STOR 635 Notes (S13)

Jimmy Jin
UNC-Chapel Hill
Last updated: 1/14/14

Contents

1 Measure theory and probability basics

1.1 Algebras and measure

Definition. (Probability measure)

Let $(S, \mathcal{S})$ be a measurable space. If, for a measure μ, $\mu(S) = 1$, then we say μ is a **probability measure**.

Definition. (Random variable)

A **measurable map** X is a function from one measurable space $(\Omega, \mathcal{F})$ to another measurable space $(S, \mathcal{S})$ such that

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathcal{S}$$

If the target space $(S, \mathcal{S}) = (\mathbb{R}, B(\mathbb{R}))$, then X is called a **random variable**. If the target space is $(\mathbb{R}^d, B(\mathbb{R}^d))$, then X is called a **random vector**.

Remark. We often write $\{X \in B\}$ as shorthand for $\{\omega : X(\omega) \in B\}$.

Theorem 1.1. If $\{\omega : X(\omega) \in A\} \in \mathcal{F}$ for all $A \in \mathcal{A}$ and $\sigma(\mathcal{A}) = \mathcal{S}$, then X is measurable.

Proof. Note that

$$\{\omega : X(\omega) \in \cup_{i=1}^{\infty} B_i\} = \cup_{i=1}^{\infty}\{\omega : X(\omega) \in B_i\}$$

$$\{\omega : X(\omega) \in B^c\} = \{\omega : X(\omega) \in B\}^c$$

Therefore the class of sets $\mathcal{B} = \{B : \{\omega : X(\omega) \in B\} \in \mathcal{F}\}$ is a σ-field. Then since $\mathcal{B} \supset \mathcal{A}$ and $\mathcal{A}$ generates $\mathcal{S}$, then $\mathcal{B} \supset \mathcal{S}$.

□

Definition. (π-class, λ-class)

Let S be some space and let $\mathcal{L}, \mathcal{A}$ be collections of subsets of S.

$\mathcal{A}$ is a **π-class** if it is closed under intersections.

$\mathcal{L}$ is a **λ-class** if

1. $S \in \mathcal{L}$
2. If $A, B \in \mathcal{L}$ with $A \supset B$, then $A \setminus B \in \mathcal{L}$
3. If $\{A_n\}$ is increasing with $A_i \in \mathcal{L}$, then $\lim_{n\to\infty} A_n \in \mathcal{L}$.

Theorem 1.2. (Dynkin's $\pi - \lambda$ theorem)

For $\mathcal{L}$ and $\mathcal{A}$ given above, $\mathcal{L} \supset \mathcal{A} \Rightarrow \mathcal{L} \supset \sigma(\mathcal{A})$.

Lemma 1.3. (Identification lemma)

Suppose p_1, p_2 are probability measures on $(S, \mathcal{S})$ and $p_1(A) = p_2(A)$ for all $A \in \mathcal{A}$. If $\mathcal{A}$ is a π-class and $\sigma(\mathcal{A}) = \mathcal{S}$, then $p_1(A) = p_2(A)$ for all $A \in \mathcal{S}$.

Proof. Define $\mathcal{L} = \{A \in \mathcal{S} \,|\, p_1(A) = p_2(A)\}$. Note that:

1. $\mathcal{L} \subset \mathcal{A}$ is a π-class by assumption.
2. If we show that $\mathcal{L}$ is a λ-class, then the result follows.

To show the three properties of λ-classes:

1. $S \in \mathcal{L}$ because p_1, p_2 are both probability measures.
2. To show that $A \supset B \in \mathcal{L} \Rightarrow A \setminus B \in \mathcal{L}$, use countable additivity.
3. To show that $\lim A_n \in \mathcal{L}$ for increasing $\{A_n\} \in \mathcal{L}$, use the continuity property of measures.

□

Remark. (Lebesgue measure)

There exists a unique σ-finite measure Λ on $(\mathbb{R}, B(\mathbb{R}))$ such that

$$\Lambda((a,b]) = b - a$$

Thus to get the Uniform probability measure, simply restrict Λ above to $[0,1]$. Uniqueness is shown by considering the π-class

$$\mathcal{A} = \{(a,b], a \leq b \in [0,1]\}$$

and applying the identification lemma.

Definition. (Induced measure)

Let $(S_1, \mathcal{S}_1, \mu_1)$ be a measure space and let $(S_2, \mathcal{S}_2)$ be another measurable space. Let $f : S_1 \to S_2$ be a measurable function. Then we can construct a measure μ_2 on $\mathcal{S}_2$ (called the **measure induced on $\mathcal{S}_2$ by** f) by:

$$\mu_2(B) = \mu_1\{s \in S_1 : f(s) \in B\}, \quad B \in \mathcal{S}_2$$

If μ_1 is a probability measure, then μ_2 is also a probability measure.

Definition. (Law/distribution, distribution function)

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

1. The **law** or **distribution** of a r.v. X is the induced measure on $(\mathbb{R}, B(\mathbb{R}))$ using the original probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ and the function X:

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

2. The **distribution function** of a r.v. X is a function $F : \mathbb{R} \to [0,1]$ defined by

$$F(x) = \mu((-\infty, x]) = \mathbb{P}(\omega \in \Omega : X(\omega) \in (-\infty, x])$$

that is, it is the law of x evaluated on the Borel set $(-\infty, x]$.

Remark. The distribution function has three basic properties:

1. $0 \leq F(x) \leq 1$ and F is non-decreasing.

2. F is right-continuous: $x_n \searrow x \Rightarrow F(x_n) \searrow F(x)$.
3. $\lim_{x\to\infty} F(x) = 1, \ \lim_{x\to-\infty} F(x) = 0$

Theorem 1.4. (Probability integral transform)

Suppose F is a function satisfying the three properties of distribution functions above. Then there exists a unique probability measure on $(\mathbb{R}, B(\mathbb{R}))$ whose distribution function is F.

Proof. We know that the measure space $([0,1], B([0,1]), \Lambda)$, where Λ is the uniform probability measure, exists.

Given F satisfying the three properties, define the map $G : [0,1] \to \mathbb{R}$ by

$$G(y) = \inf\{x \,|\, F(x) \geq y\}$$

From G, obtain the induced measure μ on $(\mathbb{R}, B(\mathbb{R}))$ using Λ:

$$\mu(B) = \Lambda\{s \in [0,1] \,|\, G(s) \in B\}$$

μ is guaranteed to be a probability measure since Λ is a probability measure. Then we see that

$$\begin{aligned} F_G(x) = \mu((-\infty, x)) &= \Lambda(\{y \,|\, G(y) \leq x\}) \\ &= \Lambda(\{y \,|\, y \leq F(x)\}) \\ &= F(x) \end{aligned}$$

□

1.2 Integration

STUPID NOTATION NOTE:

Let $(S, \mathcal{S}, \mu)$ be a measure space. Then the following are equivalent notations for the integral of f with respect to μ:

$$\int_S f(s)\,\mu(ds) = \int_S f(s)\,\mathrm{d}\mu(s) = \int_S f\,\mathrm{d}\mu$$

Now let $(S, \mathcal{S}, \mu)$ be a measure space where μ is σ-finite. Define $\mathcal{H}^+$ to be the space of all non-negative measurable functions $f : S \longrightarrow [0, \infty]$.

Theorem 1.5. There exists a unique map $I : \mathcal{H}^+ \to [0, \infty]$ such that:

1. $I(\mathbb{1}_A) = \mu(A)$
2. $f, g \in \mathcal{H}^+ \ \Rightarrow \ I(f+g) = I(f) + I(g)$
3. $f \in \mathcal{H}^+$ and $c \geq 0 \ \Rightarrow \ I(cf) = cI(f)$.

4. Let $\{f_n\}$ be a sequence of functions in $\mathcal{H}^+$ such that $f_n(x) \leq f_{n+1}(x)$ for each n. Also, let f be a function in $\mathcal{H}^+$ such that $f_n(x) \to f(x)$. Then $I(f_n) \to I(f)$.

   Proof. (Sketch)

   First, we define some notation:

$$I(f) = \int f \, \mathrm{d}\mu = \int_S f(s) \, \mathrm{d}s$$

$$I(f \mathbb{1}_A) = \int f \mathbb{1}_A \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu$$

   We proceed by the Nike method, also known as the Eminem method: to prove the first property, we "just do it" and define $I(\mathbb{1}_A) = \mu(A)$. Furthermore, for a simple function $f_n = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$ where the A_i's partition S, we define $I(f_n) = \sum_{i=1}^n c_i \mu(A_i)$.

   To show linearity, prove an intermediate result. Let $0 \leq f \leq L$ be a bounded function. Then for a sequence of simple functions $\{f_k\}$ such that $f_k \nearrow f$, then $I(f_k) \nearrow \lim I(f_k)$ also. Then define $I(f)$ by $\lim I(f_k)$ and show uniqueness.

   To show monotone convergence, note that each f_n has an increasing sequence of non-negative simple functions $\{f_{n,k}\}_{k=1}^{\infty}$ with $\lim_{k\to\infty} f_{n,k}(x) = f_n(x)$ for all $x \in X$.

   Also note that the sequence $\{g_k\}$ defined by $g_k(x) = \max_{n \leq k} f_{n,k}(x)$ is similarly simple and increasing. Establish thte fact that for $n \leq k$,

$$f_{n,k}(x) \leq g_k(x) \leq f_k(x)$$

   and take limits in the correct order to show that $\{g_k\}$ converges to f. Then take integrals of the above expression and then take limits again to show the main result (using the definition of an integral as the limit of integrals of simple functions).

Example. (Counting measure)

Let the σ-field $\mathcal{F} = 2^{\mathbb{N}}$ and let $f : \mathbb{N} \to [0, \infty)$. Also, let μ be the counting measure on $\mathcal{F}$, that is:

$$\mu(A) = |A| = \text{ the number of elements in A}, \quad \forall A \in \mathcal{F}$$

The question is: how do we calculate $\int f \, \mathrm{d}\mu$? First, define g_n by:

$$g_n(i) = \begin{cases} f(i), & i \leq n \\ 0, & otherwise \end{cases}$$

Note that $g_n \nearrow f$ and that $g_n(s) = \sum_{i=1}^n f(i) \mathbb{1}_{(\{i\})}(s)$, so g_n is simple. We know how to integrate simple functions:

$$\begin{aligned}\int g_n(s)\,\mathrm{d}\mu(s) &= \sum_{i=1}^{n} f(i)\mu(\{i\}) \\ &= \sum_{i=1}^{n} f(i)\end{aligned}$$

Then, invoking our theorem, we send $n \to \infty$ in the expression above to obtain the integral $\int f\,\mathrm{d}\mu$.

Lemma 1.6. Suppose $f \in \mathcal{H}^+$ and suppose there exists a set A, $\mu(A) = 0$ such that $f(s) = 0$ on $s \in A^c$ (f can be infinite on A). Then $\int f\,\mathrm{d}\mu = 0$.

Example. Let h_1, h_2 be such that $h_1 = h_2$ a.s. Then $\mu(\{s | h_1(s) \neq h_2(s)\}) = 0$.

Example. Let $\{f_n\}, f$ be such that $f_n \to f$ a.s. Then $\mu(\{s | f_n(s) \nrightarrow f(s)\}) = 0$.

Next, we discuss the notion of integrability of a function, which is simply a term for whether or not a given function has a finite integral. First, we formalize some notation.

1. For $x \in \mathbb{R}$: x is $x^+ = \max\{x, 0\}$ and $x^- = \max\{-x, 0\}$.
2. For a measurable function f: $f^+(s) = \max\{f(s), 0\}$ and $f^-(s) = \max\{-f(s), 0\}$
3. $f = f^+ - f^-$ and $|f| = f^+ + f^-$.

Definition. (Integrable function)

A measurable function f is **integrable** if:

$$\int f^+\,\mathrm{d}\mu < \infty \quad \text{and} \quad \int f^-\,\mathrm{d}\mu < \infty$$

And for a general (not necessarily non-negative) measurable function f, we define the integral as:

$$\int f\,\mathrm{d}\mu = \int f^+\,\mathrm{d}\mu - \int f^-\,\mathrm{d}\mu$$

Theorem 1.7. For integrable functions f and g on a measure space $(S, \mathcal{S}, \mu)$, the following properties hold (all integrals with respect to measure μ):

1. For $a, b \in \mathbb{R}$, $\int af + bg = a\int f + b\int g$
2. $f \geq 0$ a.e. $\Rightarrow \int f \geq 0$
3. $f = 0$ a.e. $\Rightarrow \int f = 0$
4. $f \geq g$ a.e. $\Rightarrow \int f \geq \int g$
5. $f = g$ a.e. $\Rightarrow \int f = \int g$
6. $|\int f| \leq \int |f|$

Definition. (Absolutely continuous)

Let μ and ν be measures on a σ-field $\mathcal{S}$. We say ν is **absolutely continuous** with respect to μ (written $\nu \ll \mu$) if

$$\mu(A) = 0 \quad \Rightarrow \quad \nu(A) = 0, \ A \in \mathcal{S}$$

The general setup of the theorem is the following: let μ be a σ-finite measure and let $f \geq 0$ be a measurable function. Define the set function:

$$\nu(A) = \int \mathbb{1}_A f \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu$$

It can be easily checked that ν is a measure and that $\mu(A) = 0$ implies that $\nu(A) = 0$ also, so that $\nu \ll \mu$. The Radon-Nikodym theorem gives us the following in addition:

Theorem 1.8. Let ν, μ be measures with $\nu \ll \mu$ and μ σ-finite. Then there exists an a.e. unique measurable function f such that

$$\forall A \in \mathcal{S}, \quad \nu(A) = \int_A f \, \mathrm{d}\mu$$

We call f the **Radon-Nikodym derivative** or **density**, written $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.

Lemma 1.9. Suppose μ is σ-finite, $f \geq 0$, $g \geq 0$ are integrable, and the following property holds:

$$\forall A \in \mathcal{S}, \quad \int_A f \, \mathrm{d}\mu = \int_A g \, \mathrm{d}\mu$$

Then $f = g$ a.e.

Theorem 1.10. (Change of measure)

Let X be a function $(\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{S})$ and let μ be the measure on $\mathcal{S}$ induced by X. Let $h : S \to \mathbb{R}$ be measurable and furthermore let $\mathbb{E}(|h(X)|) = \int |h(X)| \, \mathrm{d}\mathbb{P} < \infty$. Then:

$$\mathbb{E}(h(X)) = \int_\Omega h(X(\omega)) \, \mathrm{d}\mathbb{P} = \int_S h(s) \, \mathrm{d}\mu(s)$$

Example. Let B^∞ be our space of infinite coin tosses from before. Our underlying probability space is $(B^\infty, \sigma(\{A_\pi\}), \mathbb{P})$, where an element of B^∞ is given by

$$\boldsymbol{\omega} \in B^\infty = (\omega_1, \omega_2, \ldots)$$

Let $X : B^\infty \to \mathbb{R}$ be defined by $X(\boldsymbol{\omega}) = \omega_1$. We are interested in calculating the expectation of the function of our random variable: $h(x) = \sin(x)$.

By our theorem,

$$\mathbb{E}(\sin X) = \int_{B^\infty} \sin(X(\boldsymbol{\omega})) \, \mathrm{d}\mathbb{P} = \int_\mathbb{R} \sin(\omega_1) \, \mathrm{d}m$$

where m is Lebesgue measure on $B(\mathbb{R})$.

1.3 Inequalities

Lemma 1.11. If ψ is a convex function on an open subset $I \subset \mathbb{R}$, then for any $x_0 \in I$ there exists a support line l_0 such that

$$l_0(x) \leq \psi(x) \quad \forall x \in I \quad \text{and} \quad l_0(x_0) = \psi(x_0)$$

Proof. Convexity gives us two facts:

1. For any $h > 0$, applying the definition of convexity with $\alpha = 1/2$ gives:
$$\frac{\psi(x) - \psi(x-h)}{h} \leq \frac{\psi(x+h) - \psi(x)}{h}$$

2. For any $h_1 > h_2$, applying the definition of convexity with $\alpha = h_2/h_1$ gives:
$$\frac{\psi(x) - \psi(x-h_1)}{h_1} \leq \frac{\psi(x) - \psi(x-h_2)}{h_2}$$

By (2) the sequences are monotone so their limits as $h \to 0^+$ exist, so define:

$$\psi'_-(x) = \lim_{h \to 0^+} \frac{\psi(x) - \psi(x-h)}{h} \quad \text{and} \quad \psi'_+(x) = \lim_{h \to 0^+} \frac{\psi(x+h) - \psi(x)}{h}$$

By (1), $\psi'_-(x) \leq \psi'_+(x)$ for any x.

Then, for fixed z, choose some $a \in [\psi'_-(z), \psi'_+(z)]$ and define the line ℓ by:

$$l_z(x) = \psi(z) + a(x - z)$$

Clearly $\ell(z) = \psi(z)$ and by monotonicity of limits, it is easy to see that

$$l_z(x) \leq \psi(x) \quad \forall x \in I$$

□

Theorem 1.12. (Jensen's inequality)

Let g be convex on I, an open subset of $\mathbb{R}$. Let $x \in I$ with probability 1 and assume $X, g(X)$ have finite expectation. Then

$$g(EX) \leq E(g(X))$$

Proof. First note that $EX \in I$. So let $\ell(x) = ax + b$ be the support line for $g(\cdot)$ at EX. Then by the definition of support line we have

1. $\ell(EX) = g(EX)$
2. $\ell(x) \leq g(x) \quad \forall x \in I$

Taking expectations in (2) above, we obtain:

$$Eg(x) \;\geq\; E\ell(x) \;=\; E(ax+b) \;=\; aEX + b \;=\; \ell(EX)$$

Then noting (1) above, we are done.

□

Theorem 1.13. (Holder's inequality)

If $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, then for any random variables X, Y,

$$E(|XY|) \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}$$

Proof. (Sketch)

Fix $y > 0$. Consider $f(x) = \frac{x^p}{p} + \frac{y^q}{q} - xy$. By finding the minimum of $f(x)$, we can show that $f(x) \geq 0$ for all x. Then, choose

$$x = \frac{|X|}{(E|X|^p)^{1/p}}, y = \frac{|Y|}{(E|Y|^q)^{1/q}}$$

and take expections of both sides of the inequality.

Theorem 1.14. (Markov's inequality)

Suppose X is a real valued random variable, and $\psi : \mathbb{R} \mapsto \mathbb{R}$ a positive function. Fix any set $A \in \mathcal{B}(\mathbb{R})$ (e.g. $A = [a, \infty)$). Define $\mathbf{i}_A$ to be $\inf\{\psi(X) : X \in A\}$.

Then,

$$\mathbf{i}_A \mathbb{P}(X \in A) \leq E(\psi(X))$$

Proof. Note $\psi(X) \geq \mathbf{i}_A \cdot \mathbb{1}_{X \in A}$. Take expectations. □

Example. Suppose $X \geq 0, \psi(X) = X, A = [a, \infty)$. Then,

$$a\mathbb{P}(X \geq a) \leq EX$$

Similarly,

$$a^k \mathbb{P}(|X| \geq a) \leq E(|X|^k)$$

Example. Let $Y \geq 0$ and $\mathbb{E}Y^2 < \infty$. Then:

$$\mathbb{P}(Y > 0) \geq \frac{(\mathbb{E}Y)^2}{\mathbb{E}Y^2}$$

Proof. Apply Holder's inequality to $Y \cdot \mathbb{1}_{Y>0}$. □

1.4 Convergence notions

Theorem 1.15. (Relationships between notions of convergence)

1. $X_n \xrightarrow{a.s.} X \iff \mathbb{P}(\omega : |X_n(\omega) - X(\omega)| > \epsilon \ i.o.) = 0 \quad \forall \epsilon > 0$
2. $X_n \xrightarrow{\mathbb{P}} X \iff \mathbb{P}(\omega : |X_n(\omega) - X(\omega)| > \epsilon) \to 0$
3. $X_n \xrightarrow{L^p} X \iff \mathbb{E}|X_n - X|^p \to 0$
4. $X_n \xrightarrow{L^p} X \iff X_n \xrightarrow{\mathbb{P}} X$

Theorem 1.16. $X_n \xrightarrow{a.s.} X \;\Rightarrow\; X_n \xrightarrow{\mathbb{P}} X$

Proof. Since $X_n \xrightarrow{a.s.} X$, then by Egorov's theorem $X_n \xrightarrow{a.u.} X$. Then for any $\delta > 0$ there exists a set $B_\delta \in \mathcal{F}$ with $\mathbb{P}(B_\delta) < \delta$ and $X_n \xrightarrow{a.u.} X$ on $\Omega \setminus B_\delta$. So for any $\epsilon > 0$, there exists an $N(\epsilon, \delta) \in \mathbb{N}$ such that, for all $n \geq N$,

$$|X_n(\omega) - X(\omega)| < \epsilon \quad \forall \omega \in \Omega \setminus B$$

In other words,

$$\exists N(\epsilon, \delta) \text{ s.t. } \mathbb{P}(|X_n - X| \geq \epsilon) < \delta \;\text{ for } n > N$$

We can send $\delta \to \infty$ by letting $N \to \infty$, which gives us the definition of convergence in $\mathbb{P}$.

□

Theorem 1.17. $X_n \xrightarrow{L^p} X$ for some $p > 0 \;\Rightarrow\; X_n \xrightarrow{\mathbb{P}} X$.

Proof. By an application of Markov's inequality using $\phi(X) = |X|^k$,

$$a^k \cdot \mathbb{P}(|X| \geq a) \leq \mathbb{E}|X|^k$$

Take $k = p$. Apply to the r.v. $|X_n - X|$, with $A = [\epsilon, \infty)$ so that $a = \epsilon$:

$$\mathbb{P}(|X_n - X| \geq \epsilon) \leq \frac{\mathbb{E}|X_n - X|^p}{\epsilon^p}$$

□

Theorem 1.18. If $X_n \xrightarrow{a.s.} X$ and $X_n \xrightarrow{L^1} Y$, then $X = Y$ a.s.

Proof. Since $X_n \xrightarrow{L^1} Y$, then there is a subsequence $\{X_{n_k}\}$ that converges a.s. to Y (see section on Borel-Cantelli lemmas for proof).

□

Example. L^p convergence does not imply a.s. convergence, even if the sequence is bounded.

Example. a.s. convergence implies L^p convergence if the sequence is bounded.

Theorem 1.19. (L^2 weak law)

Suppose $X_1, X_2, \ldots, X_n$ are r.v.'s such that

1. $\mathbb{E}X_i = \mu \quad \forall i \geq 1$
2. $\mathbb{E}(X_i^2) \leq c \quad \forall i \geq 1$
3. $\mathbb{E}(X_i X_j) = \mathbb{E}X_i \cdot \mathbb{E}X_j, \quad \forall i, j$

Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{L^2} \mu$$

Proof. We want to show that $\mathbb{E}\{\left[\frac{1}{n}\sum X_i - \mu\right]^2\} \to 0, n \to \infty$.

Note that $\mathbb{E}(\frac{1}{n}\sum X_i) = \mu$. Then we have

$$\begin{aligned}\mathbb{E}\left\{\left[\frac{1}{n}\sum X_i - \mu\right]^2\right\} &= \mathrm{Var}\left(\frac{1}{n}\sum X_i\right) \\ &= \frac{1}{n^2}\mathrm{Var}\left(\sum X_i\right) \\ &= \frac{1}{n^2}\left\{\sum_i \mathrm{Var}(X_i) + \sum_{i\neq j}\left[\mathbb{E}(X_iX_j) - \mathbb{E}X_i\mathbb{E}X_j\right]\right\}\end{aligned}$$

By our covariance assumption, the rightmost term is zero. Also, note that

$$\begin{aligned}\mathrm{Var}(X_i) &= \mathbb{E}(X_i^2) - (\mathbb{E}X_i)^2 \\ &\leq c + \mu^2\end{aligned}$$

So our final expression reduces to

$$\mathbb{E}\left\{\left[\frac{1}{n}\sum X_i - \mu\right]^2\right\} \leq \frac{n(x+\mu^2)}{n^2} \to 0 \text{ as } n \to \infty$$

□

Theorem 1.20. (Bernstein's theorem)

Suppose $f : [0,1] \to \mathbb{R}$ is continuous. Then $\exists\{f_n\}_{n\geq 1},\ f_n = n^{\text{th}}$ degree polynomial such that

$$\sup_{x\in[0,1]} |f_n(x) - f(x)| \to 0 \text{ as } n \to \infty$$

Proof. Let $x_1, x_2, \ldots, x_n, \ldots$ be independent r.v.'s with $\mathbb{P}(x_i = 1) = x$ and $\mathbb{P}(x_i = 0) = 1 - x$, where $x \in [0,1]$ is fixed.

Consider the functions:

$$f_n(x) = \sum_{n=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right)$$

Note that, if $S_n =$ the sum of the first n x_i's, then

$$f_n(x) = \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right]$$

We want to show that $\lim\sup_{x\in[0,1]} |f_n(x) - f(x)| < \epsilon$ for all $\epsilon > 0$, so consider the quantity $|f_n(x) - f(x)|$:

$$\begin{aligned}
|f_n(x) - f(x)| &= \left| \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] - f(x) \right| \\
&= \left| \mathbb{E}\left[f\left(\frac{S_n}{n}\right) - f(x)\right] \right| \\
&\leq \mathbb{E}\left| f\left(\frac{S_n}{n}\right) - f(x) \right| \\
&= \mathbb{E}\left[\left| f\left(\frac{S_n}{n}\right) - f(x) \right| \cdot \mathbb{1}_{|\frac{S_n}{n} - x| < \delta} \right] \\
&\quad + \mathbb{E}\left[\left| f\left(\frac{S_n}{n}\right) - f(x) \right| \cdot \mathbb{1}_{|\frac{S_n}{n} - x| > \delta} \right]
\end{aligned}$$

for fixed $\delta > 0$. Now note two facts:

1. $\sup_{x \in [0,1]} f(x) = M < \infty$ due to continuity of f and compactness of $[0, 1]$.
2. f is continuous: for any $\epsilon > 0$, $\exists \delta > 0$ such that if $|\frac{s_n}{n} - x| < \delta$, then $|f(\frac{s_n}{n}) - f(x)| < \epsilon$

Therefore we have that

$$\mathbb{E}\left[\left| f\left(\frac{S_n}{n}\right) - f(x) \right| \cdot \mathbb{1}_{|\frac{S_n}{n} - x| < \delta} \right] < \epsilon$$

And that

$$\begin{aligned}
&\mathbb{E}\left[\left| f\left(\frac{S_n}{n}\right) - f(x) \right| \cdot \mathbb{1}_{|\frac{S_n}{n} - x| > \delta} \right] \\
&\leq \mathbb{E}\left[\left| f\left(\frac{S_n}{n}\right) \right| \cdot \mathbb{1}_{|\frac{S_n}{n} - x| > \delta} \right] + \mathbb{E}\left[\left| f\left(x\right) \right| \cdot \mathbb{1}_{|\frac{S_n}{n} - x| > \delta} \right] \\
&\qquad \text{(Triangle inequality)} \\
&\leq \mathbb{E}\left[M \cdot \mathbb{1}_{|\frac{S_n}{n} - x| > \delta} \right] + \mathbb{E}\left[M \cdot \mathbb{1}_{|\frac{S_n}{n} - x| > \delta} \right] \\
&= 2M \cdot \mathbb{P}\left(\left| \frac{S_n}{n} - x \right| > \delta \right) \\
&\leq 2M \cdot \frac{\mathrm{Var}(\frac{S_n}{n})}{\delta^2} \\
&\qquad \text{(Markov's inequality)} \\
&= 2M \cdot \frac{p(1-p)}{n\delta^2} \\
&\leq 2M \cdot \frac{1}{4n\delta^2} = \frac{M}{2n\delta^2}
\end{aligned}$$

And finally, putting the two pieces together, we have

$$|f_n(x) - f(x)| \leq \epsilon + \frac{M}{2n\delta^2}, \quad \forall x \in [0, 1]$$

Then since this is true for all n and the RHS sequence is monotone decreasing in n with limit ϵ, then

$$\lim_{n\to\infty} \sup_{x\in[0,1]} \Big| f_n(x) - f(x) \Big| \leq \epsilon$$

Note that the use of Markov's inequality in the above step is required to free the probability $\mathbb{P}\left(|\frac{S_n}{n} - x| \geq \delta\right)$ from dependence on x.

□

Theorem 1.21. (Skorohod's representation theorem)

If $X_n \xrightarrow{d} X$, then there are r.v.'s Y_n and Y defined on some joint probability triple with $F_{X_n} = F_{Y_n}$ and $F_X = F_Y$ for all n, such that $Y_n \xrightarrow{a.s.} Y$.

Proof. Let $([0,1], B([0,1]), m)$ be the uniform probability space where m is the uniform probability measure (Lebesgue measure). Define the random variables: Y and $Y_n : ([0,1], B([0,1])) \to \mathbb{R}$ by

$$Y(y) = \inf\{x \,|\, F_X(x) \geq y\}, \quad Y_n(y) = \inf\{x \,|\, F_{X_n}(x) \geq y\}$$

From Y and Y_n, obtain the induced measures μ_Y and μ_{Y_n} on $(\mathbb{R}, B(\mathbb{R}))$ by

$$\mu_Y(B) = m\{y \in [0,1] \,|\, Y(y) \in B\}$$

$$\mu_{Y_n}(B) = m\{y \in [0,1] \,|\, Y_n(y) \in B\}$$

μ_Y and μ_{Y_n} are probability measures since m is a probability measure. To show that $F_X = F_Y$ (the argument for $F_{X_n} = F_{Y_n}$ is exactly analogous), note:

$$\begin{aligned} F_Y(x) = \mu_Y((-\infty, x)) &= m(\{y \,|\, Y(y) \leq x\}) \\ &= m(\{y \,|\, y \leq F_X(x)\}) \\ &= F_X(x) \end{aligned}$$

The process to show that $Y_n \xrightarrow{a.s.} Y$ is the following. Since we used the probability integral transform to get corresponding random variables on the uniform probability space, we show a.s. convergence on the uniform probability space. So we fix $c \in [0,1]$ and show that $\liminf_n Y_n(c) \geq Y(c)$ and $\limsup_n Y_n(c) \leq Y(c)$

First note that if Y is not continuous at c, then since there are only countably many discontinuities, each discontinuity is a singleton with zero measure. Therefore we can define $Y(c) = Y_n(c) = 0$ to obtain a.s. convergence at that point without affecting the distribution function.

So assume that Y is continuous at c.

1. $\liminf_n Y_n(c) \geq Y(c)$

   Fix $\epsilon > 0$. Since Y is continuous at c, we can find an $\alpha \in \mathbb{R}$ such that $Y(c) - \epsilon < \alpha < Y(c)$. Thus $F_{Y_n}(\alpha) < c$ by definition of Y_n. Then since $F_{Y_n} \to F_Y$, for large enough n we can find $F_{Y_n}(\alpha) < c$ as well. By monotonicity, then $Y(c) - \epsilon < \alpha < Y_n(c)$ for large enough n. Thus $\liminf_n Y_n(c) \geq Y(c)$.

2. $\limsup_n Y_n(c) \le Y(c)$

   Fix $\epsilon > 0$. Let $d \in [0,1]$ be such that $c < d$ and Y is continuous at d. Then there exists α such that $Y(d) < \alpha < Y(d) + \epsilon$. By definition of Y, we have $c < d \le F_Y(Y(d)) \le F_Y(\alpha)$. Since $F_{Y_n} \to F_Y$, for large enough n we can find $c \le F_{Y_n}(\alpha)$ also. Thus by monotonicity, $Y_n(c) \le \alpha \le Y(d) + \epsilon$ for large enough n. Thus $\limsup_n Y_n(c) \le Y(c)$.

□

1.5 Independence

Definition. (σ-field generated by a family of r.v.'s)

The **σ-field generated by** $X_1, \ldots, X_n$ is the set

$$\sigma(X_1, \ldots, X_n) = \sigma\left\{\bigcup_{i=1}^{n} \sigma(X_i)\right\}$$

Or, equivalently,

$$\sigma(X_1, \ldots, X_n) = \Big\{\omega : \big(X_1(\omega), \ldots, X_n(\omega)\big) \in A,\ A \in B(\mathbb{R}^n)\Big\}$$

Remark. $\sigma(X_1, \ldots, X_n)$ is NOT the same as $\{\cap_{i=1}^n X_i^{-1}(B_i),\ B_i \in B(\mathbb{R})\}$. This is because $B(\mathbb{R}^n) = B(\mathbb{R}) \otimes \ldots \otimes B(\mathbb{R})$ where:

$$B(\mathbb{R}) \otimes \ldots \otimes B(\mathbb{R}) = \sigma\{(B_1, \ldots, B_n), B_i \in B(\mathbb{R})\}$$

by the definition of product spaces and their associated sigma-algebras, so $B(\mathbb{R}^n)$ is larger than the n-Cartesian product of $B(\mathbb{R})$'s.

Lemma 1.22. $\sigma(X_1, \ldots, X_n)$ is precisely the set

$$\sigma\left(\{\cap_{i=1}^n X_i^{-1}(B_i),\ B_i \in B(\mathbb{R})\}\right)$$

Proof. Define the following sets:

$$\mathcal{S}_1 = \cup_{i=1}^n \sigma(X_i)$$
$$\mathcal{S}_2 = \{\cap_{i=1}^n X_i^{-1}(B_i),\ B_i \in B(\mathbb{R})\}$$

We show that $\sigma(\mathcal{S}_1) = \sigma(\mathcal{S}_2)$.

1. $\mathcal{S}_2 \supset \mathcal{S}_1$:

   Let S be an arbitrary element of $\mathcal{S}_1$. Then $S \in \sigma(X_k)$ for some k. Then $S = X_k^{-1}(B)$ for some $B \in B(\mathbb{R})$. Therefore S can be written as $X_k^{-1}(B) \cap (\cap_{i \neq k} X_i^{-1}(\mathbb{R})) = X_k^{-1}(B) \cap \Omega$, so $S \in \mathcal{S}_2$.

2. $\mathcal{S}_2 \subset \sigma(\mathcal{S}_1)$:

   Note that $\mathcal{S}_2$ is precisely the set of all elements in $\mathcal{S}_1$ union all their countable intersections. Since $\sigma(\mathcal{S}_1)$ is a σ-field, then it includes all countable intersections of $\mathcal{S}_1$.

Therefore since $\mathcal{S}_2$ is contained inside $\sigma(\mathcal{S}_1)$, then $\sigma(\mathcal{S}_2) \subset \sigma(\mathcal{S}_1)$. But since $\mathcal{S}_2 \supset \mathcal{S}_1$, then $\sigma(\mathcal{S}_2) \supset \sigma(\mathcal{S}_1)$. Therefore $\sigma(\mathcal{S}_2) = \sigma(\mathcal{S}_1)$

□

Corollary 1.23. $\sigma(X_1, \ldots, X_n)$ is precisely the set: $\sigma\left(\{\cap_{i=1}^n A_i,\ A_i \in \sigma(X_i)\}\right)$.

Definition. (Independence of two objects)

1. **Events:** A, B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.
2. **σ-fields:** B_1, B_2 are independent σ-fields if (1) holds for all $A \in B_1, B \in B_2$.
3. **Random variables:** X, Y are independent random variables if (2) holds for their corresponding generated σ-fields $\sigma(X)$ and $\sigma(Y)$.

   Remark. If two random variables are independent, then they are defined on the same probability space.

Theorem 1.24. (Equivalent characterizations of independence)

Let $X : \Omega \to (S_1, \mathcal{S}_1)$ and $Y : \Omega \to (S_2, \mathcal{S}_2)$. TFAE:

1. X, Y are independent.
2. For all $A \in \mathcal{S}_1$ and $B \in \mathcal{S}_2$, $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B)$.
3. Suppose $\mathcal{A}_1 \subset \mathcal{S}_1, \mathcal{A}_2 \subset \mathcal{S}_2$ are π-classes which generate their corresponding σ-fields. Then for all $A \in \mathcal{A}_1, B \in \mathcal{A}_2$, $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B)$.
4. For all bounded measurable functions $h_1 : S_1 \to \mathbb{R}$ and $h_2 : S_2 \to \mathbb{R}$, $\mathbb{E}[h_1(X) \cdot h_2(Y)] = \mathbb{E}[h_1(X)] \cdot \mathbb{E}[h_2(Y)]$.

