Probability

Zhouyi Tan

December 19, 2016

Abstract

These notes are prepared based on *Probability: Theory and Examples, Third Edition* by *Rick Durrett.*

1 Law of Large Numbers

1.1 Basic Definitions

In this first section, we introduce the basic framework of probability theory.

Definition 1.1.1. A **σ-algebra** $\mathcal{F}$ on an arbitrary set Ω is a nonempty collection of subsets of Ω that satisfies the following conditions:

(i) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$;

(ii) if $A_n \in \mathcal{F}$ for each $n \in \mathbb{N}$, then $\cup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

Definition 1.1.2. Without $\mathbb{P}$, $(\Omega, \mathcal{F})$ is called a **measurable space** on which we can define a measure. A **measure** $\mu : \mathcal{F} \to [0, +\infty]$ is a set function that satisfies the following conditions:

(i) $\mu(\varnothing) = 0$;

(ii) if $(A_n)_{n=1}^{\infty}$ is a sequence of disjoint elements in $\mathcal{F}$, then $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.

If, in addition, $\mu(\Omega) = 1$, we call μ a **probability measure**.

Definition 1.1.3. A **probability space** is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is an arbitrary set, $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$.

Remark 1.1.4. In probability language, we call Ω the set of "outcomes", and $\mathcal{F}$ the set of "events". Clearly we want to assign probabilities to as many events as we want, but due to the axiom of choice, there exist non-measurable sets even if we simply assign as the probability to intervals its own length in $\mathbb{R}$.

Remark 1.1.5. Some direct arguments can verify monotonicity, subadditivity and continuity of probability measures.

Example 1.1.6 (Discrete probability space)**.** Let Ω be an at most countable space and let $\mathcal{F} = 2^{\Omega}$. Define $\mathbb{P}(A) = \sum_{\omega\in A} p(\omega)$ where $p(\omega) \geq 0$ and $\sum_{\omega\in\Omega} p(\omega) = 1$. This is the most general probability measure on this space.

Example 1.1.7 (Lebesgue measure on the unit interval)**.** Assume Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Consider its restriction on the the unit interval $(0, 1)$. Let $\Omega = (0, 1)$, $\mathcal{F} = \{B \cap (0, 1) : B \in \mathcal{B}(\mathbb{R})\}$ and $\mathbb{P}(B) = \lambda(B)$ for all $B \in \mathcal{F}$.

Remark 1.1.8. Given a collection of σ-algebras, one can easily check the intersection is a σ-algebra.

Definition 1.1.9. Given a set Ω and a collection $\mathcal{A}$ of subsets Ω, the smallest σ-algebra is called the **σ-algebra generated by** $\mathcal{A}$.

Example 1.1.10 (Product measures)**.** Let $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i), 1 \le i \le n$, be a collection of probability space. On the product space $\prod_{i=1}^n \Omega_i = \{(\omega_1, ..., \omega_n) : \omega_i \in \Omega_i\}$, define the product σ-algebra $\prod_{i=1}^n \mathcal{F}_i$ which is the σ-algebra generated by the collection $\mathcal{P} = \{R_1 \times \cdots \times R_n : R_i \in \mathcal{F}_i\}$. Then we can define the product measure

$$\mathbb{P}(A_i \times \cdots \times A_n) = \prod_{i=1}^n \mathbb{P}_i(A_i).$$

Definition 1.1.11. A real-valued function $X : \Omega \to \mathbb{R}$ is said to be a **random variable** if $X^{-1}(B) \in \mathcal{F}$ for all Borel sets $B \in \mathcal{B}(\mathbb{R})$.

Given a random variable, we can push forward the measure on the target space, i.e., the real line. Simply for any Borel set $B \in \mathcal{B}(\mathbb{R})$ we define the measure $\mu(B) = \mathbb{P}(X \in B)$. It is clear that μ is a bona fide measure on $\mathbb{R}$ and this measure is called the **distribution** of X. Also we can use the **distribution function**, $F(x) := \mathbb{P}(X \le x)$ to study the distribution of X.

Theorem 1.1.12. *A distribution function F has the following properties:*

(i) *F is increasing;*

(ii) *$\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$;*

(iii) *F is right-continuous, i.e. $\lim_{y\to x^+} F(y) = F(x)$;*

(iv) *If $F(x-) = \lim_{y\to x^-} F(y)$ then $F(x-) = \mathbb{P}(X < x)$;*

(v) *$\mathbb{P}(X = x) = F(x) - F(x-)$.*

Theorem 1.1.13. *Given an increasing, right-continuous function F that satisfies $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$, there is a random variable X with distribution function F.*

Corollary 1.1.14. *Suppose F satisfies the condition as before, then there is a unique probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mu((a, b]) = F(b) - F(a)$ for all $a < b$.*

Whenever the distribution function F is of the form $F(x) = \int_{-\infty}^x f(t)dt$, we say that X has the **density** function f.

