Y = \beta_j) \\
&= \left(\sum_{i=1}^{n} \alpha_i \mathbb{P}(X = \alpha_i)\right)\left(\sum_{j=1}^{m} \beta_j \mathbb{P}(Y = \beta_j)\right) = \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}$$

2. We have

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}((X+Y)^2) - (\mathbb{E}(X+Y))^2 \\
&= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - (\mathbb{E}(X))^2 - 2\mathbb{E}(X)\mathbb{E}(Y) - (\mathbb{E}(Y))^2 \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y),
\end{aligned}$$

since $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

□

Example 2.16. Let X be a $\mathrm{Binom}(n,p)$ random variable, so that

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad 0 \le k \le n.$$

We know that $X = X_1 + \cdots + X_n$ where X_i are iid Bernoulli(p). Then

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = np, \quad \mathrm{Var}(X) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = np(1-p).$$

Theorem 2.17 (Bernstein's Inequality).
let $X_1, \dots, X_n$ be iid random variables with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$. Then

$$\mathbb{P}\left(|\sum_{i=1}^{n} a_i X_i| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sum a_i^2}\right),$$

for any $a_1, \dots, a_n \in \mathbb{R}$.

Proof. Denote $c = \sum a_i^2$. For any $\lambda > 0$,

$$\mathbb{E}\left(\exp\left\{\lambda \sum a_i X_i\right\}\right) = \mathbb{E}\left(\prod e^{\lambda a_i X_i}\right) = \prod \mathbb{E}\left(e^{\lambda a_i X_i}\right) = \prod \cosh(\lambda a_i).$$

Using $\cosh(x) \leq e^{x^2/2}$ (exercise), we get

$$\mathbb{E}\left(\exp\left\{\lambda \sum a_i X_i\right\}\right) \leq \prod e^{\lambda^2 a_i^2/2} = e^{c\lambda^2/2}.$$

We have, using Markov's Inequality,

$$\begin{aligned}
\mathbb{P}(\sum a_i X_i \geq t) &= \mathbb{P}(\exp\{\lambda \sum a_i X_i\} \geq e^{\lambda t}) \\
&\leq e^{-\lambda t}\mathbb{E}\left(\exp\{\lambda \sum a_i X_i\}\right) \\
&\leq \exp\left\{c\lambda^2/2 - \lambda t\right\}.
\end{aligned}$$

Differentiating, we find optimal $\lambda = t/c$. Substituting this gives

$$\mathbb{P}(\sum a_i X_i \geq t) \leq e^{-t^2/2c}.$$

Similarly,

$$\mathbb{P}(\sum a_i X_i \leq -t) \leq e^{-t^2/2c}.$$

$\square$

Theorem 2.18 (SLLN for Bernoulli random variables).
Let (X_i) be iid with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$. Then

$$\frac{X_1 + \cdots + X_n}{n} \to 0 \quad a.s.$$

Proof.
Denote $S_n = X_1 + \cdots + X_n$. We want to show that $\mathbb{P}(S_n \to 0)$ i.e.

$$\mathbb{P}(|S_n| \geq n/k \text{ i.o.}) = 0 \quad \forall k.$$

This follows from BC1 and Bernstein's Inequality with $a_1 = \cdots = a_n = 1$:

$$\sum_{n=1}^{\infty} \mathbb{P}(|S_n| \geq n/k) \leq 2\sum_{n=1}^{\infty} e^{-n/2k^2} < \infty.$$

$\square$

Exercise: Using this method, prove SLLN for bounded random variables.

Definition 2.19.
Let X and Y be random variables on $(\Omega, \Sigma, \mathbb{P})$. The *joint law* of (X, Y) is the probability measure $\mu_{X,Y}$ on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ given by

$$\mu_{X,Y}(B) = \mathbb{P}((X, Y) \in B), \quad B \in \mathcal{B}(\mathbb{R}^2).$$

The *joint distribution function* $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$ is defined by

$$F_{X,Y}(x, y) = \mathbb{P}(X \le x, Y \le y) = \mu_{X,Y}((-\infty, x] \times (-\infty, y]).$$

This definition naturally extends to any finite number of random variables.

Theorem 2.20.

1. *X and Y are independent if and only if $F_{X,Y}(x, y) = F_X(x)F_Y(y)$.*
2. *If X and Y are independent and have densities f and g then $\mu_{X,Y}$ has density $(x, y) \mapsto f(x)g(y)$, i.e.*
$$\mu_{X,Y}(B) = \int_B f(x)g(y)\, dxdy, \quad B \in \mathcal{B}(\mathbb{R}^2).$$
3. *If X and Y are independent and have densities f and g, then $X + Y$ has density*
$$(f * g)(t) = \int_{-\infty}^{\infty} f(s)g(t - s)\, ds,$$
which is called the convolution of f and g.

Proof.

1. $\sigma(X)$ is generated by the π-system $\{X \le x, x \in \mathbb{R}\}$ and $\sigma(Y)$ is generated by the π-system $\{Y \le y, y \in \mathbb{R}\}$. These are independent iff
$$\mathbb{P}(X \le x, Y \le y) = \mathbb{P}(X \le x)\mathbb{P}(Y \le y).$$
2. For $B = (-\infty, x] \times (-\infty, y]$,
$$\begin{aligned}\mu_{X,Y}(B) &= \mu_{X,Y}((-\infty, x] \times (-\infty, y]) = F_{X,Y}(x, y) = F_X(x)F_Y(y)\\ &= \int_{-\infty}^{x} f(u)\, du \int_{-\infty}^{y} g(v)\, dv = \int_B f(u)g(v)\, dudv.\end{aligned}$$
These two measures agree on the π-system
$$\{(-\infty, x] \times (-\infty, y] : x, y \in \mathbb{R}\}$$
and hence on $\mathcal{B}(\mathbb{R}^2)$.

3. For any t,

$$\begin{aligned} F_{X+Y}(t) &= \mathbb{P}(X+Y \le t) = \mu_{X,Y}(\{(x,y) : x+y \le t\}) \\ &= \int_{\{(x,y) : x+y\le t\}} f(x)g(y)\, dxdy \\ &= \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{t-x} g(y)\, dydx. \end{aligned}$$

Substituting $y = v - x$, gives

$$F_{X+Y}(t) = \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{t} g(v-x)\, dvdx = \int_{-\infty}^{t} \int_{-\infty}^{\infty} f(x)g(v-x)\, dxdv.$$

Hence

$$f_{X+Y}(v) = \int_{-\infty}^{\infty} f(x)g(v-x)\, dx.$$

□

Definition 2.21.
Let (X_i) be random variables. For each $n \ge 0$, set $\tau_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$. τ_n is called the *n-th tail σ-algebra* of (X_i) and $\tau := \bigcap_{n=0}^{\infty} \tau_n$ is called the *tail σ-algebra* of (X_i). An event E is a *tail event* if $E \in \tau$.

Remark 2.22.

- τ_n contains events which do not depend on $X_1, \ldots, X_n$.
- τ contains events which do not depend on any finite number of the X_i.

Example 2.23.

- $\{X_n \to a\}$ is a tail event: for each m,

$$\{X_n \to a\} = \{X_{m+n} \to a\} = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{|X_{m+n} - a| < 1/k\} \in \tau_m,$$

  and so $\{X_n \to a\} \in \tau$.

- $\{\lim_{n\to\infty} X_n \text{ exists}\}$, $\{\sum X_n < \infty\}$ are tail events.
- $\{\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} \text{ exists}\}$ is a tail event: for each m,

$$\frac{X_1 + \cdots + X_n}{n} = \frac{X_1 + \cdots + X_m}{n} + \frac{X_{m+1} + \cdots + X_n}{n}.$$

  The first term always converges to 0, so convergence of $(X_1 + \cdots + X_n)/n$ is equivalent to convergence of $(X_{m+1} + \cdots + X_n)/n$ and so does not depend on $X_1, \ldots, X_m$. Since this is true for each m, this is a tail event.

- $\{\sup X_n > 0\}$ is not a tail event. Let $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = 1/2$ and $X_n = 0$ for all $n \geq 2$. Then $\tau_n = \{\varnothing, \Omega\}$ for all $n \geq 2$ and so $\tau = \{\varnothing, \Omega\}$. But $\{\sup X_n > 0\} = \{X_n = 1\} \notin \tau$ since $\mathbb{P}(X_1 = 1) = 1/2$.

Theorem 2.24 (Kolmogorov's 0-1 law)**.**
If (X_n) is a sequence of independent random variabes then every tail event has probability 0 or 1.

Proof. Let $\sigma_n = \sigma(X_1, \ldots, X_n)$ and recall $\tau_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$. σ_n is generated by the π-system

$$\{X_1 \in B_1, \ldots, X_n \in B_n : B_1, \ldots, B_n \in \mathcal{B}\}.$$

τ_n is generated by the π-system

$$\{X_{n+1} \in B_{n+1}, \ldots, X_{n+m} \in B_{n+m} : m \in \mathbb{N}, B_{n+1}, \ldots, B_{n+m} \in \mathcal{B}\}$$

These π-systems are independent and so σ_n and τ_n are independent. Since $\tau \subset \tau_n$, σ_n and τ are independent for each n. Set $\sigma_\infty = \sigma(X_1, X_2, \ldots)$. σ_∞ is generated by the π-system $\sigma_1 \cup \sigma_2 \cup \cdots$ which is independent of τ. Hence τ and σ_∞ are independent. Since $\tau \subset \sigma_\infty$, this implies τ is independent of itself. So each $A \in \tau$ satisfies

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2,$$

and so $\mathbb{P}(A) \in \{0, 1\}$. □

Example 2.25. If (X_n) are independent, then

1. $\mathbb{P}(X_n \to a)$, $\mathbb{P}(\lim X_n \text{ exists})$, $\mathbb{P}(\sum X_n < \infty), \ldots \in \{0, 1\}$.
2. $$\mathbb{P}(X_n/n \to 0) = \begin{cases} 1, & \text{if } \mathbb{E}|X_1| < \infty \text{ (follows from SLLN seen)}, \\ 0, & \text{if } \mathbb{E}|X_1| = \infty \text{ (seen)}. \end{cases}$$

3 Weak Convergence

When are two random variables "close"? There are many ways in which two random variables can be considered similar, for example:

- Strong sense: $X = Y$ almost surely, i.e. $\mathbb{P}(X = Y) = 1$
- Weak sense: $\mu_X = \mu_Y$

Note that for almost-sure convergence to make sense, the two random variables X and Y must lie in the same probability space. This requirement is not required for the weak sense of convergence. Let X and Y be two independent Bernoulli random variables. They are not close in the almost-sure sense since

$$\mathbb{P}(X = Y) = \mathbb{P}(X = 0, Y = 0) + \mathbb{P}(X = 1, Y = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

But they are close (in fact equal) in the weak sense as they have the same distribution. Later in the course we shall prove the Central Limit Theorem which states that for i.i.d. random variables with mean μ and variance σ^2:

$$\text{the law of} \quad \frac{X_1 + \cdots + X_n - \mu n}{\sigma\sqrt{n}} \approx N(0,1).$$

We make this notion of weak convergence precise in the following definition.

Definition 3.1 (Weak convergence)**.**
Let (X_n) and X be random variables with laws (μ_n) and μ and distribution functions (F_n) and F. (X_n) is said to *converge weakly* to X (or μ_n converges weakly to μ) if $F_n(t) \to F(t)$ for all t where F is continuous.

Remark 3.2.

1. Weak convergence is often called convergence in law or in distribution. It is usually denoted
$$\xrightarrow{d} \text{ or } \xrightarrow{w} \text{ or } \Rightarrow$$

2. Why does the definition avoid points of discontinuity? Consider the example $X_n(\omega) = 1/n$ for all ω and n (so that the distribution functions are step functions). According to our definition $X_n \xrightarrow{d} 0$ which is natural. But note that F_n converges to a function which is not right-continuous at 0. So if we did not exclude $t = 0$ then the limiting function would not be a distribution function of any random variable.

3. Let (X_n) be a sequence of independent Bernoulli random variables (a sequence of coin tosses). Then (X_n) does not converge almost surely, but it does converge weakly to a Bernoulli random variable (trivially).

Theorem 3.3 (Relation between weak and almost-sure convergence)**.**

1. *If $X_n \to X$ almost surely, then $X_n \xrightarrow{d} X$.*

2. *If $\mu_n \xrightarrow{d} \mu$ then there exist random variables (X_n) and X defined on the same probability space such that each X_n has law μ_n, X has law μ and $X_n \to X$ almost surely.*

Theorem 3.4 (Useful definition of weak convergence)**.**
$\mu_n \xrightarrow{d} \mu$ if and only if $\int h\, d\mu_n \to \int h\, d\mu$ for every continuous bounded function $h : \mathbb{R} \to \mathbb{R}$.

We shall prove these theorems in the following order:
Theorem 3.3 (2), Theorem 3.4 ($\Rightarrow$), Theorem 3.4 ($\Leftarrow$), Theorem 3.3 (1).

Proof of Theorem 3.3 (2).
Denote by (F_n) and F the distribution functions associated to laws (μ_n) and μ (i.e. $F_n(t) = \mu_n((-\infty, t])$ and $F(t) = \mu((-\infty, t])$). We shall use the Skorokhod Representation to construct the random variables (X_n) and X. Let the probability space be $([0,1], \mathcal{B}, \text{Leb})$ and define (X_n) and X as:

$$X_n(\omega) = \inf\{t : F_n(t) > \omega\}, \quad X(\omega) = \inf\{t : F(t) > \omega\}.$$

We shall show that $X_n \to X$ except on a set of (Lebesgue) measure 0. This set is defined as

$$B := \{\omega \in [0,1] : \exists x \neq y \in \mathbb{R},\, F(x) = F(y) = \omega\}.$$

We claim that $\mathrm{Leb}(B) = 0$ and $X_n(\omega) \to X(\omega)$ for all $\omega \in [0,1] \setminus B$, i.e. $X_n \to X$ almost surely.

1. $\mathrm{Leb}(B) = 0$:
   We prove this by showing that B has countably many elements. To every $\omega \in B$, we associate an interval (x_ω, y_ω) where we have $F(x_\omega) = F(y_\omega) = \omega$. By monotonicity of F, the intervals do not overlap (and hence each such interval contains only one $\omega \in B$). Furthermore, every interval contains a rational. Since the set of rationals, $\mathbb{Q}$, is countable, there can only be countably many intervals and so B is countable. Thus $\mathrm{Leb}(B)$=0.

2. F has countably many discontinuity points:
   To every discontinuity point t, associate the interval $(\lim_{s\uparrow t} F(s), \lim_{s\downarrow t} F(s))$. By monotonicity of F the intervals do not overlap. Furthermore, every interval contains a rational. Hence as $\mathbb{Q}$ is countable, there are countably many discontinuity points.

3. Fix an $\omega \in [0,1] \setminus B$. Then $X_n(\omega) \to X$:
   For ease of notation, set $x = X(\omega)$ and let $\varepsilon > 0$. Since there are only countably many points of discontinuity of F, we can find a δ with $0 < \delta \leq \varepsilon$ such that $x - \delta$ and $x + \delta$ are continuity points of F. Since $\omega \notin B$, $F(x-\delta) < \omega < F(x+\delta)$. Also, since $x-\delta$ and $x+\delta$ are continuity points, by the definition of weak convergence, $F_n(x\pm\delta) \to F(x\pm\delta)$. Hence for all n sufficiently large

   $$F_n(x - \delta) < \omega < F_n(x + \delta).$$

   Then, by the definition of X_n, for such n we have

   $$x - \delta < X_n(\omega) \leq x + \delta.$$

   Hence for all n sufficiently large,

   $$|X_n(\omega) - X(\omega)| < \delta \leq \varepsilon.$$

$\square$

Proof of Theorem 3.4 ($\Rightarrow$).
As $\mu_n \xrightarrow{d} \mu$, we use Theorem 3.3 (2) to choose (X_n) and X such that $X_n \to X$ almost surely. As h is continuous $h(X_n) \to h(X)$ and further, since h is bounded we can apply DOM to deduce that

$$\int h \, d\mu = \mathbb{E}(h(X_n)) \to \mathbb{E}(h(X)) = \int h \, d\mu. \qquad \square$$

Proof of Theorem 3.4 ($\Leftarrow$).
Let x be a point of continuity of F. We want to show that $F_n(x) \to F(x)$ as $n \to \infty$. Let $\delta > 0$ and consider any continuous function h satisfying

$$h(t) = \begin{cases} 1 & \text{if } t \leq x - \delta \\ 0 & \text{if } t \geq x \\ \in [0,1] & \text{otherwise.} \end{cases}$$

Then

$$F_n(x) = \int \mathbf{1}_{\{(-\infty,x]\}} \, d\mu_n \geq \int h \, d\mu_n,$$

and by assumption

$$\liminf F_n(x) \geq \lim \int h \, d\mu_n = \int h \, d\mu \geq \int \mathbf{1}_{\{(-\infty,x-\delta]\}} \, d\mu = F(x - \delta).$$

By continuity of F at x, letting $\delta \downarrow 0$, gives $\liminf F_n(x) \geq F(x)$.

Now consider any continuous function h satisfying

$$h(t) = \begin{cases} 1 & \text{if } t \leq x \\ 0 & \text{if } t \geq x + \delta \\ \in [0,1] & \text{otherwise.} \end{cases}$$

Then similarly to above we have

$$F_n(x) = \int \mathbf{1}_{\{(-\infty,x]\}} \, d\mu_n \leq \int h \, d\mu_n,$$

and so

$$\limsup F_n(x) \leq \lim \int h \, d\mu_n = \int h \, d\mu \leq \int \mathbf{1}_{\{(-\infty,x+\delta]\}} \, d\mu = F(x + \delta).$$

By continuity of F at x, letting $\delta \downarrow 0$, gives $\limsup F_n(x) \leq F(x)$.
Hence $\lim F_n(x) = F(x)$. $\square$

Proof of Theorem 3.3 (1). Suppose $X_n \to X$ almost surely. Then $h(X_n) \to h(X)$ for any bounded continuous function h and so for such h,

$$\int h \, d\mu_n = \mathbb{E}(h(X_n)) \overset{\text{DOM}}{\to} \mathbb{E}(h(X)) = \int h \, d\mu.$$

Hence $X_n \xrightarrow{d} X$ by Theorem 3.4 ($\Leftarrow$). $\square$

Definition 3.5.
$h : \mathbb{R} \to \mathbb{R}$ is called a C^2 *function* if h'' exists and is continuous.
It is called a C^2 *test function* if it is C^2 and equals zero outside a bounded region.