   Remark. Note that if X, Y are independent, then the last characterization (4) holds if

   $$\mathbb{E}[h_1^2(X)] < \infty \text{ and } \mathbb{E}[h_2^2(Y)] < \infty$$

   Proof. We prove (1) $\Rightarrow$ (2) $\Rightarrow$ (3) and then (2) $\iff$ (4).

   (1) implies (2):

   Consider any $A_1 \in \mathcal{S}_1,\ A_2 \in \mathcal{S}_2$. Note that

   $$\begin{aligned}\mathbb{P}(X \in A_1) &= \mathbb{P}(\{\omega : X(\omega) \in A_1\}) \\ \mathbb{P}(Y \in A_2) &= \mathbb{P}(\{\omega : Y(\omega) \in A_2\}) \\ \mathbb{P}(X \in A_1, Y \in A_2) &= \mathbb{P}(\{\omega : X(\omega) \in A_1\} \cap \{\omega : Y(\omega) \in A_2\})\end{aligned}$$

   By assumption, X and Y are independent, so that $\sigma(X)$ and $\sigma(Y)$ satisfy the condition that

   $$\forall A_1 \in \sigma(X), A_2 \in \sigma(Y),\ \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2)$$

   Changing notation and using the definition of $\sigma(X)$ and $\sigma(Y)$, this is equivalent to the statement that, for any $B_1 \in \mathcal{S}_1, B_2 \in \mathcal{S}_2$,

   $$\begin{aligned}&\mathbb{P}(\{\omega : X(\omega) \in B_1\} \cap \{\omega : Y(\omega) \in B_2\}) \\ &\quad = \mathbb{P}(\{\omega : X(\omega) \in B_1\}) \cdot \mathbb{P}(\{\omega : X(\omega) \in B_2\})\end{aligned}$$

Letting $B_1 = A_1$ and $B_2 = A_2$ shows that

$$\mathbb{P}(X \in A_1, Y \in A_2) = \mathbb{P}(X \in A_1) \cdot \mathbb{P}(Y \in A_2)$$

(2) implies (3): Trivial.

(3) implies (2): By the π-λ theorem.

Define:

$$L_1 = \{A \in \mathcal{S}_1 \,|\, \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)\ \forall B \in \mathcal{A}_2\}$$

Note that $L_1 \supset \mathcal{A}_1$. If we can show that L_1 is a Λ-class, then since $\mathcal{A}_1$ is a π-class, we will have shown that $L_1 \supset \sigma(\mathcal{A}_1)$, or that $L_1 \supset \mathcal{S}_1$. Since $L_1 \subset \mathcal{S}_1$, then we will have shown that $L_1 = \mathcal{S}_1$.

To show that L_1 is a Λ-class:

1. L_1 contains the whole space S:

   Note that $\mathbb{P}(X \in S) = \mathbb{P}(X^{-1}(S)) = \mathbb{P}(\Omega) = 1$. Therefore since X, Y are defined on the same space, then $\mathbb{P}(X \in S, Y \in B) = \mathbb{P}(\Omega \cap Y^{-1}(B)) = \mathbb{P}(Y \in B)$.

2. L_1 is closed under proper differences:

   Let $A_1, A_2 \in L_1$ such that $A_1 \supset A_2$. Note that $A_1 \setminus A_2 \in \mathcal{S}_1$ since $\mathcal{S}_1$ is a σ-field. Note:

$$\begin{aligned}
\mathbb{P}(X \in (A_1 \setminus A_2), Y \in B) &= \mathbb{P}([X^{-1}(A_1) \setminus X^{-1}(A_2)] \cap Y^{-1}(B)) \\
&= \mathbb{P}([X^{-1}(A_1) \cap Y^{-1}(B)] \setminus [X^{-1}(A_2) \cap Y^{-1}(B)]) \\
&= \mathbb{P}([X^{-1}(A_1) \cap Y^{-1}(B)]) - \mathbb{P}([X^{-1}(A_2) \cap Y^{-1}(B)]) \\
&= \mathbb{P}(A_1)\mathbb{P}(B) - \mathbb{P}(A_2)\mathbb{P}(B) \\
&= \mathbb{P}(A_1 \setminus A_2) \cdot \mathbb{P}(B)
\end{aligned}$$

   Where in the above steps we have used the facts that X, Y are measurable functions defined on the same space Ω, and that $\mathbb{P}$ is a measure on a σ-field.

3. L_1 is closed under monotone increasing limit

   Let $\{A_n\}_{n=1}^{\infty}$ be an increasing sequence in L_1. Since $\mathcal{S}_1$ is a σ-field, then $\cup_{n=1}^{\infty} A_n \in \mathcal{S}_1$ and also we can re-write $\cup_{n=1}^{\infty} A_n$ as a disjoint union $\cup_{n=1}^{\infty} G_n \in \mathcal{S}_1$ where

$$G_n = A_n \setminus \cup_{i=1}^{n} G_{n-1}, \quad G_1 = A_1$$

Then note:

$$\begin{aligned}
\mathbb{P}(X \in \cup^\infty A_n, Y \in B) &= \mathbb{P}(X \in \cup^\infty G_n, Y \in B) \\
&= \mathbb{P}(X^{-1}(\cup^\infty G_n) \cap Y^{-1}(B)) \\
&= \mathbb{P}(\cup^\infty [X^{-1}(G_n) \cap Y^{-1}(B)]) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(X^{-1}(G_n) \cap Y^{-1}(B)) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(X \in G_n) \cdot \mathbb{P}(Y \in B) \\
&= \mathbb{P}(X \in \cup^\infty G_n) \cdot \mathbb{P}(Y \in B) \\
&= \mathbb{P}(X \in \cup^\infty A_n) \cdot \mathbb{P}(Y \in B)
\end{aligned}$$

Where in going from the fourth to the fifth line we have used the fact that L_1 is closed under proper differences, and the sequence $\{G_n\}$ is a sequence of proper differences of sets in L_1.

Thus we have shown that L_1 is a Λ-class and that, therefore, $Ł_1 = \mathcal{S}_1$. Next, define:

$$L^2 = \{B \in \mathcal{S}_2 \,|\, \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)\ \forall A \in \mathcal{S}_1\}$$

Note that, by the result shown at the top, $L^2 \supset \mathcal{A}_2$. By a similar proof to the one above, it can be shown that L^2 is a Λ-class. Then since $\mathcal{A}_2$ is a π-class, $L^2 \supset \sigma(\mathcal{A}_2)$, or that $L^2 \supset \mathcal{S}_2$. Since $L^2 \supset \mathcal{S}_2$, then $L^2 = \mathcal{S}_2$, as desired.

□

(4) implies (2)

Fix $A \in \mathcal{S}_1$ and $B \in \mathcal{S}_2$, and let $h_1(X) = \mathbb{1}_A(X),\ h_2(Y) = \mathbb{1}_B(Y)$.

(2) implies (4)

Independence of X and Y implies that (4) holds at least for indicator functions $h_1 = \mathbb{1}_A$ and $h_2 = \mathbb{1}_B$. This can be extended to simple functions, and then to bounded measurable functions using approximation by simple functions.

□

Definition. (Pairwise independence)

Let $B_1, \ldots, B_2$ be σ-fields such that, for all $i \neq j$, B_i and B_j are independent. Then we say $\{B_i\}_{1 \leq i \leq n}$ is **pairwise independent**.

Definition. (Independence of countable collections)

Three characterizations:

1. Suppose $B_1, \ldots, B_n \subset \mathcal{F}$ are σ-fields. Then $\{B_i\}_{1 \leq i \leq n}$ is **independent** if, for any $A_1 \in B_1, \ldots, A_n \in B_n$,

$$\mathbb{P}(\cap^n A_i) = \prod^n \mathbb{P}(A_i)$$

2. Suppose $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are arbitrary subcollections of $\mathcal{F}$. We say $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are **independent** if, for any $I \subset \{1, 2, \ldots, n\}$ and $i \in I$, $A_i \in \mathcal{A}_i$,

$$\mathbb{P}(\cap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i)$$

3. Suppose $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are arbitrary subcollections of $\mathcal{F}$. Define $\overline{\mathcal{A}_i} = \{\Omega, A \in \mathcal{A}_i\}$. Then $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are **independent** if, for all $A_i \in \overline{\mathcal{A}_i}$,

$$\mathbb{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i)$$

Remark. What is exactly the difference between the second and third characterization? In the second, we do not force the intersection to be over all n in each I. In the third characterization, we always force the intersection to be over all n. But since we can always set $A_j = \Omega$ for any $j = 1, \ldots, n$ in the third characterization, then the two characterizations are equivalent.

Definition. (Independence of arbitrary collections)

Suppose $\{B_\alpha : \alpha \in J\}$ where $B_\alpha \subset \mathcal{F}$ and J is an arbitrary index set. We say that the B_α's are **independent** if any subcollection is independent.

That is, for $n \geq 2$ and $\alpha_1 \neq \alpha_2 \neq \ldots \neq \alpha_n \in J$, $\{B_{\alpha_i}\}_{1 \leq i \leq n}$ are independent.

Theorem 1.25. Suppose $\{B_i\}_{1 \leq i \leq n}$ are sub-σ-fields of $\mathcal{F}$ and $\mathcal{A}_i \subset B_i$ where $\mathcal{A}_i$ is a π-class that generates B_i.

If the $\{\mathcal{A}_i\}_{1 \leq i \leq n}$ are independent, then the $\{B_i\}_{1 \leq i \leq n}$ are independent.

Theorem 1.26. If $X_1, \ldots, X_n$ are independent, then $X_1, \ldots, X_j$ is independent of $X_{j+1}, \ldots, X_n$ for $1 < j < n$.

Proof. By a previous result,

$$\sigma(X_1, \ldots, X_n) = \sigma\left(\{\cap_{i=1}^n X_i^{-1}(B_i),\ B_i \in B(\mathbb{R})\}\right)$$

Note that $\{\cap_{i=1}^n X_i^{-1}(B_i),\ B_i \in B(\mathbb{R})\}$ is a π-class that generates $\sigma(X_1, \ldots, X_n)$ and apply the setup to $X_1, \ldots, X_j$ and $X_{j+1}, \ldots, X_n$.

□

2 Zero-one laws and stopping times

2.1 Borel-Cantelli lemmas

Definition. (lim sup and lim inf of a sequence)

Let A_n be a sequence of subsets of Ω. Then

1. $\limsup A_n = \lim_{m\to\infty} \cup_{n=m}^{\infty} A_n = \{\omega \text{ that are in infinitely many } A_n\}$
2. $\liminf A_n = \lim_{m\to\infty} \cap_{n=m}^{\infty} A_n = \{\omega \text{ that are in all but finitely many } A_n\}$

   Remark. Why are these sets given the name "lim sup" and "lim inf"? Because:

   $$\limsup_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\limsup A_n} \quad \text{and} \quad \liminf_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\liminf A_n}$$

Definition. (Infinitely often)

Let A_n be a sequence of subsets of Ω. Then $\limsup A_n = \{\omega : \omega \in A_n \ i.o.\}$, where i.o. stands for **infinitely often**.

Lemma 2.1. (Very weak lemmas)

1. $\mathbb{P}(A_n \ i.o.) \geq \limsup_{n\to\infty} \mathbb{P}(A_n)$
2. $\mathbb{P}(A_n \text{ occurs ultimately}) \leq \liminf_{n\to\infty} \mathbb{P}(A_n)$

   Proof. We prove the first. The proof for the second is analogous.

   Define $B_n = \cup_{n=m}^{\infty} A_n$. Note that this is a decreasing sequence with limit A_n *i.o.*.

   □

Lemma 2.2. (Borel-Cantelli 1)

Let A_n be a sequence of elements of $\mathcal{F}$. If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \ i.o.) = 0$.

Proof. By definition of lim sup, we have that $\limsup A_n \subset \cup_{m=n}^{\infty} A_m$. Therefore:

$$\begin{aligned} \mathbb{P}(\limsup A_n) &\leq \mathbb{P}(\cup_{m=n}^{\infty} A_m) \\ &\leq \sum_{m=n}^{\infty} \mathbb{P}(A_m) \quad \text{for any } n \end{aligned}$$

Since $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, its sequence of partial sums $\sum_{m=1}^{n} \mathbb{P}(A_m)$ converges. Therefore $\sum_{m=n}^{\infty} \mathbb{P}(A_m) \to 0$ as $n \to \infty$.

□

Lemma 2.3. (Borel-Cantelli 2)

Let A_n be an independent sequence of elements of $\mathcal{F}$. If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \ i.o.) = 1$.

Proof. Let $M < N < \infty$. By independence and the identity $1 - x \le e^{-x}$, we have:

$$\begin{aligned}
\mathbb{P}(\cap_{n=M}^{N} A_n^c) &= \prod_{n=M}^{N} (1 - \mathbb{P}(A_n)) \\
&\le \prod_{n=M}^{N} \exp(-\mathbb{P}(A_n)) \\
&= \exp\left(- \sum_{n=M}^{N} \mathbb{P}(A_n) \right) \to 0 \quad \text{as} \quad N \to \infty
\end{aligned}$$

Therefore since $\mathbb{P}(\cap_{n=M}^{\infty} A_n^c) = 0$, then $\mathbb{P}(\cup_{n=M}^{\infty} A_n) = 1$. Since this holds for arbitrary M and $\cup_{n=M}^{\infty} A_n$ decreases (in M) to $\limsup A_n$, then $\mathbb{P}(\limsup A_n) = 1$ also.

□

Remark. If $\{A_n\}$ is an independent sequence, then applying the contrapositive of Borel-Cantelli 2 shows that the **converse** of Borel-Cantelli 1 holds.

To see that it does not hold in general, let the measure space be given by $([0,1], B([0,1]), m)$ where m is Lebesgue measure. Consider the sequence of events $A_n = [0, 1/n]$ for $n = 1, 2, \ldots$.

How can we use the Borel-Cantelli lemmas to show convergence a.s.?

Lemma 2.4. $X_n \xrightarrow{a.s.} X$ if and only if for every $\epsilon > 0$,

1. $\lim_{m\to\infty} \mathbb{P}(|X_n - X| \le \epsilon \text{ for all } n \ge m) = 1$
2. $\lim_{m\to\infty} \mathbb{P}(|X_n - X| > \epsilon \text{ for some } n \ge m) = 0$

Proof. We show each direction separately.

$(\Rightarrow)$

Suppose that $X_n \xrightarrow{a.s.} X$. Fix $\epsilon > 0$. Define $\Omega_0 = \{\omega : X_n(\omega) \to X(\omega)\}$. By assumption, for every $\omega_0 \in \Omega_0$ there exists $N(\omega_0, \epsilon) \in \mathbb{N}$ such that

$$n \ge N(\omega_0, \epsilon) \;\Rightarrow\; |X_n(\omega_0) - X(\omega_0)| \le \epsilon$$

Thus for any such $\omega_0 \in \Omega_0$, there exists some corresponding $N(\omega_0, \epsilon)$ such that $\omega_0 \in \cap_{n=N(\omega_0,\epsilon)}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| \le \epsilon\}$. Therefore,

$$\Omega_0 = \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| \le \epsilon\}$$

The union is over a sequence of sets increasing to Ω_0 which has probability 1, so by continuity the union has probability 1. And finally note:

$$\cup_{N=1}^{\infty} \cap_{n=N}^{\infty} \{|X_n - X| \le \epsilon\} = \lim_{N\to\infty} \cap_{n=N}^{\infty} \{|X_n - X| \le \epsilon\}$$

$(\Leftarrow)$

Define $A(\epsilon) = \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} \{|X_n - X| \le \epsilon\}$.

By assumption, this set has probability 1 for all $\epsilon > 0$. For any $\omega \in A(\epsilon)$, there exists some $N(\omega, \epsilon)$ such that $n \ge N(\omega, \epsilon) \Rightarrow |X_n(\omega) - X(\omega)| \le \epsilon$.

To show $X_n \to X$ at some point ω_0 with $\mathbb{P}(\omega_0) > 0$, we only need to show that we can select an increasing sequence of such N's corresponding to $\epsilon \downarrow 0$, e.g. $\epsilon = 1/n$:

$$N(\omega_0, 1),\ N(\omega_0, 1/2),\ N(\omega_0, 1/3), \ldots$$

We can do this only for $\omega_0 \in \cap_{n=1}^{\infty} A(1/n)$. But since $\mathbb{P}(A(\epsilon)) = 1$ for all $\epsilon > 0$, then $\cap_{n=1}^{\infty} A(1/n)$ must also have probability 1 and we are done.

□

Lemma 2.5. $X_n \xrightarrow{a.s.} X$ if and only if for all $\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon\ i.o.) = 0$.

Proof. Note that:

$$\{|X_n - X| > \epsilon\ i.o.\} = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{|X_n - X| > \epsilon\}$$

And use the second iff condition in the previous lemma.

□

Theorem 2.6. Suppose $X_1, \ldots, X_n \stackrel{iid}{\sim} \exp(1)$. So we have $\mathbb{P}(X_i \le x_i) = 1 - e^{-x}$ and $f(x) = e^{-x}$ for $x \ge 0$. Then:

1. $\frac{X_n}{\log n} \xrightarrow{\mathbb{P}} 0$
2. $\limsup_{n\to\infty} \frac{X_n}{\log n} \xrightarrow{a.s.} 1$
3. Let $M_n = \max\{X_1, \ldots, X_n\}$. Then $\frac{M_n}{\log n} \xrightarrow{a.s.} 1$

Proof. In order.

Proof of (1)

Note simply that:

$$\begin{aligned}\mathbb{P}\left(\left|\frac{X_n}{\log n} - 0\right| > \epsilon\right) &= \mathbb{P}(X_n > \epsilon \log n) \\ &= \exp(-\epsilon \log n) \\ &= \frac{1}{n\epsilon}\end{aligned}$$

Proof of (2)

We show $\limsup_{n\to\infty} \frac{X_n}{\log n} \le 1$ and $\limsup_{n\to\infty} \frac{X_n}{\log n} \ge 1$. This is equivalent to showing that, for any $\epsilon > 0$,

$$\mathbb{P}\left(\frac{X_n}{\log n} \ge 1 + \epsilon\ i.o.\right) = 0 \quad \text{and} \quad \mathbb{P}\left(\frac{X_n}{\log n} \ge 1 - \epsilon\ i.o.\right) = 1$$

To show the first, observe that:

$$\mathbb{P}(X_n \geq (1+\epsilon)\log n) = \frac{1}{n^{1+\epsilon}}$$

which is summable, so the result follows from Borel-Cantelli 1. To show the second, note that

$$\mathbb{P}(X_n \geq (1+\epsilon)\log n) = \frac{1}{n^{1-\epsilon}}$$

is not summable, so the result follows from Borel-Cantelli 2.

Proof of (3)

We show the result by showing that $\liminf \frac{M_n}{\log n} \geq 1$ and $\limsup \frac{M_n}{\log n} \leq 1$.

For the first, we show $\mathbb{P}\left(\frac{M_n}{\log n} \leq 1-\epsilon \; i.o.\right) = 0$:

$$\begin{aligned}
\mathbb{P}\left(\frac{M_n}{\log n} \leq 1-\epsilon\right) &= \mathbb{P}\left(\frac{X_1}{\log n} \leq 1-\epsilon\right)^n \\
&= \left(1 - e^{-(1-\epsilon)\log n}\right)^n \\
&= \left(1 - \frac{1}{n^{1-\epsilon}}\right)^n \\
&\leq \left(e^{-\frac{1}{n^{1-\epsilon}}}\right) = e^{-n^{\epsilon}}
\end{aligned}$$

Which is summable, so the result follows by Borel-Cantelli 1.

The second follows immediately from a deterministic fact:

Let $\{x_n\}_{n\geq 1}$, $x_n \geq 1$ be a sequence of real numbers. Suppose $\{b_n\}_{n\geq 1}$ is another sequence which increases to ∞. Then

$$\limsup_{n\to\infty} \frac{x_n}{b_n} = \alpha \quad \Rightarrow \quad \limsup_{n\to\infty} \frac{m_n}{b_n} = \alpha$$

where $m_n = \max\{x_1, x_2, \ldots\}$.

To prove this, fix $j \geq 1$ and note that, since $b_n \uparrow \infty$,

$$\begin{aligned}
\limsup_{n\to\infty}\left\{\frac{\max\{x_1,\ldots,x_n\}}{b_n}\right\} &= \limsup_{n\to\infty}\left\{\frac{\max\{x_j, x_{j+1},\ldots,x_n\}}{b_n}\right\} \\
&\leq \limsup_{n\to\infty}\left\{\max\left[\frac{x_k}{b_k} : j \leq k \leq n\right]\right\} \\
&\leq \sup_{k\geq j}\left\{\frac{x_k}{b_k}\right\}
\end{aligned}$$

Letting $j \to \infty$, we have

$$\begin{aligned}
\limsup_{n\to\infty}\left\{\frac{\max\{x_1,\ldots,x_n\}}{b_n}\right\} &\leq \lim_{j\to\infty}\sup_{k\geq j}\left\{\frac{x_k}{b_k}\right\} \\
&= \limsup_{n\to\infty}\left\{\frac{x_n}{b_n}\right\}
\end{aligned}$$

Observing that $\limsup \frac{m_n}{b_n} \geq \limsup \frac{x_n}{b_n}$ completes the proof.

□

Remark. The point of this whole example is to show that convergence in probability does not necessarily imply convergence a.s.

Lemma 2.7. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of measurable functions (random variables). If, for any $\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon \; i.o.) = 0$, then $X_n \xrightarrow{\mathbb{P}} X$.

Proof. Note that $\mathbb{P}(|X_n - X| > \epsilon \; i.o.) = \mathbb{P}(\limsup |X_n - X| > \epsilon)$.

Also note $\{\lim |X_n - X| > \epsilon\} \subset \{\limsup |X_n - X| > \epsilon\}$ for any n, so

$$\lim \mathbb{P}(|X_n - X| > \epsilon) \;\leq\; \mathbb{P}(\limsup |X_n - X| > \epsilon)$$

□

Theorem 2.8. $X_n \xrightarrow{\mathbb{P}} X$ if and only if for every subsequence $\{X_{n_j}\}_{j\geq 1}$ there is a further subsequence $\{X_{n_{j_k}}\}_{k\geq 1}$ that converges a.s. to X.

To prove this theorem, we first prove a deterministic lemma:

Lemma 2.9. Let y_n be a sequence of elements of a topological space. If every subsequence y_{n_m} has a further subsequence $y_{n_{m_k}}$ that converges to y, then y_n converges to y also.

Proof. Assume, on the contrary, that $y_n \nrightarrow y$. Then there exists

1. an open set I containing the limit y and
2. a subsequence y_{n_m} such that $y_{n_m} \notin G$ for all m.

but then obviously no subsequence of y_{n_m} can converge to y, a contradiction.

□

We now prove the main result.

Proof. The proof is a simple application of the first Borel-Cantelli lemma.

$(\Rightarrow)$

Assume $X_n \xrightarrow{\mathbb{P}} X$. Fix a subsequence $\{X_{n_j}\}_{j\geq 1}$. Clearly $X_{n_j} \xrightarrow{\mathbb{P}} X$ as well, so that for any $\epsilon > 0$,

$$\mathbb{P}(|X_{n_j} - X| \geq \epsilon) \to 0 \quad \text{as } j \to \infty$$

Let $\{\epsilon_k\}_{k\geq 1}$ be a positive sequence which $\downarrow 0$. Since the above holds for any $\epsilon > 0$, then we can sequentially find a subsequence $n_{j_k} \geq n_{j_{k-1}} \geq \ldots$ such that

$$\mathbb{P}(|X_{n_{j_k}} - X| \geq \epsilon_k) \;\leq\; \frac{1}{2^k}$$

Therefore we have

$$\sum^{\infty} \mathbb{P}(|X_{n_{j_k}} - X| \geq \epsilon_k) \ \leq \ \sum^{\infty} \frac{1}{2^k} < \infty$$

Thus by the first Borel-Cantelli lemma, $\mathbb{P}(|X_{n_{j_k}} - X| > \epsilon_k \ i.o.) = 0$ and therefore $X_{n_{k_j}} \xrightarrow{a.s.} X$.

($\Leftarrow$)

Assume that, for every subsequence $\{X_{n_j}\}_{j\geq 1}$ there exists a further subsequence $\{X_{n_{j_k}}\}_{k\geq 1}$ such that $X_{n_{j_k}} \xrightarrow{a.s.} X$ as $k \to \infty$.

Fix $\epsilon > 0$. Define $y_n = \mathbb{P}(|X_n - X| > \epsilon)$. By assumption and the fact that convergence a.s. implies convergence in probability, there exists a further subsequence (of some intermediate subsequence) $\{X_{n_{j_k}}\}_{k\geq 1}$ such that

$$y_{n_{j_k}} = \mathbb{P}(|X_{n_{j_k}} - X| > \epsilon) \to 0 \quad \text{as } k \to \infty$$

Since every further subsequence (of some intermediate subsequence) is convergent to 0, then by our deterministic lemma, y_n itself is convergent to 0. Thus $X_n \xrightarrow{\mathbb{P}} X$.

□

Theorem 2.10. (Dominated convergence theorem)

If $X_n \xrightarrow{\mathbb{P}} X$ and $|X_n| \leq Y$ with $E(Y) < \infty$, then

1. $\mathbb{E}X_i \to \mathbb{E}X$
2. $\mathbb{E}(|X_n - X|) \to 0$

Proof. Suppose $X_n \xrightarrow{\mathbb{P}} X$. Given any subsequence $\{n_j\}$, there exists a further subsequence $\{n_{j_k}\}$ such that $X_{n_{j_k}} \xrightarrow{a.s.} X$. By DCT, then we have

$$\mathbb{E}X_{n_{j_k}} \to \mathbb{E}X$$

Define the sequence of real numbers $\{y_n\}$ by $y_n = \mathbb{E}X_n$. Since, given any subsequence $\{y_{n_j}\}_{j\geq 1}$, there exists a further subsequence $\{y_{n_{j_k}}\}_{k\geq 1}$ that converges to $\mathbb{E}X$, then by our previous deterministic lemma we have

$$y_n = \mathbb{E}X_n \to \mathbb{E}X$$

□

Remark. Note that this result only requires convergence in **probability**, not convergence a.s. as is normally required in the statement of DCT.

Theorem 2.11. (Miscellaneous note)

Let $\{X_n\}_{n\geq 0}$ be a sequence of random variables. Then:

$$\bigcup_{L=1}^{\infty} \{X_n > -L \ \forall n\} = \{\inf X_n > -\infty\}$$

Proof. If some ω is in the LHS, then it is in $\{X_n > -L_0 \ \forall n\}$ for some $L_0 \in \mathbb{N}$. Thus the sequence $\{X_n(\omega)\}$ is bounded below at that value and so is in the RHS.

The exact opposite argument works to showing that some ω in the RHS is also in the LHS.

□

2.2 Filtrations and stopping times

Consider an infinite sequence of random variables $X_1, X_2, \dots$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The σ-field generated by the first n random variables is:

$$\mathcal{F}_n = \{B = \{\omega : X_1(\omega), \dots, X_n(\omega) \in A\},\ A \in B(\mathbb{R}^n)\}$$

This is a σ-field since the X_i's are measurable and all intersections and unions between elements in each induced σ field are included. Since sets of those form generate the Borel σ-field on $\mathbb{R}$ and all X_i's are measurable, we can also write:

$$\mathcal{F}_n = \sigma\{\{\omega : X_1(\omega) \leq x_1, \dots, X_n(\omega) \leq x_n\},\ x_1, \dots, x_n \in R\}$$

From the first formulation, it is easy to see that if we take $A = A_1 \times \mathbb{R}$ and $A_1 \in B(\mathbb{R}^{n-1})$, then

$$\mathcal{F}_{n-1} \subset \mathcal{F}_n$$

Definition. (Filtration)

A collection of sub-σ-fields $\{\mathcal{F}_n\}_{n\geq 1}$ is a **filtration** if it satisfies

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \dots$$

Remark. In general, we want to think of these sub-σ-fields as partial information. A σ-algebra defines the set of events that can be measured. Therefore a filtration is often used to represent the change in the set of events that can be measured, through gain or loss of information.

Another interpretation of a filtration involves times. Under a filtration, we can view the set of events in $\mathcal{F}_t$ as the set of "questions that can be answered at time t," which naturally carries the ascending/descending structure of the filtration.

Lemma 2.12. Fix $n \geq 1$. Under the usual filtration, a random variable Y is $\mathcal{F}_n$-measurable if and only if:

$$Y = g(X_1, \dots, X_n)$$

where $g : \mathbb{R}^n \to \mathbb{R}$ is a deterministic measurable function.

Proof. (Sketch) Consider an arbitrary $\mathcal{F}_n$-measurable indicator function $\mathbb{1}_B(\omega)$. By definition, it looks like:

$$\mathbb{1}_B(\omega) = \mathbb{1}_A(X_1(\omega), \dots, X_n(\omega)) = g(X_1, \dots, X_n),\quad A \in B(\mathbb{R}^n)$$

Therefore simple functions also have this general form, and thus general measurable functions also have this form.

□

Definition. (Stopping time)

A random variable $\tau : \Omega \to \{1, 2, \ldots\} \cup \{\infty\}$ is a **stopping time with respect to the filtration** $\{\mathcal{F}_n\}_{n\geq 1}$ if, for any n,

$$\{\omega : \tau(\omega) = n\} \in \mathcal{F}_n \ \ \forall n$$

Corollary 2.13. The above condition is equivalent to the condition:

$$\{\omega : \tau(\omega) \leq n\} \in \mathcal{F}_n \ \ \forall n$$

Proof. Fix $n \geq 1$.

Assume the first definition. We have that: $\{\tau \leq n\} = \cup_{i=1}^{n} \{\tau = i\}$. The RHS is a union of events which are all individually $\in \mathcal{F}_n$ since the probability space is filtered. Thus the entire RHS $\in \mathcal{F}_n$.

Assume the second definition. Then the event $\{\tau \leq n - 1\} \in \mathcal{F}_{n-1}$ and thus is also $\in \mathcal{F}_n$ since the probability space is filtered. Since $\mathcal{F}_n$ is closed under set difference, then

$$\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n - 1\} \in \mathcal{F}_n$$

□

Corollary 2.14. $\{T = \infty\} \in \mathcal{F}_\infty$ where $\mathcal{F}_\infty$ is defined by $\mathcal{F}_\infty = \sigma(\cup^\infty \mathcal{F}_n)$.

Proof. Trivial.

Example. (Hitting time)

Let a filtration be given by $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. The hitting time of $(0, \infty)$ defined below is a stopping time:

$$\tau = \inf\{n \geq 1 : X_n > 0\}$$

To see this, fix $n \geq 1$ and note that $\{\tau = n\} = \cap_{j=1}^{n-1} \{X_j \leq 0\} \cap \{X_n > 0\}$.

Example. The following is not a stopping time:

$$\tau = \sup\{n \geq 1 : X_n > 0\}$$

To see this, fix $n \geq 1$ and note that $\{\tau = n\} = \{X_n > 0\} \cap \left\{\cap_{j=n+1}^{\infty} \{X_j \leq 0\}\right\}$.

Example. (Warning)

Even the most innocuous finite a.s. stopping time can have a surprisingly infinite expected value. Consider a symmetric simple random walk $\{X_n\}$ with $S_n = \sum^n X_i$.

Let $T = \inf\{n \geq 0 : S_n = 1\}$, i.e. the hitting time to 1. We show that $\mathbb{E}T = \infty$.

Since X_n is symmetric, with probability $\frac{1}{2}$ we have $T = 1$. Also with probability $\frac{1}{2}$, we have $T = 1 + T' + T''$ where T', T'' are iid copies of T. To see why this is, if $S_1 = -1$, then to get back up to $+1$ we need to first return to 0 which takes T' steps, and then return to $+1$ which takes T'' steps. Since stopping times are non-negative, we can take expectations and use linearity regardless of finiteness:

$$\mathbb{E}T = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (1 + \mathbb{E}T' + \mathbb{E}T'')$$

Using $\mathbb{E}T = \mathbb{E}T' = \mathbb{E}T''$, we obtain that $\mathbb{E}T$ must satisfy $\mathbb{E}T = 1 + \mathbb{E}T$, which has a unique solution $= \infty$.

Theorem 2.15. (Wald's lemma)

Let $X_1, X_2, \ldots$ be iid with $\mathbb{E}(|X_i|) < \infty$ and $\mathbb{E}(X_i) = \mu$. Let τ be a stopping time with respect to the filtration $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ such that $\mathbb{E}(\tau) < \infty$. Define the random sum: $S_\tau = \sum_{i=1}^{\tau} X_i$. Then:

$$\mathbb{E}S_\tau = \mathbb{E}(\tau) \cdot \mathbb{E}(X_1)$$

Proof. Observe that:

$$S_\tau = \sum_{i=1}^{\infty} X_i \cdot \mathbb{1}_{i \leq \tau}$$

Consider $\{i \leq \tau\}$ for fixed i. Because $\{i \leq \tau\} = \{\tau \leq i-1\}^c$ and τ is a stopping time,

$$\{\tau \leq i-1\} \in \mathcal{F}_{i-1} \quad \Rightarrow \quad \{i \leq \tau\} \in \mathcal{F}_{i-1}$$

The X_i's are independent, so X_i is independent of $\mathcal{F}_{i-1}$. In particular, X_i and $\mathbb{1}_{i \leq \tau}$ are independent. Therefore:

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{\infty} |X_i| \cdot \mathbb{1}_{i \leq \tau}\right] &= \sum_{i=1}^{\infty} \mathbb{E}\left[|X_i| \cdot \mathbb{1}_{i \leq \tau}\right] \quad \text{(by MCT)} \\
&= \sum_{i=1}^{\infty} \mathbb{E}|X_i| \cdot \mathbb{P}(i \leq \tau) \\
&= \mathbb{E}|X_1| \cdot \sum_{i=1}^{\infty} \mathbb{P}(i \leq \tau) \\
&= \mathbb{E}|X_1| \cdot \mathbb{E}(\tau) < \infty
\end{aligned}$$

Noting that $\sum_{i=1}^{\infty} X_i \cdot \mathbb{1}_{i \leq \tau}$ is dominated by $\sum_{i=1}^{\infty} |X_i| \cdot \mathbb{1}_{i \leq \tau}$, we can now show the full result by dominated convergence:

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{\infty} X_i \cdot \mathbb{1}_{i \leq \tau}\right] &= \sum_{i=1}^{\infty} \mathbb{E}\left[X_i \cdot \mathbb{1}_{i \leq \tau}\right] \quad \text{(by DCT)} \\
&= \sum_{i=1}^{\infty} \mathbb{E}X_i \cdot \mathbb{P}(i \leq \tau) \\
&= \mathbb{E}X_1 \cdot \mathbb{E}(\tau) < \infty
\end{aligned}$$

□

Example. (Usage of Wald's lemma)

Suppose X_i are iid with

$$X_i = \begin{cases} +1, & p = 1/2 \\ -1, & p = 1/2 \end{cases}$$

Let $\tau = \inf\{n : S_n = 30 \text{ or } -10\}$. Is this a stopping time?

It is easily checked that $\mathbb{E}(\tau) < \infty$. Now note that $\mathbb{E}S_\tau = 0$ since $\mathbb{E}X_i = 0$ (to show this more rigorously, use the stopped process $S_{\tau \wedge n}$ and apply DCT). Therefore:

$$0 = 30 \cdot \mathbb{P}(S_\tau = 30) + (-10) \cdot (1 - \mathbb{P}(S_\tau = -10)$$

And solving gives us that $\mathbb{P}(S_\tau = 30) = 1/4$.

2.3 Kolmogorov zero-one law

Definition. (Tail-after-time-n σ-field)

Let $X_1, X_2, \ldots$ be random variables. Define $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ as usual. Then:

1. The **σ-field generated by** $X_1, X_2, \ldots$ is:

$$\mathcal{F}_\infty = \sigma\left(\cup_{j=1}^\infty \sigma(X_1, \ldots, X_j)\right)$$

2. The **tail-after-time-n σ-field** is:

$$\mathcal{L}_n = \sigma(X_{n+1}, X_{n+2}, \ldots) = \sigma\left(\cup_{j=1}^\infty \sigma(X_{n+1}, \ldots, X_{n+j})\right)$$

Definition. (Tail σ-field)

Observe that $\mathcal{L}_n$ defined above is a decreasing sequence in n. We call the limit the **tail σ-field of** $X_1, X_2, \ldots$:

$$T = \cap_{n=0}^\infty \mathcal{L}_n = \cap_{n=1}^\infty \sigma(X_n, X_{n+1}, \ldots)$$

Example. The event $\{\limsup X_n \geq 2\} \in T$.