Example 1.1.15 (Normal distribution)**.** We say X is a standard normal distribution if its density function is $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$.

There is a useful tail bound for normal random variables.

Theorem 1.1.16. *If X is a standard Gaussian random variable, then for each $x > 0$,*

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \exp(-x^2/2) \le \sqrt{2\pi}\mathbb{P}(X > x) \le \frac{1}{x} \exp(-x^2/2).$$

1.2 Random Variables

1.3 Expected Value

1.4 Independence

Definition 1.4.1. Two events A, B are **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Definition 1.4.2. A finite number of collection of events $\mathcal{A}_1, ..., \mathcal{A}_n$ are independent if for any $I \subset \{1, 2, ..., n\}$,

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

whenever $A_i \in \mathcal{A}_i$.

We shall given several criteria to determine independence. To state the main result, we need Dynkin's π-λ theorem.

Definition 1.4.3. A collection $\mathcal{C}$ is called a **π-system** if it is closed under intersection.

Definition 1.4.4. A collection $\mathcal{C}$ is called a **λ-system** if
(i) $\Omega \in \mathcal{C}$;
(ii) $A, B \in \mathcal{C}$ and $A \subset B$ imply $B \setminus A \in \mathcal{C}$;
(iii) $A_n \in \mathcal{C}$ for each n and $A_n \uparrow A$ imply $A \in \mathcal{C}$.

Theorem 1.4.5 (Dynkin)**.** *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system that contains $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

The main result is the following:

Theorem 1.4.6. *If $\mathcal{A}_1, \mathcal{A}_2, ...\mathcal{A}_n$ are independent and each $\mathcal{A}_i$ is a π-system, then $\sigma(\mathcal{A}_1), ..., \sigma(\mathcal{A}_n)$ are independent.*

Proof. Let $A_i \in \mathcal{A}_i$ for $1 \le i \le n$. Denote $F = A_2 \cap \cdots \cap A_n$. Consider the following collection

$$\mathcal{L} = \{A \subset \Omega : \mathbb{P}(A \cap F) = \mathbb{P}(A)\mathbb{P}(F)\}.$$

Clearly (i) $\Omega \in \mathcal{L}$.
(ii) Suppose $A, B \in \mathcal{L}$ and $A \subset B$. Note that $(B \setminus A) \cap F = (B \cap F) \setminus (A \cap F)$, which implies

$$\mathbb{P}((B \setminus A) \cap F) = \mathbb{P}(B \cap F) - \mathbb{P}(A \cap F) = (\mathbb{P}(B) - \mathbb{P}(A))\mathbb{P}(F) = \mathbb{P}(B \setminus A)\mathbb{P}(F).$$

This shows $B \setminus A \in \mathcal{F}$.
(iii) Suppose $B_n \uparrow B \in \mathcal{L}$. We calculate

$$\mathbb{P}(B \cap F) = \lim_{n\to\infty} \mathbb{P}(B_n \cap F) = \lim_{n\to\infty} \mathbb{P}(B_n)\mathbb{P}(F) = \mathbb{P}(B)\mathbb{P}(F),$$

which shows $B \in \mathcal{L}$.

Therefore, $\mathcal{L}$ is a λ-system containing $\mathcal{A}_1$. As F is the intersection of an arbitrary finite collection of sets from $\mathcal{A}_2, ..., \mathcal{A}_n$, we see that $\sigma(\mathcal{A}_1), \mathcal{A}_2, ..., \mathcal{A}_n$ are independent. Then repeating the argument to the collection $\mathcal{A}_2, ..., \mathcal{A}_n, \sigma(\mathcal{A}_1)$ yields that $\sigma(\mathcal{A}_2), ..., \mathcal{A}_n, \sigma(\mathcal{A}_1)$ are independent. Iterate this procedure $n-1$ times will conclude the proof. □

We can apply this theorem to give several sufficient conditions on the independence of random r.v.'s.