Theorem 3.6 (Useful defintion of weak convergence 2)**.**
$\mu_n \xrightarrow{d} \mu$ if and only if $\int h \, d\mu_n \to \int h \, d\mu$ for every C^2 test function h.

Definition 3.7.
A sequence of probability measures (μ_n) is *tight* if for every $\varepsilon > 0$, there is $M > 0$ such that

$$\mu_n([-M, M]) \geq 1 - \varepsilon \qquad \text{for all } n.$$

Example 3.8.

- If each μ_n has one unit atom at n, i.e. $\mu_n(A) = \delta_n(A)$, then (μ_n) is not tight.
- If each μ_n has an atom of size $1 - a_n$ at 0 and an atom of size a_n at n, i.e.

  $$\mu_n(A) = (1 - a_n)\delta_0(A) + a_n\delta_n(A),$$

  then (μ_n) is tight if and only if $a_n \to 0$.

Proof of Theorem 3.6 ($\Rightarrow$).
This follows from Theorem 3.4 since every C^2 test function is continuous and bounded. $\square$

Proof of Theorem 3.6 ($\Leftarrow$).

1. We first show that the RHS implies tightness of (μ_n). Let $\varepsilon > 0$ and choose M_1 so that $\mu([-M_1, M_1]) \geq 1 - \varepsilon/2$. Consider a C^2 function h such that

   $$h(t) = \begin{cases} 1 & \text{if } |t| \leq M_1 \\ 0 & \text{if } |t| \geq M_1 + 1 \\ \in [0,1] & \text{otherwise.} \end{cases}$$

   We have

   $$\mu_n([-M_1 - 1, M_1 + 1]) = \int \mathbf{1}_{\{[-M_1-1, M_1+1]\}}\, d\mu_n \geq \int h\, d\mu_n,$$

   but by our assumption there exists N such that for all $n > N$, $\int h\, d\mu_n \geq \int h\, d\mu - \varepsilon/2$ and so for such n,

   $$\mu_n([-M_1 - 1, M_1 + 1]) \geq \int h\, d\mu - \varepsilon/2 \geq \int \mathbf{1}_{\{[-M_1, M_1]\}}\, d\mu - \varepsilon/2 \geq 1 - \varepsilon.$$

   For the finitely many $1 \leq n \leq N$, choose M_2 so that $\mu_n([-M_2, M_2]) \geq 1 - \varepsilon$. Now choose $M = \max\{M_1 + 1, M_2\}$.

2. Let x be a point of continuity of F. Fix $\varepsilon > 0$ and $\delta > 0$ and choose M as above. Consider the function h defined as

   $$h(t) = \begin{cases} 1 & \text{on } [-M, x - \delta] \\ 0 & \text{on } (-\infty, -M - 1] \text{ and } [x, \infty) \\ C^2 \text{ and } \in [0,1] & \text{otherwise.} \end{cases}$$

We have

$$
\begin{aligned}
F_n(x) &= \int \mathbf{1}_{\{(-\infty,x]\}}\, d\mu_n \geq \int h\, d\mu_n \\
\liminf_{n\to\infty} F_n(x) &\geq \lim_{n\to\infty} \int h\, d\mu_n = \int h\, d\mu \geq \int h\mathbf{1}_{\{(-\infty,-M]\}}\, d\mu + \int \mathbf{1}_{\{(M,x-\delta]\}}\, d\mu \\
&= \int \mathbf{1}_{\{(-\infty,x-\delta]\}}\, d\mu - \int (1-h)\mathbf{1}_{\{(-\infty,-M]\}}\, d\mu\,.
\end{aligned}
$$

Now using that $(1-h) \leq 1$, we have

$$
\liminf_{n\to\infty} F_n(x) \geq F(x-\delta) - \mu((-\infty,-M]) \geq F(x-\delta) - \varepsilon.
$$

Letting ε, $\delta \to 0$ by continuity we have $\liminf_{n\to\infty} F_n(x) \geq F(x)$.
Now consider the function h defined as

$$
h(t) = \begin{cases} 1 & \text{on } [-M,x] \\ 0 & \text{on } (-\infty,-M-1] \text{ and } [x+\delta,\infty) \\ C^2 \text{ and } \in [0,1] & \text{otherwise.} \end{cases}
$$

Similarly to above we have

$$
\begin{aligned}
F_n(x) &= \mu_n((-\infty,-M]) + \int \mathbf{1}_{\{[-M,x]\}}\, d\mu_n \leq \varepsilon + \int h\, d\mu_n \\
\limsup_{n\to\infty} F_n(x) &\leq \varepsilon + \lim_{n\to\infty} \int h\, d\mu_n = \varepsilon + \int h\, d\mu \leq \varepsilon + \int \mathbf{1}_{\{(-\infty,x+\delta]\}}\, d\mu = \varepsilon + F(x+\delta).
\end{aligned}
$$

Letting ε, $\delta \to 0$ by continuity we have $\limsup_{n\to\infty} F_n(x) \leq F(x)$.
Thus $\lim_{n\to\infty} F_n(x) = F(x)$. □

Definition 3.9.
The *Fourier transform* of a probability measure μ on $(\mathbb{R},\mathcal{B})$ is

$$
\hat{\mu} : \mathbb{R} \to \mathbb{R}, \quad \hat{\mu}(t) = \int_{\mathbb{R}} e^{itx}\, d\mu(x).
$$

The *characteristic function* of a random variable X is

$$
\varphi : \mathbb{R} \to \mathbb{R}, \quad \varphi(t) = \mathbb{E}(e^{itX}).
$$

The *Fourier transform* of a Lebesgue integrable function $h : \mathbb{R} \to \mathbb{R}$ is

$$
\hat{h} : \mathbb{R} \to \mathbb{R}, \quad \hat{h}(t) = \int_{\mathbb{R}} e^{itx} h(x)\, dx.
$$

Remark 3.10.

1. The characteristic function of a random variable is the Fourier transform of its law.
2. If μ has density h then $\hat{\mu} = \hat{h}$.

3. The Fourier transforms and characteristic functions are always well-defined since $|e^{itx}| = 1$.

4. We will prove that X and Y have the same distribution if and only if they have the same characteristic functions.

5. We will prove that $X_n \xrightarrow{d} X$ if and only if $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbb{R}$.

Theorem 3.11 (Properties of characteristic functions)**.**

1. $\varphi(0) = 1$.

2. $\varphi_{\lambda X}(t) = \varphi_X(\lambda t)$.

3. If X *and* Y *are independent then* $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.

4. φ *is continuous.*

Proof.
(1) and (2) are obvious.
(3) $\varphi_{X+Y}(t) = \mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX})\mathbb{E}(e^{itY}) = \varphi_X(t)\varphi_Y(t)$.
(4) If $t_n \to t$ then $e^{it_nX} \to e^{itX}$ and this sequence is dominated by 1 so by DOM,

$$\varphi(t_n) = \mathbb{E}(e^{it_nX}) \to \mathbb{E}(e^{itX}) = \varphi(t).$$

$\square$

Example 3.12.

1. If $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ then

$$\varphi(t) = \frac{1}{2}e^{it} + \frac{1}{2}e^{-it} = \cos t.$$

2. If X is uniform on $[-1, 1]$ then

$$\varphi(t) = \int_{-1}^{1} \frac{1}{2}e^{itx}\,dx = \frac{\sin t}{t}.$$

3. If X is Cauchy, that is, with density $f(x) = \frac{1}{\pi(1+x^2)}$, then

$$\varphi(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{itx}}{1+x^2}\,dx = e^{-|t|}$$

(this integral can be computed using contour integrals).

4. If X is N(0,1) then

$$\begin{aligned}\varphi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx-\frac{x^2}{2}}\,dx &= e^{-t^2/2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x-it)^2}{2}}\,dx \\ &= e^{-t^2/2}\frac{1}{\sqrt{2\pi}}\lim_{R\to\infty}\int_{\Gamma_1(R)} e^{-z^2/2}\,dz,\end{aligned}$$

where $\Gamma_1(R) = [-R - it, R - it]$.

Denote by $\Gamma_2(R) = [-R - it, -R]$, $\Gamma_3(R) = [-R, R]$ and $\Gamma_4(R) = [R, R - it]$. Since $e^{-z^2/2}$ is analytic, we have (Residue Theorem)

$$\int_{\Gamma_1(R)} e^{-z^2/2}\,dz + \int_{\Gamma_2(R)} e^{-z^2/2}\,dz - \int_{\Gamma_3(R)} e^{-z^2/2}\,dz + \int_{\Gamma_4(R)} e^{-z^2/2}\,dz = 0.$$

However, notice that

$$\int_{\Gamma_3(R)} e^{-z^2/2}\,dz = \sqrt{2\pi}.$$

Further, on Γ_2 and Γ_4, we have $z = \pm R - iy$ and so $\mathrm{Re}(z^2) = R^2 - y^2$ which gives

$$|e^{-z^2/2}| = e^{-\frac{R^2-y^2}{2}} \le e^{-\frac{R^2-t^2}{2}}.$$

This implies

$$\Big|\int_{\Gamma_{2,4}(R)} e^{-z^2/2}\,dz\Big| \le |t|e^{-\frac{R^2-t^2}{2}} \to 0$$

as $R \to \infty$. Thus

$$\int_{\Gamma_1(R)} e^{-z^2/2}\,dz \to \sqrt{2\pi}$$

as $R \to \infty$, giving $\varphi(t) = e^{-t^2/2}$.

Lemma 3.13.

$$\int_0^\infty \frac{\sin t}{t}\,dt = \frac{\pi}{2}$$

and the function $\rho(x) = \int_0^x \frac{\sin t}{t}\,dt$, $x \ge 0$ *is bounded.*

Proof. The value of the integral was computed in Complex Analysis.
The function is bounded since it is continuous and converges at ∞ (minimum/maximum theorem). □

Lemma 3.14 (Decay of the Fourier transform)**.**
If h is a C^2 test function then $|\hat{h}(t)| \le c/t^2$ for some $c > 0$ and all $t \neq 0$. In particular,

$$\int_{-\infty}^{\infty} |\hat{h}(t)|\,dt < \infty.$$

Proof. Using integration by parts (twice) and that h is zero outside some bounded region,

$$\hat{h}(t) = \int_{-\infty}^{\infty} e^{itx}h(x)\,dx = \left[\frac{h(x)e^{itx}}{it}\right]_{-\infty}^{\infty} - \frac{1}{it}\int_{-\infty}^{\infty} h'(x)e^{itx}\,dx = -\frac{1}{t^2}\int_{-\infty}^{\infty} h''(x)e^{itx}\,dx,$$

and so

$$|\hat{h}(t)| \le \frac{1}{t^2}\int_{-\infty}^{\infty} |h''(x)|\,dx \le \frac{c}{t^2},$$

for some $c > 0$. Then

$$\int_{-\infty}^{\infty} |\hat{h}(t)|\,dt \le \int_{-\infty}^{\infty} \frac{c}{t^2}\,dt < \infty.$$

□

Theorem 3.15 (Parseval-Plancherel)**.** *Let μ be a probability measure on $(\mathbb{R}, \mathcal{B})$ with Fourier transform φ. Then for any C^2 test function*

$$\int_{\mathbb{R}} h\, d\mu = \frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{\hat{h}(t)} \varphi(t)\, dt.$$

Proof. The RHS makes sense since $|\varphi(t)| \leq 1$ and by Lemma 3.14 $|\overline{\hat{h}(t)}| \leq c/t^2$. We have

$$\begin{aligned}
\frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{\hat{h}(t)} \varphi(t)\, dt &= \frac{1}{2\pi} \lim_{T\to\infty} \int_{-T}^{T} \overline{\hat{h}(t)} \varphi(t)\, dt \\
&= \frac{1}{2\pi} \lim_{T\to\infty} \int_{-T}^{T} \left(\int_{-\infty}^{\infty} e^{-itx} h(x)\, dx \right) \left(\int_{-\infty}^{\infty} e^{ity}\, d\mu(y) \right) dt \\
&\overset{\text{Fubini}}{=} \frac{1}{2\pi} \lim_{T\to\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x) \int_{-T}^{T} e^{it(y-x)}\, dt\, dx\, d\mu(y) \\
&= \lim_{T\to\infty} \int_{-\infty}^{\infty} \left(\frac{1}{\pi} \int_{-\infty}^{\infty} h(x) \frac{\sin(T(x-y))}{x-y}\, dx \right) d\mu(y).
\end{aligned}$$

We can use Fubini's Theorem here since $|h(x)e^{it(y-x)}| = |h(x)|$ is integrable (since h is zero outside a bounded region) with respect to $\mu \times \text{Leb} \times \text{Leb}$ on $\mathbb{R} \times \mathbb{R} \times [-T, T]$. We now claim that as $T \to \infty$,

$$\frac{1}{\pi} \int_{-\infty}^{\infty} h(x) \frac{\sin(T(x-y))}{x-y}\, dx \to h(y)$$

for each y, and in a bounded way. If so we are done by DOM as we can integrate both parts wth respect to μ and see that the limit is $\int_{\mathbb{R}} h(y)\, d\mu(y)$. So it suffices to prove this claim. We split the integral into two:

$$\frac{1}{\pi} \int_{-\infty}^{\infty} h(x) \frac{\sin(T(x-y))}{x-y}\, dx = \frac{1}{\pi} \int_{-\infty}^{y} [\cdots]\, dx + \frac{1}{\pi} \int_{y}^{\infty} [\cdots]\, dx.$$

It suffices to prove that each term converges to $\frac{h(y)}{2}$ in a bounded way. We will only show it for the second term (the first is similar).

$$\begin{aligned}
&\frac{1}{\pi} \int_{-\infty}^{\infty} h(x) \frac{\sin(T(x-y))}{x-y}\, dx \qquad \text{(integrate by parts)} \\
&= \frac{1}{\pi} \left(\left[h(x) \int_{y}^{x} \frac{\sin(T(u-y))}{u-y}\, du \right]_{x=y}^{x=\infty} - \int_{y}^{\infty} h'(x) \int_{y}^{x} \frac{\sin(T(u-y))}{u-y}\, du\, dx \right) \\
&\overset{v=T(u-y)}{=} -\frac{1}{\pi} \int_{y}^{\infty} h'(x) \int_{0}^{T(x-y)} \frac{\sin v}{v}\, dv\, dx = -\frac{1}{\pi} \int_{y}^{\infty} h'(x) \rho(T(x-y))\, dx.
\end{aligned}$$

Since $h(y) = -\int_y^\infty h'(x)\, dx$ it suffices to prove that, as $T \to \infty$,

$$\int_{y}^{\infty} h'(x) \rho(T(x-y))\, dx \to \frac{\pi}{2} \int_{y}^{\infty} h'(x)\, dx,$$

in a bounded way.

Convergence:
By Lemma 3.13, as $T \to \infty$,

$$\rho(T(x-y)) \to \frac{\pi}{2},$$

for each x, and in a bounded way by c. Hence

$$h'(x)\rho(T(x-y)) \to h'(x)\frac{\pi}{2},$$

for each x and in a dominated way by $c|h'(x)|$. Since $c|h'(x)|$ vanishes outside a bounded region it is Lebesgue measurable. By applying DOM we obtain

$$\int_y^\infty h'(x)\rho(T(x-y))\,dx \to \frac{\pi}{2}\int_y^\infty h'(x)\,dx.$$

In a bounded way:

$$\Big|\int_y^\infty h'(x)\rho(T(x-y))\,dx\Big| \le c\int_{-\infty}^\infty |h'(x)|\,dx < \infty.$$

$\square$

Theorem 3.16 (Weak convergence = convergence of characteristic functions)**.**
Let (X_n) and X be random variables with characteristic functions (φ_n) and φ. Then $X_n \xrightarrow{d} X$ if and only if $\varphi_n(t) \to \varphi(t)$ for each $t \in \mathbb{R}$.

Proof. Denote by (μ_n) and μ the laws of (X_n) and X.
($\Rightarrow$) Suppose $X_n \xrightarrow{d} X$.
For each t, the function $h(x) = e^{itx}$ is continuous and bounded. Hence

$$\varphi_n(t) = \mathbb{E}(e^{itX_n}) = \int h\,d\mu_n \overset{\text{Thm 3.4}}{\to} \int h\,d\mu = E(e^{itX}) = \varphi(t).$$

($\Leftarrow$) Suppose $\varphi_n(t) \to \varphi(t)$ as $n \to \infty$ for each t.
By Theorem 3.6 it suffices to check that $\int h\,d\mu_n \to \int h\,d\mu$ for every C^2 test function h. In turn, by Parseval-Plancherel it suffices to check that

$$\int_{-\infty}^\infty \overline{\hat{h}(t)}\varphi_n(t)\,dt \to \int_{-\infty}^\infty \overline{\hat{h}(t)}\varphi(t)\,dt. \tag{2}$$

Since $\varphi_n(t) \to \varphi(t)$, we have $\overline{\hat{h}(t)}\varphi_n(t) \to \overline{\hat{h}(t)}\varphi(t)$ for each t.