To see this, define the set

$$A_q = \{\omega : X_n(\omega) \geq 2 - 1/q \;\; i.o.\}$$

An intuitive interpretation of A_q is as the set where X_n goes above and stays above $2 - 1/q$. Now note that $\{\limsup X_n \geq 2\} = \cap_{q=1}^\infty A_q$. Therefore to show that the limsup is in the tail field, it is sufficient to show that $A_q \in T$ for all q.

Note that A_q can be written: $\cap_{m=1}^\infty \cup_{n \geq m} \{X_n \geq 2 - 1/q\}$. Fix some $k \geq 1$. Then:

$$A_q = \cap_{m=k}^\infty \cup_{n \geq m} \{X_n \geq 2 - 1/q\}$$

as well. Thus for any $k \geq 1$, A_q belongs to $\mathcal{L}_k$.

Example. The event $\{\lim S_n < \infty\} \in T$.

Apply the Cauchy criterion of convergence.

Example. The event $\{\omega : \limsup S_n > 2\} \notin T$.

i.e. this event "depends" on the first few r.v.'s.

To see this, let $X_1 = 3, X_2 = -3, X_3 = 3, X_4 = -3, \ldots$ Then S_n alternates between 0 and 3 and so $\limsup S_n = 3 > 2$ which is OK.

However, if we change $X_1 = -10$ and leave the rest of the sequence alone, then $\limsup S_n = -10 < 2$.

Theorem 2.16. (Kolmogorov zero-one law)

Suppose $X_1, X_2, \ldots$ are independent. Let $A \in T$, the tail field of $X_1, X_2, \ldots$. Then either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

> *Remark.* A more elegant proof than the one presented below can be found in the section on martingale convergence theorems.

Lemma 2.17. With the setup of the Kolmogorov zero-one law, fix $n \geq 1$. Then $\sigma(X_1, \ldots, X_n)$ and $\mathcal{L}_n$ are independent.

Proof. The definition of $\mathcal{L}_n$ is:

$$\mathcal{L}_n = \sigma\left(\cup_{j=1}^{\infty} \sigma(X_{n+1}, \ldots, X_{n_j})\right)$$

Note that $\cup_{j=1}^{\infty} \sigma(X_{n+1}, \ldots, X_{n+j})$ is a π-class which generates $\mathcal{L}_n$. Therefore, we only need to show that $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2)$ for

$$A_1 \in \sigma(X_1, \ldots, X_n) \quad \text{and} \quad A_2 \in \cup_{j=1}^{\infty} \sigma(X_{n+1}, \ldots, X_{n+j})$$

By this definition, then $A_2 \in \sigma(X_{n+1}, \ldots, X_{n+k})$ for some $k \geq 1$.

□

Lemma 2.18. $\sigma(X_1, X_2, \ldots)$ is independent of T.

Proof. Write $\sigma(X_1, X_2, \ldots) = \sigma(\cup_{j=1}^{\infty} \sigma(X_1, \ldots, X_j))$. By the π-class argument, we only need to show independence of:

$$A_1 \in \cup_{j=1}^{\infty} \sigma(X_1, \ldots, X_j) \quad \text{and} \quad A_2 \in T$$

Observe that $A_1 \in \sigma(X_1, \ldots, X_k)$ for some $k \geq 1$. Also, by definition of T, $A_2 \in \sigma(X_{k+1}, \ldots)$ for the same k. Invoke the previous lemma.

□

Proof. (of Kolmogorov zero-one law)

Fix $A \in T$. Since $\sigma(X_1, X_2, \ldots) \supset \mathcal{L}_n$ for all n, then $\sigma(X_1, X_2, \ldots) \supset \cap_{n=1}^{\infty} \mathcal{L}_n = T$. Since $A \in \sigma(X_1, X_2, \ldots)$ and $A \in T$ also, then A is independent of A by the previous lemma.

□

3 Laws of Large Numbers

3.1 4^{th} moment/L^2 strong laws, Glivenko-Cantelli

Theorem 3.1. (4^{th} moment strong law)

Let $\{X_i\}_{i\geq 1}$ be iid with $\mathbb{E}X_i = \mu$ and $\mathbb{E}(X_i^4) < \infty$. Then $\frac{S_n}{n} \xrightarrow{a.s.} \mu$.

Lemma 3.2. If $\mathbb{E}X_i = 0$, then $\mathbb{E}(S_n^4) \leq 3n^2\mathbb{E}(X_i^4)$.

Proof. Note that:

$$\mathbb{E}(S_n^4) = \mathbb{E}\left[\left(\sum X_i\right)^4\right] = \sum_{i,j,k,l} \mathbb{E}(X_i X_j X_k X_l)$$

Note that the RHS expression is zero if there is at least one index which does not equal one of the others. For example, $\mathbb{E}(X_1, X_2^2, X_3) = 0$. Therefore,

$$\begin{aligned}\mathbb{E}(S_n^4) &= \sum^n \mathbb{E}(X_i^4) + \binom{4}{2}\sum_{i\neq j}\mathbb{E}(X_i^2 X_j^2)\\ &= n\mathbb{E}(X_i^4) + 3n(n-1)\mathbb{E}(X_1^2 X_2^2)\\ &\leq n\mathbb{E}(X_1^4) + 3n(n-1)\mathbb{E}(X_1^4)\end{aligned}$$

Where the last step follows from the Cauchy-Schwarz inequality.

□

Proof. (of 4^{th} moment strong law)

Let $\mathbb{E}X_i = 0$. We show that $\frac{S_n}{n} \xrightarrow{a.s.} 0$ or, equivalently, $\mathbb{P}(|\frac{S_n}{n}| > \epsilon\ i.o.) = 0$. Note:

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \epsilon\right) &\leq \frac{\mathbb{E}\left(\left[\frac{S_n}{n}\right]^4\right)}{\epsilon^4}\\ &\leq \frac{3\mathbb{E}(X_1^4)}{n^2\epsilon^4}\end{aligned}$$

Which is summable, so we can invoke Borel-Cantelli to obtain the result.

□

Lemma 3.3. Suppose $\{A_n\}_{n\geq 1}$ is an independent sequence of events with $\mathbb{P}(A_n) = p$ for all $n \geq 1$. Then $\frac{1}{n}\sum \mathbb{1}_{A_i} \xrightarrow{a.s.} p$.

Proof. Let $X_n(\omega) = \mathbb{1}_{A_n}(\omega)$. Check assumptions and apply the 4^{th} moment strong law.

□

Theorem 3.4. (L^2 strong law)

Let $\{X_i\}_{i\geq 1}$ be a sequence of random variables such that

1. $\mathbb{E}X_i = 0\ \forall i$

2. $\mathbb{E}X_i^2 \leq c \; \forall i$

3. $\mathbb{E}(X_i X_j) = 0 \; \forall i \neq j$

Then $\sum^n X_i/n = S_n/n \xrightarrow{a.s.} 0$.

Lemma 3.5. Let $\{S_n\}_{n\geq 1}$ be a sequence in $\mathbb{R}$. If:

1. there exists a subsequence $\{n(j)\}_{j\geq 1}$ with $n(j) \uparrow \infty$ such that $\frac{S_{n(j)}}{n(j)} \to 0$

2. for $d_j = \max_{n(j)\leq n \leq n(j+1)} |S_n - S_{n(j)}|$ we have $\frac{d_j}{n(j)} \to 0$ as $j \to \infty$

Then $S_n/n \to 0$

Proof. Fix an $n(j) \leq n$.

$$\begin{aligned}\frac{|S_n|}{n} &= \frac{|S_n - S_{n(j)} + S_{n(j)}|}{n} \\ &\leq \frac{|S_{n(j)}| + |S_n - S_{n(j)}|}{n} \\ &\leq \frac{|S_{n(j)}|}{n(j)} + \frac{d_j}{n(j)} \to 0 \quad \text{as } j \to \infty\end{aligned}$$

□

Proof. (of main result) Fix $n(j) = j^2$, $j \geq 1$. We proceed in two steps:

Step 1: show that

$$\frac{S_{n(j)}}{n(j)} = \frac{S_{j^2}}{j^2} \xrightarrow{a.s.} 0$$

This follows immediately from Borel-Cantelli. Note that

$$\mathbb{P}[|S_{j^2}| > \epsilon j^2] \leq \frac{\text{Var}(S_{j^2})}{j^4\epsilon^2} \leq \frac{cj^2}{j^4\epsilon^2} = \frac{c}{j^2\epsilon^2}$$

which is summable over j.

Step 2: show that

$$\frac{D_{j^2}}{j^2} \xrightarrow{a.s.} 0, \quad \text{where} \quad D_{j^2} = \max_{j^2 \leq n \leq (j+1)^2} \left|S_n - S_{j^2}\right|$$

We show this again by Borel-Cantelli. First observe that, by assumption, $\mathbb{E}(S_n - S_i) = 0$ for all $0 \leq i \leq n$. Also, note $\mathbb{E}[(S_n - S_i)^2] = \text{Var}(\sum_{i+1}^n X_i) \leq c(n-i)$.

Second, observe that

$$D_{j^2}^2 = \max_{j^2 \leq n \leq (j+1)^2} \left|S_n - S_{j^2}\right|^2 \leq \sum_{n=j^2}^{(j+1)^2} \left|S_n - S_{j^2}\right|^2$$

Therefore we have

$$
\begin{aligned}
\mathbb{P}(D_{j^2} > \epsilon j^2) &\le \frac{\mathbb{E}[(D_{j^2})^2]}{\epsilon^2 j^4} \\
&\le \frac{\mathbb{E}\left[\sum_{n=j^2}^{(j+1)^2} \left|S_n - S_{j^2}\right|^2\right]}{\epsilon^2 j^4} \\
&\le \frac{\sum_{n=j^2}^{(j+1)^2} c(n-j^2)}{\epsilon^2 j^4} \\
&= \frac{c\sum_{k=1}^{2j+1} k}{\epsilon^2} j^4 = \frac{c(2j+1)(2j+2)}{2\epsilon^2 j^4} \\
\mathbb{P}(D_{j^2} > \epsilon j^2) &\le \frac{c'}{\epsilon^2} j^2
\end{aligned}
$$

which is summable, so $\frac{D_{j^2}}{j^2} \xrightarrow{a.s.} 0$.

Finally, observe that the two conditions for Lemma 3.5 are satisfied. Since the two conditions hold a.s., then the convergence result of the lemma is also a.s.

□

Theorem 3.6. (Glivenko-Cantelli theorem)

Suppose $\{X_i\}_{i\ge1}$ is an iid sequence. Define the empirical distribution function as

$$G_n(\omega, x) = \frac{1}{n}\sum^{n} \mathbb{1}_{X_i(\omega)\le x},\ x \in \mathbb{R}$$

Define F as the distribution function of X_i, i.e.

$$F(x) = \mathbb{P}(X_1 \le x),\ x \in \mathbb{R}$$

Then $|G_n(\omega, x) - F(x)| \xrightarrow{a.s.} 0$ for any fixed x **and**

$$\sup_{x\in\mathbb{R}} \left|G_n(\omega, x) - F(x)\right| \xrightarrow{a.s.} 0$$

Remark. Note that the first (weaker) result is akin to pointwise convergence of the distribution functions, and follows immediately from the strong law because indicator functions have finite moments. The second result gives uniform convergence of the distribution functions and is much stronger.

Lemma 3.7. (Deterministic)

Suppose $\{F_n\}_{n\ge1}$ and F are distribution functions. If

1. For all $x \in \mathbb{Q}$, $F_n(x) \to F(x)$
2. For all atoms x of F, $F_n(x) \to F(x)$ and $F_n(x_-) \to F(x_-)$

Then $sup_{x\in R}|F_n(x) - F(x)| \to 0$.

Proof. (Exercise)

Lemma 3.8. Any probability measure μ on $\mathbb{R}$ has only countably many atoms.

Proof. Let $\mathcal{A}$ be the set of atoms of μ. Consider any $x \in \mathcal{A}$. By the definition of atom, $F(x_-) < F(x_+)$, so we can pick a rational q_x such that $F(x_-) < q_x < F(x_+)$.

Then, since μ is non-decreasing, then for $y \neq z \in A$ we must have $q_y \neq q_z$. Otherwise one of the points is not an atom. Therefore the mapping from $\mathcal{A} \to \mathbb{Q}$ is one-to-one and therefore $|\mathcal{A}| = |\mathbb{Q}|$.

□

Proof. (of main result) Fix $x \in \mathbb{R}$ and define:

$$A_n = \{X_n \leq x\} = \{\omega : X_n(\omega) \leq x\}$$

This sequence $\{A_n\}$ satisfies the conditions of Lemma 3.3, so we immediately obtain the first result:

$$\frac{1}{n}\sum^{n} \mathbb{1}_{A_i} = G_n(\omega, x) \xrightarrow{a.s.} F(x)$$

To show the second result, we show that G_n and F satisfy the two conditions of Lemma 3.7.

1. First, note that if we redefine the sequence $\{A_n\}$ to be X_n strictly less than X, then we immediately obtain

$$\frac{1}{n}\sum^{n} \mathbb{1}_{A_i} = G_n(\omega, x_-) \xrightarrow{a.s.} F(x_-)$$

   Also, since we know $G_n(\omega, x) \xrightarrow{a.s.} F(x)$ for all $x \in \mathbb{R}$, then it also holds for all $q \in \mathbb{Q}$. Therefore if we define

$$B_q = \{\omega : G_n(\omega, q) \to F(q)\}$$

   then $\mathbb{P}(B_q) = 1$.

2. Now let $\mathcal{A}$ denote the set of all atoms of F. For any $x \in \mathcal{A}$, $x \in \mathbb{R}$ also and we have already shown that $G_n(\omega, x) \xrightarrow{a.s.} F(x)$ and $G_n(\omega, x_-) \xrightarrow{a.s.} F(x_-)$ for all $x \in \mathbb{R}$. Therefore if we define

$$C_x = \{\omega : G_n(\omega, x) \to F(x),\, G_n(\omega, x_-) \to F(x_-)\}$$

   then $\mathbb{P}(C_x) = 1$.

   Now note, by definition of B_q and C_x above,

$$\{\omega : \sup_{x\in\mathbb{R}} |G_n(\omega, x) - F(x)| \to 0\} \supset \bigcap_{q\in\mathbb{R}} B_q \cap \bigcap_{x\in\mathcal{A}} C_x$$

   The result follows by observing that $\{A_n\}_{n\geq 1}$ and $\mathbb{P}(A_n) = 1\ \forall n$, then $\mathbb{P}(\cap^{\infty} A_n) = 1$.

□

3.2 2^{nd} moment results

Theorem 3.9. (Kolmogorov's maximal inequality)

Let $\{X_i\}$ be a sequence of independent r.v.'s with $\mathbb{E}X_i = 0$ and $\mathbb{E}(X_i^2) < \infty\ \forall i$. Define

$$S_n = \sum^n X_i \quad \text{and} \quad S_n^* = \max_{1\le k\le n} |S_k|$$

Then

$$\mathbb{P}(S_n^* \ge x) \le \frac{\mathbb{E}(S_n^2)}{x^2}$$

Remark. Note the similarity to the Chebyshev's inequality bound on S_n:

$$\mathbb{P}(|S_n| \ge x) \le \frac{\mathbb{E}(S_n^2)}{x^2}$$

Proof. Fix k. Define the random variable

$$A_k = \{|S_i| < x \text{ and } |S_k| \ge x \text{ for } 1 \le i \le k-1\}$$

In other words, A_k is the event that the first crossing of partial sums over x occurs at the k^{th} step. Observe that the A_k's are disjoint and also that: $\{S_n^* > x\} = \cup_{k=1}^n A_k$.

Furthermore, note that the joint event (A_k, S_k) is independent of the event $S_n - S_k = \sum_{k+1}^n X_i$ since the former deals only with partial sums up to k while the latter deals with sums of X_i's from $k+1$ onwards, and the X_i's are independent.

$$\begin{aligned}
\mathbb{E}(S_n^2) &= \int_{\cup_{k=1}^n A_k} S_n^2 \,\mathrm{d}\mathbb{P} \\
&= \sum_{k=1}^n \int_{A_k} S_n^2 \,\mathrm{d}\mathbb{P} \\
&= \sum_{k=1}^n \int_{A_k} (S_n - S_k + S_k)^2 \,\mathrm{d}\mathbb{P} \\
&= \sum_{k=1}^n \int_{A_k} (S_n - S_k)^2 + S_k^2 + 2S_k(S_n - S_k) \,\mathrm{d}\mathbb{P} \\
&= \sum_{k=1}^n \int_{A_k} (S_n - S_k)^2 \,\mathrm{d}\mathbb{P} + \sum_{k=1}^n \int_{A_k} S_k^2 \,\mathrm{d}\mathbb{P} \\
&\quad + \sum_{k=1}^n 2\mathbb{E}[S_k(S_n - S_k)\mathbb{1}_{A_k}] \\
&\ge \sum_{k=1}^n \int_{A_k} S_k^2 \,\mathrm{d}\mathbb{P} + \sum_{k=1}^n 2\mathbb{E}[S_k(S_n - S_k)\mathbb{1}_{A_k}] \\
&= \sum_{k=1}^n \int_{A_k} S_k^2 \,\mathrm{d}\mathbb{P} \quad \text{(by independence of } S_k,\ (S_n - S_k))
\end{aligned}$$

Note that, by definition of A_k, $S_k \geq x$ in the the last step, so the integrand is bounded below by x^2 and we have:

$$\mathbb{E}(S_n^2) \;\geq\; \sum_{k=1}^{n} \int x^2 \mathbb{1}_{A_k}\, \mathrm{d}\mathbb{P} \;=\; x^2 \sum_{k=1}^{n} \mathbb{P}(A_k)$$

By disjointness and definition of the A_k's, we then have

$$\mathbb{E}(S_n^2) \;\geq\; x^2 \cdot \mathbb{P}(\cup_{k=1}^{n} A_k) \;=\; x^2 \cdot \mathbb{P}(S_n^* > x)$$

□

Theorem 3.10. (Cauchy criterion)

A series $\sum_{n=1}^{\infty} a_n$ in a complete metric space M converges iff for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for any $n > k > N$,

$$\Big|\sum_{j=k}^{n} a_j\Big| < \epsilon$$

Or, equivalently, that

$$\sup_{n \geq k} \Big|\sum_{j=k}^{n} a_j\Big| \to 0 \quad \text{ as } k \to \infty$$

Proof. The Cauchy criterion is equivalent to the condition that the sequence of partial sums S_n is a Cauchy sequence. Since M is complete, then S_n converges. Since an infinite series converges iff the sequence of partial sums converges, then the result is shown.

□

Theorem 3.11. ("Random series theorem")

Let $\{X_i\}_{i \geq 1}$ be independent with $\mathbb{E}(X_i) = 0$ and $\mathbb{E}(X_i^2) = \sigma_i^2 < \infty$ for all i. Suppose that $\sum_{i=1}^{\infty} \sigma_i^2 < \infty$. Then $\sum_{i=1}^{\infty} X_i(\omega)$ converges a.s.

Proof. We want to show that the Cauchy criterion is satisfied. Chebyshev's inequality allows us to control the size of the partial sum for given k, but that is not enough to satisfy the criterion. Instead, we must use Kolmogorov's maximal inequality to control the size of any given sub-series in the tail.

Define $M_k(\omega) = \sup_{n \geq k} |\sum_{j=k}^{n} X_j(\omega)|$. By the Cauchy criterion, it will be sufficient to show that $M_k(\omega) \to 0$ a.s.

Fix $\epsilon > 0$ and $N > k$. Note that, by Kolmogorov's maximal inequality (for the series starting from index k),

$$\mathbb{P}\left(\sup_{k \leq n \leq N} \Big|\sum_{i=k}^{n} X_i(\omega)\Big| > \epsilon\right) \leq \sum_{i=k}^{n} \frac{\sigma_i^2}{\epsilon^2}$$

Letting $N \to \infty$, and using continuity of $\mathbb{P}$, we have

$$\mathbb{P}\left(\sup_{n \geq k} \Big|\sum_{i=k}^{n} X_i(\omega)\Big| > \epsilon\right) = \mathbb{P}(M_k(\omega) > \epsilon) \leq \sum_{i=k}^{\infty} \frac{\sigma_i^2}{\epsilon^2}$$

Now we send $k \to \infty$ to obtain the limit of M_k on the left side:

$$\lim_{k\to\infty} \mathbb{P}(M_k(\omega) > \epsilon) = 0$$

where the RHS $\to 0$ as $k \to \infty$ since $\sum \sigma_i^2 < \infty$ and $\sigma_i^2 > 0$ for all i, so the series is convergent.

So we have shown $M_k(\omega) \xrightarrow{\mathbb{P}} 0$. We now show that $M_k(\omega) \xrightarrow{a.s.} 0$ as well. Fix $k \geq 1$ and define

$$W_k(\omega) = \sup_{k\leq n_1 \leq n_2} \Big| \sum_{i=n_1}^{n_2} X_i(\omega) \Big|$$

By the triangle inequality, we have $M_k \leq W_k \leq 2M_k$.

Now note that W_k is decreasing (a.s.) in k. So let $W_\infty = \lim_{k\to\infty} W_k$. Since $W_k \leq 2M_k$ and $M_k \xrightarrow{\mathbb{P}} 0$, then $W_k \xrightarrow{\mathbb{P}} 0$ also. Since $W_k \xrightarrow{a.s.} W_\infty$, then also $W_k \xrightarrow{\mathbb{P}} W_\infty$. But we just showed that $W_k \xrightarrow{\mathbb{P}} 0$, so $W_\infty = 0$.

Finally, note that since $M_k \leq W_k$, then $M_k \xrightarrow{a.s.} 0$ as well.

□

Theorem 3.12. (Kolmogorov 3-series theorem)

Let $\{X_n\}_{n\geq 1}$ be an independent sequence of $\mathbb{R}$-valued r.v.'s with $\mathbb{E}|X_n| < \infty$ and $\mathbb{E}(X_n^2) < \infty$. Define, for some $A > 0$:

$$Y_n = X_n \cdot \mathbb{1}_{(|X_n|\leq A)}$$

Then $\sum_{n=1}^\infty X_n$ converges a.s. if and only if:

1. $\sum_{n=1}^\infty \mathbb{P}(|X_n| > A) < \infty$
2. $\sum_{n=1}^\infty \mathbb{E}(Y_n)$ converges
3. $\sum_{n=1}^\infty \mathrm{Var}(Y_n) < \infty$

Remark. The conditions for 1) and 3) only require the series to be bounded, but since the terms of those series are non-negative, then it is equivalent to requiring the series to be convergent.

Proof. Define $X_n' = Y_n - \mathbb{E}Y_n$. Thus $\mathbb{E}X_n' = 0$ and $\mathrm{Var}(X_n') = \mathrm{Var}(Y_n)$.

Note that the sequence of r.v.'s $\{X_n'\}$ satisfies the requirements of the "Random series theorem." Therefore $\sum_{n=1}^\infty X_n' = \sum_{n=1}^\infty (Y_n - \mathbb{E}Y_n)$ converges a.s. Therefore, since $\sum_{n=1}^\infty \mathbb{E}Y_n$ converges by assumption, the difference of the series $\sum_{n=1}^\infty Y_n$ must converge a.s. also.

Note $\mathbb{P}(X_n \neq Y_n) = \mathbb{P}(|X_n| > A)$. By assumption, $\sum_{n=1}^\infty \mathbb{P}(|X_n| > A) < \infty$ so that by Borel-Cantelli, $\mathbb{P}(X_n \neq Y_n \ i.o.) = 0$.

Equivalently, this means that $\mathbb{P}(X_n = Y_n \textit{ eventually}) = 1$. So for a.e. ω, there exists $N(\omega)$ such that for $n \geq N(\omega)$, $X_n(\omega) = Y_n(\omega)$. Therefore,

$$\sum_{n=N(\omega)}^{\infty} Y_n = \sum_{n=N(\omega)}^{\infty} X_n$$

And since $\sum Y_n$ converges a.s. as shown above, its tail converges a.s. Thus the tail of $\sum X_n$ converges a.s., so that the entire series converges a.s.

□

3.3 Strong law of large numbers

We begin by proving the important Kronecker's lemma and give some motivation for the proof.

Lemma 3.13. (Cesaro's lemma)

Let $\{a_n\}$ be a sequence of strictly positive real numbers with $a_n \uparrow \infty$. Let $\{X_n\}$ be a convergent sequence of real numbers. Then:

$$\frac{1}{X_n}\sum_{k=1}^{n}(a_k - a_{k-1})X_k \to \lim X_n$$

Proof. Let $\epsilon > 0$. Since X_n is convergent, we can choose N such that

$$X_k > \lim X_n - \epsilon \quad \text{whenever } k \geq N$$

Then split the sum into the portion up to N and the portion beyond N, and apply the above inequality:

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{X_n}\sum_{i=1}^{n}(a_k - a_{k-1})X_k \\
&\quad \geq \liminf_{n\to\infty}\left\{\frac{1}{X_n}\sum_{k=1}^{N}(a_k - a_{k-1})X_k + \frac{a_n - a_N}{X_n}(\lim X_n - \epsilon)\right\} \\
&\quad \geq \liminf_{n\to\infty}\left\{\frac{1}{X_n}\sum_{k=1}^{N}(a_k - a_{k-1})X_k + \frac{a_n - a_N}{a_n}(\lim X_n - \epsilon)\right\} \\
&\quad \geq 0 + \lim X_n - \epsilon
\end{aligned}$$

Where the last step follows from the fact that $a_n \uparrow \infty$ and a_N is finite.

This is true for all $\epsilon > 0$, so $\liminf \geq \lim X_n$. To show that $\limsup \leq \lim X_n$, follow the same argument except choose N such that $X_k < \lim X_n + \epsilon$ whenever $k \geq N$.

□

Lemma 3.14. (Kronecker's lemma)

Let $\{Y_n\}_{n\geq 1}$ be a real-valued sequence and let $\{a_n\}_{n\geq 1}$ be a sequence of strictly positive real numbers with $a_n \uparrow \infty$. If $\sum Y_n/a_n < \infty$, then $S_n/a_n \to 0$.

Proof. Define $X_n = \sum_{i=1}^n Y_n/a_n$. By assumption, X_n converges so $\lim X_n$ exists. Note $X_k - X_{k-1} = Y_k/a_k$ so that:

$$\begin{aligned} S_n &= \sum_{k=1}^n a_k(X_k - X_{k-1}) \\ &= a_n X_n - \sum_{k=1}^n (a_n - a_{n-1})X_{k-1} \end{aligned}$$

Divide by a_n, send $n \to \infty$, and apply Cesaro's lemma.

□

Theorem 3.15. (Kolmogorov criterion)

Suppose $\{X_n\}_{n\geq 1}$ is a sequence of independent r.v.'s with $\mathbb{E}X_n = 0$ and $\text{Var}(X_n) = \mathbb{E}X_n^2 < \infty$.

If, for some $0 < a_n \uparrow \infty$, $\sum \mathbb{E}(X_n^2)/a_n^2 < \infty$, then $S_n/a_n \xrightarrow{a.s.} 0$.

Proof. Consider the sequence of random variables $Y_n = X_n/a_n$. Note that $\mathbb{E}Y_n = 0$ and $\text{Var}(Y_n) = \mathbb{E}X_n^2/a_n^2$. By independence, $\text{Var}(Y_n) < \infty$.

By the "random series theorem", then $\sum Y_n(\omega) = \sum X_n(\omega)/a_n$ converges a.s. and thus, by Kronecker's lemma, $S_n/a_n \to 0$ a.s.

□

Example. (Rates of convergence)

Another way of interpreting the result that $S_n/a_n \to 0$ is that our choice of a_n grows faster than S_n. To illustrate how this can be used, consider the following setup:

Let $\{X_n\}_{n\geq 1}$ be iid with $\mathbb{E}(X_n^2) < \infty$ and $\mathbb{E}X_i = 0$. Let $a_n^2 = n(\log n)^{1+\epsilon}$, $\epsilon > 0$. Then note:

$$\sum^{\infty} \frac{\mathbb{E}(X_n^2)}{n(\log n)^{1+\epsilon}} = c\sum^{\infty} \frac{1}{n(\log n)^{1+\epsilon}} < \infty$$

And therefore by the Kolmogorov criterion,

$$\frac{S_n}{a_n} = \frac{S_n}{\sqrt{n \log n}(\log n)^{\epsilon/2}} \xrightarrow{a.s.} 0$$

To compare, the CLT gives us that $\limsup S_n/\sqrt{n} = \infty$ a.s. and the law of the interated logarithm gives us the most optimal bound $\limsup S_n/\sqrt{2n \log\log n} \xrightarrow{a.s.} 1$. So our choice of a_n grows slightly faster than optimally, but not by much.

Theorem 3.16. (Strong law of large numbers)

Suppose $\{X_i\}_{i\geq 1}$ are iid with $\mathbb{E}|X_i| < \infty$ with $\mathbb{E}X_i = \mu$. Then $S_n/n \xrightarrow{a.s.} \mu$.

Proof. We use truncation to obtain a sequence of r.v.'s with finite 2^{nd} moment. We then show a.s. convergence of the sequence of partial sums of that series by Kolmogorov's criterion, and then use dominated convergence to to show convergence of the original sequence.

Step 1: show that truncating is OK

Let $Y_k = X_k \cdot \mathbb{1}_{(|X_k| \le k)}$. Note $\mathbb{E}(Y_k^2) \le k^2 < \infty$ so that we can apply our 2nd moment results to this sequence. Observe that:

$$\begin{aligned}\sum_{k=1}^{\infty} \mathbb{P}(Y_k \neq X_j) &= \sum_{k=1}^{\infty} \mathbb{P}(|X_k| \ge k) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(|X_1| \ge k) \\ &\le \int_0^{\infty} \mathbb{P}(|X_1| > x) \, \mathrm{d}x \\ &= \mathbb{E}|X_1| < \infty\end{aligned}$$

Therefore, by Borel-Cantelli, we have $\mathbb{P}(Y_k \neq X_k \ i.o.) = 0$. So for a.e. ω, there exists $N(\omega)$ s.t. $\forall k \ge N(\omega), X_k(\omega) = Y_k(\omega)$. Thus the tail behavior is identical after a certain point, so it will be sufficient to show that

$$\frac{1}{n} \sum^{n} Y_k \xrightarrow{a.s.} \mu$$

Step 2: show $\frac{1}{n} \sum^n (Y_k - \mathbb{E}Y_k) \xrightarrow{a.s.} 0$

Let $X'_k = Y_k - \mathbb{E}Y_k$. Note $\mathbb{E}X'_k = 0$ and $\mathrm{Var}(X'^2_k) = \mathrm{Var}(Y_k) < \infty$.

By the Kolmogorov criterion, to show $\frac{1}{n} \sum^n X'_k \xrightarrow{a.s.} 0$, it will be sufficient to show that $\sum^{\infty} \mathbb{E}(X'^2_n)/n^2 < \infty$.

Note that $\mathbb{E}(X'^2_n) \le \mathbb{E}(Y_n^2)$. Also note the following identity for a non-negative random variable X:

$$\mathbb{E}(X^p) = p \int_0^{\infty} x^{p-1} \cdot \mathbb{P}(X > x) \, \mathrm{d}x$$

Applying this to the random variable $|Y_n|$ with $p = 2$, we have:

$$\begin{aligned}\mathbb{E}(Y_n^2) &= 2 \int_0^{\infty} x \cdot \mathbb{P}(|Y_k| > x) \, \mathrm{d}x \\ &= 2 \int_0^{\infty} x \cdot \mathbb{P}(|X_k| > x) \cdot \mathbb{1}_{(x \le n)} \, \mathrm{d}x\end{aligned}$$

Thus we have:

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{\mathbb{E}(X'^2_n)}{n^2} &\le \sum_{n=1}^{\infty} \frac{2 \int_0^{\infty} x \cdot \mathbb{P}(|X_k| > x) \cdot \mathbb{1}_{(x \le n)} \, \mathrm{d}x}{n^2} \\ &= \int_0^{\infty} \sum_{n=1}^{\infty} \frac{2x \cdot \mathbb{P}(|X_k| > x) \cdot \mathbb{1}_{(x \le n)}}{n^2} \, \mathrm{d}x \\ &= \int_0^{\infty} \mathbb{P}(|X_1| > x) \cdot G(x) \, \mathrm{d}x \\ &\text{where } G(x) = 2x \sum_{n=1}^{\infty} \frac{\mathbb{1}_{x \le n}}{n^2}\end{aligned}$$

We show that $G(x)$ is upper bounded so that we can pull it outside of the integral. Then we will have shown that the LHS sum is $\leq$ a constant times $\mathbb{E}|X_1|$, which is finite by assumption.

We show that $G(x)$ is upper bounded by 4. First, suppose that $x \leq 1$. Then:

$$G(x) \leq 2\sum^{\infty} \frac{1}{n^2} = 2 \cdot \frac{\pi^2}{6}$$

Therefore for general x, we have that

$$G(x) \leq 2x \sum_{n=\lceil x \rceil}^{\infty} \frac{1}{n^2}$$

Next, using the identity $\frac{1}{n^2} \leq \int_{n-1}^{n} \frac{1}{y^2}\,\mathrm{d}y$, we have

$$\begin{aligned} G(x) &\leq 2x \int_{\lceil x \rceil - 1}^{\infty} \frac{1}{y^2}\,\mathrm{d}y \\ &\leq 2x \left[\frac{1}{y}\right]\Bigg|_{\lceil x \rceil - 1}^{\infty} \\ &= \frac{2x}{\lceil x \rceil - 1} \leq 4 \end{aligned}$$

Thus the $\sum \mathbb{E}(X_n'^2)/n^2 < \infty$, so $\frac{1}{n}\sum^n (Y_k - \mathbb{E}Y_k)$ converges to 0 a.s.

Step 3: show $\frac{1}{n}\sum^n \mathbb{E}Y_k \xrightarrow{a.s.} 0$

At this stage we have that $\frac{1}{n}\sum^n (Y_k - \mathbb{E}Y_k) \xrightarrow{a.s.} 0$.

If we can show that $\frac{1}{n}\sum^n \mathbb{E}Y_k \xrightarrow{a.s.} \mu$, then it will imply that $\frac{1}{n}\sum^n Y_k \xrightarrow{a.s.} \mu$ as well.

To see this, note simply that

$$\begin{aligned} \mathbb{E}Y_k &= \mathbb{E}(X_k \cdot \mathbb{1}_{|X_k| \leq k}) \\ &= \mathbb{E}(X_1 \cdot \mathbb{1}_{|X_1| \leq k}) \\ &\xrightarrow{DC} \mathbb{E}X_1 \quad \text{as } k \to \infty \end{aligned}$$

Thus since $\mathbb{E}Y_k \xrightarrow{a.s.} \mu$, then $\frac{1}{n}\sum^n \mathbb{E}Y_k \xrightarrow{a.s.} \mu$ also.

□

Corollary 3.17. Suppose $X_1, X_2, \ldots$ are iid with $\mathbb{E}X_i^+ = \infty$ and $\mathbb{E}X_i^- < \infty$. Then $\frac{1}{n}\sum^n X_k \xrightarrow{a.s.} \infty$.

Proof. Fix $B > 0$. Let $Y_k = X_k \cdot \mathbb{1}_{X_k \leq B}$.

Note that the $\{Y_k\}_{k \geq 1}$ are also iid with

$$\mathbb{E}|Y_k| = \mathbb{E}(X_1^+ \cdot \mathbb{1}_{X_1^+ \leq B}) + \mathbb{E}X_1^- < \infty$$

$$\mathbb{E}Y_k = \mathbb{E}(X_1^+ \cdot \mathbb{1}_{X_1^+ \leq B}) - \mathbb{E}X_1^- < \infty$$

By the strong law,

$$\frac{1}{n}\sum^n Y_k \xrightarrow{a.s.} \mathbb{E}(X_1^+ \cdot \mathbb{1}_{X_1^+ \leq B}) - \mathbb{E}X_1^-$$

And since $Y_k \leq X_k$ for all k, then

$$\liminf \frac{1}{n}\sum^n X_k \geq \lim \frac{1}{n}\sum^n Y_k = \mathbb{E}(X_1^+ \cdot \mathbb{1}_{X_1^+ \leq B}) - \mathbb{E}X_1^-$$

This is true for any B. So sending $B \to \infty$, by monotone convergence we can see that the RHS diverges to infinity. Thus the limit of $\frac{1}{n}\sum^n X_k = \infty$.