Corollary 1.4.7. *Let $X_1, ..., X_n$ be independent r.v.'s and let F_i be the corresponding distribution function. If $\mathbb{P}(X_1 \le x_1, ...X_n \le x_n) = \prod_{i=1}^n F_i(x_i)$ for any $x_i \in \mathbb{R}$, then $X_1, ..., X_n$ are independent.*

Proof. Define $A_i(x) = \{\omega : X_i(\omega) \le x\}$. For each $1 \le i \le n$, consider $\mathcal{A}_i = \{A_i(x) : x \in \mathbb{R}\}$. Clearly $\sigma(X_i) = \sigma(\mathcal{A}_i)$. Note that each $\mathcal{A}_i$ is a π-system because $A_i(x) \cap A_i(y) = \{\omega : X_i(\omega) \le x \wedge y\} \in \mathcal{A}_i$. Hence by Theorem 1.4.6, $X_1, ..., X_n$ are independent. □

Corollary 1.4.8. *If $\mathcal{F}_{i,j}, 1 \le i \le n, 1 \le j \le n(i)$ are independent, then $\mathcal{G}_i = \sigma(\cup_{j=1}^{n(i)} \mathcal{F}_{i,j})$ are independent.*

Proof. For each $1 \le i \le n$, define $\mathcal{A}_i = \{\cap_{j=1}^{n(i)} A_{i,j} : A_{i,j} \in \mathcal{F}_{i,j}\}$. Clearly $\mathcal{A}_i$ is a π-system, and moreover $\cup_{j=1}^{n(i)} \mathcal{F}_{i,j} \subset \mathcal{A}_i$. Thus $\sigma(\mathcal{A}_1), ..., \sigma(\mathcal{A}_n)$ are independent and $\mathcal{G}_i$'s are independent. □

Corollary 1.4.9. *If $X_{i,j}, 1 \le i \le n, 1 \le j \le n(i)$ are independent and $f_i : \mathbb{R}^{n(i)} \to \mathbb{R}$ are measurable, then $f_i(X_{i,1}, ..., X_{i,n(i)})$ are independent.*

Proof. Clearly $\sigma(f_i(X_{i,1}, ..., X_{i,n(i)})) \subset \sigma((X_{i,1}, ..., X_{i,n(i)})) \subset \mathcal{G}_i$ where $\mathcal{G}_i = \sigma(\cup_{j=1}^{n(i)} \sigma(X_{i,j}))$. Indeed, if $R_1 \times \cdots \times R_{n(i)}$ is a measurable rectangle, then

$$\{\omega : (X_{i,1}(\omega), ..., X_{i,n(i)}(\omega)) \in R_1 \times \cdots \times R_{n(i)}\} = \cap_{j=1}^{n(i)} \{\omega : X_{i,j}(\omega) \in R_j\} \in \mathcal{G}_i.$$

By the previous corollary $\mathcal{G}_i$'s are independent. It follows that $f_i(X_{i,1}, ..., X_{i,n(i)})$'s are independent. This concludes the proof. □

1.5 Weak Laws of Large Numbers

Let us investigate various limit phenomena in the sense of convergence in probability.

Definition 1.5.1. $Y_n \to Y$ **in probability** if for all $\varepsilon > 0$, $\mathbb{P}(|Y_n - Y| > \varepsilon) \to 0$ as $n \to \infty$.

Definition 1.5.2. A family of random variables $(X_i)_{i \in I}$ having finite second moment is said to be **uncorrelated** if $\mathbb{E}X_iX_j = 0$ for any distinct $i, j \in I$.

Lemma 1.5.3. *Let $X_1, ..., X_n$ have $\mathbb{E}X_i^2 < \infty$ and be uncorrelated. Then*

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$$

Proof. WLOG, assume that $X_1, ..., X_n$ are centered r.v.'s. We can compute

$$\begin{aligned}
\mathrm{Var}(X_1 + \cdots + X_n) &= \mathbb{E}\left(\sum_{i=1}^n X_i\right)^2 \\
&= \mathbb{E}\sum_{i,j}^n X_iX_j \\
&= \sum_{i=1}^n \mathbb{E}X_i^2 + 2\sum_{i=1}^n \sum_{j<i} \mathbb{E}X_iX_j \\
&= \sum_{i=1}^n \mathbb{E}X_i^2 = \sum_{i=1}^n \mathrm{Var}(X_i).
\end{aligned}$$