Since $|\overline{\hat{h}(t)}| \le c/t^2$ by Lemma 3.14 and $|\varphi_n(t)| \le 1$ for all n and t we have

$$|\overline{\hat{h}(t)}\varphi_n(t)| \le c/t^2$$

for all t. Since c/t^2 is integrable (with respect to the Lebesgue measure), (2) follows now from DOM. $\square$

Summary

$$
\begin{tikzcd}
 & \boxed{\int h\,d\mu_n \to \int h\,d\mu \text{ for continuous and bounded } h} \arrow[rr, Rightarrow, "\text{Thm 3.16}"] & & \boxed{\varphi_n(t) \to \varphi(t) \text{ for all } t} \arrow[d, Rightarrow, "\text{Thm 3.16}"] \\
\boxed{X_n \xrightarrow{d} X} \arrow[ur, Leftrightarrow, "\text{Thm 3.4}"] \arrow[dr, Leftrightarrow, "\text{Thm 3.6}"'] & & & \\
 & \boxed{\int h\,d\mu_n \to \int h\,d\mu \text{ for } C^2 \text{ test } h} \arrow[rr, Leftrightarrow, "\text{Thm 3.15}"] & & \boxed{\int \bar{\hat{h}}\varphi_n\,dt \to \int \bar{\hat{h}}\varphi\,dt \text{ for } C^2 \text{ test } h}
\end{tikzcd}
$$

Theorem 3.17 (Distributions are determined by their characteristic functions)**.** *If random variables X and Y have the same characteristic functions $\varphi_X = \varphi_Y$ then X and Y have the same distribution.*

Proof. Let $X_n = X$ for all n. Then $\varphi_{X_n} = \varphi_X = \varphi_Y$. So $\varphi_{X_n} \to \varphi_Y$ and $X_n \xrightarrow{d} Y$ by Theorem 3.16.

Hence $F_X(t) = F_{X_n}(t) \to F_Y(t)$ at the continuity points of F_Y. So $F_X(t) = F_Y(t)$ at the continuity points.

If t is not a continuity point, it can be approximated by continuity points $t_n \downarrow t$ (since there are only countably many discontinuity points). By right continuity,

$$F_X(t) = \lim_{n\to\infty} F_X(t_n) = \lim_{n\to\infty} F_Y(t_n) = F_Y(t).$$

$\square$

Example 3.18. Let X and Y be independent Cauchy random variables, i.e. both have density $(\pi(1+x^2))^{-1}$. What is the distribution of $X+Y$?

We know the formula for the density of $X+Y$:

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} \frac{1}{\pi(1+(t-v)^2)} \cdot \frac{1}{\pi(1+v^2)}\,dv.$$

However, it would be tricky to compute this. Instead, we can use characteristic functions:

$$\varphi_X(t) = \varphi_Y(t) = e^{-|t|} \quad \Rightarrow \quad \varphi_{X+Y}(t) = e^{-2|t|} \quad \Rightarrow \quad \varphi_{\frac{X+Y}{2}}(t) = e^{-|t|}.$$

Hence $\frac{X+Y}{2}$ is Cauchy. Hence $F_{X+Y}(t) = F_{X+Y}2(t/2)$ and

$$f_{X+Y}(t) = \frac{1}{2} f_{\frac{X+Y}{2}}(t/2) = \frac{1}{2\pi(1+(t/2)^2)}.$$

Example 3.19. Binomial distribution $\mathrm{Bin}(n,p)$ converges weakly to Poisson with mean 1.

- Compute the characteristic function of $\mathrm{Bin}(n,p)$: $\varphi_{n,p}(t) = (1-p+pe^{it})^n$.
- Substitute $p = 1/n$: $\varphi_{n,1/n}(t) = \left(1 + \frac{e^{it}-t}{n}\right)^n$.
- Observe that $\varphi_{n,1/n}(t) \to \exp(e^{it}-1)$ as $n \to \infty$.
- Compute the characteristic function of Poisson(1): $\varphi(t) = \exp(e^{it}-1)$.

Theorem 3.20 (Derivatives of characteristic functions)**.**
Let X be a square-integrable random variable with characteristic function φ. Then

1. *φ is twice differentiable and φ'' is continuous at 0,*
2. *$\varphi'(0) = i\,\mathbb{E}(X)$,*
3. *$\varphi''(0) = -\mathbb{E}(X^2)$.*

Proof.
Observe that

$$\lim_{h\to 0}\frac{e^{ihX}-1}{h} = (e^{ihX})'\big|_{h=0} = iX \quad \text{and} \quad \Big|\frac{e^{ihX}-1}{h}\Big| = \Big|\frac{1}{h}\int_0^{hX} e^{is}\,ds\Big| \le |X|.$$

Therefore we have

$$\begin{aligned}\varphi'(t) &= \lim_{h\to 0}\frac{\varphi(t+h)-\varphi(t)}{h} = \lim_{h\to 0}\mathbb{E}\left(\frac{e^{i(t+h)X}-e^{itX}}{h}\right)\\ &= \lim_{h\to 0}\mathbb{E}\left(e^{itX}\frac{e^{ihX}-1}{h}\right) \overset{\text{DOM}}{=} \mathbb{E}(e^{itX}iX),\end{aligned}$$

since the convergence is dominated by $|X|$ (which is integrable by assumption). Hence φ is differentiable and $\varphi'(0) = i\mathbb{E}(X)$. Furthermore, we have

$$\begin{aligned}\varphi''(t) &= \lim_{h\to 0}\frac{\varphi'(t+h)-\varphi'(t)}{h} = \lim_{h\to 0}\mathbb{E}\left(\frac{e^{i(t+h)X}iX - e^{itX}iX}{h}\right)\\ &= \lim_{h\to 0}\mathbb{E}\left(iXe^{itX}\frac{e^{ihX}-1}{h}\right) \overset{\text{DOM}}{=} \mathbb{E}\left(iXe^{itX}\lim_{h\to 0}\frac{e^{ihX}-1}{h}\right)\\ &= -\mathbb{E}(X^2e^{itX}),\end{aligned}$$

since the convergence is dominated by X^2. Hence φ is twice differentiable and

$$\varphi''(0) = -\mathbb{E}(X^2).$$

The continuity of φ'' at zero follows from

$$\varphi''(t) = -\mathbb{E}(X^2e^{itX}) \overset{\text{DOM}}{\to} -\mathbb{E}(X^2) = \varphi''(0), \quad \text{as } t \to 0. \qquad \square$$

Example 3.21.
If X is Cauchy then $\varphi(t) = e^{-|t|}$ which is not differentiable at zero. Why does this not invalidate the theorem?

Lemma 3.22.

$$|\log(1+z) - z| \le |z|^2, \quad \textit{if } z \in \mathbb{C},\ |z| \le 1/2.$$

Proof. Denote by Γ the straight path from 0 to z. Then

$$|\log(1+z)-z| = \Big|\int_\Gamma \Big(\frac{1}{1+w}-1\Big)\,dw\Big| \le \int_\Gamma \Big|\frac{w}{1+w}\Big|\,dw.$$

Using $|a+b| \ge \big||a|-|b|\big|$ (reverse triangle inequality), and that $|z| \le 1/2$, we have

$$|\log(1+z)-z| \le 2\int_\Gamma |w|\,dw = 2\int_0^{|z|} u\,du = |z|^2.$$

$\square$

Theorem 3.23 (Central Limit Theorem)**.**
Let $(X_i)_{i\in\mathbb{N}}$ *be a sequence of i.i.d. random variables in* $\mathcal{L}^2$ *with mean* $\mathbb{E}X_1 = \mu$ *and variance* $\mathrm{Var}X_1 = \sigma^2$*. Set* $S_n = X_1 + \cdots + X_n$*. Then, as* $n \to \infty$*,*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1).$$

Proof. Set $Y = (X_1-\mu)/\sigma$ so that $\mathbb{E}[Y] = 0$ and $\mathrm{Var}[Y] = 1$. Let φ_Y and φ be the characteristic functions of Y and X_1, respectively. By Theorem 3.11,

$$\varphi(t) = e^{it\mu}\varphi_Y(\sigma t).$$

Let φ_n denote the characteristic function of $(S_n - n\mu)/\sigma\sqrt{n}$. Again by Theorem 3.11,

$$\varphi_n(t) = e^{-it\sqrt{n}\mu/\sigma}\varphi(t/\sqrt{n}\sigma)^n = \varphi_Y(t/\sqrt{n})^n. \tag{3}$$

Since φ'' is twice differentiable (Theorem 3.20), we have by Taylor's Theorem,

$$\varphi_Y(s) = \varphi_Y(0) + s\varphi_Y'(0) + \frac{s^2}{2}\varphi_Y''(0) + o(s^2),$$

as $s \to 0$. Using Theorem 3.20, this is

$$\varphi_Y(s) = 1 - \frac{s^2}{2} + o(s^2),$$

which we plug into (3) by setting $s = t/\sqrt{n}$ to give

$$\varphi_n(t) = \Big(1 - \frac{t^2}{2n} + o\Big(\frac{t^2}{n}\Big)\Big)^n,$$

as $n \to \infty$. By Lemma 3.22 (which we can apply for n sufficiently large), for each t,

$$n\log\Big(1 - \frac{t^2}{2n} + o\Big(\frac{t^2}{n}\Big)\Big) = n\Big(-\frac{t^2}{2n} + o\Big(\frac{t^2}{n}\Big)\Big) \to -\frac{t^2}{2},$$

as $n \to \infty$, which gives that, for all t,

$$\varphi_n(t) \to e^{-\frac{t^2}{2}}.$$

However, $e^{-\frac{t^2}{2}}$ is the characteristic function of the $N(0,1)$ distribution and so the proof is complete by Theorem 3.16. $\square$

Remark 3.24.
The Central Limit Theorem implies the Weak Law of Large Numbers $S_n/n \to \mu$, but there is a much simpler proof (see homework).

4 Martingales

4.1 Conditional expectation

Example 4.1. Imagine we roll a fair 6-sided die. Let $(\Omega, \Sigma, \mathbb{P})$ be the corresponding probability space, that is, $\Omega = \{1, 2, 3, 4, 5, 6\}$. Σ =set of all subsets of Ω, and $\mathbb{P}(\{i\}) = 1/6$ for $i \in \{1, \ldots, 6\}$. Let $X : \Omega \to \mathbb{R}$ be the Σ-measurable random variable satisfying $X(i) = i$ for $i \in \{1, \ldots, 6\}$. Consider another σ-algebra $\mathcal{F} = \{\varnothing, \{1, 2, 3\}, \{4, 5, 6\}, \Omega\}$. Note that X is not $\mathcal{F}$-measurable, e.g. $X^{-1}(\{1\}) = \{1\} \notin \mathcal{F}$.
Question: What is the best model for X in the $\mathcal{F}$ world? (i.e. among all $\mathcal{F}$-measurable functions)
Let Y be a random variable with $Y(1) = Y(2) = Y(3) = 2$ and $Y(4) = Y(5) = Y(6) = 5$ (average over the values X takes in each set of $\mathcal{F}$). Note that Y is $\mathcal{F}$-measurable.
Set $A = \{1, 2, 3\}$ and observe that

$$\mathbb{E}[X\mathbf{1}_{\{A\}}] = \int_A X \mathrm{d}\mathbb{P} = \Big(\frac{1}{6}\Big)(1 + 2 + 3) = 1,$$
$$\mathbb{E}[Y\mathbf{1}_{\{A\}}] = \int_A Y \mathrm{d}\mathbb{P} = \Big(\frac{1}{6}\Big)(2 + 2 + 2) = 1.$$

Theorem 4.2 (Definition, existence and uniqueness of conditional expectation)**.**
Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, and $X : \Omega \to \mathbb{R}$ be an integrable random variable. Let $\mathcal{F}$ be a sub-σ-algebra of Σ. Then there exists a random variable Y such that

1. *Y is $\mathcal{F}$-measurable,*
2. *Y is integrable,*
3. *$\int_A X \mathrm{d}\mathbb{P} = \int_A Y \mathrm{d}\mathbb{P}$ for all $A \in \mathcal{F}$.*

If $\tilde{Y}$ is another random variable with these properties then $\tilde{Y} = Y$ almost surely. Any random variable with the above three properties is called (a version of) the conditional expectation *$\mathbb{E}[X|\mathcal{F}]$ of X given $\mathcal{F}$.*

Remark 4.3. Suppose $\mathcal{F}$ is the trivial σ-algebra, $\mathcal{F} = \{\varnothing, \Omega\}$. Then the conditional expectation of an integrable random variable X with respect to $\mathcal{F}$ is the usual expectation. Indeed, integrability of Y follows from the integrability of X, and Y is $\mathcal{F}$-measurable since

$$Y^{-1}(x) = \begin{cases} \Omega, & \text{if } x = \mathbb{E}[X], \\ \varnothing, & \text{otherwise.} \end{cases}$$

Condition 3. is satisfied since:

$$\int_\varnothing X \mathrm{d}\mathbb{P} = \int_\varnothing \mathbb{E}[X] \mathrm{d}\mathbb{P} = 0, \quad \int_\Omega X \mathrm{d}\mathbb{P} = \mathbb{E}[X] = \mathbb{E}[X]\mathbb{P}(\Omega) = \int_\Omega \mathbb{E}[X] \mathrm{d}\mathbb{P}.$$

Recall the following theorem:

Theorem 4.4 (Radon-Nikodým derivative).
Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and suppose that $\mathbb{Q}$ is a finite measure on (Ω, Σ) which is absolutely continuous with respect to $\mathbb{P}$ (for any $S \in \Sigma$, $\mathbb{P}(S) = 0$ implies $\mathbb{Q}(S) = 0$). Then there exists an integrable random variable X on $(\Omega, \Sigma, \mathbb{P})$ such that

$$\mathbb{Q}(S) = \int_S X \mathrm{d}\mathbb{P} = \mathbb{E}[X\mathbf{1}_{\{S\}}] \qquad \forall S \in \Sigma.$$

The random variable X is called a version of the Radon-Nikodým derivative of $\mathbb{Q}$ relative to $\mathbb{P}$ on (Ω, Σ).

Proof of Theorem 4.2.
Decompose X into $X = X_+ - X_-$ where $X_+ = \max\{X, 0\}$ and $X_- = \max\{-X, 0\}$. Note that $|X| = X_+ + X_-$ and thus both X_+ and X_- are integrable. For $A \in \mathcal{F}$ define

$$\nu_\pm(A) := \int_A X_\pm \mathrm{d}\mathbb{P}.$$

By Fubini, for any non-negative random variable Z,

$$\mathbb{E}(Z) = \int Z(\omega)\, d\mathbb{P} = \int \int_0^\infty \mathbf{1}_{\{x \leq Z(\omega)\}}\, dx\, d\mathbb{P} = \int_0^\infty \int \mathbf{1}_{\{x \leq Z(\omega)\}}\, d\mathbb{P}\, dx = \int_0^\infty \mathbb{P}(X \geq x)\, d\mathbb{P}.$$

Next, we claim that ν_+ and ν_- are both absolutely continuous with respect to $\mathbb{P}$ on $(\Omega, \mathcal{F}, \mathbb{P})$. Indeed this is true since, by Fubini (using the fact that $X_\pm \geq 0$),

$$\mathbb{E}[X_\pm \mathbf{1}_{\{A\}}] = \int_0^\infty \mathbb{P}(X_\pm \mathbf{1}_{\{A\}} \geq y)\mathrm{d}y = \int_0^\infty \mathbb{P}(X_\pm \geq y, A)\mathrm{d}y \leq \int_0^\infty \mathbb{P}(A)\mathrm{d}y = 0,$$

if $\mathbb{P}(A) = 0$. Furthermore ν_+ and ν_- are finite measures on $\mathcal{F}$ (since $\mathbb{E}|X_\pm| < \infty$) and thus we can apply the Radon-Nikodým Theorem to deduce the existence of $\mathcal{F}$-measurable integrable random variables $Y_\pm$ such that

$$\int_A Y_\pm \mathrm{d}\mathbb{P} = \nu_\pm(A),$$

for all $A \in \mathcal{F}$. Furthermore since

$$\int_A Y_\pm \mathrm{d}\mathbb{P} = \nu_\pm(A) = \int_A X_\pm \mathrm{d}\mathbb{P}$$

for all $A \in \mathcal{F}$, we see that Y_+ and Y_- satisfy conditions 1, 2 and 3 of Theorem 4.2 for X_+ and X_- respectively. Set $Y = Y_+ - Y_-$. Then Y is integrable as $\mathbb{E}|Y| = \mathbb{E}Y_+ + \mathbb{E}Y_- < \infty$ and also

$$\int_A Y \mathrm{d}\mathbb{P} = \int_A (Y_+ - Y_-)\mathrm{d}\mathbb{P} = \int_A (X_+ - X_-)\mathrm{d}\mathbb{P} = \int_A X \mathrm{d}\mathbb{P}.$$

To complete the proof of existence it remains to show that Y is $\mathcal{F}$-measurable. However it is a standard result that the sum of two measurable functions is measurable. Indeed

$$Y_+(\omega) - Y_-(\omega) > c \iff \exists q \in \mathbb{Q} : Y_-(\omega) + c < q < Y_+(\omega),$$

and thus

$$\{\omega : Y_+(\omega) - Y_-(\omega) > c\} = \bigcup_{q\in\mathbb{Q}} (\{Y_+(\omega) > q\} \cap \{Y_-(\omega) < q - c\}) \in \mathcal{F}$$

since $\mathcal{F}$ is closed under countable unions. Let $\mathcal{C}$ be the class of intervals of the form (c, ∞) for $c \in \mathbb{R}$. Note that $\sigma(\mathcal{C}) = \mathcal{B}$. We have just shown that $Y^{-1} : \mathcal{C} \to \mathcal{F}$. Let $\mathcal{E}$ be the class of elements $B \in \mathcal{B}$ such that $Y^{-1}(B) \in \mathcal{F}$. We want to show that $\mathcal{B} \subseteq \mathcal{E}$. However, $\mathcal{E}$ is a σ-algebra (check!) and by definiton $\mathcal{C} \subseteq \mathcal{E}$ and thus $\sigma(\mathcal{C}) \subseteq \mathcal{E}$, i.e. $\mathcal{B} \subseteq \mathcal{E}$.

To show uniqueness, suppose $\tilde{Y}$ also satisfies the three conditions. We want to show that $\mathbb{P}(Y \neq \tilde{Y}) = 0$. Suppose this is not the case, i.e. $P(Y \neq \tilde{Y}) > 0$. We have

$$\{Y \neq \tilde{Y}\} = \bigcup_{k\in\mathbb{N}} \{Y - \tilde{Y} > 1/k\} \cup \bigcup_{k\in\mathbb{N}} \{\tilde{Y} - Y > 1/k\},$$

and at least one of these events has positive measure. Suppose $\mathbb{P}(A) > 0$ where $A = \{Y - \tilde{Y} > 1/n\}$. Then $A \in \mathcal{F}$ since Y and $\tilde{Y}$ are both $\mathcal{F}$-measurable. We have

$$0 = \int_A (X - X)\mathrm{d}\mathbb{P} = \int_A (Y - \tilde{Y})\mathrm{d}\mathbb{P} > \int_A \frac{1}{n}\mathrm{d}\mathbb{P} = \frac{1}{n}\mathbb{P}(A) > 0.$$

This is a contradiction which completes the proof. □

Theorem 4.5 (Basic properties of conditional expectation)**.**
Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and $\mathcal{F}$ a sub-σ-algebra of Σ. We assume all random variables are integrable.