□

4 Martingales

4.1 Conditional expectation

Definition. (Conditional expectation)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a random variable with $\mathbb{E}|X| < \infty$.

Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field of $\mathcal{F}$ (e.g. $\mathcal{G} = \sigma(Y)$ for some random variable Y defined on the same probability space).

Then there exists a random variable Z called the **conditional expectation of** X **given** $\mathcal{G}$, written $\mathbb{E}(X \,|\, \mathcal{G})$, with the following two properties:

1. Z is $\mathcal{G}$-measurable.
2. For any set $A \in \mathcal{G}$,
$$\int_A X \,\mathrm{d}\mathbb{P} = \int_A Z \,\mathrm{d}\mathbb{P}$$

Proof. (of existence)

Consider the probability space $(\Omega, \mathcal{G}, \mathbb{P})$. Since $\mathbb{P}$ is defined on $\mathcal{F}$ and $\mathcal{G} \subset \mathcal{F}$, then its restriction is a measure on $\mathcal{G}$.

First suppose that $X \geq 0$. Define the function $\nu(A) = \int_A X \,\mathrm{d}\mathbb{P}$, $A \in \mathcal{G}$. It is easily checked that ν is a measure on $\mathcal{G}$ such that $\nu \ll \mathbb{P}$. Then by the Radon-Nikodym theorem, there exists an a.e. unique $\mathcal{G}$-measurable function $Z : (\Omega, \mathcal{G}, \mathbb{P}) \to \mathbb{R}$ such that
$$\nu(A) = \int_A Z \,\mathrm{d}\mathbb{P} = \int_A X \,\mathrm{d}\mathbb{P}$$

Furthermore, Z is integrable since $\mathbb{E}|X| < \infty$.

For the case of X not necessarily ≥ 0, write $X = X^+ - X^-$ and let $Z_1 = \mathbb{E}(X^+ \,|\, \mathcal{G})$ and $Z_2 = \mathbb{E}(X^- \,|\, \mathcal{G})$. Then $Z_1 - Z_2$ is integrable and:
$$\begin{aligned}\int_A X \,\mathrm{d}\mathbb{P} &= \int_A X^+ \,\mathrm{d}\mathbb{P} - \int_A X^- \,\mathrm{d}\mathbb{P} \\ &= \int_A Z_1 \,\mathrm{d}\mathbb{P} - \int_A Z_2 \,\mathrm{d}\mathbb{P} \\ &= \int_A (Z_1 - Z_2) \,\mathrm{d}\mathbb{P}\end{aligned}$$

□

Note that the a.e. uniqueness of the CE follows immediately from Radon-Nikodym. But proving it explicitly is easy:

Proof. Assume there exists Z, Z' such that, for $A \in \mathcal{G}$,

$$\int_A X \, d\mathbb{P} = \int_A Z_1 \, d\mathbb{P} = \int_A Z_2 \, d\mathbb{P}$$

Fix $\epsilon > 0$ and let $E = \{Z_1 - Z_2 > \epsilon\}$. Then

$$0 = \int_E X - X \, d\mathbb{P} = \int_E Z - Z' \, d\mathbb{P} \geq \epsilon \cdot \mathbb{P}(E)$$

So $Z \geq Z'$ a.e. Flip Z, Z' to show the opposite direction.

□

Remark. Note that we cannot conclude that $Z = X$ from the fact that $\int_A Z \, d\mathbb{P} = \int_A X \, d\mathbb{P}$, since X may not be $\mathcal{G}$-measurable.

Theorem 4.1. The second condition in the definition of CE is equivalent to the following conditions:

1. Let D be a π-class with $\mathcal{G} = \sigma(D)$. Then $\int_A X \, d\mathbb{P} = \int_A Z \, d\mathbb{P}$ for all $A \in D$.
2. For any bounded $\mathcal{G}$-measurable r.v. V, $\mathbb{E}(XV) = \mathbb{E}(ZV)$.

   Proof. For the first, define $L = \{A \in \mathcal{G} : \int_A X \, d\mathbb{P} - \int_A Z \, d\mathbb{P}\}$. Show that L is a λ-class.

   For the second, note that the condition obviously holds for $V = \mathbb{1}_A$. Thus it must hold for simple functions, and, by monotone convergence, for general measurable functions.

Example. (Full information)

If $\mathcal{G} = \mathcal{F}$, then $\mathbb{E}(X \mid \mathcal{G}) = X$. To see this, note that X always satisfies condition (2) so the only thing keeping X from equalling $\mathbb{E}(X \mid \mathcal{G})$ is condition (1), which is satisfied.

Intuitively, if we know exactly what happened in Ω, then our best guess of the value of $X(\omega)$ is $X(\omega)$ itself.

Example. (No information)

If $\mathcal{G} = \{\emptyset, \Omega\}$, then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}X$. To see this, note that $\mathcal{G}$ is independent of X. Therefore:

$$\int_A X \, d\mathbb{P} = \mathbb{E}(X \mathbb{1}_A) = \mathbb{E}X \cdot \mathbb{E}\mathbb{1}_A = \mathbb{E}X$$

since $A = \emptyset$ or $A = \Omega$.

Intuitively, if we don't know anything about what happened in Ω, then our best guess of the value of $X(\omega)$ is simply the unconditional expected value.

Example. (Partition information)

Let $\{B_n\}_{n \geq 1}$ be a measurable partition of Ω. Let $\mathcal{G} = \sigma(B_1, B_2, \ldots)$. Then

$$\mathbb{E}(X \mid \mathcal{G}) = \frac{\int_{B_i} X \, d\mathbb{P}}{\mathbb{P}(B_i)} = \frac{\mathbb{E}(X; B_i)}{\mathbb{P}(B_i)} \quad \text{on } B_i$$

Where $\mathbb{E}(X; B_i)$ is the expected value of X restricted to B_i:

$$\mathbb{E}(X; B_i) = \int_{B_i} X \, \mathrm{d}\mathbb{P}$$

To see this, note that the RHS is constant on each B_i, so it is measurable w.r.t. $\mathcal{G}$ and condition (1) is satisfied. To show condition (2), note that the set $\{\emptyset, B_1, B_2, \ldots\}$ generates $\mathcal{G}$ and also is a π-class since the B_i's constitute a partition. Therefore we only need to check equality for a given B_i by the generating π-class property:

$$\int_{B_i} \frac{\mathbb{E}(X; B_i)}{\mathbb{P}(B_i)} \, \mathrm{d}\mathbb{P} = \frac{\mathbb{E}(X; B_i)}{\mathbb{P}(B_i)} \int_{B_i} \mathrm{d}\mathbb{P} = \mathbb{E}(X; B_i) = \int_{B_i} X \, \mathrm{d}\mathbb{P}$$

Intuitively, the information in B_i tells us which element of the partition our outcome lies in, and, given this information, our best guess for X is the average value of X over that partition.

Note that the example of no information is a special case of this example.

Theorem 4.2. (Bayes' formula)

Define $\mathbb{P}(A \,|\, B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ and $\mathbb{P}(A \,|\, \mathcal{G}) = \mathbb{E}(\mathbb{1}_A \,|\, \mathcal{G})$.

Let $G \in \mathcal{G}$. Then

$$\mathbb{P}(G \,|\, A) = \frac{\int_G \mathbb{P}(A \,|\, \mathcal{G}) \, \mathrm{d}\mathbb{P}}{\int_\Omega \mathbb{P}(A \,|\, \mathcal{G}) \, \mathrm{d}\mathbb{P}}$$

Proof. Note that:

$$\int_G \mathbb{P}(A \,|\, \mathcal{G}) \, \mathrm{d}\mathbb{P} = \int_G \mathbb{E}(\mathbb{1}_A \,|\, \mathcal{G}) \, \mathrm{d}\mathbb{P} = \int_G \mathbb{1}_A \, \mathrm{d}\mathbb{P} = \mathbb{P}(A \cap G)$$

And by a similar argument,

$$\int_G \mathbb{P}(A \,|\, \mathcal{G}) \, \mathrm{d}\mathbb{P} = \mathbb{P}(A)$$

Therefore we have:

$$\frac{\int_G \mathbb{P}(A \,|\, \mathcal{G}) \, \mathrm{d}\mathbb{P}}{\int_\Omega \mathbb{P}(A \,|\, \mathcal{G}) \, \mathrm{d}\mathbb{P}} = \frac{\mathbb{P}(A \cap G)}{\mathbb{P}(A)} = \mathbb{P}(G \,|\, A)$$

□

Theorem 4.3. (Two basic properties)

1. If V is $\mathcal{G}$-measurable, then $\mathbb{E}(V \,|\, \mathcal{G}) = V$.
2. $\mathbb{E}(X + Y \,|\, \mathcal{G}) = \mathbb{E}(X \,|\, \mathcal{G}) + \mathbb{E}(Y \,|\, \mathcal{G})$

Proof. Trivial.

□

Theorem 4.4. Let V be $\mathcal{G}$-measurable with $\mathbb{E}|VX| < \infty$. Then

$$\mathbb{E}(VX \,|\, \mathcal{G}) = V \cdot \mathbb{E}(X \,|\, \mathcal{G})$$

Proof. We show the result for indicator functions: $V = \mathbb{1}_B,\ B \in \mathcal{G}$. Then the result will hold for simple and general measurable functions by MCT.

First note that, since V is $\mathcal{G}$-measurable and $\mathbb{E}(X \mid \mathcal{G})$ is $\mathcal{G}$-measurable, then $V \cdot \mathbb{E}(X \mid \mathcal{G})$ is $\mathcal{G}$-measurable as well.

Next note that, for $A \in \mathcal{G}$,

$$\int_A VX \,\mathrm{d}\mathbb{P} = \int_A V \cdot \mathbb{E}(X \mid \mathcal{G}) \,\mathrm{d}\mathbb{P}$$
$$\int_A \mathbb{1}_B \cdot X \,\mathrm{d}\mathbb{P} = \int_A \mathbb{1}_B \cdot \mathbb{E}(X \mid \mathcal{G}) \,\mathrm{d}\mathbb{P}$$
$$\int_{A\cap B} X \,\mathrm{d}\mathbb{P} = \int_{A\cap B} \mathbb{E}(X \mid \mathcal{G}) \,\mathrm{d}\mathbb{P}$$

And the result follows from the fact that $A \cap B \in \mathcal{G}$.

□

Remark. **Important:** note that the assumption $\mathbb{E}|VX| < \infty$ is necessary so that the CE is well-defined.

Theorem 4.5. Let X, Y be r.v.'s with $X \leq Y$ a.s. and $\mathbb{E}|X| < \infty$, $\mathbb{E}|Y| < \infty$. Then:

$$\mathbb{E}(X \mid \mathcal{G}) \leq \mathbb{E}(Y \mid \mathcal{G}) \text{ a.s.}$$

Proof. Fix $\epsilon > 0$. Let $A = \{\mathbb{E}(X \mid \mathcal{G}) - \mathbb{E}(Y \mid \mathcal{G}) > \epsilon\}$. Note A is $\mathcal{G}$-measurable so we can write:

$$\int_A \mathbb{E}(X \mid \mathcal{G}) \,\mathrm{d}\mathbb{P} = \int_A X \,\mathrm{d}\mathbb{P} \leq \int_A Y \,\mathrm{d}\mathbb{P} = \int_A \mathbb{E}(Y \mid \mathcal{G}) \,\mathrm{d}\mathbb{P}$$

Where the center inequality follows because $X \leq Y$ a.s., which shows that A must have measure zero. This is true for any $\epsilon > 0$, so $\mathbb{E}(X \mid \mathcal{G}) \leq \mathbb{E}(Y \mid \mathcal{G})$ a.s.

□

Theorem 4.6. Let $X_n \geq 0$ be a sequence of r.v.'s with $X_n \uparrow X$ and $\mathbb{E}|X| < \infty$. Then

$$\mathbb{E}(X_n \mid \mathcal{G}) \uparrow \mathbb{E}(X \mid \mathcal{G})$$

Remark. If we consider $X_n = Y_1 - Y_n$ for random variables Y_n with $Y_n \downarrow Y$ and $\mathbb{E}|Y_1|, \mathbb{E}|Y| < \infty$, then using linearity we can obtain the corresponding result $\mathbb{E}(Y_n \mid \mathcal{G}) \downarrow \mathbb{E}(Y \mid \mathcal{G})$.

Also, compare this to the similar-looking result given by the Levy zero-one law (convergence theorems).

Proof. Let $A \in \mathcal{G}$. Define $Z_n = \mathbb{E}(X_n \mid \mathcal{G})$. Note that Z_n is measurable so that Z_∞ is also measurable. Also note that $Z_n \uparrow$, so that by MCT we have:

1. $\int_A Z_n \,\mathrm{d}\mathbb{P} = \int Z_n \mathbb{1}_A \,\mathrm{d}\mathbb{P} \uparrow \int Z_\infty \mathbb{1}_A \,\mathrm{d}\mathbb{P}$
2. $\int_A Z_n \,\mathrm{d}\mathbb{P} = \int X_n \mathbb{1}_A \,\mathrm{d}\mathbb{P} \uparrow \int X_\infty \mathbb{1}_A \,\mathrm{d}\mathbb{P}$

□

Theorem 4.7. Let X be independent of $\mathcal{G}$. Then $\mathbb{E}(X \,|\, \mathcal{G}) = \mathbb{E}X$.

Lemma 4.8. Let X be independent of $\mathcal{G}$ and let $g : \mathbb{R} \to \mathbb{R}$ be such that $\mathbb{E}|g(X)| < \infty$. Let Y be a r.v. which is $\mathcal{G}$-measurable with $\mathbb{E}|Y| < \infty$ and $\mathbb{E}|g(X) \cdot Y| < \infty$. Then

$$\mathbb{E}(g(X) \cdot Y) = \mathbb{E}g(X) \cdot \mathbb{E}Y$$

Proof. **RESOLVE THE BELOW QUESTIONS FIRST:**

QUESTION 1: If X is independent of $\mathcal{G}$ and Y is $\mathcal{G}$-measurable, then that implies that X is independent of Y, right?

QUESTION 2: In the above lemma, the requirement on g is that $\mathbb{E}|g(X)| < \infty$. Is this equivalent to saying that g is a bounded measurable function? If not, then why did we require that functions be bounded and measurable in the previous section on independence?

Theorem 4.9. Let $X : \Omega \to S_1$ and $Y : \Omega \to S_2$ by independent. Also let $\phi : S_1 \times S_2 \to \mathbb{R}$ be such that $\mathbb{E}|\phi(X, Y)| < \infty$. Then,

$$\mathbb{E}(\phi(X, Y) \,|\, X) = g(X)$$

where g is defined by $g(x) = \mathbb{E}(\phi(x, Y))$.

Proof. **READ UP ON PRODUCT SPACES FIRST**

Theorem 4.10. (Tower property)

Let $\mathcal{G} \subset \mathcal{H}$ be sub-σ-fields of $\mathcal{F}$. Then,

$$\mathbb{E}\Big[\mathbb{E}(X \,|\, \mathcal{H}) \,|\, \mathcal{G}\Big] = \mathbb{E}(X \,|\, \mathcal{G})$$

Remark. The undergraduate analogue to this property is:

$$\mathbb{E}\Big[\mathbb{E}(X \,|\, (Y, Z)) \,|\, Y\Big] = \mathbb{E}(X \,|\, Y)$$

Additionally, if $\mathcal{G} = \{\emptyset, \Omega\}$, then $\mathcal{G} \subset \mathcal{H}$ obviously and:

$$\mathbb{E}\Big[\mathbb{E}(X \,|\, \mathcal{H}) \,|\, \mathcal{G}\Big] = \mathbb{E}(X \,|\, \{\emptyset, \Omega\}) = \mathbb{E}X$$

Proof. $\mathcal{G}$-measurability is obvious. Let $A \in \mathcal{G}$. We want to show:

$$\int_A \mathbb{E}(X \,|\, \mathcal{H}) \, \mathrm{d}\mathbb{P} = \int_A \mathbb{E}(X \,|\, \mathcal{G}) \, \mathrm{d}\mathbb{P}$$

This follows from the definition of conditional expectation and the fact that if $A \in \mathcal{G}$, then $A \in \mathcal{H}$. Thus:

$$\int_A \mathbb{E}(X \,|\, \mathcal{H}) \, \mathrm{d}\mathbb{P} = \int_A X \, \mathrm{d}\mathbb{P} = \int_A \mathbb{E}(X \,|\, \mathcal{G}) \, \mathrm{d}\mathbb{P}$$

□

Corollary 4.11. Keeping $\mathcal{G} \subset \mathcal{H}$, we also have:

$$\mathbb{E}\Big[\mathbb{E}(X \,|\, \mathcal{G}) \,|\, \mathcal{H}\Big] = \mathbb{E}(X \,|\, \mathcal{G})$$

Proof. $\mathbb{E}(X \,|\, \mathcal{G})$ is $\mathcal{H}$-measurable since $\mathbb{E}(X \,|\, \mathcal{G})$ is $\mathcal{G}$-measurable and $\mathcal{G} \subset \mathcal{H}$. □

Theorem 4.12. (Two important inequalities)

1. $\big|\mathbb{E}(X \,|\, \mathcal{G})\big| \leq \mathbb{E}(|X| \,|\, \mathcal{G})$
2. Let I be an open set in $\mathbb{R}$ and let $\phi : I \to \mathbb{R}$ be convex. Let X be a random variable that only takes values in I and such that $\mathbb{E}(\phi(X)) < \infty$. Then:

$$\phi\big[\mathbb{E}(X \,|\, \mathcal{G})\big] \leq \mathbb{E}(\phi(X) \,|\, \mathcal{G})$$

Proof. To prove the first, note that $-|X| \leq X \leq |X|$. Take conditional expectations.

The proof of the second is exactly analogous to the proof of the unconditional Jensen's inequality. □

4.2 L^2 interpretation of conditional expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Define $L^2(\Omega, \mathcal{F}, \mathbb{P}) = \{Y : \mathbb{E}Y^2 < \infty\}$. It is easily checked that this is a Hilbert space with the inner product

$$\langle X,\ Y\rangle = \mathbb{E}(XY)$$

The corresponding space generated by the sub-σ-field $\mathcal{G}$, $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed linear subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Furthermore, if $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, then $\mathbb{E}(X \,|\, \mathcal{G})$ is the **projection of X onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$.**

This suggests two interesting facts:

Theorem 4.13. $\langle \mathbb{E}(X \,|\, \mathcal{G}),\ X - \mathbb{E}(X \,|\, \mathcal{G})\rangle = 0$

Proof. Let $Z = \mathbb{E}(X \,|\, \mathcal{G})$. Note that $\mathbb{E}\Big[\mathbb{E}(Z(X-Z) \,|\, \mathcal{G})\Big] = \mathbb{E}(Z(X-Z))$.

We show that $\mathbb{E}(Z(X-Z) \,|\, \mathcal{G}) = 0$ so that the entire LHS expectation is 0. Since Z is $\mathcal{G}$-measurable, then:

$$\begin{aligned}
\mathbb{E}(Z(X-Z) \,|\, \mathcal{G}) &= Z \cdot \mathbb{E}(X - Z \,|\, \mathcal{G}) \\
&= Z \cdot \Big[\mathbb{E}(X \,|\, \mathcal{G}) - \mathbb{E}(Z \,|\, \mathcal{G})\Big] \\
&= Z \cdot \Big[\mathbb{E}(X \,|\, \mathcal{G}) - Z\Big] \\
&= 0
\end{aligned}$$

□

Theorem 4.14. $\mathbb{E}(X \,|\, \mathcal{G}) = \arg\min\{\mathbb{E}(X - V)^2 : V \text{ is } \mathcal{G}\text{-measurable}\}$

Proof. Let $Z = \mathbb{E}(X \,|\, \mathcal{G})$. Note that

$$\begin{aligned}\mathbb{E}(X - V)^2 &= \mathbb{E}\Big[(X - Z + Z - V)^2\Big] \\ &= \mathbb{E}(X - Z)^2 + \mathbb{E}(Z - V)^2 - 2\mathbb{E}\Big[(X - Z)(Z - V)\Big]\end{aligned}$$

If we can show that the cross term is zero, then:

$$\mathbb{E}(X - V)^2 = \mathbb{E}(X - Z)^2 + \mathbb{E}(Z - V)^2 \geq \mathbb{E}(X - Z)^2$$

Note that:

$$\mathbb{E}\Big[(X - Z)(Z - V)\Big] = \mathbb{E}\Big[Z(X - Z)\Big] - \mathbb{E}\Big[V(X - Z)\Big]$$

By the previous theorem, the first term in the RHS is zero. For the second term,

$$\begin{aligned}\mathbb{E}\Big[V(X - Z)\Big] &= \mathbb{E}\Big[\mathbb{E}(V(X - Z) \,|\, \mathcal{G})\Big] \\ &= \mathbb{E}\Big[V \cdot \mathbb{E}(X - Z \,|\, \mathcal{G})\Big] \quad V \text{ is } \mathcal{G}\text{-measurable}\end{aligned}$$

Noting that $\mathbb{E}(X - Z \,|\, \mathcal{G}) = \mathbb{E}(X \,|\, \mathcal{G}) - \mathbb{E}(Z \,|\, \mathcal{G})$ and observing that Z is $\mathcal{G}$-measurable shows that this term is zero.

□

Definition. (Conditional variance)

The **conditional variance** of X given $\mathcal{G}$ is

$$\mathrm{Var}(X \,|\, \mathcal{G}) = \mathbb{E}\Big(\big[X - \mathbb{E}(X \,|\, \mathcal{G})\big]^2 \,|\, \mathcal{G}\Big)$$

Theorem 4.15. $\mathrm{Var}(X) = \mathbb{E}\Big[\mathrm{Var}(X \,|\, \mathcal{G})\Big] + \mathbb{E}\Big[\big(\mathbb{E}(X \,|\, \mathcal{G}) - \mathbb{E}X\big)^2\Big]$

Proof.

$$\mathbb{E}\Big(X - \mathbb{E}X\Big)^2 = \mathbb{E}\Big(X - \mathbb{E}(X \,|\, \mathcal{G}) + \mathbb{E}(X \,|\, \mathcal{G}) - \mathbb{E}X\Big)^2$$

□

4.3 Martingale basics

Definition. (adapted random variables)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $G \subset \mathcal{F}$ be a sub-σ-field. Then we say a random variable X is **adapted to** $\mathcal{G}$ if X is measurable w.r.t. $\mathcal{G}$.

Definition. (Martingale)

Fix a filtration of $\{\mathcal{F}_n\}_{n\geq 1}$. Let $\{X_n\}_{n\geq 1}$ be a sequence of r.v.'s.

$\{X_n\}$ is a **martingale** with respect to $\{\mathcal{F}_n\}$ if:

1. $\mathbb{E}|X_n| < \infty \ \forall n$
2. X_n is adapated to $\mathcal{F}_n$ (i.e. $\mathcal{F}_n \subset \sigma(X_1, \ldots, X_n)$)
3. $\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) = X_n$ a.s. $\forall n$

If $\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) \geq X_n$ a.s. $\forall n$, then $\{X_n\}$ is a **submartingale**

If $\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) \leq X_n$ a.s. $\forall n$, then $\{X_n\}$ is a **supermartingale**

Example. Suppose $\xi_1, \xi_2, \ldots$ are indepedent. Let $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$. Define

$$X_n = \xi_1 + \ldots + \xi_n$$

Assum $\mathbb{E}|\xi_i| < \infty$ for all $i \geq 1$. *When is this a martingale?*

1. $\mathbb{E}|X_n| < \infty$ by the triangle inequality.
2. X_n is adapted to $\mathcal{F}_n$ (Y is a deterministic function of $\xi_1, \ldots, \xi_n$)
3. To check the last martingale condition, note that:

$$\begin{aligned}
\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) &= \mathbb{E}(\xi_1 + \ldots + \xi_{n+1} \,|\, \mathcal{F}_n) \\
&= \sum^{n} \mathbb{E}(\xi_i \,|\, \mathcal{F}_n) + \mathbb{E}(\xi_{n+1} \,|\, \mathcal{F}_n) \\
&= \sum^{n} \xi_i + \mathbb{E}(\xi_{n+1}) \\
&= X_n + \mathbb{E}(\xi_{n+1})
\end{aligned}$$

Therefore if $\mathbb{E}(\xi_i) = 0$ for all i, then $\{X_n\}_{n\geq 1}$ is a martingale. If $\mathbb{E}(\xi_i) \geq 0$, then it is a submartingale. If $\mathbb{E}(\xi_i) \leq 0$, then it is a supermartingale.

Example. Let $\{\xi_i\}_{n\geq 1}$ be independent with $\mathbb{E}\xi_i = 0$ and $\mathbb{E}(\xi_i)^2 = \sigma_i^2 < \infty \ \forall i$. Let $\{\mathcal{F}_n\}_{n\geq 1}$ be a filtration given by $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$. Define

$$Q_n = X_n^2 - \sum^{n} \sigma_i^2 \quad \text{and} \quad X_n = \sum^{n} \xi_i$$

Then $\{Q_n\}_{n\geq 1}$ is a martingale.

1. $\mathbb{E}|Q_n| = 0 < \infty$ for all $n \geq 1$.
2. Q_n is $\mathcal{F}_n$-measurable.
3. We want to show that $\mathbb{E}(Q_{n+1} \,|\, \mathcal{F}_n) = Q_n$

Note that:

$$
\begin{aligned}
\mathbb{E}(X_{n+1}^2 \,|\, \mathcal{F}_n) &= \mathbb{E}\Big((X_n + \xi_{n+1})^2 \,|\, \mathcal{F}_n\Big) \\
&= \mathbb{E}\Big(X_n^2 + \xi_{n+1}^2 + 2X_n\xi_{n+1} \,|\, \mathcal{F}_n\Big) \\
&= X_n^2 + \mathbb{E}(\xi_{n+1}^2) + 2X_n\mathbb{E}(\xi_{n+1}) \\
&= X_n^2 + \sigma_{n+1}^2
\end{aligned}
$$

Therefore we have:

$$
\begin{aligned}
\mathbb{E}(Q_{n+1} \,|\, \mathcal{F}_n) &= \mathbb{E}(X_{n+1}^2 - \sum^{n+1} \sigma_i^2 \,|\, \mathcal{F}_n) \\
&= X_n^2 + \sigma_{n+1}^2 - \sum^{n+1} \sigma_i^2 \\
&= X_n^2 + \sum^{n} \sigma_i^2 \\
&= Q_n
\end{aligned}
$$

Remark. If the ξ_i's are iid with mean zero and variance 1, then:

$$
\left\{ \left(\sum^n \xi_i \right)^2 - n \right\}_{n \geq 1} \qquad \text{is a martingale}
$$

Theorem 4.16. (Two martingale inequalities)

In both inequalities, assume the martingales are w.r.t some filtration $\{\mathcal{F}_n\}_{n\geq 1}$.

1. Suppose $\{X_n\}_{n\geq 1}$ is a martingale and let ϕ be convex. If $\mathbb{E}|\phi(X_n)| < \infty$ for all n, then $\{\phi(X_n)\}_{n\geq 1}$ is a submartingale.
2. Suppose $\{X_n\}_{n\geq 1}$ is a submartingale and let ϕ be convex and increasing. Then $\{\phi(X_n)\}_{n\geq 1}$ is a submartingale.

Proof. The proof for both parts is by conditional Jensen.

1. $\mathbb{E}\Big[\phi(X_{n+1}) \,|\, \mathcal{F}_n\Big] \geq \phi\Big[\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n)\Big] = \phi(X_n)$
2. $\mathbb{E}\Big[\phi(X_{n+1}) \,|\, \mathcal{F}_n\Big] \geq \phi\Big[\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n)\Big] \geq \phi(X_n)$

Example. (A multiplicative martingale)

Suppose $\{\xi_i\}_{i\geq 1}$ are independent with $\mathbb{E}\xi_i = 1$ for all i. Let the filtration be given by $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$.

$M_n = \prod^n \xi_i$ is a martingale w.r.t. $\{\mathcal{F}_n\}$.

Showing the first two conditions is obvious. To show the third, note that:

$$
\begin{aligned}
\mathbb{E}(M_{n+1} \,|\, \mathcal{F}_n) &= \mathbb{E}(M_n \cdot \xi_{n+1} \,|\, \mathcal{F}_n) \\
&= M_n \cdot \mathbb{E}(\xi_{n+1} \,|\, \mathcal{F}_n) \\
&= M_n \cdot \mathbb{E}(\xi_{n+1}) \\
&= M_n
\end{aligned}
$$

Example. Fix $t \in \mathbb{R}$. Let $\{X_i\}_{i\geq 1}$ be independent and let $\phi_i(t) = \mathbb{E}(e^{tX_i})$. Define:

$$M_n = \frac{\exp(t \cdot \sum^n X_i)}{\prod^n \phi_i(t)}$$

M_n is a martingale.

To see this, we just need to show that $\{M_n\}_{n\geq 1}$ satisfies the conditions of the previous example. In this case,

$$\xi_i = \frac{\exp(t\xi_i)}{\phi_i(t)} \quad \Rightarrow \quad \mathbb{E}\xi_i = 1$$

Example. (Likelihood ratio martingale)

Let $\xi_1, \xi_2, \ldots$ be iid and fix a filtration $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$. Also let f, g be densities and let the likelihood ratio be given by:

$$L_n(\xi_1, \ldots, \xi_n) = \frac{g(\xi_1) \cdot \ldots \cdot g(\xi_n)}{f(\xi_1) \cdot \ldots \cdot f(\xi_n)}$$

1. **Case 1:** ξ_i has density f

   $\{L_n\}_{n\geq 1}$ is a martingale. It is easily checked that L_n is $\mathcal{F}_n$-measurable and has finite expected value (the joint density factors by independence). To check the last condition, note:

$$\begin{aligned}
\mathbb{E}(L_{n+1} \mid \mathcal{F}_n) &= \mathbb{E}\left[\left(\prod^n \frac{g(\xi_i)}{f(\xi_i)}\right) \cdot \frac{g(\xi_{n+1})}{f(\xi_{n+1})} \;\Big|\; \mathcal{F}_n\right] \\
&= \prod^n \frac{g(\xi_i)}{f(\xi_i)} \cdot \mathbb{E}\left[\frac{g(\xi_{n+1})}{f(\xi_{n+1})}\right] \\
&= \prod^n \frac{g(\xi_i)}{f(\xi_i)} \cdot 1 \quad (\xi_{n+1} \text{ has density } f) \\
&= L_n
\end{aligned}$$

2. **Case 2:** ξ_i has density g

   If $\mathbb{E}|L_n| < \infty$, then $\{L_n\}_{n\geq 1}$ is a submartingale.

   To see this, note that $\{1/L_n\}_{n\geq 1}$ is a martingale and the function $\phi(L_n) = 1/L_n$ is convex.

Definition. (Alternate definition of martingale)

Let $\{\mathcal{F}_n\}_{n\geq 0}$ be a filtration and let $\{X_n\}_{n\geq 0}$ be a sequence of random variables. Define:

$$\Delta_n = X_n - X_{n-1}, \quad n \geq 1$$

$\{X_n\}$ is a **martingale** with respect to $\{\mathcal{F}_n\}$ if:

1. $\mathbb{E}|X_0| < \infty$ and $\mathbb{E}|\Delta_n| < \infty, \quad \forall n \geq 1$.
2. X_0 is $\mathcal{F}_0$-measurable and Δ_n is $\mathcal{F}_n$-measurable $\forall n \geq 1$.

3. $\mathbb{E}(\Delta_n \,|\, \mathcal{F}_{n-1}) = 0$ a.s. $\forall n$

If $\mathbb{E}(\Delta_n \,|\, \mathcal{F}_{n-1}) \geq 0$ a.s. $\forall n$, then $\{X_n\}$ is a **submartingale**

If $\mathbb{E}(\Delta_n \,|\, \mathcal{F}_{n-1}) \leq 0$ a.s. $\forall n$, then $\{X_n\}$ is a **supermartingale**

Definition. (Predictable process)

$\{H_n\}_{n\geq 1}$ is a **predictable process with respect to filtration** $\{\mathcal{F}_n\}_{n\geq 0}$ if, for all $n \geq 1$, H_n is $\mathcal{F}_{n-1}$-measurable.

Example. Let H_n be the amount of stock bought on day $n-1$ at price x_{n-1}, and sold on day n at price x_n. The amount of total profit or loss from this action is $H_n(x_n - x_{n-1})$.

> *Remark.* Note that, for a process $\{X_n\}_{n\geq 0}$, if we have X_0 and $\Delta_1, \Delta_2, \ldots$ where $\Delta_n = X_n - X_{n-1}$, then
>
> $$X_n = X_0 + \sum_{j=1}^{n} \Delta_n$$

Lemma 4.17. (Doob decomposition)

Let $\{X_n\}_{n\geq 0}$ be a sequence of r.v.'s adapted to filtration $\{\mathcal{F}_n\}_{n\geq 0}$ with $\mathbb{E}|X_n| < \infty$ for all $n \geq 0$. Let $\Delta_n^X = X_n - X_{n-1}$.

Define $\{Y_n\}_{n\geq 0}$ by:

$$Y_0 = X_0, \quad \Delta_n^Y = \Delta_n^X - \mathbb{E}(\Delta_n^X \,|\, \mathcal{F}_{n-1})$$

Define $\{Z_n\}_{n\geq 0}$ by:

$$Z_0 = 0, \quad \Delta_n^Z = \mathbb{E}(\Delta_n^X \,|\, \mathcal{F}_{n-1})$$

Then the following hold:

1. $X_n = Y_n + Z_n$
2. $\{Y_n\}_{n\geq 1}$ is a martingale.
3. $\{Z_n\}_{n\geq 1}$ is a predictable process.

> *Proof.* Z_n is $\mathcal{F}_{n-1}$-measurable by defintion of conditional expectation $\forall n$, and therefore is a predictable process.

□

4.4 Discrete stochastic integral

Lemma 4.18. Recall that, if $\{\mathcal{F}_n\}_{n\geq 0}$ is a filtration, then:

$$\mathcal{F}_\infty = \sigma\left(\cup_{n=0}^{\infty} \mathcal{F}_n\right)$$

If T is a stopping time, then $\{T = \infty\} \in \mathcal{F}_\infty$.

Proof. $\{T = \infty\} = \cap_{n=0}^{\infty}\{T > n\}$. Note that $\{T > n\} = \{T \leq n\}^c \in \mathcal{F}_n$ for each n. Since $\mathcal{F}_\infty$ is the σ-field generated by the union of these $\mathcal{F}_n$'s, then it is closed under countable intersection. Thus $\{T = \infty\} \in \mathcal{F}_\infty$. □

Definition. (σ-field up to time T)

Suppose T is a stopping time. The **σ-field up to time T** is:

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T = n\} \in \mathcal{F}_n,\ n \geq 0\}$$

Proof. We show the three conditions to be a σ-field are met:

1. $\Omega \in \mathcal{F}_T$: obvious.
2. Closed under complements: $A^c \cap \{T = n\} = \{T = n\} \setminus [A \cap \{T = n\}]$
3. Closed under countable union: $\cup^\infty A_i \cap \{T = n\} = \cup^\infty [A_i \cap \{T = n\}]$

□

Corollary 4.19. Note that the σ-field up to time T can also be characterized by:

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq n\} \in \mathcal{F}_n,\ n \geq 0\}$$

Proof. Note that the requirement that $\{T = n\} \in \mathcal{F}_n$ for a stopping time T is equivalent to the requirement that $\{T \leq n\} \in \mathcal{F}_n$.

Theorem 4.20. (Stopping time fact, part 1)

Let T be a stopping time and let $\{X_n\}_{n\geq 0}$ be a sequence of r.v.'s adapted to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$. Define:

$$Y = \begin{cases} X_T & \text{if } T < \infty \\ 0 & \text{if } T = \infty \end{cases}$$

Then Y is $\mathcal{F}_T$-measurable.