□

Theorem 1.5.4 (L^2 weak law)**.** *Let $(X_n)_{n=1}^\infty$ be uncorrelated random variables with $\mathbb{E}X_i = \mu$ and $\mathrm{Var}(X_i) \le C < \infty$. If $S_n = X_1 + \cdots + X_n$, then as $n \to \infty$,*

$$\frac{S_n}{n} \to \mu \text{ in } L^2 \text{ and in probability.}$$

Proof. We observe $\mathbb{E}(S_n/n - \mu)^2 = \mathrm{Var}(S_n/n) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) \le \frac{C}{n} \to 0$ as $n \to \infty$. The conclusion then follows from the following Lemma 1.5.5. □

Lemma 1.5.5. *If $p > 0$ and $\mathbb{E}|Z_n|^p \to 0$ then $Z_n \to 0$ in probability.*

Proof. Using Markov's inequality to estimate

$$\mathbb{P}(|Z_n| > \varepsilon) \le \frac{\mathbb{E}|Z_n|^p}{\varepsilon^p} \to 0 \text{ as } n \to \infty.$$

□

Example 1.5.6. Berstein polynomial approximation to continuous function on $[0, 1]$.

Limit theorems for triangular arrays. Consider random arrays $X_{n,k}, 1 \le k \le n$. We would like to obtain some limiting results on $S_n = \sum_{k=1}^n X_{n,k}$. A trivial but useful one is the following.

Theorem 1.5.7. *Let* $\mu_n = \mathbb{E}S_n, \sigma_n^2 = \mathrm{Var}(S_n)$*. If* $\sigma_n^2/b_n^2 \to 0$*, then*

$$\frac{S_n - \mu_n}{b_n} \to 0 \text{ in } L^2 \text{ and in probability.}$$

Proof. It simply follows from the computation

$$\mathbb{E}\left(\frac{S_n - \mu_n}{b_n}\right)^2 = \frac{\sigma_n^2}{b_n^2} \to 0 \text{ as } n \to \infty.$$

Convergence in probability follows from Lemma 1.5.5. □

Example 1.5.8 (Coupon collector's problem)**.** Let $X_1, X_2, ...$ be i.i.d. uniform on $\{1, 2, ..., n\}$. Define $\tau_k^n = \inf\{m : |\{X_1, ..., X_m\}| = k\}$ which is the first time we have collected k different object. We are interested in the asymptotic behavior of $T_n := \tau_n^n$. Consider $X_{n,k} := \tau_k^n - \tau_{k-1}^n$. Clearly $X_{n,k}$ has geometric distribution with parameter $(1 - (k-1)/n)$ and is independent of $X_{n,1}, ..., X_{n,k-1}$. Note that $\mathbb{E}X_{n,k} = (1-(k-1)/n)^{-1}$ and $\mathrm{Var}(X_{n,k}) \le (1-(k-1)/n)^{-2}$. Thus

$$\mathbb{E}T_n = \sum_{k=1}^n \mathbb{E}X_{n,k} = \sum_{k=1}^n (1 - (k-1)/n)^{-1} = n \sum_{k=1}^n k^{-1}$$

$$\mathrm{Var}(T_n) \le \sum_{k=1}^n (1 - (k-1)/n)^{-2} = n^2 \sum_{k=1}^\infty k^{-2}.$$

Taking $b_n = n \log n$ in the previous theorem, we see that

$$\frac{T_n - n\sum_{m=1}^n 1/m}{n \log n} \to 0 \text{ in probability,}$$

which implies that $T_n/n\log n \to 1$ in probability.

Example 1.5.9 (Random permutations)**.**

Example 1.5.10 (An occupancy problem)**.**

Theorem 1.5.11 (Weak law for arrays)**.** *For each* n*, let* $X_{n,k}, 1 \le k \le n$*, be independent. Let* $b_n > 0$ *and* $b_n \to \infty$ *as* $n \to \infty$ *and let* $\bar{X}_{n,k} = X_{n,k}\mathbb{1}_{\{|X_{n,k}| \le b_n\}}$*. Suppose as* $n \to \infty$*,*

(i) $\sum_{k=1}^n \mathbb{P}(|X_{n,k}| > b_n) \to 0$,

(ii) $b_n^{-2}\sum_{k=1}^n \mathbb{E}\bar{X}_{n,k}^2 \to 0$.