(a) $\mathbb{E}(\mathbb{E}[X|\mathcal{F}]) = \mathbb{E}X$*,*

(b) If X is $\mathcal{F}$-measurable then $\mathbb{E}[X|\mathcal{F}] = X$ almost surely,

(c) Linearity: $\mathbb{E}[a_1X_1 + a_2X_2|\mathcal{F}] = a_1\mathbb{E}[X_1|\mathcal{F}] + a_2\mathbb{E}[X_2|\mathcal{F}]$ *almost surely,*

(d) Positivity: *If $X \geq 0$ almost surely then $\mathbb{E}[X|\mathcal{F}] \geq 0$ almost surely,*

(e) Conditional MON*: If $0 \leq X_n \uparrow X$ almost surely, then $\mathbb{E}[X_n|\mathcal{F}] \uparrow \mathbb{E}[X|\mathcal{F}]$ almost surely,*

(f) Taking out what is known: *If Z is $\mathcal{F}$-measurable and bounded then $\mathbb{E}[ZX|\mathcal{F}] = Z\mathbb{E}[X|\mathcal{F}]$ almost surely,*

(g) Independence: *If X is independent of $\mathcal{F}$ then $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}X$ almost surely,*

(h) Tower property: *If $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$ then $\mathbb{E}[\mathbb{E}[X|\mathcal{F}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]$ almost surely,*

(i) Conditional Jensen: *If $\phi : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}|\phi(X)| < \infty$ then $\mathbb{E}[\phi(X)|\mathcal{F}] \geq \phi(\mathbb{E}[X|\mathcal{F}])$ almost surely.*

Proof.

(a) $\mathbb{E}(\mathbb{E}[X|\mathcal{F}]) = \int_\Omega \mathbb{E}[X|\mathcal{F}]\mathrm{d}\mathbb{P} = \int_\Omega X\mathrm{d}\mathbb{P} = \mathbb{E}X$ by part 3 of the definition setting $A = \Omega$.

(b) X satisfies 1, 2 and 3 of the definition.

(c) This is obvious once we clarify the meaning: If Y_1 is a version of $\mathbb{E}[X_1|\mathcal{F}]$ and Y_2 is a version of $\mathbb{E}[X_2|\mathcal{F}]$ then $a_1Y_1 + a_2Y_2$ is a version of $\mathbb{E}[a_1X_1 + a_2X_2|\mathcal{F}]$. Indeed integrability can be shown using triangle inequality and linearity of expectation, $\mathcal{F}$-measurablity follows as sums of $\mathcal{F}$-measurable functions are $\mathcal{F}$-measurable and condition 3 follows by linearity of expectation.

(d) Let $Y = \mathbb{E}[X|\mathcal{F}]$ and suppose $\mathbb{P}(Y < 0) > 0$. Then there exists $n \in \mathbb{N}$ such that $\mathbb{P}(Y < -1/n) > 0$. Set $A = \{Y < -1/n\}$, then $A \in \mathcal{F}$ and $\mathbb{P}(A) > 0$. Since $X \geq 0$ a.s. we have that $\mathbb{E}[X\mathbf{1}_{\{A\}}] \geq 0$ but also

$$\mathbb{E}[X\mathbf{1}_{\{A\}}] = \mathbb{E}[Y\mathbf{1}_{\{A\}}] < -\frac{1}{n}\mathbb{P}(A) < 0,$$

which is a contradiction.

(e) Denote $Y_n = \mathbb{E}[X_n|\mathcal{F}]$ By (d) $0 \leq Y_n \uparrow$. Denote $Y = \liminf Y_n$. Then Y is $\mathcal{F}$-measurable as it is the lim inf of $\mathcal{F}$-measurable functions. Y is also integrable (by Fatou, for example). For $A \in \mathcal{F}$,

$$\int_A Y \mathrm{d}\mathbb{P} \overset{(\mathrm{MON})}{=} \lim_{n\to\infty} \int_A Y_n \mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_A X_n \mathrm{d}\mathbb{P} \overset{(\mathrm{MON})}{=} \int_A X \mathrm{d}\mathbb{P}.$$

Thus $Y = \mathbb{E}[X|\mathcal{F}]$ almost surely.

(f) We first show this is true for Z an indicator, $Z = \mathbf{1}_{\{B\}}$ for $B \in \mathcal{F}$. For each $A \in \mathcal{F}$, we have

$$\int_A Z\mathbb{E}[X|\mathcal{F}]\mathrm{d}\mathbb{P} = \int_{A\cap B} \mathbb{E}[X|\mathcal{F}]\mathrm{d}\mathbb{P} = \int_{A\cap B} X\mathrm{d}\mathbb{P} = \int_A ZX\mathrm{d}\mathbb{P}.$$

We can extend the result to simple functions by linearity. We can then extend to positive functions using conditional MON. Finally we use linearity to extend to all functions.

(g) Again we start by showing the result for $X = \mathbf{1}_{\{B\}}$ for B independent of $\mathcal{F}$ and then use standard machinery. For $A \in \mathcal{F}$, we have

$$\int_A X\mathrm{d}\mathbb{P} = P(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{E}[X] = \int_A \mathbb{E}[X]\mathrm{d}\mathbb{P}.$$

(h) $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable and integrable by definition, so it remains to show that for each $B \in \mathcal{G}$,

$$\int_B \mathbb{E}[X|\mathcal{F}]\mathrm{d}\mathbb{P} = \int_B \mathbb{E}[X|\mathcal{G}]\mathrm{d}\mathbb{P}.$$

But $B \in \mathcal{F}$ and so both integrals are equal to $\int_B X\mathrm{d}\mathbb{P}$.

(i) We omit the proof. □

Example 4.6.

(1) Let $(X_n)_{n\in\mathbb{N}}$ be independent, identically distributed Bernoulli(p) random variables (so that $\mathbb{P}(X_1 = 1) = p = 1 - \mathbb{P}(X_1 = 0)$) and set $S_n = X_1 + \cdots + X_n$. Let $m < n$ and $\sigma_m = \sigma(X_1, \ldots, X_m)$. Then

$$\mathbb{E}[S_n|\sigma_m] \overset{(\mathrm{c})}{=} \mathbb{E}[X_1+\cdots+X_m|\sigma_m]+\mathbb{E}[X_{m+1}+\cdots+X_n|\sigma_m] \overset{(\mathrm{b})+(\mathrm{g})}{=} X_1+\cdots+X_m+(n-m)p.$$

(2) Fix $0 < p_1 \neq p_2 < 1$ and let X be the random variable with

$$\mathbb{P}(X = p_1) = q = 1 - \mathbb{P}(X = p_2).$$

Now let Y be Bernoulli(X). What is $\mathbb{E}[Y|\sigma(X)]$? Problem: Y is not independent of X. Let Y_1 be Bernoulli(p_1) and Y_2 be Bernoulli(p_2) independent of X so that $Y = Y_1 \mathbf{1}_{\{X=p_1\}} + Y_2 \mathbf{1}_{\{X=p_2\}}$. We have

$$\begin{aligned}
\mathbb{E}[Y|\sigma(X)] &= \mathbb{E}[Y_1 \mathbf{1}_{\{X=p_1\}} + Y_2 \mathbf{1}_{\{X=p_2\}} | \sigma(X)] \\
&\overset{\text{(c)}}{=} \mathbb{E}[Y_1 \mathbf{1}_{\{X=p_1\}} | \sigma(X)] + \mathbb{E}[Y_2 \mathbf{1}_{\{X=p_2\}} | \sigma(X)] \\
&\overset{\text{(f)}}{=} \mathbf{1}_{\{X=p_1\}} \mathbb{E}[Y_1|\sigma(X)] + \mathbf{1}_{\{X=p_2\}} \mathbb{E}[Y_2|\sigma(X)] \\
&\overset{\text{(g)}}{=} \mathbf{1}_{\{X=p_1\}} \mathbb{E}[Y_1] + \mathbf{1}_{\{X=p_2\}} \mathbb{E}[Y_2] \\
&= p_1 \mathbf{1}_{\{X=p_1\}} + p_2 \mathbf{1}_{\{X=p_2\}} \\
&= X.
\end{aligned}$$

(3) Galton-Watson Process

Consider a stochastic process which models the number of individuals in a population. Every individual alive at time n has a random number of children independently of each other and distributed according to an offspring distribution which we denote by the random variable N. The parent individual then dies. Suppose that at time 0 there is just one individual in the population. Let Z_n denote the number of individuals in generation n, so that $Z_0 = 1$ and suppose that N has finite mean.

Question: What is $\mathbb{E}[Z_n|\sigma(Z_{n-1})]$? Intuitively it should be $\mathbb{E}[N] Z_{n-1}$. We prove this rigorously. Let $\{N_i^{(j)}\}_{i,j\in\mathbb{N}}$ be iid copies of N. We have

$$\begin{aligned}
\mathbb{E}[Z_n|\sigma(Z_{n-1})] &= \mathbb{E}\Big[\sum_{i=0}^{\infty} Z_n \mathbf{1}_{\{Z_{n-1}=i\}}|\sigma(Z_{n-1})\Big] \\
&= \mathbb{E}\Big[\lim_{M\to\infty}\sum_{i=0}^{M} Z_n \mathbf{1}_{\{Z_{n-1}=i\}}|\sigma(Z_{n-1})\Big] \\
&\overset{(e)}{=} \lim_{M\to\infty}\mathbb{E}\Big[\sum_{i=0}^{M} Z_n \mathbf{1}_{\{Z_{n-1}=i\}}|\sigma(Z_{n-1})\Big] \\
&\overset{(c)}{=} \lim_{M\to\infty}\sum_{i=0}^{M}\mathbb{E}[Z_n \mathbf{1}_{\{Z_{n-1}=i\}}|\sigma(Z_{n-1})] \\
&= \lim_{M\to\infty}\sum_{i=0}^{M}\mathbb{E}[(N_1^{(n-1)}+\cdots+N_i^{(n-1)})\mathbf{1}_{\{Z_{n-1}=i\}}|\sigma(Z_{n-1})] \\
&\overset{(f)}{=} \lim_{M\to\infty}\sum_{i=0}^{M}\mathbf{1}_{\{Z_{n-1}=i\}}\mathbb{E}[(N_1^{(n-1)}+\cdots+N_i^{(n-1)})|\sigma(Z_{n-1})] \\
&\overset{(g)}{=} \lim_{M\to\infty}\sum_{i=0}^{M}\mathbf{1}_{\{Z_{n-1}=i\}}\mathbb{E}[N_1^{(n-1)}+\cdots+N_i^{(n-1)}] \\
&= \lim_{M\to\infty}\sum_{i=0}^{M}\mathbf{1}_{\{Z_{n-1}=i\}} i\mathbb{E}[N] \\
&= \lim_{M\to\infty}\mathbb{E}[N]\sum_{i=0}^{M} i\mathbf{1}_{\{Z_{n-1}=i\}} \\
&= \lim_{M\to\infty}\mathbb{E}[N]Z_{n-1}\mathbf{1}_{\{Z_{n-1}\le M\}} \\
&= \mathbb{E}[N]Z_{n-1}.
\end{aligned}$$

4.2 Martingales

Definition 4.7. Given a probability space $(\Omega, \Sigma, \mathbb{P})$, we define a *filtration* on $(\Omega, \Sigma, \mathbb{P})$ to be an increasing family of sub-σ-algebras of Σ:

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \Sigma.$$

Intuition: The information about ω in Ω available to us at time n consists of the values of $Z(\omega)$ for all $\mathcal{F}_n$-measurable functions Z.

Example 4.8. Given random variables (X_n), the filtration $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ is called the *natural filtration* of (X_n).

Definition 4.9. A sequence of random variables (X_n) is *adapted* to the filtration if X_n is $\mathcal{F}_n$-measurable for each n.
Intuition: If (X_n) is adapted, the value $X_n(\omega)$ is known to us at time n.

Remark 4.10. (X_n) is always adapted to the natural filtration. Let $S_n = X_1 + \cdots + X_n$. Then (S_n) is also adapted to the natural filtration.

Example 4.11 (Motivating the idea of a martingale)**.**

(1) Let $X_0 = 0$, $X_1 = \pm 1$ with equal probability, and set $X_2 = 2X_1$ with probability $1/2$, otherwise $X_2 = 3X_1$. Is this game fair?
In some sense yes, since $\mathbb{E}[X_2] = \mathbb{E}[X_1] = \mathbb{E}[X_0] = 0$. On the other hand, it is not fair on a day to day basis: if you win on day 1, you know you will win on day 2. A *martingale* is a process which is fair from day to day.

(2) Let $(\varepsilon_k)_{k\in\mathbb{N}}$ be iid random variables such that $\mathbb{P}(\varepsilon_k = 1) = \mathbb{P}(\varepsilon_k = -1) = 1/2$. Play a game: at time $k \in \mathbb{N}$, bet amount Y_k that $\varepsilon_k = 1$ (coin lands heads). The amount bet Y_k is a random variable and is allowed to depend on the history, i.e. on $\varepsilon_1, \dots, \varepsilon_{k-1}$ but not on $\varepsilon_k, \varepsilon_{k+1}, \dots$. At time k you win $\varepsilon_k Y_k$ (which is a loss if $\varepsilon_k = -1$). Your total winnings after n such rounds is $\sum_{k=1}^n \varepsilon_k Y_k$.
This game is fair: conditioned on your total winnings at time $n-1$, your expected total winnings at time n is what you had at time $n-1$:

$$\begin{aligned}
\mathbb{E}\Big[\sum_{k=1}^{n} \varepsilon_k Y_k \big| \sigma(\varepsilon_1, \dots, \varepsilon_{n-1})\Big] &= \mathbb{E}\Big[\varepsilon_n Y_n + \sum_{k=1}^{n-1} \varepsilon_k Y_k \big| \sigma(\varepsilon_1, \dots, \varepsilon_{n-1})\Big] \\
&= \mathbb{E}[\varepsilon_n Y_n | \sigma(\varepsilon_1, \dots, \varepsilon_{n-1})] + \sum_{k=1}^{n-1} \varepsilon_k Y_k \\
&= Y_n \mathbb{E}[\varepsilon_n] + \sum_{k=1}^{n-1} \varepsilon_k Y_k \\
&= \sum_{k=1}^{n-1} \varepsilon_k Y_k.
\end{aligned}$$

Definition 4.12. Let $(X_n)_{n\geq 0}$ be a sequence of random variables and $(\mathcal{F}_n)_{n\geq 0}$ a filtration such that

(1) (X_n) is adapted to $(\mathcal{F}_n)_{n\geq 0}$,

(2) $\mathbb{E}|X_n| < \infty$ for each n.

Then

- (X_n) is a *martingale* with respect to $(\mathcal{F}_n)$ if $\mathbb{E}[X_n|\mathcal{F}_{n-1}] = X_{n-1}$ almost surely, for each $n \geq 1$.
- (X_n) is a *submartingale* with respect to $(\mathcal{F}_n)$ if $\mathbb{E}[X_n|\mathcal{F}_{n-1}] \geq X_{n-1}$ almost surely, for each $n \geq 1$.
- (X_n) is a *supermartingale* with respect to $(\mathcal{F}_n)$ if $\mathbb{E}[X_n|\mathcal{F}_{n-1}] \leq X_{n-1}$ almost surely, for each $n \geq 1$.

Remark 4.13. If (X_n) is a martingale (respectively, submartingale) then

- $\mathbb{E}[X_n|\mathcal{F}_m] = X_m$ (respectively, $\mathbb{E}[X_n|\mathcal{F}_m] \geq X_m$) for all $m < n$.
  This comes from repeated use of the tower property. For martingales (similar for submartingales) we have
  $$\mathbb{E}[X_n|\mathcal{F}_m] = \mathbb{E}\big[\mathbb{E}[X_n|\mathcal{F}_{n-1}]\big|\mathcal{F}_m\big] = \mathbb{E}[X_{n-1}|\mathcal{F}_m] = \cdots = \mathbb{E}[X_{m+1}|\mathcal{F}_m] = X_m.$$
- $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ for all n.
  We know that $\mathbb{E}[X_n|\mathcal{F}_0] = X_0$ and taking expectations and using (a) we are done.

Example 4.14.

(1) The fair game in Example 4.11 (2) is a martingale (assuming the Y_i are bounded) with respect to the natural filtration of (ε_k). Taking $Y_k = 1$ for all k we see that a simple random walk is a martingale.

(2) Let (U_i) be independent integrable random variables with mean 0. Then $X_n = U_1 + \cdots + U_n$ is a martingale with respect to the natural filtration of (U_i):
$$\mathbb{E}[X_n|\sigma(U_1,\ldots,U_{n-1})] = \mathbb{E}[U_1+\cdots+U_n|\sigma(U_1,\ldots,U_{n-1})] = U_1+\cdots+U_{n-1}+\mathbb{E}[U_n] = X_{n-1}.$$

(3) Let (V_i) be non-negative independent random variables with mean 1. Then $X_n = \prod_{i=1}^n V_i$ is a martingale with respect to the natural filtration of (V_i):
$$\begin{aligned} \mathbb{E}[X_n|\sigma(V_1,\ldots,V_{n-1})] &= \mathbb{E}\Big[V_n \prod_{i=1}^{n-1} V_i|\sigma(V_1,\ldots,V_n)\Big] \\ \text{(“taking out known”)} \quad &= \mathbb{E}[V_n|\sigma(V_1,\ldots,V_n)] \prod_{i=1}^{n-1} V_i \\ \text{(indep.)} \quad &= \mathbb{E}[V_n] \prod_{i=1}^{n-1} V_i = X_{n-1}. \end{aligned}$$
The non-negativity is used to show integrability.