Proof. Fix $B \in B(\mathbb{R})$. We need to check that $\{Y \in B\} \in \mathcal{F}_T$. Naturally we split the problem into two cases:

1. $T < \infty$:

   Note that the condition $\{Y \in B\} \in \mathcal{F}_T$ is equivalent to the condition that $\{Y \in B\} \cap \{T = n\} \in \mathcal{F}_n$ for all $n \geq 0$. Then note that

   $$\{Y \in B\} \cap \{T = n\} = \{X_n \in B\} \cap \{T = n\}$$

   and that $\{X_n\}_{n\geq 0}$ is adapted to the filtration $\{\mathcal{F}_n\}_{n\geq 0}$, so that $\{X_n \in B\} \in \mathcal{F}_n$, and that T is a stopping time.

2. $T = \infty$:

   Note that the condition $\{Y \in B\} \in \mathcal{F}_T$ is equivalent to the condition that $\{Y \in B\} \cap \{T = \infty\} \in \mathcal{F}_\infty$.

   Since $Y = 0$, then $\{Y \in B\}$ is either Ω or $\emptyset$ for any $B \in B(\mathbb{R})$, which is automatically in $\mathcal{F}_\infty$. Also, by our lemma above, $\{T = \infty\} \in \mathcal{F}_\infty$.

□

Theorem 4.21. (Stopping time fact, part 2)

Suppose S and T are stopping times with respect to the same filtration **and** with $S \leq T$. Then $\mathcal{F}_S \subset \mathcal{F}_T$.

Proof. Let $A \in \mathcal{F}_S$. We show that $A \in \mathcal{F}_T$ also.

By the definition of σ-field up to time S, we have

$$A \cap \{S \leq n\} \in \mathcal{F}_n \quad \text{for all } n \geq 0$$

Then since $S \leq T$, we have:

$$A \cap \{T \leq n\} = A \cap \{S \leq n\} \cap \{T \leq n\}$$

Noting $\{T \leq n\} \in \mathcal{F}_n$ completes the proof.

□

Theorem 4.22. (Stopping time fact, part 3)

Suppose S and T are stopping times with respect to the same filtration. Then $\{S = T\} \in \mathcal{F}_S \cap \mathcal{F}_T$.

Proof. Recall that if S and T are stopping times, then $\min(S,T)$ is also a stopping time. Furthermore, by the definition of σ-field after time $\min(S,T)$,

$$\begin{aligned}\mathcal{F}_{\min(S,T)} &= \{A \in \mathcal{F}_\infty : A \cap \{\min(S,T) \leq n\} \in \mathcal{F}_n,\ n \geq 0\} \\ &= \mathcal{F}_S \cap \mathcal{F}_T\end{aligned}$$

Now note that: $\{S = T\} = \cup_{n=0}^{\infty}\Big\{\{S = n\} \cap \{T = n\}\Big\}$.

Clearly, $\{S = n\} \cap \{T = n\} \in \mathcal{F}_S \cap \mathcal{F}_T$ for fixed n. But since the intersection of two σ-fields is again a σ-field, then $\mathcal{F}_S \cap \mathcal{F}_T$ is closed under countable union, so that

$$\{S = T\} = \cup_{n=0}^{\infty}\Big\{\{S = n\} \cap \{T = n\}\Big\} \in \mathcal{F}_S \cap \mathcal{F}_T$$

□

Definition. (Process stopped at time n)

Let T be a stopping time and let $\{X_n\}_{n \geq 0}$ be a sequence of r.v.'s adapted to a filtration $\{\mathcal{F}_n\}_{n \geq 0}$. Define $T \wedge n = \min(T, n)$ for some fixed n.

We call the sequence of r.v.'s $\{X_{T \wedge n} : n \geq 0\}$ a **process stopped at time** n.

Example. Let $\{X_n\}$ be as above with $T = \inf\{n \geq 0 : X_n \geq 30\}$.

$\{X_{T \wedge n}\}_{n \geq 0}$ will evolve as $\{X_n\}$ until n reaches $T = t_0$, the first time at which X_n goes above 30. Then $\{X_{T \wedge n}\}$ will take the value X_{t_0} (i.e. the value $\{X_n\}$ takes when it goes above 30 for the first time) for the rest of its lifetime.

Definition. (Discrete stochastic integral)

Let $\{\mathcal{F}_n\}_{n \geq 0}$ be a filtration, let $\{X_n\}_{n \geq 0}$ be a process adapted to that filtration, and let $\{H_n\}_{n \geq 1}$ be a predictable process.

The sequence $\{Y_n\}_{n\geq 0}$ defined by $Y_0 = 0$ and:

$$\begin{aligned} \Delta_n^Y &= Y_n - Y_{n-1} = H_n \Delta_n^X = H_n(X_n - X_{n-1}) \\ Y_n &= Y_0 + \sum_{j=1}^{n} \Delta_n^Y = \sum_{j=1}^{n} H_j(X_j - X_{j-1}) \end{aligned}$$

is called the **discrete stochastic integral**, denoted $Y = HX$ or $Y_n = (HX)_n$.

Theorem 4.23. (Martingale properties of discrete stochastic integrals)

1. If $\{X_n\}_{n\geq 0}$ is a martingale and $\{H_n\}_{n\geq 1}$ is bounded and predictable, then $\{Y_n\}_{n\geq 1}$ is a martingale also.
2. If $\{X_n\}_{n\geq 0}$ is a sub(super)martingale and $\{H_n\}_{n\geq 1}$ is bounded, predictable and non-negative, then $\{Y_n\}_{n\geq 1}$ is a sub(super)martingale also.

Remark. If we interpret the predictable process $\{H_n\}$ as a betting strategy, then the fact that $\{Y_n\}$ must be a martingale tells us that we cannot come up with a betting strategy that will "consistently" make a net profit/loss so long as $\{X_n\}$ is a martingale.

Proof. Checking the first two conditions (in the alternate definition of a martingale) is trivial.

For the third condition, consider the case of a martingale:

$$\begin{aligned} \mathbb{E}(\Delta_n^Y \,|\, \mathcal{F}_{n-1}) &= \mathbb{E}(H_n(X_n - X_{n-1}) \,|\, \mathcal{F}_{n-1}) \\ &= H_n \cdot \mathbb{E}(X_n - X_{n-1} \,|\, \mathcal{F}_{n-1}) \\ &= 0 \end{aligned}$$

Where we can pull out H_n since it is bounded, so that the expectation of $H_n(X_n - X_{n-1})$ is finite. The argument for a sub(super)martingale is exactly analogous, except for the additional requirement that $H_n \geq 0$ so that the sign of the expected value does not switch.

□

Theorem 4.24. Suppose $\{X_n\}_{n\geq 0}$ is a (sub)martingale and T is a stopping time. Then $\{X_{T\wedge n}\}_{n\geq 0}$ is also a (sub)martingale.

Proof. Let $H_k = \mathbb{1}_{k\leq T}$. Thus is the strategy where, every day until day T, you buy 1 unit of stock and sell it immediately the next day.

Since $\{X_n\}_{n\geq 0}$ is a (sub)martingale, then the discrete stochastic integral $\{Y_n\}_{n\geq 0}$ defined by

$$Y_n = (HX)_n = \begin{cases} \sum_{k=1}^{n} H_k(X_k - X_{k-1}), & n \geq 1 \\ 0, & n = 0 \end{cases}$$

is a (sub)martingale if $\{H_k\}_{n\geq 0}$ is predictable, bounded, and non-negative. Clearly it is bounded and non-negative. To see that it is predictable, note that $\{T \geq k\} = \{T \leq k-1\}^c \in \mathcal{F}_{k-1}$.

Since $H_k = \mathbb{1}_{k \le T}$, then Y_n is a sum of Δ_k^X from $k = 0$ to $k = n$ if $T \ge n$. Similarly, Y_n is a sum of Δ_k^X from $k = 0$ to $k = T$ if $T < n$. In other words,

$$Y_n = \begin{cases} X_n - X_0, & T \ge n \\ X_T - X_0, & T < n \end{cases}$$

But this is precisely the definition of $X_{T \wedge n} - X_0$, and we already shown that $\{Y_n\}_{n \ge 0}$ is a (sub)martingale.

□

Proof. (Alternate)

This proof does not use the discrete stochastic integral. We show only the conditional expectation property of a martingale:

$$\begin{aligned} \mathbb{E}\Big[X_{(n+1) \wedge T} \,|\, \mathcal{F}_n\Big] &= \mathbb{E}\Big[X_T \cdot \mathbb{1}_{T \le n} + X_{n+1} \cdot \mathbb{1}_{T > n} \,|\, \mathcal{F}_n\Big] \\ &= X_T \cdot \mathbb{1}_{T \le n} + \mathbb{1}_{T > n} \cdot \mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) \\ &= X_T \cdot \mathbb{1}_{T \le n} + X_n \cdot \mathbb{1}_{T > n} \\ &= X_{n \wedge T} \end{aligned}$$

□

4.5 Optional sampling theorems

We begin by noting two simple facts:

1. (Follows from tower property)
   (a) If $\{X_n\}_{n \ge 0}$ is a martingale, then $\mathbb{E}X_0 = \mathbb{E}X_1 = \dots$.
   (b) If $\{X_n\}_{n \ge 0}$ is a submartingale, then $\mathbb{E}X_0 \le \mathbb{E}X_1 \le \dots$.
   (c) If $\{X_n\}_{n \ge 0}$ is a supermartingale, then $\mathbb{E}X_0 \ge \mathbb{E}X_1 \ge \dots$.
2. If $\{X_n\}_{n \ge 0}$ is a submartingale, then $Z_n = -X_n \;\Rightarrow\; \{Z_n\}_{n \ge 0}$ is a supermartingale

Theorem 4.25. (Optional sampling theorem)

Let $\{X_n\}_{n \ge 0}$ be a submartingale and let T_1, T_2 be stopping times such that

1. $T_1 \le T_2$ a.s.
2. $T_1 < \infty$, $T_2 < \infty$

Under certain conditions $(*)$, then:

$$\mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1}) \ge X_{T_1} \quad \text{and so} \quad \mathbb{E}X_{T_2} \ge \mathbb{E}X_{T_1}$$

and if $\{X_n\}_{n \ge 0}$ is a martingale, then equality holds:

$$\mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1}) = X_{T_1} \quad \text{and so} \quad \mathbb{E}X_{T_2} = \mathbb{E}X_{T_1}$$

Theorem 4.26. (”Bounded” OST)

If $T_1 \le T_2 \le t_0$ where t_0 is a constant, then the OST holds.

Proof. Fix $A \in \mathcal{F}_{T_1}$. It is sufficient to show:

$$\int_A \mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1})\,\mathrm{d}\mathbb{P} \ge \int_A X_{T_1}\,\mathrm{d}\mathbb{P} \quad \forall A \in \mathcal{F}_{T_1}$$

Our approach is to construct a martingale using a discrete stochastic integral with an appropriate predictable process.

We find our predictable process by considering the stock-buying strategy of buying 1 unit of stock at the end of day T_1 and continuing to buy each day up to day T_2 if the event A happens:

$$H_k = \mathbb{1}_A \cdot \mathbb{1}_{T_1 < k \le T_2}, \quad k \ge 1$$

To see that $\{H_k\}_{k\ge 0}$ is predictable, write $H_k = \mathbb{1}_{A \cap \{T_1 \le k-1\}} \cdot \mathbb{1}_{T_2 \ge k}$ and note that the first indicator function is $\mathcal{F}_{k-1}$-measurable by the definition of $\mathcal{F}_{T_1}$ (since $A \in \mathcal{F}_{T_1}$).

Then the discrete stochastic integral $\{Y_n\}_{n\ge 0}$ is given by $Y_0 = 0$ and:

$$\begin{aligned} Y_n &= \sum_{k=1}^{n} H_k (X_k - X_{k-1}) \\ &= (X_{T_2 \wedge n} - X_{T_1 \wedge n}) \cdot \mathbb{1}_A \end{aligned}$$

Therefore since H_k is bounded and non-negative and $\{X_n\}_{n\ge 0}$ is a submartingale, then $\{Y_n\}_{n\ge 0}$ is also a submartingale. Thus since $\mathbb{E}Y_0 = 0$, then we have that for all $n > 0$,

$$\mathbb{E}Y_n \ge 0 \quad \text{and so} \quad \mathbb{E}\Big[(X_{T_2 \wedge n} - X_{T_1 \wedge n})\mathbb{1}_A\Big] \ge 0$$

The proof is completed by taking $n = t_0$ and applying the definition of CE to the LHS.

□

Example. (Counterexample: simple random walk)

To illustrate the importance of **bounded** stopping times in this theorem, consider an iid sequence $\{\xi_i\}_{i\ge 1}$ such that

$$\mathbb{P}(\xi_i = +1) = \frac{1}{2}, \quad \mathbb{P}(\xi_i = -1) = \frac{1}{2}$$

Let a filtration $\{\mathcal{F}_n\}$ be given by $\mathcal{F}_n = \sigma(\xi_1, \ldots \xi_n)$ and define $\{S_n\}_{n\ge 0}$ by:

$$S_0 = 1, \quad S_n = 1 + \sum_{i=1}^{n} \xi_i$$

$\{S_n\}_{n\ge 0}$ is obviously adapted to $\{\mathcal{F}_n\}$, $\mathbb{E}S_n = 1 < \infty$ for all n, and:

$$\mathbb{E}(S_{n+1} \,|\, \mathcal{F}_n) = 1 + \mathbb{E}\left(\sum_{i=1}^{n+1} \xi_i \,|\, \mathcal{F}_n\right) = 1 + \sum_{i=1}^{n} \xi_i + \mathbb{E}(\xi_{n+1} \,|\, \mathcal{F}_n)$$

so that it is a martingale. Now consider the stopping time $T = \inf\{n \geq 1 : S_n = 0\}$. Note that $\mathbb{P}(T < \infty) = 1$ but there does **not** exist $t_0 < \infty$ such that $\mathbb{P}(T < t_0) = 1$.

Suppose we try to apply OST using $T_1 = 0$ and $T_2 = T$. Then we would have $\mathbb{E}S_{T_2} = \mathbb{E}S_{T_1}$, but clearly $\mathbb{E}S_{T_2} = 0$ and $\mathbb{E}S_{T_1} = 1$.

Theorem 4.27. (”Unbounded” OST)

If $T_1 \leq T_2 < \infty$ a.s. and, additionally,

$$\lim_{k\to\infty} \mathbb{E}\big(|X_{T_i} - X_{T_i \wedge k}|\big) = 0 \quad \text{for } i = 1, 2$$

then the OST holds.

Proof. Fix $k_1, k_2 \in \{0, 1, \ldots\}$ such that $k_1 < k_2$. Consider the bounded versions of T_1 and T_2 given by:

$$T_1 \wedge k_1 \leq k_1, \quad T_2 \wedge k_2 \leq k_2$$

Applying the bounded OST with $T_1 \wedge k_1$ and $T_2 \wedge k_2$, we obtain:

$$\mathbb{E}(X_{T_2 \wedge k_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) \geq X_{T_1 \wedge k_1}$$

To extend this result to the unbounded the case, we will take $k_2 \to \infty$ first and then $k_1 \to \infty$. To justify taking $k_2 \to \infty$, we show L^1 convergence of $\mathbb{E}(X_{T_2 \wedge k_2} \,|\, \mathcal{F}_{T_1 \wedge k_1})$ to $\mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1 \wedge k_1})$:

$$\begin{aligned}
\mathbb{E}\Big[\big|\mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) - \mathbb{E}(X_{T_2 \wedge k_2} \,|\, \mathcal{F}_{T_1 \wedge k_1})\big|\Big] \\
&= \mathbb{E}\Big[\big|\mathbb{E}(X_{T_2} - X_{T_2 \wedge k_2} \,|\, \mathcal{F}_{T_1 \wedge k_1})\big|\Big] \\
&\leq \mathbb{E}\Big[\mathbb{E}(|X_{T_2} - X_{T_2 \wedge k_2}| \,|\, \mathcal{F}_{T_1 \wedge k_1})\Big] \\
&\qquad \text{(Conditional Jensen)} \\
&= \mathbb{E}(|X_{T_2} - X_{T_2 \wedge k_2}|) \to 0 \text{ as } k_2 \to \infty
\end{aligned}$$

We show that this implies $\mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) \geq X_{T_1 \wedge k_1}$. Fix some $A \in \mathcal{F}$ (?) and note that:

$$\begin{aligned}
&\int \Big|\mathbb{E}(X_{T_2 \wedge k_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) - \mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1 \wedge k_1})\Big| \, \mathrm{d}\mathbb{P} \to 0 \text{ as } k_2 \to \infty \\
\Rightarrow &\int \Big|\mathbb{E}(X_{T_2 \wedge k_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) \cdot \mathbb{1}_A - \mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) \cdot \mathbb{1}_A\Big| \, \mathrm{d}\mathbb{P} \to 0 \\
\Rightarrow &\Big|\int_A \mathbb{E}(X_{T_2 \wedge k_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) \, \mathrm{d}\mathbb{P} - \int_A \mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) \, \mathrm{d}\mathbb{P}\Big| \to 0
\end{aligned}$$

Now since $\mathbb{E}(X_{T_2 \wedge k_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) \geq X_{T_1 \wedge k_1}$, we also have that

$$\int_A \mathbb{E}(X_{T_2 \wedge k_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) \, \mathrm{d}\mathbb{P} \geq \int_A X_{T_1 \wedge k_1} \, \mathrm{d}\mathbb{P} \quad \forall k_2$$

This is true for all k_2 and since $\int_A \mathbb{E}(X_{T_2 \wedge k_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) \, \mathrm{d}\mathbb{P}$ converges to $\int_A \mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) \, \mathrm{d}\mathbb{P}$ as just shown, then the inequality must still hold for the limit:

$$\int_A \mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1 \wedge k_1}) \, \mathrm{d}\mathbb{P} \geq \int_A X_{T_1 \wedge k_1} \, \mathrm{d}\mathbb{P} \ \text{ for arbitrary } A$$

Thus we have $\mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1\wedge k_1}) \geq X_{T_1\wedge k_1}$. Equivalently, we can write this as

$$\mathbb{E}(X_{T_2} \,|\, \mathcal{F}_{T_1\wedge k_1}) \geq X_{T_1} \text{ on } \{T_1 \leq k_1\}$$

Send $k_1 \to \infty$.

□

Note that it is not obvious how to check the condition $\lim \mathbb{E}(|X_{T_i} - X_{T_i\wedge k}|) = 0$. The following lemma gives two sufficient conditions for that condition to hold:

Lemma 4.28. Let $\{X_n\}_{n\geq 0}$ be a martingale and T be a stopping time such that

1. $\mathbb{E}(|X_k| \cdot \mathbb{1}_{T>k}) \to 0$ a.s.
2. $\mathbb{E}(|X_T|) < \infty$

Then $\lim \mathbb{E}(|X_T - X_{T\wedge k}|) = 0$.

Proof.

$$\begin{aligned}\mathbb{E}(|X_T - X_{T\wedge k}|) &= \mathbb{E}(|X_T - X_{T\wedge k}| \cdot \mathbb{1}_{T\leq k}) + \mathbb{E}(|X_T - X_{T\wedge k}| \cdot \mathbb{1}_{T>k}) \\ &= \mathbb{E}(|X_T - X_{T\wedge k}| \cdot \mathbb{1}_{T>k}) \\ &\leq \mathbb{E}(|X_T| \cdot \mathbb{1}_{T\geq k}) + \mathbb{E}|X_k| \cdot \mathbb{1}_{T>k} \\ &\qquad \text{(by triangle inequality)}\end{aligned}$$

The rightmost term $\to 0$ by assumption, so what is left is to show that $\lim \mathbb{E}(|X_T| \cdot \mathbb{1}_{T\geq k}) = 0$.

Define $Z_k = |X_T| \cdot \mathbb{1}_{T\geq k}$. Since $\mathbb{P}(T < \infty) = 1$, then $Z_k \xrightarrow{a.s.} 0$. Also, $|Z_k| \leq |X_T|$ for all k. Thus since $\mathbb{E}(|X_T|) < \infty$ by assumption, we can apply DCT and the result follows.

□

4.6 Maximal inequalities and upcrossing lemma

Definition. (Some notation)

For a sequence of random variables $\{X_n\}_{n\geq 0}$, we define:

1. $X_N^* = \sup_{0\leq n\leq N} X_n$
2. $X^* = \sup_{n\geq 0} X_n$

In these equalities we will make clearer the connection between sub(super)martingales and increasing(decreasing) sequences. Note that, for a deterministic decreasing sequence $\{X_n\}_{n\geq 0}$,

$$\sup_{n\geq 0} X_n = X_0$$

An analogous result holds for supermartingales:

Theorem 4.29. (Supermartingale maximal inequality)

Let $\{X_n\}_{n\geq 0}$ be a supermartingale such that $X_n \geq 0$ a.s. Then

$$\lambda \cdot \mathbb{P}(X^* \geq \lambda) \leq \mathbb{E}X_0$$

Remark. Note the relation to Markov's inequality applied to the first term:

$$\lambda \cdot \mathbb{P}(X_0 \geq \lambda) \leq \mathbb{E}X_0$$

Proof. Fix $N > 0$ and $\lambda > 0$ and define $T = \inf\{n \geq 0 : X_n \geq \lambda\}$. Then:

$$\{X_N^* \geq \lambda\} = \{T \leq N\}$$

The proof is by "bounded" OST.

Note that since T is a stopping time, then $T \wedge N$ is a bounded stopping time. Let $T_1 = 0 \ \leq \ T_2 = T \wedge N$. Then by OST,

$$\begin{aligned}
\mathbb{E}X_0 &\geq \mathbb{E}X_{T\wedge N} \\
&= \mathbb{E}(X_{T\wedge N} \cdot \mathbb{1}_{T\leq N}) + \mathbb{E}(X_{T\wedge N} \cdot \mathbb{1}_{T>N}) \\
&\geq \mathbb{E}(X_{T\wedge N} \cdot \mathbb{1}_{T\leq N}) \quad \text{(since } X_n \geq 0) \\
&= \mathbb{E}(X_T \cdot \mathbb{1}_{T\leq N}) \\
&= \mathbb{E}(X_T \cdot \mathbb{1}_{X_N^* \geq \lambda}) \\
&\geq \lambda \cdot \mathbb{P}(X_N^* \geq \lambda)
\end{aligned}$$

Thus we have the inequality for any given N. Now note that as $N \to \infty$, then $\{X_N^* \geq \lambda\} \uparrow A$, some A. Now note that $A \subset \{X^* \geq \lambda\}$, so we do not have equality. However,

$$\{X^* > \lambda\} \subset A \subset \{X^* \geq \lambda\}$$

So that we now have the inequality $\lambda \cdot \mathbb{P}(X^* > \lambda) \leq \mathbb{E}X_0$. To complete the proof, take a sequence $\lambda_i \uparrow \lambda$ and note that:

$$\mathbb{P}(X^* \geq \lambda) \leq \mathbb{P}(X^* > \lambda_i) \leq \frac{\mathbb{E}X_0}{\lambda_i}$$

And let $i \to \infty$.

□

Theorem 4.30. (Doob's submartingale maximal inequality)

Let $\{X_n\}_{n\geq 0}$ is a submartingale with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$. For any $N \in \mathbb{N}$ and $\lambda > 0$, then

$$\begin{aligned}
\lambda \cdot \mathbb{P}(X_N^* \geq \lambda) &\leq \mathbb{E}\Big[X_N \cdot \mathbb{1}_{X_N^* \geq \lambda}\Big] \\
&\leq \mathbb{E}X_N^+
\end{aligned}$$

Remark. Note the relation to Markov's inequality applied to the N^{th} term:

$$\lambda \cdot \mathbb{P}(X_N^+ \geq \lambda) \leq \mathbb{E}X_N^+$$

Proof. Again, the proof is by "bounded" OST.

Define $T = \inf\{n \geq 0 : X_n \geq \lambda\}$. Let $T_2 = N$ and $T_1 = T \wedge N$. Note:

1. $T_1 \leq T_2 \leq N$
2. $\{T \leq N\} \in \mathcal{F}_{T_1}$

Where (2) follows from $\{T \leq N\} \cap \{T \wedge N \leq n\} = \{T \leq n\}$. Therefore,

$$\begin{aligned}\int_{\{T\leq N\}} \mathbb{E}(X_{T_2} \mid \mathcal{F}_{T_1})\,\mathrm{d}\mathbb{P} &= \int_{\{T\leq N\}} X_{T_2}\,\mathrm{d}\mathbb{P}\\ &\geq \int_{\{T\leq N\}} X_{T_1}\,\mathrm{d}\mathbb{P} \quad \text{(by OST)}\end{aligned}$$

Now using the fact that $\{T \leq N\} = \{X_N^* \geq \lambda\}$, we have

$$\begin{aligned}\int_{\{X_N^*\geq\lambda\}} X_N\,\mathrm{d}\mathbb{P} &\geq \int_{\{X_N^*\geq\lambda\}} X_{T\wedge N}\,\mathrm{d}\mathbb{P}\\ \int_{\{X_N^*\geq\lambda\}} X_N\,\mathrm{d}\mathbb{P} &\geq \int_{\{X_N^*\geq\lambda\}} X_T\,\mathrm{d}\mathbb{P}\\ \int X_N \cdot \mathbb{1}_{\{X_N^*\geq\lambda\}}\,\mathrm{d}\mathbb{P} &\geq \int X_T \cdot \mathbb{1}_{\{X_N^*\geq\lambda\}}\,\mathrm{d}\mathbb{P}\\ &\geq \int \lambda \cdot \mathbb{1}_{\{X_N^*\geq\lambda\}}\,\mathrm{d}\mathbb{P}\end{aligned}$$

Where the last step follows from the definition of T.

□

Remark. Using the fact that $\mathbb{P}(X \geq t) \leq \mathbb{P}(X^2 \geq t^2)$ for any r.v. X and $t > 0$, and the fact that S_n^2 is a submartingale if S_n is a martingale, this gives Kolmogorov's maximal inequality.

Theorem 4.31. (Doob's L^2 maximal inequality)

Let $\{X_n\}_{n\geq 0}$ be a submartingale with $\mathbb{E}(X_n^2) < \infty$ for all n.

For fixed $N \geq 0$,

$$\mathbb{E}\Big[(\max(X_N^*, 0))^2\Big] \leq 4 \cdot \mathbb{E}\Big[(X_N^+)^2\Big]$$

Remark. If $\{Y_n\}_{n\geq 0}$ is a martingale, then $\{X_n\}_{n\geq 0}$ defined by $X_n = |Y_n|$ is a positive submartingale and we have:

$$\mathbb{E}\Big[(\max_{0\leq n\leq N} |Y_n|)^2\Big] \leq 4 \cdot \mathbb{E}(Y_N^2)$$

Proof.

$$\begin{aligned}
\mathbb{E}\Big[(\max(X_N^*, 0))^2\Big] &= 2\int_0^\infty \lambda \cdot \mathbb{P}(X_N^* \geq \lambda)\,\mathrm{d}\lambda \\
&\leq 2\int_0^\infty \mathbb{E}\Big[X_N \cdot \mathbb{1}_{X_N^* \geq \lambda}\Big]\,\mathrm{d}\lambda \\
&\qquad \text{(by sub-MG max'l inequality)} \\
&\leq 2\int_0^\infty \mathbb{E}\Big[X_N^+ \cdot \mathbb{1}_{X_N^* \geq \lambda}\Big]\,\mathrm{d}\lambda \\
&= 2\int_0^\infty \int_\Omega X_N^+ \cdot \mathbb{1}_{X_N^* \geq \lambda}\,\mathrm{d}\mathbb{P}\,\mathrm{d}\lambda \\
&= 2\int_\Omega X_N^+ \int_0^\infty \mathbb{1}_{X_N^* \geq \lambda}\,\mathrm{d}\lambda\,\mathrm{d}\mathbb{P} \\
&\qquad \text{(by Fubini)} \\
&= 2\int_\Omega X_N^+ \cdot \max(X_N^*, 0)\,\mathrm{d}\mathbb{P} \\
&= 2\,\mathbb{E}\Big[X_N^+ \cdot \max(X_N^*, 0)\Big]
\end{aligned}$$

Apply Cauchy-Schwarz and rearrange.

□

We now introduce some motivation for the main result of this section–the upcrossing lemma. Consider the following two sequences. How many times do they "upcross" a given threshold?

1. $X_n = 1/n$

   Crosses only once for any threshold. Key observation: if a sequence converges, then the number of upcrossings will always be finite.

2. $X_{2n} = 1/2n$ and $X_{2n+1} = 1 + 1/2n$

   This sequence has two subsequences that converge to different limits. There exist thresholds (e.g. 1) where the number of upcrossings is infinite.

Definition. (Setting for upcrossing lemma)

1. Let $\{X_n\}_{n\geq 0}$ be a sequence of random variables.
2. Fix some $a < b$. Define a sequence of stopping times by:

$$\begin{aligned}
S_1 &= \inf\{n \geq 0 : X_n \leq a\} \\
T_1 &= \inf\{n \geq 0 : X_n \geq b\} \\
S_2 &= \inf\{n > T_1 : X_n \leq a\} \\
T_2 &= \inf\{n > S_1 : X_n \geq b\} \\
&\vdots
\end{aligned}$$

3. Define the number of upcrossings of $[a, b]$ up until time n by:

$$U_n[a,b] = \max\{i : T_i \le n\}$$

Theorem 4.32. (Upcrossing lemma)

Suppose that $\{X_n\}_{n\ge 0}$ is a submartingale. Then

$$\begin{aligned}\mathbb{E}\Big[U_n[a,b]\Big] &\le \frac{\mathbb{E}[(X_n-a)^+] - \mathbb{E}[(X_0-a)^+]}{b-a} \\ &\le \frac{\mathbb{E}X_n^+ + |a|}{b-a}\end{aligned}$$

Proof. Note that $U(a, b)$ for a general subMG $\{X_n\}_{n\ge0}$ is equal to $U(0, b-a)$ for the shifted and truncated subMG $\{(X_n - a)^+\}_{n\ge0}$. So WLOG assume that $X_n \ge a$ for all n (i.e. $X_{S_i} = a$) and show:

$$\mathbb{E}\Big[U_n[a,b]\Big] \le \frac{\mathbb{E}X_n - \mathbb{E}X_0}{b-a}$$

The general idea of the proof is to use the discrete stochastic integral with **two** different predictable processes, to obtain two bounds which can then be combined into our final bound:

1. "Buy low, sell high:" $H_n = \mathbb{1}_{S_1<n\le T_1} + \mathbb{1}_{S_2<n\le T_2} + \ldots$

   $\{H_n\}_{n\ge1}$ is predictable because each indicator function can be rewritten as: $\mathbb{1}_{\{S_i\le n-1\}} \cdot \mathbb{1}_{\{T_i \le n-1\}^c}$. Consider the discrete stochastic integral $Y_n = (HX)_n$:

   $$\begin{aligned}Y_n &= (X_{T_1} - X_{S_1}) + \ldots + (X_{T_{U_n}} - X_{S_{U_n}}) + (X_n - X_{S_{U_{n+1}}}) \cdot \mathbb{1}_{n>S_{U_{n+1}}} \\ &= \sum_{i=1}^{U_n} (X_{T_i} - X_{S_i}) + (X_n - X_{S_{U_{n+1}}}) \cdot \mathbb{1}_{n>S_{U_{n+1}}} \\ &\ge (b-a)\cdot U_n + (X_n - X_{S_{U_{n+1}}}) \cdot \mathbb{1}_{n>S_{U_{n+1}}} \\ &\ge (b-a)\cdot U_n \quad \text{(since } X_n \ge a = X_{S_{U_{n+1}}}\text{)}\end{aligned}$$

   Taking expectations, we have

   $$\mathbb{E}Y_n \ge (b-a)\cdot \mathbb{E}U_n$$

2. "Buy high, sell low:" $K_n = 1 - H_n$

   Define $\{Z_n\}_{n\ge0}$ by $Z_n = X_n - Y_n$. Then $Z_n = (KX)_n$. Note that $\{K_n\}_{n\ge1}$ is predictable and bounded because $\{H_n\}_{n\ge1}$ is predictable and bounded. Thus $\{Z_n\}_{n\ge0}$ is a submartingale and:

   $$\mathbb{E}Z_n \ge \mathbb{E}Z_0 = \mathbb{E}X_0 - \mathbb{E}Y_0 = \mathbb{E}X_0$$

   Therefore $\mathbb{E}X_n - \mathbb{E}Y_n \ge \mathbb{E}X_0$ and so $\mathbb{E}Y_n \le \mathbb{E}X_n - \mathbb{E}X_0$.

Combining the two conclusions yields the desired result.

□

4.7 Convergence theorems

The theorems in this section all deal with the general question of: If a sequence is a martingale, then what conditions must be imposed to guarantee convergence?

Lemma 4.33. (Deterministic lemma)

Let $\{x_n\}_{n\geq 1}$ be a sequence of real numbers. Then $\{x_n\}_{n\geq 1}$ converges if and only if the number of upcrossings $U[a,b]$ is finite for any $a < b$.

Theorem 4.34. (Martingale convergence theorem)

Let $\{X_n\}_{n\geq 0}$ be a submartingale. If $\sup_n \mathbb{E}X_n^+ < \infty$, then $X_n \xrightarrow{a.s.} X_\infty$ with $\mathbb{E}|X_\infty| < \infty$.

Proof. This proof has two parts. First we show that $X_n \xrightarrow{a.s.} X_\infty$ by the upcrossing lemma. Then we show that X_∞ has finite expectation by an application of Fatou's lemma.

1. Note that by our deterministic lemma, if X_n does not converge to X a.s., then there must exist some $a < b \in \mathbb{Q}$ such that $U_\infty[a,b] = \infty$. To show that this is not possible, write:

   $$\{\limsup X_n = \liminf X_n\} = \bigcap_{q>r} \{U_\infty[r,q] < \infty\},\ r,q \in \mathbb{Q}$$

   So fix $r, q \in \mathbb{Q}$. To show $U_\infty[r,q] < \infty$, we show that $\mathbb{E}(U_\infty[r,q]) < \infty$. To show $\mathbb{E}(U_\infty[r,q]) < \infty$, we show that $\mathbb{E}(U_n[r,q]) < \infty$ for fixed n and then apply MCT.

   By the upcrossing lemma,

   $$\begin{aligned}\mathbb{E}\Big(U_n[r,q]\Big) &\leq \frac{\mathbb{E}X_n^+ + |r|}{q-r} \\ &\leq \frac{\sup_n \mathbb{E}X_n^+ + |r|}{q-r}\end{aligned}$$

   Therefore by MCT,

   $$\begin{aligned}\mathbb{E}\Big(U_\infty[r,q]\Big) &\leq \frac{\sup_n \mathbb{E}X_n^+ + |r|}{q-r} \\ &< \infty \quad \text{by assumption}\end{aligned}$$

   Thus since $\mathbb{E}X < \infty \;\Rightarrow\; X < \infty$ a.s., then $U_\infty[r,q] < \infty$ a.s. for any $q > r \in \mathbb{Q}$. Therefore:

   $$\mathbb{P}\Big(\{\limsup X_n = \liminf X_n\}\Big) = \mathbb{P}\Big(\bigcap_{q>r}\{U_\infty[r,q] < \infty\}\Big) = 1$$

   And we can define $X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega)$.

2. Write $\mathbb{E}|X_\infty| = \mathbb{E}X_\infty^+ + \mathbb{E}X_\infty^-$. We show that $\mathbb{E}X_\infty^+ < \infty$ and $\mathbb{E}X_\infty^- < \infty$.

First consider the positive part:

$$\begin{aligned}\mathbb{E}X_\infty^+ &= \int \lim X_n^+ \, \mathrm{d}\mathbb{P} \\ &\leq \liminf \int X_n^+ \, \mathrm{d}\mathbb{P} \quad \text{(Fatou's lemma)} \\ &\leq \sup_n \mathbb{E}X_n^+ \\ &< \infty \quad \text{(by assumption)}\end{aligned}$$

Now note that since $\{X_n\}_{n\geq 0}$ is a submartingale, then $\mathbb{E}X_n \geq \mathbb{E}X_0$ and:

$$\begin{aligned}\mathbb{E}X_n^- &= \mathbb{E}X_n^+ - \mathbb{E}X_n \\ &\leq \mathbb{E}X_n^+ - \mathbb{E}X_0 \\ &\leq \sup_n \mathbb{E}X_n^+ - \mathbb{E}X_0\end{aligned}$$

Since the RHS is constant over n, then $\mathbb{E}X_n^-$ is bounded so $\sup_n \mathbb{E}X_n^- < \infty$. Now following the above argument,

$$\begin{aligned}\mathbb{E}X_\infty^- &= \int \lim X_n^+ \, \mathrm{d}\mathbb{P} \\ &\leq \liminf \int X_n^- \, \mathrm{d}\mathbb{P} \\ &\leq \sup_n \mathbb{E}X_n^- \\ &< \infty\end{aligned}$$

□

Example. (Counterexample: simple random walk)

This example shows that the conditions of the above theorem do not guarantee convergence in L^1.