Set $S_n = X_{n,1} + \cdots + X_{n,n}$ *and put* $a_n = \sum_{k=1}^n \mathbb{E}\bar{X}_{n,k}$*, then*

$$\frac{S_n - a_n}{b_n} \to 0 \text{ in probability.}$$

Proof. First we observe that

$$\mathbb{P}\left(\left|\frac{S_n - a_n}{b_n}\right| > \varepsilon\right) \le \mathbb{P}(S_n \ne \bar{S}_n) + \mathbb{P}\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \varepsilon\right).$$

Then we estimate terms on the right hand side:

$$\mathbb{P}(S_n \ne \bar{S}_n) \le \sum_{k=1}^n \mathbb{P}(\bar{X}_{n,k} \ne X_{n,k}) = \sum_{k=1}^n \mathbb{P}(|X_{n,k}| > b_n),$$

and

$$\mathbb{P}\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \varepsilon\right) \le \frac{\mathrm{Var}(\bar{S}_n)}{\varepsilon^2 b_n^2} \le \frac{\sum_{k=1}^n \mathbb{E}\bar{X}_{n,k}^2}{\varepsilon^2 b_n^2}.$$

Thus by condition (i) and (ii), we have

$$\limsup_{n\to\infty} \mathbb{P}\left(\left| \frac{S_n - a_n}{b_n} \right| > \varepsilon \right) = 0,$$

as desired. □

Theorem 1.5.12. *Let $X_1, X_2, \ldots$ be i.i.d. with $x\mathbb{P}(|X_i| > x) \to 0$ as $x \to \infty$. Let $S_n = X_1 + \cdots + X_n$, and let $\mu_n = \mathbb{E}X_i \mathbb{1}_{\{|X_i|\le n\}}$. Then $\frac{S_n - \mu_n}{n} \to 0$ in probability.*

Remark 1.5.13. The condition $x\mathbb{P}(|X_i| > x) \to 0$ is also necessary for the existence of a_n such that $S_n/n - a_n \to 0$.

Proof. We wish to apply Theorem 1.5.11. Take $X_{n,k} = X_k$ for $1 \le k \le n$ and $b_n = n$. Clearly the assumption $x\mathbb{P}(|X_i| > x) \to 0$ implies $\sum_{k=1}^n \mathbb{P}(|X_{n,k}| > b_n) = n\mathbb{P}(|X_k| > n) \to 0$ as $n \to \infty$. To check (ii), we need the following lemma:

Lemma 1.5.14. *If $Y \ge 0$ and $p > 0$, then $\mathbb{E}Y^p = \int_0^\infty py^{p-1}\mathbb{P}(Y > y)dy$.*

Proof of Lemma. Consider the function $f(t) = t^p$ and clearly $f(t) = \int_0^t py^{p-1}dy$. Then we apply Fubini's theorem in the following computation

$$\begin{aligned}
\mathbb{E}Y^p &= \mathbb{E}f(Y) \\
&= \mathbb{E}\int_0^Y py^{p-1}dy \\
&= \mathbb{E}\int_0^\infty py^{p-1}\mathbb{1}_{\{y<Y\}}dy \\
&= \int_0^\infty \mathbb{E}\left(py^{p-1}\mathbb{1}_{\{y<Y\}}\right)dy \\
&= \int_0^\infty py^{p-1}\mathbb{P}(Y > y)dy.
\end{aligned}$$

This completes the proof. □

Using Lemma 1.5.14, we see

$$\mathbb{E}\bar{X}_{n,k}^2 = \int_0^\infty 2y\mathbb{P}(|\bar{X}_1| > y)dy = \int_0^n 2y\mathbb{P}(|\bar{X}_1| > y)dy \le \int_0^n 2y\mathbb{P}(|X_1| > y)dy.$$

To get an upper bound on the last term, we denote $g(y) = 2y\mathbb{P}(|X_1| > y)$. Since $g(y) \to 0$ as $y \to \infty$ and $0 \le g(y) \le 2y$, $g(y)$ is bounded above by some constant $C > 0$. Fix $m \in (0, n)$ and denote $\varepsilon_m = \sup\{g(y) : y > m\}$. Then

$$\int_0^n g(y)dy = \int_0^m g(y)dy + \int_m^n g(y)dy \le Cm + (n-m)\varepsilon_m.$$

It follows that

$$\varlimsup_{n\to\infty} \frac{\sum_{k=1}^n \mathbb{E}\bar{X}_{n,k}^2}{n^2} = \varlimsup_{n\to\infty} \frac{\mathbb{E}\bar{X}_1^2}{n} \le \varlimsup_{n\to\infty} \frac{1}{n}\int_0^n g(y)dy \le \varlimsup_{n\to\infty} \frac{Cm + (n-m)\varepsilon_m}{n} \le \varepsilon_m.$$

Since m is arbitrary, and $\varepsilon_m \to 0$ as $m \to \infty$, the result follows. □