(4) Recall the Galton-Watson process. We have already shown that $\mathbb{E}[Z_n|\sigma(Z_{n-1})] = \mathbb{E}[N]Z_{n-1}$. Thus the process (Z_n) is a martingale with respect to the natural filtration iff $\mathbb{E}[N] = 1$ (integrability is clear since the process is non-negative). The process $X_n = Z_n/(\mathbb{E}[N])^n$ is also a martingale (check!).

Definition 4.15. A map $T : \Omega \to \{0\} \cup \mathbb{N} \cup \{\infty\}$ is a *stopping time* with respect to a filtration $(\mathcal{F}_n)$ if
$$\{\omega : T(\omega) = n\} \in \mathcal{F}_n, \ \forall n \leq \infty.$$
Intuition: Can judge if random time T equals n from the information you have by time n.

Remark 4.16. Equivalently, $\{\omega : T(\omega) \leq n\} \in \mathcal{F}_n, \ \forall n \leq \infty$.
Proof of equivalence:
$$\{\omega : T(\omega) \leq n\} = \bigcup_{m \leq n} \{\omega : T(\omega) = m\}$$
and thus if $\{\omega : T(\omega) = m\} \in \mathcal{F}_m \subseteq \mathcal{F}_n$ then $\{\omega : T(\omega) \leq n\} \in \mathcal{F}_n$. Also
$$\{T = n\} = \{T \leq n\} \cap \{T \leq n-1\}^{\complement}$$
and thus if $\{T \leq n\} \in \mathcal{F}_n$ then $\{T = n\} \in \mathcal{F}_n$.

Example 4.17.

(1) Suppose $\mathcal{F}_n$ is weather information up to time n and T is the first rainy day. Then T is a stopping time. On the other hand if T is the last rainy day then T is not a stopping time.

(2) Let (X_n) be a sequence adapted to $(\mathcal{F}_n)$ and $T = \min\{n : X_n > 1\}$. This is a stopping time:

$$\{T = n\} = \{X_m \leq 1 \, \forall m < n, X_n > 1\} = \{X_1 \leq 1\} \cap \cdots \cap \{X_{n-1} \leq 1\} \cap \{X_n > 1\}$$

and note that since (X_n) is adapted we have $\{X_i \leq 1\} \in \mathcal{F}_i \subseteq \mathcal{F}_n$ and thus $\{T = n\} \in \mathcal{F}_n$. On the other hand the map $T = \max\{n : X_n > 1\}$ is not a stopping time since it could be the case that, for example, X_{n+1} is not $\mathcal{F}_n$-measurable.

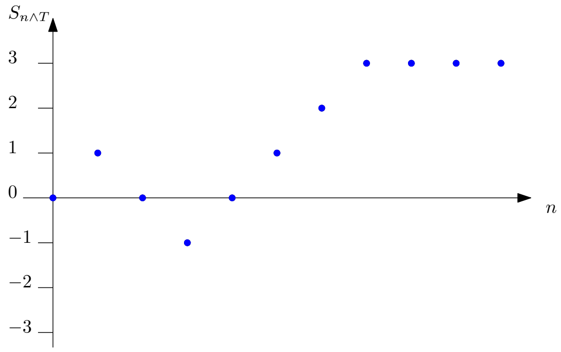
(3) Let (ε_n) be an iid sequence of random variables such that $\mathbb{P}(\varepsilon_n = 1) = \mathbb{P}(\varepsilon_n = -1) = 1/2$ and suppose $(\mathcal{F}_n)$ is the natural filtration. Let $S_n = \varepsilon_1 + \cdots + \varepsilon_n$ (set $S_0 = 0$) so that (S_n) is a simple random walk on $\mathbb{Z}$. Let T be the first time that S_n hits either level $-a$ or level b, for $a, b > 0$, i.e. $T = \min\{n : S_n = -a \text{ or } S_n = b\}$. Then T is a stopping time:

$$\{T = n\} = \{-a < S_1 < b\} \cap \cdots \cap \{-a < S_{n-1} < b\} \cap \{S_n \in \{-a, b\}\} \in \mathcal{F}_n.$$

Let (X_n) be a process adapted to $(\mathcal{F}_n)$ and let T be an a.s. finite stopping time. We will be interested in the random variable X_T $(= X_{T(\omega)}(\omega))$ and the process $(X_{n\wedge T})_{n\geq 0}$ called a *stopped* process.

Example 4.18. Let (S_n) be a simple random walk on $\mathbb{Z}$.

- If $T = \min\{n : S_n = 3\}$ then $S_T = 3$.



- If $T = \min\{n : S_n \in \{-2, 2\}\}$ then (by symmetry) $S_T = 2$ or -2 with equal probability $1/2$.

Theorem 4.19 (Stopped martingales are martingales)**.**
If (X_n) is a martingale (respectively, submartingale) and T a stopping time, then $(X_{n\wedge T})$ is a martingale (respectively, submartingale). In particular, $\mathbb{E}[X_{n\wedge T}] = \mathbb{E}[X_0]$ (respectively, $\mathbb{E}[X_{n\wedge T}] \geq \mathbb{E}[X_0]$) for all $n \in \mathbb{N}$.

Proof. We can write $X_{n\wedge T}$ as

$$X_{n\wedge T} = \sum_{i=1}^{n-1} \mathbf{1}_{\{T=i\}} X_i + \mathbf{1}_{\{T\geq n\}} X_n.$$

From this decomposition we see that $X_{n\wedge T}$ is $\mathcal{F}_n$-measurable: indeed, this follows from (X_n) being adapted and T being a stopping time. Note that in fact $\mathbf{1}_{\{T\geq n\}}$ is $\mathcal{F}_{n-1}$-measurable since

$$\{T \geq n\} = \{T \leq n-1\}^{\complement} \in \mathcal{F}_{n-1}.$$

Also, by the triangle inequality and the above decomposition (and linearity of expectation), we have

$$\mathbb{E}|X_{n\wedge T}| \leq \mathbb{E}\Big[\sum_{i=1}^{n-1} |X_i| + |X_n|\Big] < \infty.$$

For the martingale case (submartingale argument is similar), the final condition to check is that $\mathbb{E}[X_{n\wedge T}|\mathcal{F}_{n-1}] = X_{(n-1)\wedge T}$ almost surely. We have

$$\begin{aligned}
\mathbb{E}[X_{n\wedge T}|\mathcal{F}_{n-1}] &= \mathbb{E}\Big[\sum_{i=1}^{n-1} \mathbf{1}_{\{T=i\}} X_i + \mathbf{1}_{\{T\geq n\}} X_n \big| \mathcal{F}_{n-1}\Big] \\
(\mathbf{1}_{\{T=i\}} X_i \text{ is } \mathcal{F}_i\text{-measurable}) \quad &= \sum_{i=1}^{n-1} \mathbf{1}_{\{T=i\}} X_i + \mathbb{E}[\mathbf{1}_{\{T\geq n\}} X_n \big| \mathcal{F}_{n-1}] \\
(\text{“taking out known”}) \quad &= \sum_{i=1}^{n-1} \mathbf{1}_{\{T=i\}} X_i + \mathbf{1}_{\{T\geq n\}} \mathbb{E}[X_n | \mathcal{F}_{n-1}] \\
&= \sum_{i=1}^{n-1} \mathbf{1}_{\{T=i\}} X_i + \mathbf{1}_{\{T\geq n\}} X_{n-1} \\
&= \sum_{i=1}^{n-2} \mathbf{1}_{\{T=i\}} X_i + (\mathbf{1}_{\{T=n-1\}} + \mathbf{1}_{\{T\geq n\}}) X_{n-1} \\
&= \sum_{i=1}^{n-2} \mathbf{1}_{\{T=i\}} X_i + \mathbf{1}_{\{T\geq n-1\}} X_{n-1} \\
&= X_{(n-1)\wedge T}.
\end{aligned}$$

□

Remark 4.20. As $n \to \infty$, $n\wedge T \to T$ almost surely (since T is a.s. finite) and so $X_{n\wedge T} \to X_T$ almost surely. Is it true that $\mathbb{E}[X_T] = \lim_{n\to\infty} \mathbb{E}[X_{n\wedge T}] = \mathbb{E}[X_0]$?
No: consider the SRW on $\mathbb{Z}$ and $T = \min\{n : S_n = 1\}$ (it can be shown that such a T is a.s. finite). Then $\mathbb{E}[S_T] = 1 \neq 0 = \mathbb{E}[S_0]$. We would like to know when we can say that $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Theorem 4.21 (Doob's Optional Stopping/Sampling Theorem (OST))**.**
Let (X_n) be a martingale (respectively, submartingale) and T a stopping time. Then X_T is integrable and $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ (respectively, $\mathbb{E}[X_T] \geq \mathbb{E}[X_0]$) in each of the following situations:

(i) T is bounded,

(ii) T is a.s. finite and (X_n) is uniformly bounded (that is, $\exists c > 0 : |X_n(\omega)| < c \,\forall n, \omega$),

(iii) $\mathbb{E}[T] < \infty$ and for some $c > 0$, $|X_n(\omega) - X_{n-1}(\omega)| \le c \;\forall n, \omega$.

Proof. For the case where (X_n) is a submartingale we know by Theorem 4.19 that $(X_{n\wedge T})$ is a submartingale, thus $\mathbb{E}[X_{n\wedge T}] \ge \mathbb{E}[X_0]$ for all $n \in \mathbb{N}$.

(i) If $T(\omega) \le N \;\forall \omega \in \Omega$, then $\mathbb{E}[X_T] = \mathbb{E}[X_{N\wedge T}] \ge \mathbb{E}[X_0]$. Further, $\mathbb{E}|X_T| = \mathbb{E}|X_{N\wedge T}| < \infty$ as $(X_{n\wedge T})$ is integrable.

(ii) Integrability follows from uniform boundedness of (X_n):

$$|X_T| \le \sum_{n=0}^{\infty} |X_n| \mathbf{1}_{\{T=n\}} \le c \sum_{n=0}^{\infty} \mathbf{1}_{\{T=n\}} = c.$$

The $X_{n\wedge T}$ are also uniformly bounded so by (DOM) $\mathbb{E}[X_T] = \lim_{n\to\infty} \mathbb{E}[X_{n\wedge T}] \ge \mathbb{E}[X_0]$.

(iii) We have

$$|X_{n\wedge T} - X_0| = \Big| \sum_{k=1}^{n\wedge T} (X_k - X_{k-1}) \Big| \le \sum_{k=1}^{n\wedge T} |X_k - X_{k-1}| \le c(n \wedge T) \le cT.$$

By the triangle inequality, for all n,

$$|X_{n\wedge T}| \le |X_{n\wedge T} - X_0| + |X_0| \le cT + |X_0|,$$

and so, since $\mathbb{E}T < \infty$ and X_0 is integrable, we can apply (DOM) to deduce that $\mathbb{E}[X_T] = \lim_{n\to\infty} \mathbb{E}[X_{n\wedge T}] \ge \mathbb{E}[X_0]$. To see integrability of X_T, we write

$$|X_T| - |X_0| \le |X_T - X_0| \le \sum_{k=1}^{T} |X_k - X_{k-1}| \le cT.$$

The argument is similar for the case of martingales. □

Remark 4.22.

(1) For (ii) it suffices to have the weaker condition that $(X_{n\wedge T})$ is uniformly bounded rather than (X_n). Similarly in (iii) it suffices to have bounded increments for $(X_{n\wedge T})$ instead of (X_n).

(2) The simple random walk on $\mathbb{Z}$ (S_n) and stopping time $T = \min\{n : S_n = 1\}$ violate all three conditions (T is a.s. finite but has infinite mean).

(3) If you play a fair game (martingale) then you cannot expect to win, even if you decide to stop the game at some random time. However, you can sometimes beat the system if you have an infinite amount of time or money.

Example 4.23.

(1) Let (S_n) be the simple random walk on $\mathbb{Z}$.

- What is the probability that it will reach level $b > 0$ before reaching level $-a < 0$? Set $T = \min\{n : S_n \in \{-a, b\}\}$ and stop the martingale (S_n) at stopping time T. Then $(S_{n\wedge T})$ is uniformly bounded (by $\max\{a, b\}$) so we can apply OST with condition (ii) if we can show that T is almost surely finite. One way to show that T is a.s. finite: for each n,

$$\begin{aligned}\mathbb{P}(T = \infty) &\le \mathbb{P}(\text{no sequence of } (a+b) \text{ successive moves to the right up to time } n(a+b)) \\ &\le \mathbb{P}(\ge 1 \text{ left jump in } (0, a+b], \ldots, \ge 1 \text{ left jump in } ((n-1)(a+b), n(a+b)]) \\ &= (1 - (1/2)^{a+b})^n \le e^{-(1/2)^{a+b}} \to 0,\end{aligned}$$

as $n \to \infty$. Thus we can apply OST to deduce that

$$0 = \mathbb{E}[S_0] = \mathbb{E}[S_T] = b\mathbb{P}(S_T = b) - a\mathbb{P}(S_T = -a) = b\mathbb{P}(S_T = b) - a(1 - \mathbb{P}(S_T = b)),$$

and so we have $\mathbb{P}(S_T = b) = a/(a+b)$.

- How long do we have to wait until (S_n) hits one of the barriers? We can solve this problem by using another martingale. Set $M_n = S_n^2 - n$. Then (M_n) is clearly adapted to the natural filtration $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and is clearly integrable ($\mathbb{E}|M_n| \le n^2 + n$). Furthermore,

$$\begin{aligned}\mathbb{E}[M_{n+1}|\mathcal{F}_n] &= \mathbb{E}[(S_n + X_{n+1})^2 - (n+1)|\mathcal{F}_n] \\ &= \mathbb{E}[S_n^2 + 2X_{n+1}S_n + X_{n+1}^2 - n - 1|\mathcal{F}_n] \\ &= S_n^2 + 2S_n\mathbb{E}[X_{n+1}] + \mathbb{E}[X_{n+1}^2] - n - 1 \\ &= S_n^2 - n = M_n.\end{aligned}$$

Thus (M_n) is a martingale. We cannot direcly apply OST since the increments of M_n are not bounded and M_n is not uniformly bounded. However,

$$0 = \mathbb{E}[M_0] = \mathbb{E}[M_{n\wedge T}] = \mathbb{E}[S_{n\wedge T}^2] - \mathbb{E}[n \wedge T].$$

Thus $\mathbb{E}[n \wedge T] = \mathbb{E}[S_{n\wedge T}^2]$. By MON, $\mathbb{E}[n \wedge T] \to \mathbb{E}[T]$ (T is a.s. finite). Also $S_{n\wedge T}^2 \le \max\{a^2, b^2\}$. So by DOM, $\mathbb{E}[S_{n\wedge T}^2] \to \mathbb{E}[S_T^2]$. Thus $\mathbb{E}[T] = \mathbb{E}[S_T^2]$. However,

$$\mathbb{E}[S_T^2] = a^2\frac{b}{a+b} + b^2\frac{a}{a+b} = \frac{ab(a+b)}{a+b} = ab,$$

i.e. $\mathbb{E}[T] = ab$.

(2) We repeatedly roll a die and wait to see the sequence 456 or 666. How long do we expect to have to wait in each case? Idea: construct a martingale!

- 666
Just before each roll of the die, a new gambler arrives and bets 1 that the next roll will be a 6. If he loses he leaves the game forever. If he wins he receives 6 and bets it all on the event that the next roll is a 6. Again, if he loses he leaves forever and if he wins he receives 6^2 and bets it all on the next roll being a 6. If he wins he get 6^3 and regardless of the outcome of the roll then leaves.
Since this game is fair, we should think that the total investment minus total winnings (of all gamblers) at time n is a martingale.

If we stop the game at time T when 666 first occurs, the total winnings will be $6^3 + 6^2 + 6$ (the 6^3 is winnings for gambler who arrived at time $T-2$, the 6^2 for gambler who arrived at time $T-1$ and 6 for gambler who arrived at time T). The total investment at time T is T (since 1 put into the system at each time step by the new gambler at that time). Thus it seems like we can use the OST to show that $\mathbb{E}[T] = 6^3 + 6^2 + 6$. We show this rigorously.