Let $\{\xi_i\}_{i\geq 1}$ be an iid sequence such that

$$\mathbb{P}(\xi_i = +1) = \frac{1}{2}, \quad \mathbb{P}(\xi_i = -1) = \frac{1}{2}$$

Let a filtration $\{\mathcal{F}_n\}$ be given by $\mathcal{F}_n = \sigma(\xi_1, \dots \xi_n)$ and define $\{S_n\}_{n\geq 0}$ by:

$$S_0 = 1, \quad S_n = 1 + \sum_{i=1}^n \xi_i$$

Let $N = \inf\{n \geq 1 : S_n = 0\}$ be a stopping time. Then since $\{S_n\}_{n\geq 0}$ is a martingale, the sequence $\{X_n\}_{n\geq 0}$ defined by $X_n = S_{N\wedge n}$ is also a martingale.

X_n must converge to 0. To see this, note that if $X_n = k \neq 0$, then the next term in the sequence is $k \pm 1$ so it does not converge. In other words, X_n can only converge by having S_n hit 0, i.e. by having the process $X_{N\wedge n}$ stop at time N.

Now note that $\mathbb{E}X_n = 1$ for all n while $X_\infty = 0 \Rightarrow \mathbb{E}X_\infty = 0$ as just shown, so that convergence cannot occur in L^1.

Corollary 4.35. (Two easy corollaries of the MG convergence theorem)

1. If $\{X_n\}_{n\geq 0}$ is a positive supermartingale, then there exists X_∞ such that

$$X_n \xrightarrow{a.s.} X_\infty, \quad \mathbb{E}X_\infty \leq \lim \mathbb{E}X_n$$

2. If $\{X_n\}_{n\geq 0}$ is a non-negative martingale, then there exists X_∞ such that

$$X_n \xrightarrow{a.s.} X_\infty, \quad \mathbb{E}X_\infty \leq \lim \mathbb{E}X_0$$

Proof. For the first, note that $\{-X_n\}_{n\geq 0}$ is a submartingale and

$$\sup_n \mathbb{E}\Big[(-X_n)^+\Big] = 0$$

□

Theorem 4.36. (Levy zero-one law)

Let $\{\mathcal{F}_n\}_{n\geq 0}$ be a filtration with $\mathcal{F}_\infty = \sigma(\cup_{j=1}^{\infty}\mathcal{F}_j)$. Let Z be a $\mathcal{F}_\infty$-measurable random variable with $\mathbb{E}|Z| < \infty$. Define $Z_n = \mathbb{E}(Z \,|\, \mathcal{F}_n)$. Then

$$Z_n \xrightarrow{a.s.} Z \quad \text{and} \quad Z_n \xrightarrow{L^1} Z$$

Remark. Why is this theorem called a "zero-one" law? Because it can be applied to the case when Z is an $\mathcal{F}_\infty$-measurable indicator function $\mathbb{1}_A$. Then the sequence of conditional expectations $\mathbb{E}(\mathbb{1}_A \,|\, \mathcal{F}_n)$ converges to a random variable which can only take value 0 or 1.

Proof. The general outline of the proof is to show (1) $Z_n \xrightarrow{a.s.} Z_\infty$ and then (2) $Z_n \xrightarrow{L^1} Z$, from which it follows that $Z = Z_\infty$ a.s.

1. We show that $\{Z_n\}_{n\geq 0}$ satisfies the conditions of the martingale convergence theorem. Obviously, Z_n is $\mathcal{F}_n$-measurable and has finite expectation. Also,

$$\begin{aligned}\mathbb{E}(Z_{n+1} \,|\, \mathcal{F}_n) &= \mathbb{E}\Big[\mathbb{E}(Z \,|\, \mathcal{F}_{n+1}) \,|\, \mathcal{F}_n\Big] \\ &= \mathbb{E}(Z \,|\, \mathcal{F}_n)\end{aligned}$$

Also, note that

$$\begin{aligned}\mathbb{E}(Z_n^+) &\leq \mathbb{E}|Z_n| \\ &= \mathbb{E}(|\mathbb{E}(Z \,|\, \mathcal{F}_n)|) \\ &\leq \mathbb{E}|Z| \\ &< \infty\end{aligned}$$

Therefore we have that $\mathbb{E}(Z_n^+)$ is bounded and so $\sup_n \mathbb{E}(Z_n^+) < \infty$. Therefore, by the martingale convergence theorem, there exists X_∞ such that

$$Z_n \xrightarrow{a.s.} Z_\infty$$

2. To show $Z_n \xrightarrow{L^1} Z$, we first prove a lemma:

Lemma 4.37. Let Z be a $\mathcal{F}_\infty$-measurable r.v. with $\mathbb{E}|Z| < \infty$. Then, given any $m \geq 1$, there exists $0 \leq k_m < \infty$ and an $\mathcal{F}_{k_m}$-measurable r.v. Y_m such that:

(a) $\mathbb{E}|Y_{k_m}| < \infty$

(b) $\mathbb{E}|Y_{k_m} - Z| \leq \frac{1}{2^m}$

Proof. (Sketch)

We show that this holds for $Z = \mathbb{1}_A$, $A \in \mathcal{F}_\infty$. Then it will hold for simple functions, and then general measurable functions by MCT.

Define the set G by the set of all sets $A \in \mathcal{F}_\infty$ such that, given any $m \geq 1$, there exists $0 \leq k_m < \infty$ and $B_m \in \mathcal{F}_{k_m}$ such that

$$\mathbb{E}|\mathbb{1}_A - \mathbb{1}_B| = \mathbb{P}(A \triangle B_m) \leq \frac{1}{2^m}$$

Now note two facts:

(a) G contains $\cup_{n=1}^{\infty} \mathcal{F}_n$, a π-class.

To see this, let $A \in \mathcal{F}_J$ for some J. Then for any $m \geq 1$, let $k_m = J$ and $B_m = A$.

(b) G is a λ-class.

□

Returning to the proof of the main result, we apply our lemma immediately: given $m \geq 1$, find $0 \leq k_m < \infty$ and $Y_m \in \mathcal{F}_{k_m}$ such that $\mathbb{E}|Z - Y_m| \leq 2^{-m}$.

Let $X_n = \mathbb{E}(Z - Y_m \,|\, \mathcal{F}_n)$. For $n > k_m$, $\{X_n\}$ is a martingale:

$$\mathbb{E}(Z - Y_m \,|\, \mathcal{F}_{n+1}) = \mathbb{E}\Big[\mathbb{E}(Z - Y_m \,|\, \mathcal{F}_{n+1}) \,|\, \mathcal{F}_n\Big] = \mathbb{E}(Z - Y_m \,|\, \mathcal{F}_n)$$

Now note that:

$$\begin{aligned}
\mathbb{E}|X_n| &= \mathbb{E}\Big[|\mathbb{E}(Z - Y_m \,|\, \mathcal{F}_n)|\Big] \\
&\leq \mathbb{E}(|Z - Y_m|) \\
&\leq \frac{1}{2^m}
\end{aligned}$$

Therefore it follows that:

$$\begin{aligned}
\mathbb{E}(|Z - Z_n|) &= \mathbb{E}\Big[|Z - \mathbb{E}(Z \,|\, \mathcal{F}_n)|\Big] \\
&= \mathbb{E}\Big[|Z - Y_m + Y_m - \mathbb{E}(Z \,|\, \mathcal{F}_n)|\Big] \\
&\leq \mathbb{E}\Big[|Z - Y_m| + |Y_m - \mathbb{E}(Z \,|\, \mathcal{F}_n)|\Big] \\
&\qquad \text{(Triangle inequality)} \\
&= \mathbb{E}(|Z - Y_m|) + \mathbb{E}\Big[|Y_m - \mathbb{E}(Z \,|\, \mathcal{F}_n)|\Big] \\
&= \mathbb{E}(|Z - Y_m|) + \mathbb{E}\Big[|\mathbb{E}(Z - Y_m \,|\, \mathcal{F}_n)|\Big] \\
&\qquad (Y_m \text{ is } \mathcal{F}_n\text{-measurable}) \\
&\leq \frac{1}{2^{m-1}}
\end{aligned}$$

Therefore for any $m \geq 1$, we have

$$\limsup_{n\to\infty} \mathbb{E}(|Z - Z_n|) \leq \frac{1}{2^{m-1}}$$

□

Corollary 4.38. (Two easy corollaries of the Levy zero-one law)

1. Let $Y_1, Y_2, \ldots$ be a sequence of r.v.'s and let $\mathcal{F}_n = \sigma(Y_1, Y_2, \ldots)$. If Z is a $\mathcal{F}_\infty$-measurable random variable with $\mathbb{E}|Z| < \infty$, then

$$\mathbb{E}(Z \,|\, Y_1, \ldots, Y_n) \xrightarrow{a.s.} Z \quad \text{and} \quad \mathbb{E}(Z \,|\, Y_1, \ldots, Y_n) \xrightarrow{L^1} Z$$

2. (Kolmogorov zero-one law)

   Let $Y_1, Y_2, \ldots$ be a sequence of independent r.v.'s. Let $A \in T$, the tail σ-field defined by $T = \cap_{m=0}^{\infty} \sigma(Y_{m+1}, Y_{m+2}, \ldots)$.

   For any given n, A is independent of $\mathcal{F}_n$. Therefore

$$\mathbb{E}(\mathbb{1}_A \,|\, Y_1, \ldots, Y_n) = \mathbb{E}\mathbb{1}_A \;\Rightarrow\; \mathbb{P}(A \,|\, Y_1, \ldots, Y_n) = \mathbb{P}(A)$$

   And therefore we have

$$\mathbb{P}(A) \xrightarrow{a.s.} \mathbb{1}_A \quad \text{and} \quad \mathbb{P}(A) \xrightarrow{L^1} \mathbb{1}_A$$

Theorem 4.39. (Convergence or divergence theorem)

Let $\{X_n\}_{n\geq 0}$ be a martingale such that the "step size" is bounded:

$$\exists\, k < \infty \text{ s.t. } |X_n - X_{n-1}| < k \quad \forall n \text{ a.s.}$$

Define the sets C and D by:

$$\begin{aligned}
C &= \{\omega : X_n(\omega) \to \text{ some finite value}\} \\
D &= \{\omega : \limsup X_n(\omega) = \infty,\ \liminf X_n(\omega) = -\infty\}
\end{aligned}$$

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

Proof. Assume WLOG that $X_0 = 0$ (otherwise consider $\{X_n - X_0\}$). We only need to show that $C \supset D^c$ since then $\mathbb{P}(C \cup D) \geq \mathbb{P}(D^c \cup D) = 1$.

We first break the proof into two parts. First, we show that the assumptions imply that the process does not take "too large" negative values, and therefore

$$\{X_n \to \text{ some finite value}\} \supset \{\inf X_n > -\infty\}$$

Then we show that the assumptions imply that the process does not take "too large" positive values, and therefore

$$\{X_n \to \text{ some finite value}\} \supset \{\sup X_n > -\infty\}$$

1. The process does not take "too large" negative values:

   Fix some $L \geq 1$. Define the stopping time $T_L = \inf\{n \geq 0 : X_n \leq -L\}$. Note two facts about the stopped process $\{X_{T_L \wedge n}\}_{n \geq 0}$:

   (a) $X_{T_L \wedge n} \geq -(L + k)$ by definition of T_L.

   (b) $\{X_{T_L \wedge n}\}_{n \geq 0}$ is a martingale, so $\{X_{T_L \wedge n} + (L + k)\}_{n \geq 0}$ is a non-negative martingale.

   Therefore by the martingale convergence theorem,

   $$X_{T_L \wedge n} + (L + k) \xrightarrow{a.s.} \text{ something finite}$$

   In particular, if $T_L = \infty$, i.e. $X_n > -L$, then the convergence to a finite r.v. still holds. And since our choice of $L \geq 1$ was arbitrary,

   $$\begin{aligned} \{\omega : X_n(\omega) \to \text{ something finite}\} &\supset \{T_L = \infty\}, \ \forall L \geq 1 \\ &= \{X_n > -L \ \forall n\}, \ \forall L \geq 1 \\ &\supset \cup_{L=1}^{\infty} \{X_n > -L \ \forall n\} \\ &= \{\inf X_n > -\infty\} \end{aligned}$$

2. The process does not take "too large" positive values:

   Fix some $L \geq 1$. Follow the same argument above using the stopping time $T_L = \inf\{n \geq 0 : X_n \geq L\}$ and noting that $X_{T_L \wedge n} \leq (L + k)$.

Therefore we have shown:

$$C = \{\omega : X_n(\omega) \to \text{ finite}\} \supset B = \{\sup X_n < \infty\} \cup \{\inf X_n > -\infty\}$$

And since sups are decreasing and infs are increasing,

$$B \supset \{\limsup X_n < \infty \ \text{ or } \ \liminf X_n > -\infty\} = D^c$$

□

Theorem 4.40. (Conditional Borel-Cantelli lemmas)

Let $\{A_n\}_{n \geq 1}$ be a sequence of events with $A_n \in \mathcal{F}_n$ for all n. Define

1. $B_n = \cup_{m=n}^{\infty} A_m$
2. $B = A_n$ i.o. $= \cap_{n=1}^{\infty} \cup_{m=n}^{\infty} A_m = \lim B_n$

Then the following are true:

1. $\mathbb{P}(B_{n+1} \,|\, \mathcal{F}_n) \xrightarrow{a.s.} \mathbb{1}_B$
2. A_ni.o. $= \{\omega : \sum_{n=1}^{\infty} \mathbb{P}(A_n \,|\, \mathcal{F}_{n-1}) = \infty\}$

Proof. We prove the two parts separately:

1. (of $\mathbb{P}(B_{n+1} \,|\, \mathcal{F}_n) \xrightarrow{a.s.} \mathbb{1}_B$)

   Fix $n > k$. Then $\mathbb{1}_B \leq \mathbb{1}_{B_{n+1}} \leq \mathbb{1}_{B_k}$ and:

   $$\mathbb{E}(\mathbb{1}_B \,|\, \mathcal{F}_n) \leq \mathbb{E}(\mathbb{1}_{B_{n+1}} \,|\, \mathcal{F}_n) \leq \mathbb{E}(\mathbb{1}_{B_k} \,|\, \mathcal{F}_n)$$

   Thus, rewriting, we have

   $$\mathbb{P}(B \,|\, \mathcal{F}_n) \leq \mathbb{P}(B_{n+1} \,|\, \mathcal{F}_n) \leq \mathbb{P}(B_k \,|\, \mathcal{F}_n)$$

   Taking lim inf and lim sup over n and applying the Levy zero-one law to the leftmost and rightmost term, we have:

   $$\mathbb{1}_B \leq \liminf \mathbb{P}(B_{n+1} \,|\, \mathcal{F}_n) \leq \limsup \mathbb{P}(B_{n+1} \,|\, \mathcal{F}_n) \leq \mathbb{1}_{B_k}$$

   Letting $k \to \infty$, then $\mathbb{1}_{B_k} \to \mathbb{1}_B$ and we obtain:

   $$\mathbb{1}_B \leq \liminf \mathbb{P}(B_{n+1} \,|\, \mathcal{F}_n) \leq \limsup \mathbb{P}(B_{n+1} \,|\, \mathcal{F}_n) \leq \mathbb{1}_B$$

   □

2. (of A_ni.o. $= \{\omega : \sum_{n=1}^{\infty} \mathbb{P}(A_n \,|\, \mathcal{F}_{n-1}) = \infty\}$)

   Define the sequence $\{X_n\}_{n\geq 0}$ by $X_0 = 0$ and:

   $$X_n = \sum_{m=1}^{n} \Big[\mathbb{1}_{A_m} - \mathbb{P}(A_m \,|\, \mathcal{F}_{m-1})\Big]$$
   $$X_{n+1} = X_n + \Big[\mathbb{1}_{A_{n+1}} - \mathbb{P}(A_{n+1} \,|\, \mathcal{F}_n)\Big]$$

   Note two facts about this sequence:

   (a) $\{X_n\}_{n\geq 0}$ is a martingale:

   $$\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) = X_n + \mathbb{E}(\mathbb{1}_{A_{n+1}} \,|\, \mathcal{F}_n) - \mathbb{P}(A_{n+1} \,|\, \mathcal{F}_n) = X_n$$

   (b) $\{X_n\}_{n\geq 0}$ has bounded increments:

   $$\Big|X_{n+1} - X_n\Big| = \Big|\mathbb{1}_{A_{n+1}} - \mathbb{P}(A_{n+1} \,|\, \mathcal{F}_n)\Big| \leq 1$$

   Therefore, by the convergence or divergence theorem, only two things can happen:

(a) $C = \{X_n \text{ converges to some finite number}\}$

(b) $D = \{\limsup X_n = \infty,\ \liminf X_n = -\infty\}$

Now note that we can write:

$$X_n = \sum_{m=1}^{n} \mathbb{1}_{A_m} - \sum_{m=1}^{n} \mathbb{P}(A_m \,|\, \mathcal{F}_{m-1}) \triangleq A - B$$

where A_n i.o. $= \{\sum_{m=1}^{\infty} \mathbb{1}_{A_m} = \infty\} = A$.

The proof is completed by noting that if X_n converges to some finite value (i.e. $\omega \in C$), then $A = \infty$ if and only if $B = \infty$; similarly, if X_n has $\limsup = \infty$ and $\liminf = -\infty$ (i.e. $\omega \in D$), then it must be true that $A = \infty$ and $B = \infty$. □

Theorem 4.41. (Bounded OST)

Let $\{X_n\}_{n \geq 0}$ be a submartingale and let $T < \infty$ be a stopping time. If:

1. $\mathbb{E}T < \infty$
2. $\mathbb{E}(|\Delta_n^X| \,|\, \mathcal{F}_{n-1}) \leq B$ on $\{T \geq n\}$

Then:

$$\mathbb{E}(X_T \,|\, \mathcal{F}_0) \geq X_0 \quad \Rightarrow \quad \mathbb{E}X_T \geq \mathbb{E}X_0$$

Proof. We check the conditions for showing that $\mathbb{E}(|X_T - X_{T\wedge k}|) \to 0$.

1. $\mathbb{E}(|X_T|) < \infty$

   Writing $X_T = X_0 + \sum_{m=1}^{T} \Delta_m^X$, we have

$$\begin{aligned} |X_T| &= |X_0 + \sum_{m=1}^{T} \Delta_m^X| \\ &\leq |X_0| + \sum_{m=1}^{T} |\Delta_m^X| \\ &= |X_0| + \sum_{m=1}^{\infty} |\Delta_m^X| \cdot \mathbb{1}_{m \leq T} \\ &\triangleq Y \end{aligned}$$

Taking expectations, we have

$$\begin{aligned}\mathbb{E}|X_T| &\le \mathbb{E}|X_0| + \sum_{m=1}^{\infty} \mathbb{E}\Big(|\Delta_m^X| \cdot \mathbb{1}_{m\le T}\Big) \\ &\le \mathbb{E}|X_0| + \sum_{m=1}^{\infty} \mathbb{E}\Big[\mathbb{E}\Big(|\Delta_m^X| \cdot \mathbb{1}_{m\le T} \,|\, \mathcal{F}_{m-1}\Big)\Big] \\ &\le \mathbb{E}|X_0| + \sum_{m=1}^{\infty} \mathbb{E}\Big[\mathbb{1}_{m\le T} \cdot \mathbb{E}\Big(|\Delta_m^X| \,|\, \mathcal{F}_{m-1}\Big)\Big] \\ &\le \mathbb{E}|X_0| + B \cdot \sum_{m=1}^{\infty} \mathbb{P}(T \ge m) \\ &\qquad \text{(by assumption)}\end{aligned}$$

2. $\mathbb{E}(|X_n| \cdot \mathbb{1}_{T\ge n}) \to 0$

   Note that $|X_n| \cdot \mathbb{1}_{T\ge n} \le Y \cdot \mathbb{1}_{T\ge n}$. Therefore, it is sufficient to show:

$$\mathbb{E}(Y \cdot \mathbb{1}_{T\ge n}) \to 0$$

   To see this, note that $T < \infty$ and $Y \cdot \mathbb{1}_{T\ge n} \xrightarrow{a.s.} 0$. Establish the fact that $\mathbb{E}Y < \infty$, and then apply DCT to the sequence $Y \cdot \mathbb{1}_{T\ge n}$.

□

Theorem 4.42. (Generalization of Wald's lemma)

Let $\{\xi_i\}_{i\ge1}$ be an independent sequence such that

1. $\mathbb{E}|\xi_i| \le k \ \ \forall i$
2. $\mu_1 \le \mathbb{E}\xi_i \le \mu_2 \ \ \forall i$

Let T be a stopping time with $\mathbb{E}T < \infty$ and let the filtration $\{\mathcal{F}_n\}_{n\ge0}$ be given by $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$. Then:

$$\mathbb{E}T \cdot \mu_1 \le \mathbb{E}S_T \le \mathbb{E}T \cdot \mu_2$$

Remark. Compare this to the non-OST result:

$$n\mu_1 \le \mathbb{E}S_n \le n\mu_2$$

Proof. Define the sequence $\{X_n\}_{n\ge0}$ by $X_0 = 0$ and $X_n = S_n - n\mu_1$. Note that $\{X_n\}_{n\ge0}$ is a submartingale:

$$\begin{aligned}\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) &= X_n + \mathbb{E}(\xi_{n+1} - \mu_1 \,|\, \mathcal{F}_n) \\ &= X_n + \mathbb{E}(\xi_{n+1} - \mu_1) \quad \text{(by independence)} \\ &\ge X_n\end{aligned}$$

To apply the bounded OST, we need to check the step size condition:

$$\begin{aligned}\mathbb{E}\Big[|X_n - X_{n-1}| \,\Big|\, \mathcal{F}_{n-1}\Big] &= \mathbb{E}\Big[|\xi_n - \mu_1| \,\Big|\, \mathcal{F}_{n-1}\Big] \\ &= \mathbb{E}|\xi_n - \mu_1| \\ &\le \mathbb{E}|\xi_n| + |\mu_1|\end{aligned}$$

Thus each step is bounded by $k + |\mu_1|$. Therefore by OST,

$$\mathbb{E}X_T \geq \mathbb{E}X_0 \quad \Rightarrow \quad \mathbb{E}(S_T - T\mu_1) \geq 0$$

To show the upper bound, note that $\{Y_n\}_{n\geq 0}$ defined by $Y_n = S_n - n\mu_2$ is a supermartingale, so that $\{-Y_n\}_{n\geq 0}$ is a submartingale. Then follow the same argument as above.

□

4.8 Boundary crossings and Azuma-Hoeffding

This section is motivated by a class of *boundary crossing* questions: given a random variable or process, what is the probability that it will cross a certain boundary?

First consider studying whether or not a boundary crossing will exist at all (for the case of a simple mean zero iid process):

Theorem 4.43. (Existence of a boundary crossing)

Let $\{\xi_i\}_{n\geq 1}$ be an iid sequence. Assume $\mathbb{E}\xi_i = 0$ and fix $a, b > 0$. What is the probability that there exists some n such that $S_n \geq a + bn$?

$$\mathbb{P}(\exists n \geq 0 : S_n \geq a + bn) \leq \exp(-\theta a)$$

where $\theta > 0$ satisfies $\mathbb{E}(e^{\theta \xi_i}) = e^{\theta b}$.

Proof. Note that $S_n - bn = \sum_{i=1}^n (\xi_i - b)$, so:

$$\mathbb{P}\Big(\exists n : S_n \geq a + bn\Big) = \mathbb{P}\Big(\exists n : \sum_{i=1}^n (\xi_i - b) \geq a\Big)$$

Let $\hat{S}_n = \sum_{i=1}^n (\xi_i - b)$. Then for any $\theta > 0$,

$$\mathbb{P}(\exists n : \hat{S}_n \geq a) = \mathbb{P}(\exists n : \exp(\theta \hat{S}_n) \geq \exp(\theta a))$$

Now define the sequence $\{X_n\}_{n\geq 0}$ by $X_n = e^{\theta \hat{S}_n}$ and $X_0 = 1$. Note:

$$\begin{aligned}
\mathbb{E}\Big[X_{n+1} \,|\, \mathcal{F}_n\Big] &= e^{\theta \hat{S}_n} \cdot \mathbb{E}\Big[e^{\theta(\xi_{n+1} - b)} \,|\, \mathcal{F}_n\Big] \\
&= e^{\theta \hat{S}_n} \cdot \frac{\mathbb{E}(e^{\theta \xi_{n+1}})}{e^{\theta b}} \quad \text{(by independence)} \\
&= X_n \quad \text{(by assumption)}
\end{aligned}$$

Therefore $\{X_n\}_{n\geq 0}$ is a positive martingale. Thus by the supermartingale maximal inequality,

$$\lambda \cdot \mathbb{P}\left(\sup_{n\geq 0} X_n \geq \lambda\right) \leq \mathbb{E}X_0$$

Letting $\lambda = e^{\theta a}$ and noting that, in general, $\mathbb{P}(\exists n : X_n \geq a) = \mathbb{P}(\sup X_n \geq a)$ gives the desired result.

□

Now suppose are interested in the first time a process escapes out of some fixed interval (a, b). The main question of this section is: **What is the probability that the first crossing will be to above/below?**

First, we prove a theorem which gives us the crucial result regarding S_T to allow us to calculate these bounds. Then, we prove a related theorem that provides sufficient conditions for the conditions of the first theorem to hold.

Theorem 4.44. Let $\{\xi_i\}$ be iid and let T be a stopping time with $\mathbb{E}T < \infty$. Also let a filtration $\{\mathcal{F}_n\}_{n\geq 0}$ be given by $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$. Suppose

1. $\exists\, \theta > 0$ such that $\mathbb{E}(e^{\theta\xi_i}) = 1$ for all i
2. for $n < T$, $S_n \leq B$

Then $\mathbb{E}(e^{\theta S_T}) = 1$.

Proof. Define the sequence $\{X_n\}_{n\geq 0}$ by $X_0 = 1$ and $X_n = e^{\theta S_n}$. Note

$$\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) = e^{\theta S_n} \cdot \mathbb{E}(e^{\theta \xi_{n+1}} \,|\, \mathcal{F}_n) = e^{\theta S_n} \cdot \mathbb{E}(e^{\theta\xi_{n+1}})$$

Thus $\{X_n\}_{n\geq 0}$ is a martingale. If we show that the assumption of the bounded OST hold, then we will have $\mathbb{E}X_T = \mathbb{E}X_0 = 1$ and we are done.

By assumption, $\mathbb{E}T < \infty$. To show the other condition, note:

$$\begin{aligned}
\mathbb{E}(|\Delta_n^X|\big|\mathcal{F}_{n-1}) &= \mathbb{E}(|e^{\theta S_n} - e^{\theta S_{n-1}}|\big|\mathcal{F}_{n-1}) \\
&= e^{\theta S_{n-1}} \cdot \mathbb{E}(|e^{\theta\xi_n} - 1|\big|\mathcal{F}_{n-1}) \\
&\leq e^{\theta S_{n-1}} \cdot |\mathbb{E}e^{\theta\xi_n} + 1| \quad \text{(by Triangle ineq.)} \\
&\leq 2e^{\theta S_{n-1}}
\end{aligned}$$

Now note that by assumption, for $n < T$ (i.e. $n - 1 \leq T$), $e^{\theta S_{n-1}} \leq e^{\theta B}$. Therefore

$$\mathbb{E}(|\Delta_n^X|\big|\mathcal{F}_{n-1}) \leq 2e^{\theta B} \quad \text{for } n \leq T$$

So the OST holds and we are done.

□

This next theorem provides a set of convenient sufficient conditions to allow us to satisfy the conditions of the first theorem:

Theorem 4.45. Suppose $\{\xi_i\}$ are iid with $\mathbb{E}\xi_i < 0$, $\mathbb{P}(\xi_i > 0) > 0$, and $|\xi_i| \leq L$.

Fix $a < 0 < b$. Let $T = \inf\{n \geq 0 : S_n \leq a \text{ or } S_n \geq b\}$. Then:

1. $\mathbb{E}T < \infty$
2. $\exists \theta > 0$ such that $\mathbb{E}e^{\theta\xi_i} = 1$

Proof. We prove (1) first and then (2) second:

1. First note that $T < \infty$ because $S_n/n \xrightarrow{a.s.} \mathbb{E}\xi_i < 0$.

   Then using $T \wedge n$ so that $\mathbb{E}(T \wedge n) < \infty$, Wald's lemma gives

$$\mathbb{E}S_{T\wedge n} = \mathbb{E}\xi_i \cdot \mathbb{E}(T \wedge n)$$

Furthermore, since $|\xi_i| \le L$, then $|S_{T\wedge n}| \le |a|+|b|+L$ for all $n \ge 1$. Therefore, $S_{T\wedge n}$ is bounded, so by DCT we have $\mathbb{E}S_{T\wedge n} \to \mathbb{E}S_T < \infty$. Furthermore, by MCT, we have $\mathbb{E}(T \wedge n) \to \mathbb{E}T$. Plugging into the Wald's lemma equation, we obtain $\mathbb{E}T < \infty$.

2. Define $\phi(\theta) = \mathbb{E}e^{\theta\xi_i}$. Note that $\phi'(\theta)|_{\theta=0} = \mathbb{E}\xi_i < 0$ with $\phi(0) = 1$, so that ϕ is initially decreasing for $\theta > 0$.

   However, since $\mathbb{P}(\xi_i > 0) > 0$, then $\phi(\theta) = \mathbb{E}e^{\theta\xi_i} \to \infty$ as $\theta \to \infty$. Therefore ϕ must cross 1 from below at some $\theta > 0$.

□

How to calculate the actual bounds:

Since $|\xi_i| \le L$, note that:

1. If S_T crosses to above at b, then its minimum value is $S_T = b$ and its maximum value is $S_T = b + L$
2. If S_T crosses to below at a, then its minimum value is $S_T = a - L$ and its maximum value is $S_T = a$

Since $\mathbb{E}e^{\theta S_T} = 1$, then

1. Using the **maximum** values, $\mathbb{P}(S_T \ge b) \cdot e^{\theta(b+L)} + [1 - \mathbb{P}(S_T \ge b)] \cdot e^{\theta a} \ge 1$
2. Using the **minimum** values, $\mathbb{P}(S_T \ge b) \cdot e^{\theta b} + [1 - \mathbb{P}(S_T \ge b)] \cdot e^{\theta(a-L)} \ge 1$

Solve for $\mathbb{P}(S_T \ge b)$ in both to get upper and lower bounds.

Example. (Asymmetric simple random walk)

Consider the process $\{\xi_i\}_{i\ge1}$ with $\mathbb{P}(\xi_i = 1) = p$ and $\mathbb{P}(\xi_i = -1) = 1 - p$.

Let $p < 1/2$ so that $\mathbb{E}\xi_i < 0$ and $|\xi_i| \le 1$. Let $T = \inf\{n \ge 0 : S_n \le a;\ \text{or}\ \ S_n \ge b\}$ with $a < 0 < b$, $a, b \in \mathbb{Z}$ so that either $S_T = a$ or $S_T = b$.

Calculating $\mathbb{P}(S_T = b)$ directly is hard, but it is easy to check that the conditions of the previous theorem are satisfied, so that we can use the OST-derived result from the first theorem: $\mathbb{E}e^{\theta S_T} = 1$.

1. Step 1: Get $\theta > 0$ such that $\mathbb{E}e^{\theta\xi_i} = 1$.

   Rewriting the condition $\mathbb{E}e^{\theta\xi_i} = 1$, we have

$$e^{\theta}p + e^{-\theta}(1-p) = 1 \quad \Rightarrow \quad \theta = \log\left(\frac{1-p}{p}\right)$$

2. Step 2: Apply the OST result

   Using the fact that either $S_T = a$ or $S_T = b$, we have

$$\begin{aligned} \mathbb{E}e^{\theta S_T} &= 1 \\ \mathbb{P}(S_T = b)e^{\theta b} + (1 - \mathbb{P}(S_T = b))e^{\theta a} &= 1 \\ \mathbb{P}(S_T = b)\left(\frac{1-p}{p}\right)^b + (1 - \mathbb{P}(S_T = b))\left(\frac{1-p}{p}\right)^a &= 1 \end{aligned}$$

And then solve the equation for $\mathbb{P}(S_T = b)$.

A powerful extension of this idea of boundary crossings is the Azuma-Hoeffding inequality, which gives an exponential bound for the probability of a martingale with bounded differences exceeding some interval.

Theorem 4.46. (Azuma-Hoeffding)

Let $S_n = \sum_{i=1}^n X_i$, and $S_0 = 0$. Let $\{\mathcal{F}_n\}$ be defined by $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$.

If $\{S_n\}_{n\geq 0}$ is a martingale and $|X_n| \leq 1$ for all n, then for any $\lambda > 0$,

$$\mathbb{P}(S_n \geq \lambda\sqrt{n}) \leq \exp\left(\frac{-\lambda^2}{2}\right)$$

Applying this again to the martingale $\{-S_n\}_{n\geq 0}$, we obtain

$$\mathbb{P}(|S_n| \geq \lambda\sqrt{n}) \leq 2 \cdot \exp\left(\frac{-\lambda^2}{2}\right)$$

Remark. Since $S_n = \sum_{i=1}^n X_i$, the condition that $|X_n| \leq 1$ for all n is equivalent to the condition that $|S_n - S_{n-1}| \leq 1$ for all n.

Lemma 4.47. If Y is a random variable with $\mathbb{E}Y = 0$ and $|Y| \leq 1$, then

$$\mathbb{E}e^{\alpha Y} \leq \exp\left(\frac{\alpha^2}{2}\right) \quad \text{for all } \alpha > 0$$

Proof. Define the function $f(Y) = e^{\alpha Y}$ for $-1 \leq Y \leq 1$. Also, let the function $L(Y)$ for $-1 \leq Y \leq 1$ be the straight line connecting $f(-1)$ and $f(1)$.

Note f is convex, then $f(Y) \leq L(Y)$ for $|Y| \leq 1$. Also, $L(0) = \frac{1}{2}(e^{-\alpha} + e^{\alpha})$. Therefore:

$$\mathbb{E}f(Y) \leq \mathbb{E}L(Y) = L(\mathbb{E}(Y)) = L(0) = \frac{1}{2}(e^{-\alpha} + e^{\alpha})$$

The proof is completed by noting that $\frac{1}{2}(e^{-\alpha}+e^{\alpha}) \leq e^{\alpha^2/2}$ for all $\alpha > 0$ (treat both as a function of α and take derivatives).

□

Proof. (of Azuma-Hoeffding)

Note that $|X_n| \leq 1$ and $\mathbb{E}X_n = 0$ for all n since $S_0 = 0$ and $\{S_n\}_{n\geq 0}$ is a martingale. Thus our previous lemma applies.

The idea is to expand S_n using conditional expectation, and then use independence to apply the lemma:

$$\begin{aligned}\mathbb{E}\left(e^{\alpha S_n} \mid \mathcal{F}_{n-1}\right) &= e^{\alpha S_{n-1}} \cdot \mathbb{E}\left(e^{\alpha X_n} \mid \mathcal{F}_{n-1}\right) \\ &= e^{\alpha S_{n-1}} \cdot \mathbb{E}e^{\alpha X_n} \\ &\leq e^{\alpha S_{n-1}} \cdot e^{\alpha^2/2}\end{aligned}$$

Taking expectations of both sides, we obtain:

$$\mathbb{E}e^{\alpha S_n} \leq e^{\alpha^2/2} \cdot \mathbb{E}e^{\alpha S_{n-1}}$$

Expanding $\mathbb{E}e^{\alpha S_{n-1}}$ using conditional expectation as above and repeating $n-1$ times, we finally obtain:

$$\mathbb{E}e^{\alpha S_n} \leq e^{(n\alpha^2)/2}$$

Now note that

$$\begin{aligned}
\mathbb{P}(S_n \geq \lambda\sqrt{n}) &= \mathbb{P}(e^{\alpha S_n} \geq e^{\alpha\lambda\sqrt{n}}) \quad (\text{for } \alpha > 0) \\
&\leq e^{-\alpha\lambda\sqrt{n}} \cdot \mathbb{E}e^{\alpha S_n} \quad (\text{by Markov}) \\
&\leq e^{-\alpha\lambda\sqrt{n}} \cdot e^{(n\alpha^2)/2} \\
&= \exp\left(-\alpha\lambda\sqrt{n} + \frac{n\alpha^2}{2}\right)
\end{aligned}$$

Optimizing over $\alpha > 0$ gives $\alpha = \lambda/\sqrt{n}$.