Corollary 1.5.15. *Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}|X_i| < \infty$. Let $S_n = X_1 + \cdots + X_n$ and let $\mu = \mathbb{E}X_i$. Then $\frac{S_n - \mu}{n} \to 0$ in probability.*

Proof. Dominated convergence theorem shows that $\mu_n \to \mu$ as $n \to \infty$ and $x\mathbb{P}(|X_i| > x) \leq \mathbb{E}|X_i|\mathbb{1}_{\{|X_i|>x\}} \to 0$ as $x \to \infty$. Then by Theorem 1.5.12 $\mathbb{P}\left(\left|\frac{S_n - n\mu_n}{n}\right| > \varepsilon/2\right) \to 0$ as $n \to \infty$ for every $\varepsilon > 0$. Thus

$$\mathbb{P}\left(\left|\frac{S_n - n\mu}{n}\right| > \varepsilon\right) \leq \mathbb{P}(|\mu_n - \mu| > \varepsilon/2) + \mathbb{P}\left(\left|\frac{S_n - n\mu_n}{n}\right| > \varepsilon/2\right) \to 0$$

as $n \to \infty$. ☐

Example 1.5.16 (St. Petersburg paradox)**.** Consider an i.i.d. sequence $X_1, X_2, \ldots$ with distribution $\mathbb{P}(X_1 = 2^j) = 2^{-j}$. Clearly $\mathbb{E}X_1 = \infty$. Yet we could still use Theorem 1.5.11 to gain some insights about the asymptotic behavior of S_n.

1.6 Borel-Cantelli Lemmas

Let $(A_n)_{n=1}^{\infty}$ be a sequence of events. We define

$$\limsup A_n := \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} A_n = \{\omega : \omega \text{ belongs to infinitely many } A_n\text{'s}\}$$

$$\liminf A_n := \cup_{m=1}^{\infty} \cap_{n=m}^{\infty} A_n = \{\omega : \omega \text{ belongs to all but finitely many } A_n\text{'s}\}$$

Clearly $\limsup_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\limsup A_n}$ and $\liminf_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\liminf A_n}$. In this terminology, $X_n \to X$ a.s. if and only if for each $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon \text{ i.o.}) = 1$.

Theorem 1.6.1 (Borel-Cantelli)**.** *If* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$*, then* $\mathbb{P}(A_n \text{ i.o.}) = 0$.

Proof. Let $N := \sum_{n=1}^{\infty} \mathbb{1}_{A_n}$ be the number of events that occur. But $\mathbb{E}N < \infty$ implies $N < \infty$ a.s. In other words, $\mathbb{P}(A_n \text{ i.o.}) = 0$. ☐

Theorem 1.6.2. *A sequence of r.v.'s* $X_n \to X$ *in probability if and only if for every subsequence* $X_{n(m)}$ *there is a further subsequence* $X_{n(m_k)}$ *that converges a.s. to* X.

Proof. First note that $X_{n(m)} \to X$ in probability. First we pick $n(m_1)$ such that $\mathbb{P}(|X_{n(m_1)} - X| > 1/2) < 1/2$. Inductively we pick $n(m_k) > n(m_{k-1})$ such that $\mathbb{P}(|X_{n(m_k)} - X| > 1/2^k) < 1/2^k$. Given any $\varepsilon > 0$, $\#\{k : 1/2^k > \varepsilon\}$ is finite. Then we have the following bound

$$\sum_{k=1}^{\infty} \mathbb{P}(|X_{n(m_k)} - X| > \varepsilon) \leq \#\{k : 1/2^k > \varepsilon\} + 1 < \infty.$$

Thus by Borel-Cantelli lemma, $X_{n(m_k)} \to X$ a.s. Conversely, suppose the subsequence $X_{n(m)}$ has a further subsequence $X_{(m(m_k))}$ converging to X a.s. The conclusion immediately follows as a consequence of the following lemma:

Lemma 1.6.3. *Let* $(y_n)_{n=1}^{\infty}$ *be a sequence in a topological space. If for each subsequence* $(y_{n(m)})_{m=1}^{\infty}$ *there is a further subsequence* $(y_{n(m_k)})_{k=1}^{\infty}$ *converges to* y*, then* $y_n \to y$ *as* $n \to \infty$.

We apply this lemma with $y_n = \mathbb{P}(|X_n - X| > \varepsilon)$ to complete the proof. ☐

Corollary 1.6.4. *If* $X_n \to X$ *in probability and* f *is continuous, then* $f(X_n) \to f(X)$ *in probability. If* f *in addition is bounded then* $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$.