Let (X_n) denote the independent die rolls and $(\mathcal{F}_n)$ their natural filtration. The total winnings at time n is

$$W_n = \sum_{i=1}^{n-2} 6^3 \mathbf{1}_{\{X_i=6, X_{i+1}=6, X_{i+2}=6\}} + 6^2 \mathbf{1}_{\{X_{n-1}=6, X_n=6\}} + 6\mathbf{1}_{\{X_n=6\}}.$$

The total investment at time n is n. Denote

$$M_n := W_n - n,$$

so that (M_n) is adapted and integrable and furthermore:

$$\begin{aligned}
\mathbb{E}[M_{n+1}|\mathcal{F}_n] &= \sum_{i=1}^{n-2} 6^3 \mathbf{1}_{\{X_i=6, X_{i+1}=6, X_{i+2}=6\}} + \mathbb{E}[6^3 \mathbf{1}_{\{X_{n-1}=6, X_n=6, X_{n+1}=6\}}|\mathcal{F}_n] \\
&\quad + \mathbb{E}[6^2 \mathbf{1}_{\{X_n=6, X_{n+1}=6\}}|\mathcal{F}_n] + \mathbb{E}[6\mathbf{1}_{\{X_{n+1}=6\}}|\mathcal{F}_n] - (n+1) \\
&= \sum_{i=1}^{n-2} 6^3 \mathbf{1}_{\{X_i=6, X_{i+1}=6, X_{i+2}=6\}} + 6^3 \mathbf{1}_{\{X_{n-1}=6, X_n=6\}} \times \frac{1}{6} \\
&\quad + 6^2 \mathbf{1}_{\{X_n=6\}} \times \frac{1}{6} + 1 - n - 1 \\
&= \sum_{i=1}^{n-2} 6^3 \mathbf{1}_{\{X_i=6, X_{i+1}=6, X_{i+2}=6\}} + 6^2 \mathbf{1}_{\{X_{n-1}=6, X_n=6\}} + 6\mathbf{1}_{\{X_n=6\}} - n \\
&= M_n.
\end{aligned}$$

Thus (M_n) is a martingale. We set

$$T = \min\{n : X_{n-2} = 6, X_{n-1} = 6, X_n = 6\}.$$

This is clearly a stopping time and we claim it has finite mean. To see this note that we can write

$$\mathbb{E}[T] = \sum_{m=1}^{\infty} \mathbb{P}(T \geq m).$$

Set $k = 3\lfloor \frac{m-1}{3} \rfloor$. We also have

$$\begin{aligned}
\mathbb{P}(T \geq m) &= \mathbb{P}(\text{no 3 successive 6s up to time } m-1) \\
&\leq \mathbb{P}(\text{no 3 6s at times } \{1,2,3\}, \ldots, \text{no 3 6s at times } \{k-2, k-1, k\}) \\
&= (1 - (1/6)^3)^{\lfloor \frac{m-1}{3} \rfloor}.
\end{aligned}$$

Thus

$$\mathbb{E}[T] \leq \sum_{m=3}^{\infty} (1 - (1/6)^3)^{\lfloor \frac{m-1}{3} \rfloor} < \infty,$$

as claimed. Thus by OST,

$$0 = \mathbb{E}[M_0] = \mathbb{E}[M_T] = \mathbb{E}[W_T] - \mathbb{E}[T].$$

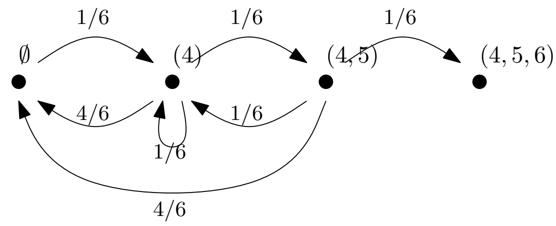
But since $\mathbb{E}[W_T] = 6^3 + 6^2 + 6$, we deduce that $\mathbb{E}[T] = 6^3 + 6^2 + 6$.
Alternatively, we can avoid using OST and argue as follows. For each n,

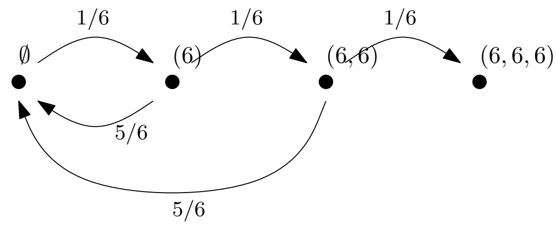
$$0 = \mathbb{E}[M_0] = \mathbb{E}[M_{n\wedge T}] = \mathbb{E}[W_{n\wedge T}] - \mathbb{E}[n \wedge T].$$

Thus $\mathbb{E}[n\wedge T] = \mathbb{E}[W_{n\wedge T}]$. By MON, $\mathbb{E}[n\wedge T] \to \mathbb{E}[T]$ and since $|W_{n\wedge T}| \leq 6^3+6^2+6$, by DOM, $\mathbb{E}[W_{n\wedge T}] \to \mathbb{E}[W_T]$. Thus $\mathbb{E}[T] = \mathbb{E}[W_T] = 6^3 + 6^2 + 6$.

- 456 Exercise!
  Show that $\mathbb{E}[T] = 6^3$ and thus that we expect to wait less time to see 456 compared to 666.

This may seem counter-intuitive at first, but consider the following representation of the problem:





From drawing the two chains, we see why 666 takes longer to occur on average.

4.3 Maximal inequalities

Reminder 4.24 (Markov's/Chebyshev's inequality)**.**
For a non-negative random variable X, and $c > 0$,

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[X]}{c}.$$

Theorem 4.25 (Doob's submartingale inequality)**.**
Let (X_n) be a non-negative submartingale. Then, for each $c > 0$,

$$\mathbb{P}\big(\max_{0\leq k\leq n} X_k \geq c\big) \leq \frac{\mathbb{E}[X_n]}{c}.$$

Remark 4.26. This is an extension of Markov's inequality, exploiting the fact that the (X_n) are growing in expectation.

Proof. Set $S = \min\{k : X_k \geq c\}$ and $T = n \wedge S$. Then T is a stopping time:

$$\{T = k\} = \bigcap_{i=1}^{k-1} \{X_i < c\} \cap \{X_k \geq c\} \in \mathcal{F}_k, \quad 0 \leq k \leq n-1,$$
$$\{T = n\} = \bigcap_{i=1}^{n-1} \{X_i < c\} \in \mathcal{F}_n,$$
$$\{T = k\} = \varnothing \in \mathcal{F}_k, \quad k > n.$$

We know $T \leq n$ and claim that $\mathbb{E}[X_T] \leq \mathbb{E}[X_n]$. We have

$$\mathbb{E}[X_T] = \mathbb{E}\Big[\sum_{k=0}^{n} X_k \mathbf{1}_{\{T=k\}}\Big] = \sum_{k=0}^{n} \mathbb{E}[X_k \mathbf{1}_{\{T=k\}}]. \tag{4}$$

Since (X_n) is a submartingale $\mathbb{E}[X_n|\mathcal{F}_k] \geq X_k$ almost surely. Since $\{T = k\} \in \mathcal{F}_k$, and by the definition of conditional expectation,

$$\mathbb{E}[X_k \mathbf{1}_{\{T=k\}}] = \int_{\{T=k\}} X_k \mathrm{d}\mathbb{P} \leq \int_{\{T=k\}} \mathbb{E}[X_n|\mathcal{F}_k]\mathrm{d}\mathbb{P} = \int_{\{T=k\}} X_n \mathrm{d}\mathbb{P} = \mathbb{E}[X_n \mathbf{1}_{\{T=k\}}].$$

Plugging this into equation (4) we obtain

$$\mathbb{E}[X_T] \leq \sum_{k=0}^{n} \mathbb{E}[X_n \mathbf{1}_{\{T=k\}}] = \mathbb{E}\Big[X_n \sum_{k=0}^{n} \mathbf{1}_{\{T=k\}}\Big] = \mathbb{E}[X_n],$$

as claimed. Now, using the fact that X_n is non-negative,

$$\mathbb{E}[X_n] \geq \mathbb{E}[X_T] = \mathbb{E}[X_S \mathbf{1}_{\{S \leq n\}}] + \mathbb{E}[X_n \mathbf{1}_{\{S > n\}}] \geq c\mathbb{P}(S \leq n).$$

But $\{S \leq n\} = \{\max_{0 \leq k \leq n} X_k \geq c\}$, and so we have

$$c\mathbb{P}\big(\max_{0 \leq k \leq n} X_k \geq c\big) \leq \mathbb{E}[X_n],$$

completing the proof. □

Theorem 4.27 (Kolmogorov's inequality)**.**
Let (X_n) be a sequence of independent random variables with $\mathbb{E}[X_n^2] < \infty$ and $\mathbb{E}[X_n] = 0$, for all n. Then, for each $c > 0$,

$$\mathbb{P}\Big(\max_{1 \leq k \leq n} \Big|\sum_{i=1}^{k} X_i\Big| \geq c\Big) \leq \frac{\sum_{i=1}^{n} \mathbb{E}[X_i^2]}{c^2}.$$

Proof. Set $S_n = X_1 + \cdots + X_n$ so that (S_n) is a martingale with respect to the natural filtration of (X_n). Then S_n^2 is integrable for each n since

$$\mathbb{E}[S_n^2] = \mathbb{E}\Big[\Big(\sum_{i=1}^{n} X_i\Big)^2\Big] = \mathbb{E}\Big[\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j\Big] = \sum_{i-1}^{n} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i]\mathbb{E}[X_j] = \sum_{i=1}^{n} \mathbb{E}[X_i^2] < \infty.$$

By conditional Jensen's inequality with $\phi(x) = x^2$, we have

$$\mathbb{E}[S_n^2 | \mathcal{F}_{n-1}] \geq (\mathbb{E}[S_n | \mathcal{F}_{n-1}])^2 = S_{n-1}^2,$$

and thus (S_n^2) is a non-negative submartingale. We can thus apply Doob's submartingale inequality to (S_n^2):

$$\mathbb{P}\Big(\max_{1 \leq k \leq n} \Big| \sum_{i=1}^k X_i \Big| \geq c \Big) = \mathbb{P}\Big(\max_{1 \leq k \leq n} \Big(\sum_{i=1}^k X_i \Big)^2 \geq c^2 \Big) \leq \frac{\mathbb{E}[S_n^2]}{c^2} = \frac{\sum_{i=1}^n \mathbb{E}[X_i^2]}{c^2}. \qquad \square$$

Example 4.28 (Durrett)**.**
Suppose (X_n) is a sequence of independent identically distributed random variables with zero mean and $\mathrm{Var}(X_1) < 1$. Let $S_n = X_1 + \cdots + X_n$ and fix $p > 1/2$.
Claim: $S_n / n^p \to 0$ almost surely, as $n \to \infty$.
Fix $N \in \mathbb{N}$; we have by Kolmogorov's inequality,

$$\mathbb{P}\Big(\max_{1 \leq k \leq N} \frac{|S_k|}{N^p} > \varepsilon \Big) \leq \varepsilon^{-2} \sum_{i=1}^N \frac{\mathbb{E}[X_i^2]}{N^{2p}} < \varepsilon^{-2} N^{1-2p}.$$

Since $p > 1/2$, we can choose $\alpha > 0$ with $\alpha(2p-1) > 1$. Now let $m = N^{1/\alpha}$ so that

$$\mathbb{P}\Big(\max_{1 \leq k \leq m^\alpha} \frac{|S_k|}{m^{\alpha p}} > \varepsilon \Big) < \varepsilon^{-2} m^{\alpha(1-2p)},$$

and so since $\alpha(1-2p) < -1$,

$$\sum_{m=1}^\infty \mathbb{P}\Big(\max_{1 \leq k \leq m^\alpha} \frac{|S_k|}{m^{\alpha p}} > \varepsilon \Big) < \infty.$$

By Borel-Cantelli 1, we therefore have

$$\mathbb{P}\Big(\max_{1 \leq k \leq m^\alpha} \frac{|S_k|}{m^{\alpha p}} > \varepsilon \text{ i.o.} \Big) = 0, \quad \forall \varepsilon > 0,$$

and thus

$$\max_{1 \leq k \leq m^\alpha} \frac{|S_k|}{m^{\alpha p}} \to 0 \quad \text{almost surely as } m \to \infty. \tag{5}$$

We want to show $|S_n|/n^p \to 0$ almost surely as $n \to \infty$, and so set $m = m(n) = n^{1/\alpha}$ and note that $(m-1)^\alpha < n$. Then

$$\frac{|S_n|}{n^p} \leq \max_{1 \leq k \leq n} \frac{|S_k|}{n^p} = \max_{1 \leq k \leq m^\alpha} \frac{|S_k|}{n^p} \leq \max_{1 \leq k \leq m^\alpha} \frac{|S_k|}{(m-1)^{\alpha p}} = \max_{1 \leq k \leq m^\alpha} \frac{|S_k|}{m^{\alpha p}} \Big(\frac{m}{m-1} \Big)^{\alpha p} \to 0,$$

almost surely as n (and hence m) goes to ∞, since $\frac{m}{m-1} \to 1$.

4.4 Strong Law of Large Numbers

Theorem 4.29 (Kolmogorov's theorem)**.**
Let (X_n) be independent random variables with mean 0 and $\sum_{n=1}^{\infty} \mathbb{E}[X_n^2] < \infty$. Then $\sum_{n=1}^{\infty} X_n < \infty$ almost surely.

Proof. We need to show that $\mathbb{P}\big(\sum_{i=1}^{\infty} X_i < \infty\big) = 1$. It thus suffices to show that

$$\begin{aligned}
&\mathbb{P}\Big(\big(\sum_{i=1}^{n} X_i\big)_{n\in\mathbb{N}} \text{ is a Cauchy sequence}\Big) = 1 \\
\text{i.e.}\quad &\mathbb{P}\Big(\bigcap_{k=1}^{\infty}\bigcup_{N=1}^{\infty}\bigcap_{n=N+1}^{\infty} \{|\sum_{i=1}^{n} X_i - \sum_{i=1}^{N} X_i| < 1/k\}\Big) = 1 \\
\text{i.e.}\quad &\mathbb{P}\Big(\bigcup_{N=1}^{\infty}\bigcap_{n=N+1}^{\infty} \{|\sum_{i=N+1}^{n} X_i| < 1/k\}\Big) = 1 \qquad \text{for each } k \\
\text{i.e.}\quad &\mathbb{P}\Big(\bigcap_{N=1}^{\infty}\bigcup_{n=N+1}^{\infty} \{|\sum_{i=N+1}^{n} X_i| \geq 1/k\}\Big) = 0 \qquad \text{for each } k.
\end{aligned}$$

We have

$$\mathbb{P}\Big(\bigcap_{N=1}^{\infty}\bigcup_{n=N+1}^{\infty} \{|\sum_{i=N+1}^{n} X_i| \geq 1/k\}\Big) \leq \lim_{N\to\infty} \mathbb{P}\Big(\bigcup_{n=N+1}^{\infty} \{|\sum_{i=N+1}^{n} X_i| \geq 1/k\}\Big). \tag{6}$$

For $M \geq N$, by Kolmogorov's inequality

$$\mathbb{P}\Big(\bigcup_{n=N+1}^{M} \{|\sum_{i=N+1}^{n} X_i| \geq 1/k\}\Big) = \mathbb{P}\Big(\max_{N+1\leq n\leq M} |\sum_{i=N+1}^{n} X_i| \geq 1/k\Big) \leq k^2 \sum_{n=N+1}^{M} \mathbb{E}[X_n^2].$$

Hence, by MON for sets, we have

$$\begin{aligned}
\mathbb{P}\Big(\bigcup_{n=N+1}^{\infty} \{|\sum_{i=N+1}^{n} X_i| \geq 1/k\}\Big) &= \lim_{M\to\infty} \mathbb{P}\Big(\bigcup_{n=N+1}^{M} \{|\sum_{i=N+1}^{n} X_i| \geq 1/k\}\Big) \\
&\leq k^2 \sum_{n=N+1}^{\infty} \mathbb{E}[X_n^2] \to 0 \quad \text{as } N \to \infty,
\end{aligned}$$

since $\sum_{n=1}^{\infty} \mathbb{E}[X_n^2] < \infty$. Plugging this into (6) gives the result. □

Example 4.30. Let (ε_n) be a sequence of independent identically distributed random variables with

$$\mathbb{P}(\varepsilon_n = 1) = \mathbb{P}(\varepsilon_n = -1) = 1/2.$$

Then

$$\sum_{n=1}^{\infty} \mathbb{E}[(\varepsilon_n/n)^2] = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

so by Kolmogorov's theorem, $\sum_{n=1}^{\infty} \varepsilon_n/n < \infty$ almost surely. Note that this is therefore more in common with the sum $\sum_{n=1}^{\infty}(-1)^n/n = -\log 2 < \infty$ than with $\sum_{n=1}^{\infty} 1/n = \infty$.

Moreover,

$$\sum_{n=1}^{\infty} \frac{\varepsilon_n}{n^{1/2+\delta}} < \infty$$

almost surely for $\delta > 0$. On the other hand (see problem sheet)

$$\sum_{n=1}^{\infty} \frac{\varepsilon_n}{\sqrt{n}} = \infty$$

almost surely.

Lemma 4.31 (Cesaro)**.**
If $a_n \to a$ then

$$\frac{a_1 + \cdots + a_n}{n} \to a.$$

Proof. Exercise in Analysis 1. □

Lemma 4.32 (Kronecker)**.**
If $\sum_{n=1}^{\infty} \frac{a_n}{n} < \infty$, then

$$\frac{a_1 + \cdots + a_n}{n} \to 0.$$

Proof. Denote $u_0 = 0$, $u_n = \sum_{i=1}^{n} \frac{a_i}{i}$ and $u = \sum_{i=1}^{\infty} \frac{a_i}{i}$. We have $u_n \to u$. Observe that $a_n = n(u_n - u_{n-1})$. We have

$$\begin{aligned}\frac{a_1 + \cdots + a_n}{n} &= \frac{1}{n}(u_1 - u_2 + 2(u_2 - u_1) + \cdots + n(u_n - u_{n-1})) \\ &= \frac{1}{n}(nu_n - u_0 - u_1 - \cdots - u_{n-1}) = u_n - \frac{u_0 + u_1 + \cdots + u_{n-1}}{n} \overset{\text{Cesaro}}{\to} u - u = 0.\end{aligned}$$

□

Remark 4.33. So far we have seen the proof of SLLN for 1/2-Bernoulli random variables (lectures) and for bounded random variables (homework). Here we extend it to square-integrable random variables, which includes Poisson, Exponential, Normal, etc. The case of integrable but not square-integrable random variables remains open (e.g $\mathbb{P}(X > x) = x^{-3/2}$).

Theorem 4.34 (SLLN for square-integrable random variables)**.**
Let (X_n) be independent identically distributed random variables with 0 mean and finite variance. Then

$$\frac{X_1 + \cdots + X_n}{n} \to 0,$$

almost surely as $n \to \infty$.

Proof. Set $Y_n = X_n/n$. Then (Y_n) is a sequence of independent random variables with 0 mean and

$$\sum_{n=1}^{\infty} \mathbb{E}[Y_n^2] = \mathbb{E}[X_1^2] \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Therefore by Kolmogorov's theorem, $\sum_{n=1}^{\infty} Y_n < \infty$ almost surely, i.e. $\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty$ almost surely. Now apply Kronecker's lemma to complete the proof. □

Remark 4.35. This result can obviously be extended to iid square-integrable random variables with a finite, non-zero mean.

We are now going to see how we can prove the general SLLN for independent identically distributed integrable random variables using Kolmogorov's Theorem plus a truncation technique.

Lemma 4.36 (Kolmogorov's truncation lemma)**.**
Let (X_n) be iid integrable random variables with mean μ. Define

$$Y_n = \begin{cases} X_n, & \text{if } |X_n| \leq n, \\ 0, & \text{otherwise.} \end{cases}$$

Then

(a) $\mathbb{P}(X_n = Y_n \text{ eventually}) \Big(= \mathbb{P}(\exists N : \forall n \geq N, X_n = Y_n)\Big) = 1,$

(b) $\mathbb{E}[Y_n] \to \mu,$

(c) $\sum_{n=1}^{\infty} \frac{\mathrm{Var}(Y_n)}{n^2} < \infty.$

Proof.