□

Corollary 4.48. (Alternate statement)

With the same setup of the Azuma-Hoeffding inequality,

$$\mathbb{P}(S_n \geq \lambda) \leq \exp\left(\frac{-\lambda^2}{2n}\right)$$

Proof. Following the above proof, we again obtain:

$$\mathbb{E}e^{\alpha S_n} \leq e^{(n\alpha^2)/2}$$

Now using Markov in the same way as above, we have:

$$\begin{aligned}
\mathbb{P}(S_n \geq \lambda) &= \mathbb{P}(e^{\alpha S_n} \geq e^{\alpha\lambda}) \quad (\text{for } \alpha > 0) \\
&\leq e^{-\alpha\lambda} \cdot \mathbb{E}e^{\alpha S_n} \quad (\text{by Markov}) \\
&\leq e^{-\alpha\lambda} \cdot e^{(n\alpha^2)/2} \\
&= \exp\left(-\alpha\lambda + \frac{n\alpha^2}{2}\right)
\end{aligned}$$

Optmizing over $\alpha > 0$ gives $\alpha = \lambda/n$.

□

The Azume-Hoeffding inequality immediately suggests a useful technique for evaluating the probability of an event of the form $|Z - \mathbb{E}Z|$:

Theorem 4.49. (Method of bounded differences)

Suppose $\xi_1, \ldots, \xi_n$ are independent. Let $Z = f(\xi_1, \ldots, x_n)$. Also, assume that Z does not change too much if any of the ξ_i's are changed, i.e.

$$|f(\xi_1, \ldots, \xi_k, \ldots, \xi_n) - f(\xi_1, \ldots, \xi_k', \ldots, \xi_n)| \leq 1$$

Where ξ_k and ξ'_k are two realizations of the same random variable. Then for $\lambda > 0$,

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq \lambda\sqrt{n}) \leq 2 \cdot \exp\left(-\frac{\lambda^2}{2}\right)$$

Remark. This proof assumes that the absolute difference is bounded by 1. However, the result still holds if the difference is bounded by another constant c. See the generalization of the Azuma-Hoeffding inequality above.

Proof. The proof is by the Azuma-Hoeffding inequality, using the trick of independent copies.

Define $S_m = \mathbb{E}(Z \,|\, \xi_1, \ldots, \xi_m) - \mathbb{E}Z$. Note that:

1. $\{S_m\}_{m \geq 0}$ is a martingale.
2. If $m = n$, then $S_m = S_n = Z - \mathbb{E}Z$.

Therefore if we can show that $|S_m - S_{m-1}| \leq 1$, then the result will follow immediately from Azuma-Hoeffding.

Let $\xi'_1, \xi'_2, \ldots, \xi'_n$ be independent copies of the original ξ_i's. Note that:

$$\begin{aligned} &S_m - S_{m-1} = \\ &\quad = \mathbb{E}\Big[f(\xi_1, \ldots, \xi_n) \,|\, \xi_1, \ldots, \xi_m\Big] - \mathbb{E}\Big[f(\xi_1, \ldots, \xi_n) \,|\, \xi_1, \ldots, \xi_{m-1}\Big] \end{aligned}$$

We would like to use linearity of expectation here to combine the two terms, but we cannot because we are conditioning on different sub-σ-algebras. To remedy this, note that, since $\xi_m \perp \xi'_m$,

$$\begin{aligned} &\mathbb{E}\Big[f(\xi_1, \ldots, \xi_m, \ldots, \xi_n) \,|\, \xi_1, \ldots, \xi_{m-1}\Big] = \\ &\quad = \mathbb{E}\Big[f(\xi_1, \ldots, \xi'_m, \ldots, \xi_n) \,|\, \xi_1, \ldots, \xi_{m-1}, \xi_m\Big] \end{aligned}$$

Therefore we have:

$$\begin{aligned} &|S_m - S_{m-1}| = \\ &\quad = \Big|\,\mathbb{E}\Big[f(\xi_1, \ldots, \xi_m, \ldots, \xi_n) - f(\xi_1, \ldots, \xi'_m, \ldots, \xi_n) \,|\, \xi_1, \ldots, \xi_m\Big]\,\Big| \\ &\quad \leq \mathbb{E}\Big[\big|f(\xi_1, \ldots, \xi_m, \ldots, \xi_n) - f(\xi_1, \ldots, \xi'_m, \ldots, \xi_n)\big| \,|\, \xi_1, \ldots, \xi_m\Big] \\ &\quad \leq \mathbb{E}\Big[1 \,|\, \xi_1, \ldots, \xi_m\Big] \end{aligned}$$

□

4.9 Uniform integrability and branching processes

Question: if we know that $X_n \xrightarrow{\mathbb{P}} X$, then under what conditions can we also conclude that $X_n \xrightarrow{L^1} X$, $\mathbb{E}|X_n - X| \to 0$?

Definition. (Uniformly integrable)

A family of random variables $\{Y_\alpha : \alpha \in I\}$ is **uniformly integrable** if

$$\lim_{b\to\infty} \sup_{\alpha\in I} \mathbb{E}\Big[|Y_\alpha| \cdot \mathbb{1}_{|Y_\alpha|\geq b}\Big] = 0$$

Remark. Why is this property called "**uniform** integrability?" Because for a single random variable X (on a finite measure space),

$$\lim_{b\to\infty} \mathbb{E}\Big[|X| \cdot \mathbb{1}_{|X|\geq b}\Big] = 0 \quad \iff \quad X \in L_1$$

This can be easily shown using DCT and by splitting the expression $\mathbb{E}|X|$ into two integrals over disjoint regions.

Theorem 4.50. (Some properties of UI)

1. If $\sup_\alpha \mathbb{E}(|Y_\alpha|^q) < \infty$ for some $q > 1$, then $\{Y_\alpha : \alpha \in I\}$ are UI.
2. If $\{Y_\alpha : \alpha \in I\}$ are UI, then $\sup_\alpha \mathbb{E}(|Y_\alpha|) < \infty$.
3. If $Y_n \xrightarrow{\mathbb{P}} Y$ and $\{Y_n : n \geq 1\}$ are UI, then $Y_n \xrightarrow{L^1} Y$ also.
4. If $Y_n \xrightarrow{d} Y$ and $\{Y_n : n \geq 1\}$ are UI, then $Y_n \xrightarrow{L^1} Y$ also.

Remark. Note that for (3), the random variables must be defined on the same probability space in order for convergence in $\mathbb{P}$ to make sense. However, for (4), the random variables may be defined on different probability spaces.

Lemma 4.51. (An "absolute continuity" property)

Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Then given $\epsilon > 0$, there exists $\delta > 0$ such that for $F \in \mathcal{F}$,

$$\mathbb{P}(F) < \delta \quad \Rightarrow \quad \mathbb{E}(|X| \cdot \mathbb{1}_F) < \epsilon$$

Remark. This result generalizes to arbitrary measurable spaces by monotone convergence (see Leadbetter notes).

Proof. Suppose not. Then for some $\epsilon_0 > 0$, we can find a sequence $\{F_n\}$ of elements of $\mathcal{F}$ such that for all n,

$$\mathbb{P}(F_n) < \frac{1}{2^n} \quad \text{and} \quad \mathbb{E}(|X| \cdot \mathbb{1}_{F_n}) \geq \epsilon_0$$

Define $H = \limsup F_n$. Then Borel-Cantelli gives $\mathbb{P}(H) = 0$ but (reverse) Fatou gives $\mathbb{E}(|X| \cdot \mathbb{1}_H) \geq \epsilon_0$, a contradiction.

□

Theorem 4.52. Suppose X is integrable. Then $\{\mathbb{E}(X \,|\, \mathcal{G}) : \mathcal{G} \subset \mathcal{F} \text{ is a sub-}\sigma\text{-field}\}$ is UI.

Proof. Let $\mathbb{E}(X \,|\, \mathcal{G})$ be an arbitrary element of the above set.

Fix $\epsilon > 0$. By the above lemma, there exists $\delta > 0$ such that for $F \in \mathcal{F}$,

$$\mathbb{P}(F) < \delta \quad \Rightarrow \quad \mathbb{E}(|X| \cdot \mathbb{1}_F) < \epsilon$$

By conditional Jensen and definition of CE, we have:

$$\begin{aligned}\mathbb{E}\Big[|\mathbb{E}(X \mid \mathcal{G})| \cdot \mathbb{1}_{|\mathbb{E}(X \mid \mathcal{G})|>M}\Big] &\leq \mathbb{E}\Big[|\mathbb{E}(X \mid \mathcal{G})| \cdot \mathbb{1}_{\mathbb{E}(|X| \mid \mathcal{G})>M}\Big] \\ &\leq \mathbb{E}\Big[\mathbb{E}(|X| \mid \mathcal{G}) \cdot \mathbb{1}_{\mathbb{E}(|X| \mid \mathcal{G})>M}\Big] \\ &= \mathbb{E}\Big[|X| \cdot \mathbb{1}_{\mathbb{E}(|X| \mid \mathcal{G})>M}\Big]\end{aligned}$$

Where the last step is because $\{\mathbb{E}(|X| \mid \mathcal{G}) > M\} \in \mathcal{G}$. Now we show that by making M large enough, we may make the measure of the set we are integrating over $< \delta$. By Chebyshev's inequality,

$$\mathbb{P}(\mathbb{E}(|X| \mid \mathcal{G}) > M) \leq \mathbb{E}[\mathbb{E}(|X| \mid \mathcal{G})]/M = \mathbb{E}|X|/M$$

X is integrable, so choosing $M > 0$ large enough makes $\mathbb{E}|X|/M < \delta$. Therefore by our lemma, we have

$$\mathbb{E}\Big[|X| \cdot \mathbb{1}_{\mathbb{E}(|X| \mid \mathcal{G})>M_0}\Big] < \epsilon$$

Therefore $\{\mathbb{E}(X \mid G) : G \subset \mathcal{F} \text{ is a sub-}\sigma\text{-field}\}$ is UI.

□

Corollary 4.53. (Uniform integrability of martingales)

If $\{X_n\}_{n\geq 0}$ is a submartingale and $\sup_n \mathbb{E}(|X_n|^q) < \infty$ for some $q > 1$ (i.e. $\{X_n\}_{n\geq 0}$ is UI), then there exists X such that $X_n \xrightarrow{a.s.} X$ **and** $X_n \xrightarrow{L^1} X$.

Proof. Since $\sup_n \mathbb{E}(|X_n|^q) < \infty$ for $q > 1$, then $\sup_n \mathbb{E}|X_n| < \infty$ as well. Therefore by the martingale convergence theorem, $\exists X$ with $\mathbb{E}X < \infty$ such that $X_n \xrightarrow{a.s.} X$. Then since $\{X_n\}_{n\geq 0}$ is UI, $X_n \xrightarrow{L^1} X$ also.

Definition. (Galton-Watson branching process)

Define the sequence of iid random variables

$$\{\xi_i^n : i \geq 1,\ n \geq 1\}, \quad \mathbb{E}\xi_i^n = \mu < \infty,\ \mathrm{Var}(\xi_i^n) < \infty$$

A **Galton-Watson branching process** $\{Z_n\}$ is defined by $Z_0 = 1$ and the recurrence relation:

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

This is a process that evolves in discrete time. At generation n, a random number of children are birthed which then go on to make up the $n + 1^{\text{th}}$ generation, where the # of births is iid.

Theorem 4.54. (Extinction probability)

Let $P = \mathbb{P}(Z_n = 0 \text{ eventually})$. What is P? Three cases:

1. $\mu < 1$. Then $P = 1$.

2. $\mu = 1$. Then $P = 1$ also.

3. What about $\mu > 1$? Answer: $P < 1$.

   Proof. It is easily shown (HW) that the sequence $\{Z_n/\mu^n\}$ is a positive martingale. Thus

   $$\sup_n \mathbb{E}\left[\left(\frac{Z_n}{\mu^n}\right)^+\right] = \sup_n \mathbb{E}\left(\frac{Z_n}{\mu^n}\right) = 1 < \infty$$

   Therefore by the MG convergence theorem, $\exists\, W$ with $\mathbb{E}W < \infty$ such that

   $$\frac{Z_n}{\mu^n} \xrightarrow{a.s.} W$$

   If we can show the convergence in L^1, then $\lim \mathbb{E}(Z_n/\mu^n) = \mathbb{E}W = 1$. It follows then that $\mathbb{P}(W = 0) < 1$, and since $W = \lim Z_n/\mu^n$ a.s., then $P < 1$.

   So to show $Z_n/\mu^n \xrightarrow{L^1} W$, we only need to show $\{Z_n/\mu^n\}_{n\geq 0}$ is UI. To show $\{Z_n/\mu^n\}_{n\geq 0}$ is UI, it is sufficient to show that $\sup_n \mathbb{E}((Z_n/\mu^n)^2) < \infty$. But since $\mathbb{E}(Z_n/\mu^n) = 1$ for all $n \geq 0$, then it suffices to show that $\sup_n \operatorname{Var}(Z_n/\mu^n) < \infty$.

   Let $\mathcal{F}_n$ be the information known up to time n:

   $$\mathcal{F}_n = \sigma\Big(\{\xi_i^m : i \geq 1, m \leq n-1\}\Big)$$

   Where m only runs up to $n-1$ since $Z_n = \sum_{i=1}^{Z_{n-1}} \xi_i^{n-1}$. Now note that:

   $$\operatorname{Var}(Z_n) = \mathbb{E}\Big[\operatorname{Var}(Z_n \,|\, \mathcal{F}_{n-1})\Big] + \operatorname{Var}\Big[\mathbb{E}(Z_n \,|\, \mathcal{F}_{n-1})\Big]$$

   Although Z_n is independent of $\mathcal{F}_{n-1}$, once we condition on $\mathcal{F}_{n-1}$ we know the value of Z_{n-1}. And since $Z_n = \sum_{i=1}^{Z_{n-1}} \xi_i^{n-1}$ as noted above, then we have:

   1. $\operatorname{Var}(Z_n \,|\, \mathcal{F}_{n-1}) = \operatorname{Var}\xi_i \cdot Z_{n-1}$
   2. $\mathbb{E}(Z_n \,|\, \mathcal{F}_{n-1}) = \mu \cdot Z_{n-1}$
   3. $\operatorname{Var}(\mathbb{E}(Z_n \,|\, \mathcal{F}_{n-1})) = \operatorname{Var}(\mu \cdot Z_{n-1}) = \mu^2 \cdot \operatorname{Var}(Z_{n-1})$

Thus finally, we have:

$$\begin{aligned}
\operatorname{Var}(Z_n) &= \sigma^2 \cdot \mu^{n-1} + \mu^2 \cdot \operatorname{Var}(Z_{n-1}) \\
\operatorname{Var}\left(\frac{Z_n}{\mu^n}\right) &= \frac{1}{\mu^{2n}}\left[\sigma^2 \cdot \mu^{n-1} + \mu^2 \cdot \operatorname{Var}(Z_{n-1})\right] \\
&= \frac{\sigma^2}{\mu^{n+1}} + \operatorname{Var}\left(\frac{Z_{n-1}}{\mu^{n-1}}\right) \\
&\vdots \\
&= \sigma^2 \cdot \sum_{j=2}^{n+1} \frac{1}{\mu^j} \\
&\leq \sigma^2 \sum_{j=1}^{\infty} \frac{1}{\mu^j} \\
&< \infty \quad (\text{since } \mu > 1)
\end{aligned}$$

Therefore $\sup_n \operatorname{Var}(Z_n/\mu^n) < \infty$ and so Z_n/μ^n is UI.

$\square$

5 Weak convergence

5.1 Basic properties, Scheffe's theorem

Recall that if μ is a probability measure on $(\mathbb{R}, B(\mathbb{R}))$, then the CDF of μ is:

$$F_\mu(x) = \mu((-\infty, x])$$

Remark. In this notation, μ is actually the *induced* probability measure of some unspecified random variable X defined on some unspecified probability space $(\Omega, \mathcal{F}, \mathbb{P})$:

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

Definition. (Continuity point)

We call x a **continuity point** of F_μ if:

$$F_\mu(x) = F_\mu(x^-) = \lim_{y \uparrow x} F_\mu(y)$$

Remark. Two observations:

1. This left-continuous condition is sufficient since all distribution functions are right-continuous.
2. If x is a continuity point, then

$$\mu(\{x\}) = F_\mu(x) - F_\mu(x^-) = 0$$

Theorem 5.1. (Convergence in distribution)

Let μ and $\{\mu_n\}_{n \geq 1}$ be (induced) probability measures on $(\mathbb{R}, B(\mathbb{R}))$. TFAE:

1. For all continuity points x of F_μ, $F_{\mu_n}(x) \to F_\mu(x)$
2. For all bounded continuous functions $g(\cdot)$,

$$\int g(x)\,\mathrm{d}\mu_n \to \int f(x)\,\mathrm{d}\mu$$

i.e. if $X_n \sim \mu_n$ and $X \sim \mu$, then $\mathbb{E}(g(X_n)) \to \mathbb{E}(g(X))$.

3. There exists $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables $\hat{X}_n : \Omega \to \mathbb{R}$ and $\hat{X} : \Omega \to \mathbb{R}$ with $\hat{X}_n \sim \mu_n$ and $\hat{X} \sim \mu$ such that $\hat{X}_n \xrightarrow{a.s.} \hat{X}$.

Proof. (3) $\Rightarrow$ (2):

If g is continuous, then by the continuous mapping theorem

$$g(\hat{X}_n(\omega)) \to g(\hat{X}(\omega)) \quad \text{for a.e. } \omega$$

Now g is bounded, so there exists M such that $|g(X)| \leq M$ for all x. Therefore by DCT,

$$\mathbb{E}(g(\hat{X}_n)) \to \mathbb{E}(g(\hat{X}))$$
$$\int_{\mathbb{R}} g(x)\,\mathrm{d}\mu_n(x) \to \int_{\mathbb{R}} g(x)\,\mathrm{d}\mu(x) \quad (\Delta \text{ of measure})$$

(2) ⇒ (1):

Fix some x. Note that (1) is equivalent to (2) with $g(y) = \mathbb{1}_{y \le x}$. However, this function is not continuous, so...

Upper bound: Let $g_j(y)$ be a function which takes value 1 for all $y \le x$, 0 for all $y > x + \frac{1}{j}$, and which decreases smoothly from 1 to 0 between x and $x + \frac{1}{j}$. Thus $g_j(y)$ is continuous and $g_j(y) \ge \mathbb{1}_{y \le x}$ for all x. So:

$$\begin{aligned} F_{\mu_n}(x) &= \int_{\mathbb{R}} \mathbb{1}_{y \le x} \, \mathrm{d}\mu_n(y) \\ &\le \int_{\mathbb{R}} g_j(y) \, \mathrm{d}\mu_n(y) \\ &\to \int_{\mathbb{R}} g_j(y) \, \mathrm{d}\mu(y) \quad \text{as } n \to \infty \\ &\le F_\mu(x + 1/j) \end{aligned}$$

Therefore, $\limsup_n F_{\mu_n}(x) \le F_\mu(x + 1/j)$ for all $j \ge 1$. Thus, sending $j \to \infty$ and noting that F_μ is right-continuous, we obtain:

$$\limsup_n F_{\mu_n}(x) \le F_\mu(x)$$

Lower bound: Let $h_j(y)$ be a function which takes value 1 for all $y \le x - \frac{1}{j}$, 0 for all $y > x$, and which decreases smoothly from 1 to 0 between $x - \frac{1}{j}$ and x. Thus $h_j(y)$ is continuous and $h_j(y) \le \mathbb{1}_{y \le x}$ for all x. So:

$$\begin{aligned} F_{\mu_n}(x) &= \int_{\mathbb{R}} \mathbb{1}_{y \le x} \, \mathrm{d}\mu_n(y) \\ &\ge \int_{\mathbb{R}} h_j(y) \, \mathrm{d}\mu_n(y) \\ &\to \int_{\mathbb{R}} h_j(y) \, \mathrm{d}\mu(y) \quad \text{as } n \to \infty \\ &\ge F_\mu(x - 1/j) \end{aligned}$$

Therefore, $\liminf_n F_{\mu_n}(x) \le F_\mu(x - 1/j)$ for all $j \ge 1$. Thus, sending $j \to \infty$, we obtain:

$$\liminf_n F_{\mu_n}(x) \ge F_\mu(x)$$

only if x is a continuous point of F_μ. Thus, combining the two bounds, we have $F_{\mu_n}(x) \to F_\mu(x)$ if x is a continuous point of F_μ.

(1) ⇒ (2)

Fill this in later (see previous statement).

□

Example. (Showing $\xrightarrow{d}$ directly)

Let X_n have a Uniform$\{1, 2, \ldots, n\}$ distribution. We show that $\frac{X_n}{n} \xrightarrow{d}$ Uniform$(0, 1)$.

$$\begin{aligned}\mathbb{P}\left(\frac{X_n}{n} \le x\right) &= \mathbb{P}(X_n \le \lfloor nx \rfloor) \\ &= \frac{\lfloor nx \rfloor}{n} \\ &\to x \quad \text{as } n \to \infty\end{aligned}$$

Example. (Showing $\xrightarrow{d}$ directly)

Let $X_\theta \sim$ Geometric(θ), i.e. $\mathbb{P}(X_\theta > i) = (1-\theta)^i$ for $i = 1, 2, \ldots$

We show that $\theta \cdot X_\theta \xrightarrow{d} \exp(1)$ as $\theta \to 0$:

$$\begin{aligned}\mathbb{P}(\theta \cdot X_\theta \le x) &= \mathbb{P}\left(X_\theta \le \left\lfloor \frac{x}{\theta} \right\rfloor\right) \\ &= 1 - \mathbb{P}\left(X_\theta > \left\lfloor \frac{x}{\theta} \right\rfloor\right) \\ &= 1 - (1-\theta)^{\lfloor \frac{x}{\theta} \rfloor} \\ &\to 1 - e^{-x}\end{aligned}$$

Corollary 5.2. (Skorohod corollary 1: continuous mapping theorem)

Fix a function g. Define the set $D_g = \{x : g \text{ is not continuous at x}\}$.

Suppose $X_n \xrightarrow{d} X$ and $\mathbb{P}(X \in D_g) = 0$. Then $g(X_n) \xrightarrow{d} g(X)$.

Proof. Since $X_n \xrightarrow{d} X$, there exist $\hat{X}_n \stackrel{d}{=} X_n$ and $\hat{X} \stackrel{d}{=} X$ such that $\hat{X}_n \xrightarrow{a.s.} \hat{X}$. Furthermore, since $\mathbb{P}(\hat{X} \in D_g) = 0$, then $g(\hat{X}_n) \xrightarrow{a.s.} g(\hat{X})$.

Thus $g(\hat{X}_n) \xrightarrow{d} g(\hat{X})$ as well. Note $g(\hat{X}_n) \stackrel{d}{=} g(X_n)$ and $g(\hat{X}) \stackrel{d}{=} g(X)$.

□

Corollary 5.3. (Skorohod corollary 2: expectation bounds)

Let $X_n \xrightarrow{d} X$ and $g : \mathbb{R} \to [0, \infty)$ be continuous. Then $\mathbb{E}g(X) \le \liminf \mathbb{E}g(X_n)$.

Proof. Since $X_n \xrightarrow{d} X$, there exist $\hat{X}_n \stackrel{d}{=} X_n$ and $\hat{X} \stackrel{d}{=} X$ such that $\hat{X}_n \xrightarrow{a.s.} \hat{X}$. Since g is continuous, $g(\hat{X}_n) \xrightarrow{a.s.} g(\hat{X})$. By Fatou's lemma,

$$\mathbb{E}g(\hat{X}) \le \liminf \mathbb{E}g(\hat{X}_n)$$

Note $g(\hat{X}_n) \stackrel{d}{=} g(X_n)$ and $g(\hat{X}) \stackrel{d}{=} g(X)$.

□

Theorem 5.4. (Scheffe's theorem)

Let $(S, \mathcal{S}, \theta)$ be a measure space and let $\{h_n\} : S \to [0, \infty)$ and $h : S \to [0, \infty)$ be $\mathcal{S}$-measurable. Suppose:

1. $\int_S h_n(s)\, \theta(\mathrm{d}s) = \int_S h(s)\, \theta(\mathrm{d}s) = 1$

2. $h_n(s) \to h(s)$ for a.e.(θ) s

Then $\int |h_n(s) - h(s)|\, \theta(\mathrm{d}s) \to 0$.

Proof. By assumption, we have:

$$\begin{aligned} 0 &= \int_S \Big[h(s) - h_n(s)\Big]\, \theta(\mathrm{d}s) \\ &= \int_S \Big[h(s) - h_n(s)\Big]^+ \theta(\mathrm{d}s) - \int_S \Big[h(s) - h_n(s)\Big]^- \theta(\mathrm{d}s) \end{aligned}$$

Thus the integral of the positive part is equal to the integral of the negative part, so we can write:

$$\int_S \Big|h(s) - h_n(s)\Big|\, \theta(\mathrm{d}s) = 2\int_S \Big[h(s) - h_n(s)\Big]^+ \theta(\mathrm{d}s)$$

Now since h and h_n take only non-negative values, then

$$0 \leq \Big[h(s) - h_n(s)\Big]^+ \leq h(s)$$

And since $\int h(s) = 1 < \infty$, we can apply dominated convergence to obtain the result.

$\square$

Theorem 5.5. Suppose $\{X_n\}, X$ are integer-valued random variables. TFAE:

1. $X_n \xrightarrow{d} X$
2. $\mathbb{P}(X_n = i) \to \mathbb{P}(X = i)\ \forall i \in \mathbb{Z}$
3. $\sum_{i\in\mathbb{Z}} \Big|\mathbb{P}(X_n = i) - \mathbb{P}(X = i)\Big| \to 0$

Proof. (1) $\Rightarrow$ (2)

Since $\{X_n\}$ and X are integer-valued, then they are not continuous at any $i \in \mathbb{Z}$. But they are continuous at $i + 1/2$. Thus:

$$\begin{aligned} \mathbb{P}(X_n = i) &= \mathbb{P}(X_n \leq i + 1/2) - \mathbb{P}(X_n \leq i - 1/2) \\ &\to \mathbb{P}(X \leq i + 1/2) - \mathbb{P}(X \leq i - 1/2) \\ &= \mathbb{P}(X = i) \end{aligned}$$

(2) $\Rightarrow$ (3)

Consider the measure space $(\mathbb{Z}, 2^{\mathbb{Z}}, \theta)$, where θ is the counting measure. Define the functions $\{h_n\} : \mathbb{Z} \to [0, \infty)$ and $h : \mathbb{Z} \to [0, \infty)$ by:

$$h_n(i) = \mathbb{P}(X_n = i), \quad h(i) = \mathbb{P}(X = i)$$

Then apply Scheffe's theorem.

(3) $\Rightarrow$ (1)

Fix some $x \notin \mathbb{Z}$ so that x is a continuous point. Then:

$$\begin{aligned}
\Big|\mathbb{P}(X_n \leq x) - \mathbb{P}(X \leq x)\Big| &= \Big|\sum_{i \leq x} \mathbb{P}(X_n = i) - \sum_{i \leq x} \mathbb{P}(X = i)\Big| \\
&\leq \sum_{i \leq x} \Big|\mathbb{P}(X_n = i) - \mathbb{P}(X = i)\Big| \\
&\leq \sum_{i \in \mathbb{Z}} \Big|\mathbb{P}(X_n = i) - \mathbb{P}(X = i)\Big| \\
&\to 0 \text{ by assumption}
\end{aligned}$$

□

Example. Binomial$(n, \lambda/n) \to$ Poisson(λ)

Theorem 5.6. Let λ be the Lebesgue measure on $B(\mathbb{R})$ and suppose $\{f_n\}, f$ are pdfs on $(\mathbb{R}, B(\mathbb{R}))$:

$$\int_{\mathbb{R}} f_n \, \mathrm{d}\lambda(x) = \int_{\mathbb{R}} f \, \mathrm{d}\lambda(x) = 1$$

Define $\{\mu_n\}$ and μ, measures on $(\mathbb{R}, B(\mathbb{R}))$ by:

$$\mu_n(B) = \int_B f_n \, \mathrm{d}\lambda(x) \quad \text{and} \quad \mu(B) = \int_B f \, \mathrm{d}\lambda(x)$$

If $f_n(x) \to f(x)$ for a.e. x, then $\mu_n \xrightarrow{d} \mu$.

Proof. By Scheffe's theorem, we have L^1 convergence of f_n to f:

$$\int_{\mathbb{R}} |f_n(x) - f(x)| \, \mathrm{d}\lambda(x) \to 0$$

So note that:

$$\begin{aligned}
\big|\mathbb{P}(X_n \leq x) - \mathbb{P}(X \leq x)\big| &= \Big|\int_{-\infty}^{x} f_n(y) \, \mathrm{d}\lambda(y) - \int_{-\infty}^{x} f(y) \, \mathrm{d}\lambda(y)\Big| \\
&\leq \int_{-\infty}^{x} |f_n(y) - f(y)| \, \mathrm{d}\lambda(y) \\
&\leq \int_{\mathbb{R}} |f_n(y) - f(y)| \, \mathrm{d}\lambda(y)
\end{aligned}$$

□

5.2 Helly's theorem

Definition. (Tightness)

We say that a family of random variables $\{X_\alpha : \alpha \in I\}$ is **tight** if:

$$\lim_{b \to \infty} \sup_{\alpha} \mathbb{P}\left(|X_\alpha| > b\right) = 0$$

Remark. Two observations:

1. We can also say that the corresponding family of *probability measures* $\{\mu_\alpha : \alpha \in I\}$ (on $(\mathbb{R}, B(\mathbb{R}))$) are tight if:

$$\lim_{b\to\infty} \sup_\alpha \mu_n([-b,b]) = 0$$

2. Compare this to the definition of uniform integrability:

$$\lim_{b\to\infty} \sup_\alpha \mathbb{E}\left[|X_\alpha| \cdot \mathbb{1}_{|X_\alpha|>b}\right] = 0$$

Theorem 5.7. (Equivalent definition of tightness)

The previous definition is equivalent to the following definition:

A family of probability measures $\{\mu_\alpha : \alpha \in I\}$ is **tight** if, for every $\epsilon > 0$, there exists a compact interval $[a,b]$ such that $\mu_\alpha([a,b]) > 1-\epsilon$ for all α.

Proof. Assume that $\lim_{b\to\infty} \sup_\alpha \mathbb{P}(|X_\alpha| > b) = 0$. Fix some $\epsilon > 0$. Then there exists B such that, for all $b > B$, $\sup_\alpha \mathbb{P}(|X_\alpha| > b) < \epsilon$. Then if μ_α is the probability measure corresponding to X_α, we have $\sup_\alpha \mu_\alpha([-b,b]) > 1-\epsilon$ for all $b > B$. Then $\mu_\alpha([-B-\epsilon, B+\epsilon]) > 1-\epsilon$ for all α.

Now assume that, for every $\epsilon > 0$, there exists a compact interval $[a,b]$ such that $\mu_\alpha([a,b]) > 1-\epsilon$ for all α. Fix $\epsilon > 0$. Then there exists $[a_0, b_0]$ such that $\sup_\alpha \mu_\alpha([a_0,b_0]) > 1-\epsilon$. If X_α is the random variable corresponding to μ_α, then this is equivalent to $\sup_\alpha \mathbb{P}(a_0 \le X_\alpha \le b_0) > 1-\epsilon$.

Let $c = \max\{|a_0|, |b_0|\}$. Then $\sup_\alpha \mathbb{P}(|X_\alpha| < c) > 1-\epsilon$, so that we have $\sup_\alpha \mathbb{P}(|X_\alpha| > c) < \epsilon$. Thus for any ϵ, we can find a corresponding c such that $\sup_\alpha \mathbb{P}(|X_\alpha| > c) < \epsilon$.

□

Example. (A collection of r.v.'s that are not tight)

Theorem 5.8. (Sufficient conditions for tightness, UI)

1. (Tightness) Suppose $\sup_\alpha \mathbb{E}|X_\alpha| < \infty$ **or** $\exists\, \phi : [0,\infty) \to [0,\infty)$ such that
   (a) ϕ is increasing
   (b) ϕ is increasing as $x \to \infty$
   (c) $\sup_\alpha \mathbb{E}[\phi|X_\alpha|] < \infty$

   Then $\{X_\alpha : \alpha \in I\}$ is tight.

2. (Uniform integrability) Suppose $\sup_\alpha \mathbb{E}(|X_\alpha|^2) < \infty$ **or** $\exists\, \phi : [0,\infty) \to [0,\infty)$ such that
   (a) $\phi \uparrow \infty$
   (b) $\phi/x \uparrow \infty$
   (c) $\sup_\alpha \mathbb{E}[\phi|X_\alpha|] < \infty$

Then $\{X_\alpha : \alpha \in I\}$ is uniformly integrable.

Proof. We prove only the first condition of the first part (tightness).

Suppose $\sup_\alpha \mathbb{E}|X_\alpha| = c < \infty$. Then:

$$\begin{aligned} \mathbb{P}(|X_\alpha| > b) &\leq \mathbb{E}|X_\alpha|/b \quad \text{(by Markov)} \\ &\leq c/b \\ \sup_\alpha \mathbb{P}(|X_\alpha| > b) &\leq c/b \end{aligned}$$

Sending $b \to \infty$ gives the definition of tightness.

□

Definition. (Extended distribution function/EDF)

An **extended distribution function** is a function $G : \mathbb{R} \to \mathbb{R}$ such that

1. G is increasing
2. G is right-continuous
3. $\lim_{x \to -\infty} G(x) \geq 0$ and $\lim_{x \to \infty} G(x) \leq 1$

Remark. Any distribution function satisfies the definition of an EDF with

$$\lim_{x \to -\infty} G(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} G(x) = 1$$

Theorem 5.9. (Helly's selection theorem)

Suppose $\{F_n\}_{n \geq 1}$ is a sequence of CDF's. Then there exists a subsequence $\{F_{n_j}\}_{j \geq 1}$ and an **extended** distribution function G such that $F_{n_j}(x) \to G(x)$ for all continuous points x of G.

Proof. The proof has two steps:

1. **Construct** G:

   Let $Q = \{q_1, q_2, \ldots\}$ be a denumeration of $\mathbb{Q}$. Define a function G_0 on Q by considering each point sequentially:

   Consider q_1. $\{F_n(q_1)\}_{n \geq 1}$ has a convergent subsequence by Bolzano-Weierstrass. Denote the corresponding subsequence of **functions** by $\{F_{n(1,j)}\}$ and define:

   $$G_0(q_1) = \lim_{j \to \infty} F_{n(1,j)}(q_1)$$

   Now consider q_2. $\{F_{n(1,j)}(q_2)\}$ **itself** has a further convergent subsequence by Bolzano-Weierstrass. Denote the corresponding subsequence of functions by $\{F_{n(2,j)}\}$ and define:

   $$G_0(q_2) = \lim_{j \to \infty} F_{n(2,j)}(q_2)$$

   and so on.

   Now set $n_i = n(i, i)$. Then $\{F_{n_i}\}$ is a subsequence of each $\{F_{n(i,j)}\}$. Thus for every $q \in Q$, we have $F_{n_i}(q) \to G_0(q)$ as $i \to \infty$.

Now G_0 is defined only on Q. So define the extension to $\mathbb{R}$ by:

$$G(x) = \inf_{q>x}\{G_0(q)\}, \;\; x \in \mathbb{R}$$

2. **Show that G is an EDF**:

   $0 \le G(x) \le 1$ since $0 \le F_n(x) \le 1$ for all n, and G is non-decreasing since each of the F_n's are non-decreasing. To see that G is right-continuous, fix $x \in \mathbb{R}$ and $\epsilon > 0$.