Proof. Suppose that $f(X_{n(m)})$ is a subsequence. By Theorem 1.6.2, we know there is a further subsequence $(X_{n(m_k)})_{k=1}^{\infty}$ such that $X_{n(m_k)} \to X$ a.s. The continuity of f implies $f(X_{n(m_k)}) \to f(X)$ almost surely. Hence Theorem 1.6.2 asserts that $f(X_n) \to f(X)$ in probability as well. To prove the remaining claim, we consider the sequence $y_n = \mathbb{E}f(X_n)$. Let $(y_{n(i)})_{i=1}^{\infty}$ be a subsequence. We can find a further subsequence $f(X_{n(i_k)}) \to f(X)$ a.s. By bounded convergence theorem, $\mathbb{E}f(X_{n(i_k)}) \to \mathbb{E}f(X_{n(i_k)})$. Thus Lemma 1.6.3 directly concludes the result. ☐

Next we apply Borel-Cantelli lemma to prove the strong law of large numbers assuming finite 4th moment.

Theorem 1.6.5. *Let $X_1, X_2, ...$ be i.i.d. random variables with $\mathbb{E}X_i = \mu$ and $\mathbb{E}X_i^4 < \infty$. Then $S_n/n \to \mu$ a.s.*

Proof. WLOG we may assume $\mu = 0$. The estimate $\mathbb{P}(|S_n| > n\varepsilon) \leq \frac{\mathbb{E}S_n^4}{n^4\varepsilon^4}$ reveals that our task is to show $\mathbb{E}S_n^4 = O(n^k)$ with $k < 3$. Then

$$\mathbb{E}S_n^4 = \sum_{1 \leq i,j,k,\ell \leq n} \mathbb{E}X_i X_j X_l X_\ell.$$

Since X_n's are independent and $\mathbb{E}X_i = 0$ for each i, we only need to count the number of terms involving $X_i^2 X_j^2$ and X_i^4. Clearly there are n terms involving X_i^4. To compute the number of terms involving $X_i^2 X_j^2$, we can first pick two indices out of n which can arise in $n(n-1)/2$ ways. For each fixed indices i, j, the term $X_i^2 X_j^2$ can arise in 6 ways. Therefore $\mathbb{E}S_n^4 = n\mathbb{E}X_1^4 + 3n(n-1)(\mathbb{E}X_1^2)^2 \leq Cn^2$ for some $C > 0$ and

$$\sum_{n=1}^{\infty} \mathbb{P}(|S_n| > n\varepsilon) \leq \sum_{n=1}^{\infty} \frac{C}{n^2\varepsilon^4} < \infty.$$

By Borel-Cantelli lemma, $S_n/n \to 0$ a.s. and the proof is complete. □

Example 1.6.6 (Converse of Borel-Cantelli lemma is false)**.** Consider uniform measure on $(0, 1)$. Let $A_n = (0, 1/n)$. Clearly $\limsup A_n = \varnothing$. But $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} 1/n = \infty$.

Remark 1.6.7. In general, we can only assert that

$$\mathbb{P}(\liminf A_n) \leq \liminf_{n\to\infty} \mathbb{P}(A_n) \leq \limsup_{n\to\infty} \mathbb{P}(A_n) \leq \mathbb{P}(\limsup A_n).$$

If, however, the events A_n's are independent, then we will have the converse implication.

Theorem 1.6.8 (The second Borel-Cantelli lemma)**.** *If the events A_n are independent then $\sum_n \mathbb{P}(A_n) = \infty$ implies* $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Proof. Pick $0 < M < N$. Using the elementary inequality $1 - x \leq e^{-x}$ and independence, we see

$$\mathbb{P}(\cap_{k=M}^N A_k^c) = \prod_{k=M}^{N} (1 - \mathbb{P}(A_k)) \leq e^{-\sum_{k=M}^N \mathbb{P}(A_k)} \to 0 \text{ as } N \to \infty.$$

Thus $\mathbb{P}(\cup_{k=M}^{\infty} A_k) = 1$ for all $M \in \mathbb{Z}_+$. Then observe $\cup_{k=M}^{\infty} A_k \uparrow \limsup A_k$ and so $\mathbb{P}(A_k \text{ i.o.}) = \lim_{M\to\infty} \mathbb{P}(\cup_{k=M}^{\infty} A_k) = 1$. □

The second Borel-Cantelli lemma is useful to establish non-existence results.