(a) $\mathbb{P}(X_n = Y_n \text{ eventually}) = \mathbb{P}(|X_n| \leq n \text{ eventually}) = 1 - \mathbb{P}(|X_n| > n \text{ i.o.})$. Note that

$$\begin{aligned}\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n) &= \sum_{n=1}^{\infty} \mathbb{P}(|X_1| > n) = \sum_{n=1}^{\infty} \mathbb{E}[\mathbf{1}_{\{|X_1|>n\}}] = \lim_{M\to\infty} \sum_{n=1}^{M} \mathbb{E}[\mathbf{1}_{\{|X_1|>n\}}] \\ &= \lim_{M\to\infty} \mathbb{E}\big[\sum_{n=1}^{M} \mathbf{1}_{\{|X_1|>n\}}\big] \leq \lim_{M\to\infty} \mathbb{E}\big[\sum_{n=1}^{M\wedge|X_1|} 1\big] = \lim_{M\to\infty} \mathbb{E}[M \wedge |X_1|] \\ &\leq \mathbb{E}|X_1| < \infty.\end{aligned}$$

Applying Borel-Cantelli 1 gives (a).

(b) $\mathbb{E}[Y_n] = \mathbb{E}[Y_n \mathbf{1}_{\{|X_n|\leq n\}}] + \mathbb{E}[Y_n \mathbf{1}_{\{|X_n|>n\}}] = \mathbb{E}[X_1 \mathbf{1}_{\{|X_1|\leq n\}}]$. However, $X_1 \mathbf{1}_{\{|X_1|\leq n\}} \to X_1$ and $|X_1 \mathbf{1}_{\{|X_1|\leq n\}}| \leq |X_1|$ so by DOM,

$$\mathbb{E}[X_1 \mathbf{1}_{\{|X_1|\leq n\}}] \to \mathbb{E}[X_1] = \mu.$$

(c)

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{\mathbb{E}[Y_n^2]}{n^2} = \sum_{n=1}^{\infty} \frac{\mathbb{E}[X_n^2 \mathbf{1}_{\{|X_n|\leq n\}}]}{n^2} &= \sum_{n=1}^{\infty} \frac{\mathbb{E}[X_1^2 \mathbf{1}_{\{|X_1|\leq n\}}]}{n^2} \\ &\overset{\text{MON}}{=} \mathbb{E}\Big[\sum_{n=1}^{\infty} \frac{X_1^2 \mathbf{1}_{\{|X_1|\leq n\}}}{n^2}\Big] = \mathbb{E}\Big[X_1^2 \sum_{n\geq|X_1|} \frac{1}{n^2}\Big].\end{aligned}$$

If $|X_1| < 1$,

$$\mathbb{E}\Big[X_1^2 \sum_{n \geq |X_1|} \frac{1}{n^2}\Big] \leq \mathbb{E}[|X_1|^2 \pi^2/6] < \infty.$$

Observe that for $m \geq 1$,

$$\sum_{n=m}^{\infty} \frac{1}{n^2} \leq \sum_{n=m}^{\infty} \frac{2}{n(n+1)} = 2\sum_{n=m}^{\infty} \Big(\frac{1}{n} - \frac{1}{n+1}\Big) = \frac{2}{m}.$$

Thus if $|X_1| \geq 1$,

$$\mathbb{E}\Big[X_1^2 \sum_{n \geq |X_1|} \frac{1}{n^2}\Big] \leq \mathbb{E}\Big[X_1^2 \frac{2}{|X_1|}\Big] = 2\mathbb{E}[|X_1|] < \infty.$$

We deduce that

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}[Y_n^2]}{n^2} < \infty.$$

By (b), there exists a $c > 0$ such that we have

$$\sum_{n=1}^{\infty} \frac{(\mathbb{E}[Y_n])^2}{n^2} \leq \sum_{n=1}^{\infty} \frac{c}{n^2} < \infty.$$

Using $\mathrm{Var}[Y_n] = \mathbb{E}[Y_n^2] - (\mathbb{E}[Y_n])^2$ completes the proof. □

Theorem 4.37 (Strong Law of Large Numbers)**.**
Let (X_n) be iid integrable random variables with mean μ. Then

$$\frac{X_1 + \cdots + X_n}{n} \to \mu \quad \textit{a.s. as } n \to \infty.$$

Proof. Define Y_n as in Kolmogorov's Truncation Lemma. By KTL(a) we know that $X_n = Y_n$ eventually and so it suffices to prove that

$$\frac{Y_1 + \cdots + Y_n}{n} \to \mu \quad \text{a.s. as } n \to \infty.$$

By KTL(b), $\mathbb{E}[Y_n] \to \mu$, so by Cesaro's Lemma,

$$\frac{\mathbb{E}[Y_1] + \cdots + \mathbb{E}[Y_n]}{n} \to \mu.$$

Note that we can write

$$\begin{aligned}
\frac{Y_1 + \cdots + Y_n}{n} &= \frac{1}{n}\Big\{(Y_1 - \mathbb{E}[Y_1]) + \cdots + (Y_n - \mathbb{E}[Y_n]) + \mathbb{E}[Y_1 + \cdots + Y_n]\Big\} \\
&= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i]) + \frac{\mathbb{E}[Y_1] + \cdots + \mathbb{E}[Y_n]}{n}.
\end{aligned}$$

Thus it suffices to show

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i]) \to 0 \quad \text{a.s. as } n \to \infty.$$

Set $Z_n = Y_n - \mathbb{E}[Y_n]$. Then (Z_n) are independent, zero-mean random variables with

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}[Z_n^2]}{n^2} = \sum_{n=1}^{\infty} \frac{\text{Var}\,[Y_n]}{n^2} < \infty$$

by KTL(c). So by Kolmogorov's Theorem,

$$\sum_{n=1}^{\infty} \frac{Z_n}{n} < \infty \quad \text{a.s.}$$

Thus by Kronecker,

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i]) = \frac{1}{n}\sum_{i=1}^{n} Z_i \to 0 \quad \text{a.s.}$$

□

4.5 Martingale Convergence Theorem

For a martingale (X_n) with $X_0 = 0$, think of $X_n - X_{n-1}$ as your net winnings per unit stake in game n in a series of fair games, played at times $n = 1, 2, \ldots$.

Definition 4.38. A process (C_n) is *previsible/predictable* with respect to a filtration $(\mathcal{F}_n)$ if C_n is $\mathcal{F}_{n-1}$-measurable for each $n \geq 1$. For a martingale (X_n) and a previsible process (C_n) the process (Y_n) defined as $Y_0 = 0$ and for all $n \in \mathbb{N}$,

$$Y_n = \sum_{i=1}^{n} C_i(X_i - X_{i-1})$$

is called the *martingale transform* of (X_n) by (C_n). It is denoted as $((C \bullet X)_n)$.

Remark 4.39. Since we can write any process (X_n) with $X_0 = 0$ as

$$X_n = \sum_{i=1}^{n} 1 \cdot (X_i - X_{i-1}),$$

it corresponds to betting £1 at every step of the game. When (X_n) is a martingale, doing the martingale transform corresponds to betting C_n at time n. You have to decide on the value of C_n based on the history up to time $n-1$ (hence the term previsible). Can we choose (C_n) in such a way so that our expected total winnings is positive?...

Theorem 4.40 (You cannot beat the system)**.**
If each C_n is bounded then $((C \bullet X)_n)$ is a martingale.

Proof. $Y_n := \sum_{i=1}^{n} C_i(X_i - X_{i-1})$ is obviously $\mathcal{F}_n$-measurable. It is integrable since each C_n is bounded. Finally,

$$\mathbb{E}[Y_n|\mathcal{F}_{n-1}] = Y_{n-1} + \mathbb{E}[C_n(X_n - X_{n-1})|\mathcal{F}_{n-1}] = Y_{n-1} + C_n(\mathbb{E}[X_n|\mathcal{F}_{n-1}] - X_{n-1}) = Y_{n-1}.$$

□

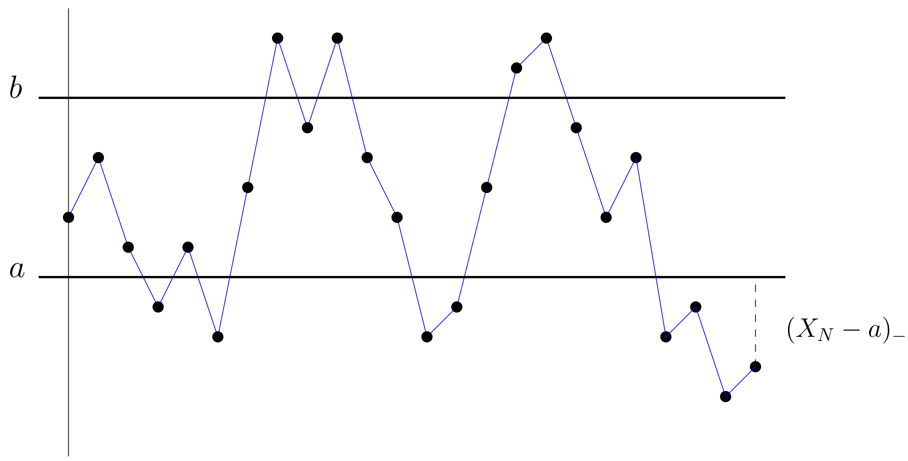


Figure 1: There are 2 upcrossings of $[a, b]$ shown.

Definition 4.41. Let (X_n) be a martingale, $N \in \mathbb{N}$, $a < b$. The *number* $U_N[a, b]$ *of upcrossings* of $[a, b]$ made by (X_n) up to time N is the largest integer k such that there are integers $0 \leq s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \leq N$ with $X_{s_i} < a$ and $X_{t_i} > b$ for all $1 \leq i \leq k$.

Recall the notation: $f_- = \max\{-f, 0\}$.

Lemma 4.42 (Doob's Upcrossing Lemma)**.**
Let (X_n) *be a martingale,* $N \in \mathbb{N}$*,* $a < b$*. Then*

$$(b - a)\mathbb{E}[U_N[a, b]] \leq \mathbb{E}[(X_N - a)_-].$$

Proof. We shall use Theorem 4.40 for a suitably chosen previsible strategy. For $n \leq s_1$ we set $C_n = 0$, for each $i \geq 1$ and $s_i + 1 \leq n \leq t_i$ set $C_n = 1$ and for each $i \geq 1$ and $t_i + 1 \leq n \leq s_{i+1}$ set $C_n = 0$. This corresponds to the gambling strategy of waiting until X goes below level a and then playing unit stakes until X gets above b and then stop playing (until X again goes below a). Then (C_n) is previsible and bounded and so $Y_n := (C \bullet X)_n$ is a martingale by Theorem 4.40. It follows that

$$0 = \mathbb{E}[Y_0] = \mathbb{E}[Y_N] = \sum_{i=1}^{N} C_i(X_i - X_{i-1}). \tag{7}$$

It is clear that for all k, $X_{t_k} - X_{s_k} \geq b - a$. We decompose Y_N in the following way:

$$\begin{aligned}
\sum_{i=1}^{N} C_i(X_i - X_{i-1}) &= \sum_{i=1}^{t_{U_N[a,b]}} C_i(X_i - X_{i-1}) + \sum_{i=t_{U_N[a,b]}+1}^{N} C_i(X_i - X_{i-1}) \\
&= \sum_{j=1}^{U_N[a,b]} \sum_{i=s_j+1}^{t_j} C_i(X_i - X_{i-1}) + \sum_{i=s_{U_N[a,b]+1}}^{N} C_i(X_i - X_{i-1}) \\
&= \sum_{j=1}^{U_N[a,b]} (X_{t_j} - X_{s_j}) + (X_N - X_{s_{U_N[a,b]}}) \mathbf{1}_{\{s_{U_N[a,b]}+1 \leq N\}} \\
&\geq (b-a) U_N[a,b] + (X_N - X_{s_{U_N[a,b]}}) \mathbf{1}_{\{s_{U_N[a,b]}+1 \leq N\}} \\
&\geq (b-a) U_N[a,b] - (X_N - a)_- .
\end{aligned}$$

Taking expectations and using (7) gives the result. $\square$

Theorem 4.43 (Martingale Convergence Theorem).
Let (X_n) be a martingale bounded in L_1 (i.e. $\exists c : \mathbb{E}|X_n| \leq c \forall n$). Then there exists an almost-surely finite random variable X on the same probability space such that $X_n \to X$ almost surely as $n \to \infty$.

Proof. Let $a < b \in \mathbb{Q}$. Denote by $U[a,b] = \lim_{N\to\infty} U_N[a,b]$. By Doob's Upcrossing Lemma, this random variable is almost-surely finite:

$$(b-a)\mathbb{E}[U[a,b]] \overset{\text{MON}}{=} \lim_{N\to\infty} (b-a)\mathbb{E}[U_N[a,b]] \leq \lim_{N\to\infty} \mathbb{E}[(X_N - a)_-] \leq \lim_{N\to\infty} \mathbb{E}[|X_N| + |a|] \leq c + |a|,$$

i.e. $\mathbb{E}[U[a,b]] < \infty$ and thus $U[a,b]$ is almost-surely finite. Notice now that

$$\left\{ \liminf_{n\to\infty} X_n \neq \limsup_{n\to\infty} X_n \right\} = \bigcup_{\substack{a<b \\ a,b\in\mathbb{Q}}} \left\{ \liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n \right\} \subset \bigcup_{\substack{a<b \\ a,b\in\mathbb{Q}}} \{U[a,b] = \infty\}.$$

But since each of the events $\{U[a,b] = \infty\}$ has probability zero, the probability of a countable union of such events also has probability zero. Thus we obtain

$$\liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n \quad \text{a.s.}$$

Denote $X = \liminf_{n\to\infty} X_n$. It remains to show that X is almost-surely finite. This follows easily by Fatou's Lemma:

$$\mathbb{E}|X| = \mathbb{E}\Big[\liminf_{n\to\infty} |X_n| \Big] \leq \liminf_{n\to\infty} \mathbb{E}|X_n| \leq c.$$

$\square$

Example 4.44.

(1) The simple symmetric random walk (S_n) is a martingale but does not converge. Indeed (S_n) is not bounded in L_1. On the other hand, the stopped random walk (stopped when reaching level b or $-a$) does converge to a random variable with two values a and b (depending on which level S_n hits first).

(2) If (X_n) is a non-negative martingale then $\mathbb{E}|X_n| = \mathbb{E}[X_n] = \mathbb{E}[X_0]$ so it is L_1-bounded and converges.

(3) Polya's Urn: the proportion of black balls is a non-negative martingale and so it converges. If we start with one black and one white ball, the limit is a uniform random variable. If we start with b black balls and w white balls, the limit has a beta distribution. (Exercise: show this! Hint: show that the probability that in n draws you pick m black balls and $n-m$ white balls in a certain order does not depend on that order).

(4) Galton-Watson process with offspring distribution N. Recall that Z_n denotes the number of individuals in generation n. What happens if $\mathbb{E}[N] < 1$? It turns out that the population becomes extinct in finite time. What about if $\mathbb{E}[N] = 1$? Then (Z_n) is a martingale. If $\mathbb{P}(N = 1) = 1$ then the process is deterministic – indeed $Z_n = 1$ for all n. But if $\mathbb{P}(N = 1) < 1$, then (Z_n) is a non-negative martingale and so by the MCT it converges to some random variable Z. Note that since $\mathbb{E}[N] = 1$, it must be that $\mathbb{P}(N = 0) > 0$:

$$\begin{aligned} 1 = \mathbb{E}[N] \geq \mathbb{E}[N\mathbf{1}_{\{N=1\}}] + \mathbb{E}[N\mathbf{1}_{\{N>1\}}] &= \mathbb{P}(N = 1) + \mathbb{E}[N\mathbf{1}_{\{N>1\}}] \\ &> \mathbb{P}(N = 1) + \mathbb{P}(N > 1) = \mathbb{P}(N \geq 1), \end{aligned}$$

i.e. $\mathbb{P}(N \geq 1) < 1$ and so $\mathbb{P}(N = 0) > 0$. It follows that if $k \neq 0$, then

$$\mathbb{P}(Z_1 = k | Z_0 = k) \leq \mathbb{P}(Z_1 \neq 0 | Z_0 = k) = 1 - \mathbb{P}(Z_1 = 0 | Z_0 = k) = 1 - \mathbb{P}(N = 0)^k < 1.$$

Now since Z_n is integer-valued,

$$\mathbb{P}(Z = k) = \mathbb{P}(Z_n \to k) = \mathbb{P}(\exists m : \forall n \geq m, Z_n = k) = \mathbb{P}\left(\bigcup_{m=1}^{\infty} \{\forall n \geq m, Z_n = k\}\right).$$

For $k \neq 0$, we have

$$\begin{aligned} \mathbb{P}(\forall n \geq m, Z_n = k) &\leq \lim_{M\to\infty} \mathbb{P}(Z_n = k \, \forall m \leq n \leq M) \\ &= \lim_{M\to\infty} \mathbb{P}(Z_n = k)\mathbb{P}(Z_1 = k | Z_0 = k)^{M-m} \\ &= 0. \end{aligned}$$

Since a countable union of measure 0 events has measure 0, it must be that $Z_n = 0$ eventually ($Z = 0$ almost surely), i.e. the process becomes extinct in finite time. However we have $\mathbb{E}[Z_n] = \mathbb{E}[Z_0] = 1$ but $\mathbb{E}[Z] = 0$ so $\mathbb{E}[Z_n] \nrightarrow \mathbb{E}[Z]$. The martingale (Z_n) converges almost surely to Z but not in L_1. When can we say that a martingale converges almost surely and in L_1?