   By definition of inf, there exists $q_0 \in Q$ such that $G_0(q_0) < G(x) + \epsilon$. Then since G is non-decreasing, if $x \le y < q_0$ then $G(y) \le G_0(q_0) < G(x) + \epsilon$. That is, for $y \ge x$, whenever $y - x < q_0$ then $G(y) - G(x) < \epsilon$. So G is right-continuous.

3. **Show convergence at all continuous points**:

   Let x be a continuous point of G. We show that $\limsup_{j\to\infty} F_{n_j}(x) \le G(x)$ and $\liminf_{j\to\infty} F_{n_j}(x) \ge G(x)$.

   Fix $q \in Q > x$. Then $F_{n_j}(x) \le F_{n_j}(q)$ and:

   $$\limsup_{j\to\infty} F_{n_j}(x) \le \limsup_{j\to\infty} F_{n_j}(q) = G_0(q)$$

   This inequality holds for all $q > x$, so it also holds for $\inf_{q>x}$. Apply the definition of $G(x)$.

   The proof for $\liminf_{j\to\infty} F_{n_j}(x)$ is exactly analogous.

□

Corollary 5.10. Suppose the $\{F_n\}_{n\ge1}$ in Helly's selection theorem are **tight**. Then G is a distribution function.

Proof. We first show $\lim_{b\to-\infty} G(b) = 0$. We have just shown that for any continuous point b of G,

$$\lim_{j\to\infty} F_{n_j}(b) = \limsup_{j\to\infty} F_{n_j}(b) = G(b)$$

This immediately implies that

$$\sup_n F_n(b) \ge \limsup_{n\to\infty} F_n(b) \ge G(b)$$

Now since $\{F_n\}_{n\ge1}$ is tight, $\lim_{b\to\infty} \sup_n \big[F_n(-b) + (1 - F_n(-b))\big] = 0$.

This implies that $\lim_{b\to-\infty} \sup_n F_n(b) = 0$, which gives the result when combined with the above inequality.

The proof for $\lim_{b\to\infty} G(b) = 1$ is exactly analogous, using the fact that tightness also implies $\lim_{b\to\infty} \inf_n \big[F_n(b) + (1 - F_n(b))\big] = 0$.

□

Corollary 5.11. (Subsequence trick)

Suppose $\{X_n\}_{n\geq 1}$ are tight and X is some fixed distribution/random variable. By Helly's theorem, for any subsequence $\{X_{m_j}\}_{j\geq 1}$ there exists a further subsequence $\{X_{n_j}\}_{j\geq 1}$ and some Y such that

$$X_{n_j} \xrightarrow{d} Y \quad \text{as } j \to \infty$$

If $Y \overset{d}{=} X$ for **any** subsequence $\{m_j\}_{j\geq 1}$, then $X_n \xrightarrow{d} X$.

Proof. (by contradiction)

If $X_n \overset{d}{\nrightarrow} X$, then there exists $\epsilon > 0$, a continuous point x of F_x, and some subsequence $m_k \uparrow \infty$ such that

$$\left| \mathbb{P}(X_{m_j} \leq x) - \mathbb{P}(X \leq x) \right| > \epsilon \quad \forall m_j$$

This leads to the obvious contradiction.

□

5.3 Characteristic functions

Fact. (Two basic facts about $z \in \mathbb{C}$)

Representation: $z = x + iy$. Then:

1. $|z| = \sqrt{x^2 + y^2}$
2. For $x \in \mathbb{R}$, $e^{ix} = \cos x + i \sin x$ (i.e. e^{ix} is on the unit circle of $\mathbb{C}$)

Fact. (Two basic facts about a $\mathbb{C}$-valued **random variable** Z)

Representation: $Z = X + iY$ where X, Y are real-valued. Then:

1. $\mathbb{E}Z = \mathbb{E}X + i\mathbb{E}Y$
2. Jensen's inequality: $|\mathbb{E}Z| \leq \mathbb{E}|Z|$

Definition. (Characteristic function)

Suppose X is a real-valued r.v. Then the **characteristic function of** X is:

$$\begin{aligned}\phi_x(t) &= \mathbb{E}e^{itx},\ t \in \mathbb{R} \\ &= \mathbb{E}(\cos(tx) + i\sin(tx))\end{aligned}$$

Corollary 5.12. (Three easy facts)

1. Since e^{itx} is on the unit circle in $\mathbb{C}$, then

$$|\mathbb{E}e^{itx}| \leq \mathbb{E}|e^{itx}| \leq 1$$

2. The CF of a sum of independent r.v.'s is the product of the CF of each of the individual r.v.'s.

3. If μ is the distribution of X, then

$$\phi_X(t) = \int_\Omega e^{itx(\omega)}\,\mathrm{d}\mathbb{P} = \int_\mathbb{R} e^{itx}\,\mathrm{d}\mu(x)$$

Theorem 5.13. (CF's are continuous everywhere)

For any r.v. X and any $t \in \mathbb{R}$, $\phi_X(t+h) \to \phi_X(t)$ as $h \to 0$.

Proof.

$$\begin{aligned}\Big|\phi_X(t+h) - \phi_X(t)\Big| &= \Big|\mathbb{E}(e^{i(t+h)X} - e^{itX})\Big| \\ &= \Big|\mathbb{E}(e^{itX}(e^{ihX} - 1))\Big| \\ &\le \mathbb{E}\Big[|e^{itX}| \cdot |e^{ihX} - 1|\Big] \\ &= \mathbb{E}\Big|e^{ihX} - 1\Big|\end{aligned}$$

Now observe that $|e^{ihX} - 1| \to 0$ as $h \to 0$ and apply DCT using:

$$|e^{ihX} - 1| \le |e^{ihX}| + 1 = 2$$

□

Theorem 5.14. (The inversion formulas)

1. Suppose X has distribution μ and CF $\phi_X(\cdot)$. Then for real $a < b$:

$$\mu((a,b)) + \frac{1}{2}\mu(\{a\}) + \frac{1}{2}\mu(\{b\}) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \cdot \phi_X(t)\,\mathrm{d}t$$

2. If X has density $f_X(\cdot)$ with respect to Lebesgue measure, then:

$$\phi_X(t) = \mathbb{E}e^{itx} = \int_{-\infty}^{\infty} e^{itx} \cdot f_X(x)\,\mathrm{d}x$$

And we have:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} \cdot \phi_X(t)\,\mathrm{d}t$$

Remark. These formulas explicitly link CF's and distributions, showing that if two r.v.'s have the same CF, then they have the same distribution.

Example. (Two easy examples)

1. (Sum of normal r.v.'s)

If $X \sim N(\mu, \sigma^2)$, then $\phi_X(t) = \exp(it\mu - (t^2\sigma^2)/2)$.

Suppose $X_1 \sim N(0, \sigma_1^2) \perp X_2 \sim N(0, \sigma_2^2)$. Then:

$$\phi_{X_1+X_2}(t) = \exp\left(-\frac{t^2\sigma_1^2}{2}\right) \cdot \exp\left(-\frac{t^2\sigma_2^2}{2}\right) = \exp\left(-\frac{t^2}{2}(\sigma_1^2 + \sigma_2^2)\right)$$

therefore $X_1 + X_2 \sim N(0, \sigma_1^2 + \sigma_2^2)$.

2. (Exponential r.v.'s)

   Let $X \sim \exp(1)$, with $\phi_X(t) = \int_0^\infty e^{itx} e^{-x} \, \mathrm{d}x = \frac{1}{1-it}$.

   Suppose Y is double exponential. Then

$$f_Y(y) = \frac{1}{2}\exp(-|y|) = \frac{1}{2}\text{dist}(X) + \frac{1}{2}\text{dist}(-X)$$

   Therefore we have:

$$\begin{aligned}
\phi_Y(t) &= \frac{1}{2}\phi_X(t) + \frac{1}{2}\phi_{-Y}(t) \\
&= \frac{1}{2}\phi_X(t) + \frac{1}{2}\phi_Y(-t) \\
&= \frac{1}{2} \cdot \frac{1}{1-it} + \frac{1}{2} \cdot \frac{1}{1+it} \\
&= \frac{1}{2} \cdot \frac{2}{1-i^2t^2} \\
&= \frac{1}{1+t^2}
\end{aligned}$$

   which is the CF of a standard Cauchy distribution.

Theorem 5.15. (Dual pairs)

Let ϕ_X be the CF of a r.v. X.

Assume $\phi_X(t) \in \mathbb{R}^+ \setminus \{0\}$ for all $t \in \mathbb{R}$ and $\int_{-\infty}^\infty \phi_X(t) \, \mathrm{d}t < \infty$. Then:

1. $\frac{\phi_X(t)}{2\pi f_X(0)}$ is a pdf
2. If Y has the above pdf $\frac{\phi_X(t)}{2\pi f_X(0)}$, then $\phi_Y(t) = \frac{f_X(t)}{f_X(0)}$.

   Proof. We prove the two parts in order:

   1. Let ϕ_X be a CF satisfying the conditions of (1). Note $\phi_X(t) \in (0,1]$. We need to show that the quantity integrates to 1. By the second inversion formula,

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^\infty e^{-itx} \cdot \phi_X(t) \, \mathrm{d}t, \quad f_X(0) = \frac{1}{2\pi}\int_{-\infty}^\infty \phi_X(t) \, \mathrm{d}t$$

      Therefore we have:

$$\int_{-\infty}^\infty \frac{1}{2\pi f_X(0)} \cdot \phi_X(t) \, \mathrm{d}t = 1$$

   2. First, note that if $\phi_X \in \mathbb{R}$, then it has no imaginary component:

$$\phi_X(t) = \mathbb{E}(\cos(tx)) + i \cdot \mathbb{E}(\sin(tx)) = \mathbb{E}(\cos(tx))$$

      Then since $\cos(tx) = \cos(-tx)$, ϕ_X is symmetric about 0. Now write ϕ_Y as a function x to avoid confusion later:

$$\begin{aligned}\phi_Y(x) &= \mathbb{E}e^{ixy} \\ &= \int_{-\infty}^{\infty} e^{ixy} \cdot f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} e^{ixy} \cdot \frac{\phi_X(y)}{2\pi f_X(0)}\,\mathrm{d}y\end{aligned}$$

Now since we are integrating over the entire real line, we can make the change of variable $t = -y$ and use symmetry of ϕ_X to obtain:

$$\phi_Y(x) = \int_{-\infty}^{\infty} e^{-ixt} \cdot \frac{\phi_X(t)}{2\pi f_X(0)}\,\mathrm{d}t$$

Applying the second inversion formula gives the result.

□

Example. (Double exponential/Cauchy)

Let $X \sim$ double exponential. Then we have

$$f_X(x) = \frac{1}{2}\exp(-|x|), \quad \phi_X(t) = \frac{1}{1+t^2}$$

Then the dual density is Cauchy:

$$f_Y(t) = \frac{\phi_X(t)}{2\pi f_X(0)} = \frac{1}{\pi(1+t^2)}$$

with corresponding CF:

$$\phi_Y(t) = \frac{f_X(t)}{f_X(0)} = e^{-|t|}$$

Also note that the sample mean of independent Cauchy r.v.'s is again Cauchy:

$$\phi_{\bar{Y}}(t) = \phi_{S_n}(t/n) = \prod_{i=1}^{n} \phi_{Y_i}(t/n) = e^{-|t|}$$

5.4 The continuity theorem, iid CLT

Fact. (Important identity for e)

If $c_n \to c \in \mathbb{C}$, then $(1 + c_n/n)^n \to e^c$.

Theorem 5.16. (Parseval identity)

Suppose μ, ν are probability measures on $\mathbb{R}$ (e.g. distributions of some random variables X, Y). The corresponding CF's are:

$$\phi_\mu(t) = \int_{-\infty}^{\infty} e^{itx}\,\mathrm{d}\mu(x), \qquad \phi_\nu(t) = \int_{-\infty}^{\infty} e^{itx}\,\mathrm{d}\nu(x)$$

Then in fact:

$$\int_{-\infty}^{\infty} \phi_\mu(t)\,\mathrm{d}\nu(t) = \int_{-\infty}^{\infty} \phi_\nu(t)\,\mathrm{d}\mu(t)$$

Remark. This tells us that we can always get information about a probability measure μ using its CF ϕ_μ and a "good" probability measure ν of one's choice.

Proof. Suppose $X \sim \mu \perp Y \sim \nu$. Then:

$$\begin{aligned}\mathbb{E}e^{iXY} &= \mathbb{E}\Big[\mathbb{E}(e^{iXY} \,|\, Y)\Big] \\ &= \mathbb{E}(\phi_X(Y)) \\ &= \int_{-\infty}^{\infty} \phi_\mu(t)\, \mathrm{d}\nu(t)\end{aligned}$$

Conditioning with respect to X and following the same steps gives the other side of the equality.

□

Corollary 5.17. If μ is some probability measure and $X \sim \mu$, then:

$$\mathbb{E}\left[\frac{\sin(cx)}{cx}\right] = \frac{1}{2c}\int_{-c}^{c} \phi_\mu(t)\, \mathrm{d}t$$

Proof. We apply Parseval's identity with $\nu =$ the Uniform probability measure on $[-c, c]$ for $c > 0$. The density is: $f_\nu(x) = 1/2c$, $-c < x < c$, so the characteristic function is:

$$\begin{aligned}\phi_\nu(t) &= \int_{-c}^{c} e^{itx} \cdot \frac{1}{2c}\, \mathrm{d}x \\ &= \int_{-c}^{c} \frac{\cos(tx)}{2c}\, \mathrm{d}x + i\int_{-c}^{c} \frac{\sin(tx)}{2c}\, \mathrm{d}x \\ &= \int_{-c}^{c} \frac{\cos(tx)}{2c}\, \mathrm{d}x \\ &= \frac{\sin(tc)}{tc}\end{aligned}$$

Now let μ be a probability measure on $\mathbb{R}$. The LHS in Parseval's identity is:

$$\int_{-\infty}^{\infty} \phi_\mu(t)\, \mathrm{d}\nu(t) = \frac{1}{2c}\int_{-c}^{c} \phi_\mu(t)\, \mathrm{d}t$$

And the RHS is:

$$\int_{-\infty}^{\infty} \phi_\nu(t)\, \mathrm{d}\mu(t) = \int_{-\infty}^{\infty} \frac{\sin(tc)}{tc}\, \mathrm{d}\mu(t) = \mathbb{E}\left[\frac{\sin(cx)}{cx}\right]$$

Where the last equality follows because $X \sim \mu$. Therefore LHS = RHS by Parseval and we are done.

□

Theorem 5.18. (Converse to continuity theorem)

Suppose $X_n \xrightarrow{d} X$. Then $\phi_{X_n}(t) \to \phi_X(t)$ for all $t \in \mathbb{R}$.

Proof. If $X_n \xrightarrow{d} X$, then $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$ for all bdd cts fns g.

□

Theorem 5.19. (Continuity theorem)

Suppose $\{X_n\}_{n\geq 1}$ are real-valued r.v.'s with corresponding CF's $\{\phi_n\}_{n\geq 1}$. Suppose there exists some function $\phi_\infty(t)$ such that $\forall t \in \mathbb{R}$, $\phi_n(t) \to \phi_\infty(t)$. If **either** of the following hold:

1. $\{X_n\}_{n\geq 1}$ is tight.
2. $\phi_\infty(t) \to 1$ as $t \to 0$

Then $X_n \xrightarrow{d} X$, where $\phi_X(t) = \phi_\infty(t)$.

Remark. Since CF's are continuous everywhere, then they are continuous at 0. Also, $\phi(0) = 1$ for any CF ϕ. Thus if one can show that ϕ_∞ is a CF of some r.v., then (2) is automatically satisfied.

Proof. We first show (1), and then show that (2) implies (1).

1. Assume $\{X_n\}_{n\geq 1}$ is tight.

   Then given any subsequence $\{m_j\} \uparrow \infty$, there exists a further subsequence $\{n_j\} \subset \{m_j\}$ such that $X_{n_j} \xrightarrow{d} Y$ as $j \to \infty$. Thus $\phi_{n_j}(t) \to \phi_Y(t)$ for some Y. But since $\phi_n(t) \to \phi_\infty(t)$, then by uniqueness of limits $\phi_Y(t) = \phi_\infty(t)$.

   Thus since this holds for arbitrary subsequences and CF's uniquely characterize distributions, then for any subsequence $\{m_j\}$, there exists a further subsequence $\{n_j\}$ such that $X_{n_j} \xrightarrow{d} Y$. Thus by the subsequence trick, $X_n \xrightarrow{d} Y$ where Y has CF $\phi_\infty(t)$.

2. We want to show that $\limsup_{k\to\infty} \sup_n \mathbb{P}(|X_n| > k) = 0$. First observe that:

   $$\mathbb{P}(|X_n| > k) \leq 2 \cdot \mathbb{E}\left[\left(1 - \frac{k}{2|X_n|}\right) \cdot \mathbb{1}_{|X_n|>k}\right]$$

   To see this, note that if $|X_n| = k$, then $(1 - \frac{k}{2|X_n|}) = \frac{1}{2}$ and the relation holds with equality. Then note that on the set $|X_n| > k$ we have $(1 - \frac{k}{2|X_n|}) > \frac{1}{2}$. Now in order to introduce CF's into the picture,

   $$\begin{aligned}\mathbb{P}(|X_n| > k) &\leq 2 \cdot \mathbb{E}\left[\left(1 - \frac{1}{c|X_n|}\right) \cdot \mathbb{1}_{|X_n|>k}\right] \quad (c = 2/k)\\ &\leq 2 \cdot \mathbb{E}\left[\left(1 - \frac{\sin(c|X_n|)}{c|X_n|}\right) \cdot \mathbb{1}_{|X_n|>k}\right]\\ &= 2 \cdot \mathbb{E}\left[\left(1 - \frac{\sin(cX_n)}{cX_n}\right) \cdot \mathbb{1}_{|X_n|>k}\right]\\ &\leq 2 \cdot \mathbb{E}\left[1 - \frac{\sin(cX_n)}{cX_n}\right]\end{aligned}$$

   Applying the corollary to Parseval's identity, we obtain

   $$\mathbb{P}(|X_n| > k) \leq \frac{1}{c}\int_{-c}^{c} (1 - \phi_n(t))\,\mathrm{d}t$$

Now $\phi_n(t) \to \phi_\infty(t)$ as $n \to \infty$ and $|\phi_n(t)| \le 1$ for all n, so by DCT,

$$\int_{-c}^{c} (1 - \phi_n(t))\,\mathrm{d}t \to \int_{-c}^{c} (1 - \phi_\infty(t))\,\mathrm{d}t$$

Thus we can take limsups of both sides in the above inequality to obtain:

$$\limsup_{n\to\infty} \mathbb{P}(|X_n| > k) \le \frac{1}{c}\int_{-c}^{c} (1 - \phi_\infty(t))\,\mathrm{d}t$$

We want to show that $\lim_{k\to\infty}$ LHS equals zero. Since $c = \frac{2}{k}$, then it is equivalent to show that $\lim_{c\to 0}$ RHS equals zero.To see this, note that the condition $\phi_\infty(t) \to 1$ as $t \to 0$ guarantees that:

$$\lim_{c\to 0} \int_{-c}^{c} (1 - \phi_\infty(t))\,\mathrm{d}t = 0$$

Therefore by an application of L'Hopital's rule:

$$\lim_{c\to \infty} \frac{1}{c}\int_{-c}^{c} (1 - \phi_\infty(t))\,\mathrm{d}t = 0$$

Thus we have shown $\limsup_k \limsup_n \mathbb{P}(|X_n| > k) = 0$.

□

Fact. (Inequality from complex analysis)

$$\left| e^{iy} - \sum_{k=0}^{m} \frac{(iy)^k}{k!} \right| \le \min\left\{ \frac{2|y|^m}{m!}, \frac{|y|^{m+1}}{(m+1)!} \right\}$$

Lemma 5.20. (Expansion lemma)

Fix $m \ge 1$ and let X be such that $\mathbb{E}(|X|^m) < \infty$. Then:

$$\phi_X(t) = \sum_{k=0}^{m} \frac{(it)^k}{k!}\mathbb{E}(X^k) + \mathrm{o}\left(|t|^m\right) \quad \text{as } t \to 0$$

Remark. What is the motivation for this? Note the deterministic identity:

$$e^{itx} = \sum_{m=0}^{\infty} \frac{(itx)^m}{m!}$$

which **suggests** possibly:

$$\mathbb{E}(e^{itx}) \;=\; \sum_{m=0}^{\infty} \mathbb{E}\left[\frac{(itx)^m}{m!}\right] \;\stackrel{?}{=}\; \sum_{m=0}^{\infty} \frac{(it)^m}{m!}\mathbb{E}(X^m)$$

Proof. Using $y = tx$ in the above inequality and applying Jensen's inequality, we can bound the remainder:

$$\left| \mathbb{E}(e^{itx}) - \sum_{k=0}^{m} \frac{(it)^k}{k!}\mathbb{E}(X^k) \right| \le \mathbb{E}\left[\min\left\{ \frac{2|t|^m|X|^m}{m!}, \frac{|t|^{m+1}|X|^{m+1}}{(m+1)!} \right\}\right]$$

$$= \frac{|t|^m}{m!} \cdot \mathbb{E}\left[\min\left\{ 2|X|^m, \frac{|t| \cdot |X|^{m+1}}{m+1} \right\}\right]$$

We want to show that the RHS is $\mathrm{o}(|t|^m)$. So if we divide the RHS by $|t|^m$, then we only need to show that the expectation $\to 0$ as $t \to 0$.

Let $z_t = \min\{\cdot, \cdot\}$. Note that $z_t \to 0$ as $t \to 0$, and there exists t_0 such that:

$$|z_t| \leq 2|X|^m \quad \forall\, |t| < |t_0|$$

And since $\mathbb{E}(|X|^m) < \infty$, then the result follows by DCT.

□

Theorem 5.21. (Weak Law of Large Numbers)

Let $\{Y_i\}_{i\geq 1}$ be iid and $\mathbb{E}|Y_i| < \infty$. Let $\mathbb{E}Y_i = \mu$. Then $S_n/n \xrightarrow{d} \mu$.

Remark. This implies $S_n/n \xrightarrow{\mathbb{P}} \mu$ since μ is a constant.

Proof. We want to show that $\phi_{S_n/n}(t) \to \phi_\mu(t) = e^{it\mu}$. Note:

$$\begin{aligned}
\phi_{S_n/n}(t) &= \mathbb{E}\left[e^{it(S_n/n)}\right] \\
&= \phi_{S_n}(t/n) \\
&= \left(\phi_{Y_i}(t/n)\right)^n \\
&= \left[1 + \frac{n(\phi_{Y_i}(t/n) - 1)}{n}\right]^n
\end{aligned}$$

So by the identity at the beginning of the section, we want to show that $n(\phi_{Y_i}(t/n) - 1) \to it\mu$. By the expansion lemma,

$$\phi_{Y_i}(t/n) = 1 + \frac{it\mu}{n} + \mathrm{o}(|t|/n)$$

Therefore:

$$\begin{aligned}
n(\phi_{Y_i}(t/n) - 1) &= it\mu + n \cdot \mathrm{o}(|t|/n) \\
&= it\mu + |t| \cdot \frac{\mathrm{o}(|t|/n)}{|t|/n} \\
&\to it\mu \quad \text{as } n \to \infty
\end{aligned}$$

□

Theorem 5.22. (iid Central Limit Theorem)

Let $\{Y_i\}_{i\geq 1}$ be iid with $\mathbb{E}Y_i = 0$ and $\mathbb{E}Y_i^2 = \sigma^2 < \infty$. Then $S_n/\sqrt{n} \to N(0, \sigma^2)$.

Proof. By the continuity theorem, sufficient to show that for fixed $t \in \mathbb{R}$,

$$\phi_{S_n/n}(t) \to \exp\left(-\frac{t^2\sigma^2}{2}\right)$$

As in the proof of the WLLN above,

$$\begin{aligned}\phi_{S_n/\sqrt{n}}(t) &= \phi_{S_n}(t/\sqrt{n}) \\ &= \left(\phi_{Y_i}(t/\sqrt{n})\right)^n \\ &= \left[1 + \frac{n(\phi_{Y_i}(t/\sqrt{n}) - 1)}{n}\right]^n\end{aligned}$$

Again, we need to show that $n(\phi_{Y_i}(t/\sqrt{n}) - 1) \to -t^2\sigma^2/2$. By the expansion lemma,

$$\begin{aligned}\phi_{Y_i}(t) &= 1 + it\mathbb{E}Y_i + \frac{i^2t^2}{2}\mathbb{E}Y_i^2 + \mathrm{o}\left(|t|^2\right) \\ &= 1 - \frac{t^2\sigma^2}{2} + \mathrm{o}\left(|t|^2\right) \\ \phi_{Y_i}(t/\sqrt{n}) &= 1 - \frac{t^2\sigma^2}{2n} + \mathrm{o}\left(|t|^2/n\right)\end{aligned}$$

This gives us that:

$$n(\phi_{Y_i}(t/\sqrt{n}) - 1) = -\frac{t^2\sigma^2}{2} + n \cdot \mathrm{o}\left(|t|^2/n\right)$$

Noting that $n \cdot \mathrm{o}\left(|t|^2/n\right) \to 0$ as $n \to \infty$ completes the proof.

□

5.5 Lindeberg-Feller CLT

Lemma 5.23. (Important fact)

Suppose $w_1, \dots, w_n$ and $z_1, \dots, z_n$ are $\in \mathbb{C}$ with $|w_i| \le 1$, $|z_i| \le 1$. Then:

$$\left|\prod_{i=1}^n w_i - \prod_{i=1}^n z_i\right| \le \sum_{i=1}^n |w_i - z_i|$$

Proof. Expand the difference of products inside the modulus like so:

$$\begin{aligned}\prod_{i=1}^n w_i - \prod_{i=1}^n z_i = &\left(\prod_{i=1}^n w_i - z_1\prod_{i=2}^n w_i\right) \\ &+ \left(z_1\prod_{i=2}^n w_i - z_1z_2\prod_{i=3}^n w_i\right) \\ &+ \left(z_1z_2\prod_{i=3}^n w_i - \dots\right) \\ &+ (\dots + z_1z_2\dots z_{n-1}w_n) - z_1z_2\dots z_n\end{aligned}$$

Now note that for the first group in the RHS,

$$\left|\prod_{i=1}^n w_i - z_1\prod_{i=2}^n w_i\right| = \prod_{i=2}^n |w_i| \cdot |w_1 - z_1| \le |w_1 - z_1|$$

And for the second group in the RHS,

$$\left| z_1 \prod_{i=2}^{n} w_i - z_1 z_2 \prod_{i=3}^{n} w_i \right| = z_1 \prod_{i=3}^{n} |w_i| \cdot |w_2 - z_2| \leq |w_2 - z_2|$$

And so on.

□

Theorem 5.24. (Lindeberg-Feller CLT)

Let $X_{n,k}$ for $k \leq n, n \geq 1$ be a triangular array of mean-zero random variables with $\mathbb{E}X_{n,i}^2 = \sigma_{n,i}^2 < \infty$. Assume that:

1. Within n, $X_{n,1}, \ldots, X_{n,n}$ are defined on the same probability space and are independent.
2. There exists $\sigma^2 < \infty$ such that $\sum_{i=1}^n \sigma_{n,i}^2 \to \sigma^2$ as $n \to \infty$.
3. For any $\epsilon > 0$, $\sum_{i=1}^n \mathbb{E}(X_{n,i}^2 \cdot \mathbb{1}_{|X_{n,i}|>\epsilon}) \to 0$.

Let $S_n = \sum_{i=1}^n X_{n,i}$ and $\text{Var}(S_n) = \sum_{i=1}^b \text{Var}(X_{n,i}) = \sum_{i=1}^n \sigma_{n,i}^2$. Then:

$$S_n \xrightarrow{d} N(0, \sigma^2)$$

Proof. Define $\phi_{n,i}(t) = \mathbb{E}(e^{itX_{n,i}})$. By the expansion lemma,

$$\phi_{S_n}(t) = \prod_{i=1}^{n} \phi_{n,i}(t) \quad \text{and} \quad \phi_{n,i}(t) = 1 - \frac{t^2 \sigma_{n,i}^2}{2} + \text{o}\left(t^2\right)$$

Step 1: show that $\sum_{i=1}^n \sigma_{n,i}^4 \to 0$ as $n \to \infty$.

First we show that $\max_{1 \leq j \leq n} \sigma_{n,i}^2 \to 0$ as $n \to \infty$:

$$\begin{aligned}
\sigma_{n,i}^2 &= \mathbb{E}(X_{n,i}^2) \\
&= \mathbb{E}(X_{n,i}^2 \cdot \mathbb{1}_{|X_{n,i}|>\epsilon}) + \mathbb{E}(X_{n,i}^2 \cdot \mathbb{1}_{|X_{n,i}|\leq\epsilon}) \\
&\leq \sum_{j=1}^{n} \mathbb{E}(X_{n,j}^2 \cdot \mathbb{1}_{|X_{n,j}|>\epsilon}) + \epsilon^2
\end{aligned}$$

The summation in the last step frees the bound from dependence on i. So:

$$\max_{1 \leq i \leq n} \sigma_{n,i}^2 \leq \sum_{j=1}^{n} \mathbb{E}(X_{n,j}^2 \cdot \mathbb{1}_{|X_{n,j}|>\epsilon}) + \epsilon^2$$

The first term in the RHS $\to 0$ as $n \to \infty$ by UAN, and since ϵ^2 is arbitrary then $\max_{1 \leq j \leq n} \sigma_{n,i}^2 \to 0$. Now note that:

$$\sum_{i=1}^{n} \sigma_{n,i}^4 \leq \max_{1 \leq j \leq n} \sigma_{n,j}^2 \cdot \sum_{i=1}^{n} \sigma_{n,i}^2$$

Send $n \to \infty$ and apply the result just shown.

Step 2: show that $\phi_{S_n}(t) \to \prod_{i=1}^n \left[1 - \frac{t^2\sigma_{n,i}^2}{2}\right]$ as $n \to \infty$

To show (2), we combine the lemma at the top of this section with the inequality from the previous section:

$$\begin{aligned}
\left| \prod_{i=1}^n \phi_{n,i}(t) - \prod_{i=1}^n \left[1 - \frac{t^2\sigma_{n,i}^2}{2}\right] \right| &\le \sum_{i=1}^n \left| \phi_{n,i}(t) - \left[1 - \frac{t^2}{2}\sigma_{n,i}^2\right]\right| \\
&\le \sum_{i=1}^n \mathbb{E}\left[\min\left\{\frac{|t|^3|X_{n,i}|^3}{3!}, \frac{2t^2X_{n,i}^2}{2!}\right\}\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[\min\left\{t^2X_{n,i}^2, \frac{|t|^3|X_{n,i}|^3}{6}\right\}\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[\min\left\{t^2X_{n,i}^2, \frac{|t|^3|X_{n,i}|^3}{6}\right\} \cdot \mathbb{1}_{|X_{n,i}|>\epsilon}\right] \\
&\quad + \sum_{i=1}^n \mathbb{E}\left[\min\left\{t^2X_{n,i}^2, \frac{|t|^3|X_{n,i}|^3}{6}\right\} \cdot \mathbb{1}_{|X_{n,i}|<\epsilon}\right] \\
&\le |t^2| \sum_{i=1}^n \mathbb{E}\Big(X_{n,i}^2 \cdot \mathbb{1}_{|X_{n,i}|>\epsilon}\Big) \\
&\quad + \sum_{i=1}^n \mathbb{E}\Big(\epsilon \cdot |t|^3 \cdot |X_{n,i}|^2\Big)
\end{aligned}$$

As $n \to \infty$, the first term $\to 0$ by UAN, and the second term $\to \epsilon \cdot |t|^3\sigma^2$. So,

$$\limsup_{n\to\infty} \left| \phi_{S_n}(t) - \prod_{i=1}^n \left[1 - \frac{t^2\sigma_{n,i}^2}{2}\right] \right| \le \epsilon \cdot |t|^3\sigma^2$$

Since this holds for any $\epsilon > 0$, then the remainder $\to 0$ as $n \to \infty$ so

$$\phi_{S_n}(t) \to \prod_{i=1}^n \left[1 - \frac{t^2\sigma_{n,i}^2}{2}\right] \quad \text{as } n \to \infty$$

Step 3: show that $\prod_{i=1}^n \left[1 - \frac{t^2\sigma_{n,i}^2}{2}\right] \to \exp\left(-\frac{t^2\sigma^2}{2}\right)$ as $n \to \infty$

Equivalently, we show $\sum_{i=1}^n \log\left(1 - \frac{t^2}{2}\sigma_{n,i}^2\right) \to -\frac{t^2\sigma^2}{2}$. Note the identity:

$$\Big|\log(1-x) - (-x)\Big| \le cx^2 \quad \text{for } 0 \le x \le \frac{1}{2}$$

Applying this to our series, we obtain:

$$\begin{aligned}
&\left| \sum_{i=1}^n \left\{ \log\left(1 - \frac{t^2\sigma_{n,i}^2}{2}\right) - \left(-\frac{t^2\sigma_{n,i}^2}{2}\right)\right\} \right| \\
&\qquad = \left| \sum_{i=1}^n \log\left(1 - \frac{t^2\sigma_{n,i}^2}{2}\right) - \left(-\frac{t^2}{2}\sum_{i=1}^n \sigma_{n,i}^2\right) \right| \le \frac{t^4}{4}\sum_{i=1}^n \sigma_{n,i}^4 \cdot c
\end{aligned}$$

The second term in the modulus $\to \frac{t^2}{2}\sigma^2$ and the RHS sum $\to 0$ by the result of part 1, and the proof is complete.

□

Corollary 5.25. (Typical application)

Let $\{Y_{n,i}\}_{1\leq i\leq n}$ be independent with $\mathbb{E}Y_{n,i} = 0$. Define:

$$S_n = \sum_{i=1}^{n} Y_{n,i}, \qquad \mathcal{S}_n^2 = \sum_{i=1}^{n} \sigma_{n,i}^2 = \mathrm{Var}(S_n)$$

Suppose that for any $\epsilon > 0$:

$$\sum_{i=1}^{n} \mathbb{E}\left(\frac{Y_{n,i}^2}{\mathcal{S}_n^2} \cdot \mathbb{1}_{|Y_{n,i}/\mathcal{S}_n|>\epsilon}\right) \to 0$$

Then $S_n/\mathcal{S}_n \to N(0,1)$.

Proof. Let $X_{n,i} = Y_{n,i}/\mathcal{S}_n$ and apply the Lindeberg-Feller CLT.

□

Theorem 5.26. (Lyapunov's condition)

Suppose $\exists\delta > 0$ such that

1. $\mathbb{E}\left(|Y_{n,i}|^{2+\delta}\right) < \infty$ for all $1 \leq i \leq n$
2. $L_n = \sum_{i=1}^{n} \mathbb{E}\left(|Y_{n,i}|^{2+\delta}\right) / \mathcal{S}_n^{2+\delta} \to 0$

Where $\mathcal{S}_n = \sqrt{\sum_{i=1}^{n} \mathbb{E}(X_{n,i}^2)} = \sqrt{\mathrm{Var}(S_n)}$. Then $S_n/\mathcal{S}_n \to N(0,1)$.

Proof. We show that the condition of the "Typical application" is satisfied.

Fix some $\epsilon > 0$. Note that:

$$\frac{Y_{n,i}^2}{\mathcal{S}_n^2} \cdot \mathbb{1}_{|Y_{n,i}|/\mathcal{S}_n>\epsilon} \leq \frac{Y_{n,i}^2}{\mathcal{S}_n^2}\left(\frac{|Y_{n,i}|}{\epsilon\cdot\mathcal{S}_n}\right)^{\delta}$$

Therefore it follows that:

$$\sum_{i=1}^{n} \mathbb{E}\left\{\frac{Y_{n,i}^2}{\mathcal{S}_n^2} \cdot \mathbb{1}_{|Y_{n,i}|/\mathcal{S}_n>\epsilon}\right\} \leq \frac{L_n}{\epsilon^{\delta}} \to 0 \quad \text{as } n \to \infty$$

□