4.6 Uniform integrability

Lemma 4.45. *Let X be an integrable random variable on $(\Omega, \Sigma, \mathbb{P})$. Then for each $\varepsilon > 0$, there exists $\delta > 0$ such that for $A \in \Sigma$, if $\mathbb{P}(A) < \delta$ then $\mathbb{E}[|X|\mathbf{1}_{\{A\}}] < \varepsilon$.*

Proof. Suppose the statement is false, so that there exists $\varepsilon_0 > 0$ and a sequence $(A_n) \in \Sigma$, such that $\mathbb{P}(A_n) < 2^{-n}$ and $\mathbb{E}[|X|\mathbf{1}_{\{A_n\}}] \geq \varepsilon_0$. Set $A = \limsup A_n$ ($= \{A_n \text{ i.o.}\}$) so that by Borel-Cantelli 1, $\mathbb{P}(A) = 0$. However, by reverse Fatou and using $\mathbf{1}_{\{\limsup A_n\}} = \limsup \mathbf{1}_{\{A_n\}}$,

$$0 = \mathbb{E}[|X|\mathbf{1}_{\{A\}}] \geq \limsup \mathbb{E}[|X|\mathbf{1}_{\{A_n\}}] \geq \varepsilon_0,$$

a contradiction. $\square$

Definition 4.46. A sequence of random variables (X_n) is *uniformly integrable (UI)* if for each $\varepsilon > 0$ there exists $c > 0$ such that

$$\mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|>c\}}] < \varepsilon \quad \forall\, n.$$

We want to exclude the case that the expectation is strongly influenced by increasingly rare but increasingly large values. For example, consider the sequence (X_n) such that

$$\mathbb{P}(X_n = n) = 1 - \mathbb{P}(X_n = 0) = 1/n.$$

Then $\mathbb{E}[X_n] = 1$ for all n and thus the sequence is integrable. But

$$\mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|>c\}}] = \mathbf{1}_{\{c<n\}} \to 1$$

as $n \to \infty$ for each c, thus (X_n) is not uniformly integrable.

Theorem 4.47. *Let (X_n) be a sequence of random variables.*

(1) If (X_n) is UI then it is bounded in L_1.

(2) If (X_n) is dominated by an integrable random variable then it is UI.

(3) If (X_n) is bounded in L_p (for $p > 1$) then it is UI.

Proof.

(1) Take $\varepsilon = 1$ and c such that $\mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|>c\}}] < 1$. Then

$$\mathbb{E}|X_n| = \mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|>c\}}] + \mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|\leq c\}}] \leq 1 + c.$$

(2) Suppose $|X_n(\omega)| \leq Y(\omega)$ for all n and ω with $\mathbb{E}[Y] < \infty$. Fix $\varepsilon > 0$ and let δ be such that for $A \in \Sigma$, if $\mathbb{P}(A) < \delta$ then $\mathbb{E}[Y\mathbf{1}_{\{A\}}] < \varepsilon$ (which exists by Lemma 4.45). By Markov's inequality, $\mathbb{P}(Y > c) < \mathbb{E}[Y]/c$ and thus we can find $c > 0$ so that $\mathbb{P}(Y > c) < \delta$. The result follows since

$$\mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|>c\}}] \leq \mathbb{E}[Y\mathbf{1}_{\{Y>c\}}] < \varepsilon.$$

(3) There exists $K > 0$ such that $\mathbb{E}[|X_n|^p] < K$ for all n. We have

$$\begin{aligned}\mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|>c\}}] = \mathbb{E}[|X_n|^{1-p}|X_n|^p\mathbf{1}_{\{|X_n|>c\}}] &\leq \mathbb{E}[c^{1-p}|X_n|^p\mathbf{1}_{\{|X_n|>c\}}] \\ &\leq c^{1-p}\mathbb{E}[|X_n|^p] \leq c^{1-p}K \to 0,\end{aligned}$$

as $c \to \infty$. $\square$

Definition 4.48. We say that a sequence of random variables (X_n) *converges in L_1* to X if $\mathbb{E}[|X_n - X|] \to 0$.

Theorem 4.49. *Let (X_n) be a sequence of integrable random variables and let X be an integrable random variable with $X_n \to X$ almost surely. Then $X_n \to X$ in L_1 iff (X_n) is UI.*

Proof omitted. $\square$

Theorem 4.50. *Let X be an integrable random variable on $(\Omega, \Sigma, \mathbb{P})$. Then the class*

$$\{\mathbb{E}[X|\mathcal{F}] : \mathcal{F} \textit{ is a sub-}\sigma\textit{-algebra of } \Sigma\}$$

is UI.

Proof. Fix $\varepsilon > 0$ and choose δ as in Lemma 4.45 (using the fact that X is integrable). Let $\lambda < \infty$ be such that $\mathbb{E}|X| \leq \lambda\delta$. For any sub-$\sigma$-algebra $\mathcal{F}$ we have, by conditional Jensen's inequality,

$$\mathbb{E}\big[|\mathbb{E}[X|\mathcal{F}]|\big] \leq \mathbb{E}\big[\mathbb{E}[|X||\mathcal{F}]\big] = \mathbb{E}|X|.$$

Set $Y = \mathbb{E}[X|\mathcal{F}]$. By Markov's inequality,

$$\mathbb{P}(|Y| \geq \lambda) \leq \frac{\mathbb{E}|Y|}{\lambda} \leq \delta.$$

Again by Jensen's inequality, we have

$$\begin{aligned}
\mathbb{E}[|Y|\mathbf{1}_{\{|Y|\geq\lambda\}}] &\leq \mathbb{E}\big[\mathbb{E}[|X||\mathcal{F}]\mathbf{1}_{\{|Y|\geq\lambda\}}\big] \\
\text{(defn of condx expectation)} \quad &= \mathbb{E}[|X|\mathbf{1}_{\{|Y|\geq\lambda\}}] \\
&\leq \varepsilon,
\end{aligned}$$

where the last step follows from Lemma 4.45 using the fact that $\{|Y| \geq \lambda\} \in \mathcal{F} \subseteq \Sigma$. $\square$

Definition 4.51. A sequence of random variables (X_n) is called a *UI martingale* if it is both a martingale and UI.

Theorem 4.52. *Let (X_n) be a martingale. The following are equivalent:*

(i) (X_n) is a UI martingale.

(ii) X_n converges a.s. and in L_1 to a limit X.

(iii) There exists an integrable random variable Z such that

$$X_n = \mathbb{E}[Z|\mathcal{F}_n] \quad \textit{a.s. } \forall\, n \geq 0.$$

Proof.
(i) $\Longrightarrow$ (ii):
Since (X_n) is UI by Theorem 4.47(1), (X_n) is bounded in L_1 and thus by MCT there exists an a.s. finite random variable X such that $X_n \to X$ a.s. Thus, since (X_n) is UI, by Theorem 4.49 $X_n \to X$ in L_1.

(ii) $\implies$ (iii):
Set $Z = X$. Then Z is integrable since

$$\mathbb{E}|Z| = \mathbb{E}|X| \leq \mathbb{E}|X_n - X| + \mathbb{E}|X_n| < \infty,$$

for n sufficiently large. We now show that $X_n = \mathbb{E}[Z|\mathcal{F}_n]$ a.s. For all $m \geq n$, by the martingale property for $F \in \mathcal{F}_n$,

$$\mathbb{E}[\mathbf{1}_{\{F\}} X_m] = \mathbb{E}\big[\mathbb{E}[\mathbf{1}_{\{F\}} X_m|\mathcal{F}_n]\big] = \mathbb{E}\big[\mathbf{1}_{\{F\}}\mathbb{E}[X_m|\mathcal{F}_n]\big] = \mathbb{E}[\mathbf{1}_{\{F\}} X_n]. \tag{8}$$

But we have by Jensen's inequality

$$\big|\mathbb{E}[\mathbf{1}_{\{F\}} X_m] - \mathbb{E}[\mathbf{1}_{\{F\}} X]\big| \leq \mathbb{E}[\mathbf{1}_{\{F\}}|X_m - X|] \leq \mathbb{E}|X_m - X| \to 0,$$

as $m \to \infty$. Thus by (8) $\mathbb{E}[\mathbf{1}_{\{F\}} X] = \mathbb{E}[\mathbf{1}_{\{F\}} X_n]$ and so X_n satisfies the three properties in the definition of conditional expectation (it is $\mathcal{F}_n$-measurable and integrable since it is a martingale).

(iii) $\implies$ (i):
Uniform integrability follows from Theorem 4.50. □

Definition 4.53. A *backwards martingale* is a martingale indexed by the negative integers: that is $(X_n)_{n\leq 0}$ is adapted to $(\mathcal{F}_n)_{n\leq 0}$ and satisfies $\mathbb{E}|X_0| < \infty$,

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n \quad \text{a.s. } \forall\, n \leq -1,$$

where $(\mathcal{F}_n)_{n\leq 0}$ is an increasing sequence of σ-algebras, $\cdots \mathcal{F}_{-2} \subseteq \mathcal{F}_{-1} \subseteq \mathcal{F}_0$.

Because the σ-algebras decrease as $n \to -\infty$, the convergence theorey for backward martingales is particularly simple.

Lemma 4.54. *Suppose $(X_n)_{n\leq 0}$ is a backwards martingale with respect to a filtration $(\mathcal{F}_n)_{n\leq 0}$. Then for all $n \leq 0$,*

$$\mathbb{E}[X_0|\mathcal{F}_n] = X_n, \ \text{a.s.}$$

Proof. By the Tower property, since $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all $n \leq 0$, we have, for each $n \leq 0$,

$$\mathbb{E}[X_0|\mathcal{F}_n] = \mathbb{E}\big[\mathbb{E}[X_0|\mathcal{F}_{-1}]\big|\mathcal{F}_n\big] = \mathbb{E}[X_{-1}|\mathcal{F}_n] = \cdots = \mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n. \qquad \square$$

Theorem 4.55. *Backwards martingales are UI.*

Proof. Since X_0 is integrable, we know by Theorem 4.50 that

$$\big\{\mathbb{E}[X_0|\mathcal{F}] : \mathcal{F} \text{ is a sub-}\sigma\text{-algebra of } \Sigma\big\}$$

is a UI class. The result follows by Lemma 4.54. □

Theorem 4.56 (Backwards MCT)**.**
Let (X_n) be a backwards martingale. Then X_n converges a.s. and in L_1 as $n \to -\infty$ to $X_{-\infty} := \mathbb{E}[X_0|\mathcal{F}_{-\infty}]$ where $\mathcal{F}_{-\infty} = \bigcap_{n\le 0} \mathcal{F}_n$.

Proof. We adapt Doob's Upcrossing Lemma to this setting. Choose $a < b \in \mathbb{Q}$ and for $N \le 0$, let $U_N[a,b]$ be the number of upcrossings of $[a,b]$ between time N and 0. For $0 \le k \le -N$, set $\mathcal{G}_k = \mathcal{F}_{N+k}$. Then $\mathcal{G}_k$ is an increasing filtration and $(X_{N+k}, 0 \le k \le -N)$ is a martingale adapted to $\mathcal{G}_k$ and $U_N[a,b]$ is the number of upcrossings of $[a,b]$ by $(X_{N+k})_k$ between times 0 and $-N$. Thus by Doob's Upcrossing Lemma,

$$(b-a)\mathbb{E}[U_N[a,b]] \le \mathbb{E}[(X_0 - a)_-]. \tag{9}$$

Letting $N \to -\infty$ we have $U_N[a,b]$ increases to the total number of upcrossings of X from a to b which by (9) is a.s. finite (using MON, see proof of MCT). We thus have (again by similar arguments to the proof of MCT) that

$$X_n \to X_{-\infty} \quad \text{a.s. as } n \to -\infty,$$

for some random variable $X_{-\infty}$. We claim that $X_{-\infty}$ is $\mathcal{F}_{-\infty}$-measurable. Indeed, since the limit of (X_n) does not depend on any finite number of the (X_n), for every $m \le 0$, $X_{-\infty}$ is the limit of $(X_{m+n})_{n\le 0}$. But each X_{m+n} is $\mathcal{F}_m$-measurable since $\mathcal{F}_{m+n} \subseteq \mathcal{F}_m$ for all $n \le 0$. As the limit of $\mathcal{F}_m$-measurable functions is also $\mathcal{F}_m$-measurable we deduce that $X_{-\infty}$ is $\mathcal{F}_m$-measurable. This is true for every m and hence $X_{-\infty}$ is $\mathcal{F}_{-\infty}$-measurable.

Since X_0 is bounded in L_1 there exists c such that $\mathbb{E}|X_0| < c$. Using $X_n = \mathbb{E}[X_0|\mathcal{F}_n]$ we have, by conditional Jensen's inequality,

$$\mathbb{E}|X_n| = \mathbb{E}\big[\big|\mathbb{E}[X_0|\mathcal{F}_n]\big|\big] \le \mathbb{E}\big[\mathbb{E}[|X_0|\big|\mathcal{F}_n]\big] = \mathbb{E}|X_0| < c,$$

i.e. X_n is bounded in L_1 for all $n \le 0$.

Also, by Fatou's Lemma,

$$\mathbb{E}|X_{-\infty}| = \mathbb{E}[\liminf_{n\to-\infty} |X_n|] \le \liminf_{n\to\infty} \mathbb{E}|X_n| < c,$$

and so $X_{-\infty}$ is bounded in L_1. Hence by the triangle inequality $\mathbb{E}|X_n - X_{-\infty}| < 2c$.

By conditional Jensen's inequality (and that $X_{-\infty}$ is $\mathcal{F}_n$-measurable), we obtain

$$|X_n - X_{-\infty}| = \big|\mathbb{E}[X_0 - X_{-\infty}|\mathcal{F}_n]\big| \le \mathbb{E}[|X_0 - X_{-\infty}|\big|\mathcal{F}_n],$$

but since $|X_n - X_{-\infty}|$ is integrable, the class of random variables

$$(\mathbb{E}[|X_0 - X_{-\infty}|\big|\mathcal{F}_n])_{n\le 0}$$

is UI. Hence also $(|X_n - X_{-\infty}|)_{n\le 0}$ is UI. Thus by Theorem 4.49 $|X_n - X_{-\infty}|$ converges to 0 in L_1 as $n \to -\infty$, i.e. X_n converges to $X_{-\infty}$ in L_1 as $n \to -\infty$.

We are left to show that $X_{-\infty} = \mathbb{E}[X_0|\mathcal{F}_{-\infty}]$ a.s. Since $X_{-\infty}$ is integrable and $\mathcal{F}_{-\infty}$-measurable, we thus need to show that for each $A \in \mathcal{F}_{-\infty}$,

$$\mathbb{E}[X_0 \mathbf{1}_{\{A\}}] = \mathbb{E}[X_{-\infty}\mathbf{1}_{\{A\}}].$$

Since $A \in \mathcal{F}_n$ and $X_n = \mathbb{E}[X_0|\mathcal{F}_n]$ for all $n \leq 0$,

$$\mathbb{E}[X_0 \mathbf{1}_{\{A\}}] = \mathbb{E}[X_n \mathbf{1}_{\{A\}}],$$

and so

$$\begin{aligned}
\mathbb{E}[X_0 \mathbf{1}_{\{A\}}] &= \lim_{n\to-\infty} \mathbb{E}[X_n \mathbf{1}_{\{A\}}] \\
&= \lim_{n\to-\infty} \mathbb{E}[(X_n - X_{-\infty} + X_{-\infty})\mathbf{1}_{\{A\}}] \\
&= \lim_{n\to-\infty} \mathbb{E}[(X_n - X_{-\infty})\mathbf{1}_{\{A\}}] + \mathbb{E}[X_{-\infty}\mathbf{1}_{\{A\}}] \\
&= \mathbb{E}[X_{-\infty}\mathbf{1}_{\{A\}}],
\end{aligned}$$

since $\mathbb{E}[(X_n - X_{-\infty})\mathbf{1}_{\{A\}}] \leq \mathbb{E}[X_n - X_{-\infty}] \leq \mathbb{E}|X_n - X_{-\infty}| \to 0$ as $n \to -\infty$ by the L_1 convergence result. $\square$

Theorem 4.57 (SLLN again)**.**
Let (X_n) be iid integrable random variables with mean μ. Then

$$\frac{X_1 + \cdots + X_n}{n} \to \mu,$$

almost surely and in L_1.

Proof. Set $S_n = X_1 + \cdots + X_n$ and $\mathcal{F}_n = \sigma(S_n, S_{n+1}, \ldots) = \sigma(S_n, X_{n+1}, \ldots)$. We claim that

$$(M_n)_{n\leq -1} := \left(\frac{S_{-n}}{-n}\right)_{n\leq -1}$$

is a $(\mathcal{G}_n)_{n\leq-1} := (\mathcal{F}_{-n})_{n\leq -1}$ backwards martingale. Indeed, setting $m = -n$ and using that X_m is independent of $X_{m+1}, X_{m+2}, \ldots$, we have

$$\mathbb{E}[M_{n+1}|\mathcal{G}_n] = \mathbb{E}\Big[\frac{S_{m-1}}{m-1}|\mathcal{F}_m\Big] = \mathbb{E}\Big[\frac{S_m - X_m}{m-1}|\mathcal{F}_m\Big] = \frac{S_m}{m-1} - \mathbb{E}\Big[\frac{X_m}{m-1}|S_m\Big].$$

By symmetry, $\mathbb{E}[X_i|S_m] = \mathbb{E}[X_1|S_m]$ for all $1 \leq i \leq m$. Clearly

$$\mathbb{E}[X_1|S_m] + \cdots + \mathbb{E}[X_m|S_m] = \mathbb{E}[S_m|S_m] = S_m,$$

and thus $\mathbb{E}[X_m|S_m] = \frac{S_m}{m}$ almost surely. Thus

$$\mathbb{E}\Big[\frac{S_{m-1}}{m-1}|\mathcal{F}_m\Big] = \frac{S_m}{m-1} - \frac{S_m}{m(m-1)} = \frac{S_m}{m} = M_n,$$

almost surely. Applying the backwards MCT we deduce that M_n converges as $n \to -\infty$ almost surely and in L_1 to a random variable $Y = \lim_{m\to\infty} \frac{S_m}{m}$. For all k,

$$Y = \lim_{m\to\infty} \frac{X_{k+1} + \cdots + X_{k+m}}{m},$$

and hence Y is $\mathcal{T}_k = \sigma(X_{k+1}, \ldots)$-measurable for all k and hence $\bigcap_k \mathcal{T}_k$-measurable. By Kolmogorov's 0-1 law, we conclude that there exists $c \in \mathbb{R}$ such that $\mathbb{P}(Y = c) = 1$. But

$$c = \mathbb{E}[Y] = \mathbb{E}\big[\lim_{m\to\infty} \frac{S_m}{m}\big] = \lim_{m\to\infty} \mathbb{E}\big[\frac{S_m}{m}\big] = \mathbb{E}[X_1] = \mu,$$

where the exchange of limit and expectation is by the L_1 convergence result. $\square$