Theorem 1.6.9. *Let $(X_n)_{n=1}^{\infty}$ be an i.i.d. sequence of random variables. If $\mathbb{E}|X_1| = \infty$, then $\mathbb{P}(|X_n| \geq n$ i.o.$) = 1$. In particular, $\mathbb{P}(\lim_{n\to\infty} S_n/n$ exists in $(-\infty, \infty)) = 0$.*

Proof. We estimate

$$\sum_{n=0}^{\infty} \mathbb{P}(|X_1| \geq n) \geq \int_1^{\infty} \mathbb{P}(|X_1| \geq t) dt = \mathbb{E}|X_1| = \infty.$$

So by Theorem 1.6.8, $\mathbb{P}(|X_n| \geq n \text{ i.o.}) = 1$. Let $C := \{\omega : \lim_{n\to\infty} S_n(\omega)/n \text{ exists in } \mathbb{R}\}$. Now observe that

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{(n+1)S_n S_{n+1}}{n(n+1)} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}.$$

Clearly on C, $\frac{S_n}{n(n+1)} \to 0$ as $n \to \infty$. Thus on $C \cap \{|X_n| \geq n \text{ i.o.}\}$, for n sufficiently large,

$$\left|\frac{S_n}{n} - \frac{S_{n+1}}{n+1}\right| \geq \left|\left|\frac{S_n}{n(n+1)}\right| - \left|\frac{X_{n+1}}{n+1}\right|\right| \geq 1 - o(1).$$

Hence $C \cap \{|X_n| \geq n \text{ i.o.}\} = \varnothing$. Because $\mathbb{P}(|X_n| \geq n \text{ i.o.}) = 1$, it obvious that $\mathbb{P}(C) = 0$. □

Even we weaken the assumption in the second Borel-Cantelli lemma, we can still get a sharper conclusion.

Theorem 1.6.10. *If events $A_1, A_2, ...$ are pairwise independent and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then*

$$\frac{\sum_{k=1}^{n} \mathbb{1}_{A_k}}{\sum_{k=1}^{n} \mathbb{P}(A_k)} \to 1 \text{ almost surely as } n \to \infty.$$

Proof. Let $X_n = \mathbb{1}_{A_n}$ and it is clear that X_i, X_j are uncorrelated whenever $i \neq j$ due to pairwise independence of A_n's. Next note that $X_n \in \{0, 1\}$, the variance $\mathrm{Var} X_n \leq \mathbb{E}X_n^2 = \mathbb{E}X_n$. By the linearity and uncorrelatedness we have $\mathrm{Var} S_n \leq \mathbb{E}S_n$. Then Chebyshev's inequality reveals that the following bound holds for every $\varepsilon > 0$,

$$\mathbb{P}(|S_n - \mathbb{E}S_n| > \varepsilon \mathbb{E}S_n) \leq \frac{\mathbb{E}S_n}{\varepsilon^2 (\mathbb{E}S_n)^2} = \frac{1}{\varepsilon^2 \mathbb{E}S_n}.$$

Clearly this implies $S_n / \mathbb{E}S_n \to 1$ in probability as $n \to \infty$. To get the a.s. convergence, we first seek a convergent subsequence. Let $n(k) = \inf\{n : \mathbb{E}S_n \geq k^2\}$. Using the above estimate, we get

$$\sum_{k=1}^{\infty} \mathbb{P}(|S_{n(k)} - \mathbb{E}S_{n(k)}| > \varepsilon \mathbb{E}S_{n(k)}) \leq \sum_{k=1}^{\infty} \frac{1}{\varepsilon^2 \mathbb{E}S_{n(k)}} \leq \sum_{k=1}^{\infty} \frac{1}{\varepsilon^2 k^2} < \infty.$$

Thus by Borel-Cantelli lemma, $S_{n(k)} / \mathbb{E}S_{n(k)} \to 1$ almost surely. To handle the convergence of the whole sequence, we observe that $\mathbb{E}S_{n(k)} \leq k^2 + 1$ for otherwise, $\mathbb{E}S_{n(k)-1} > k^2$ yields the contradiction $n(k) - 1 < n(k)$. Hence, for $n(k) \leq n < n(k+1)$, we can bound

$$\frac{S_{n(k)}}{\mathbb{E}S_{n(k)}} \frac{\mathbb{E}S_{n(k)}}{\mathbb{E}S_{n(k+1)}} = \frac{S_{n(k)}}{\mathbb{E}S_{n(k+1)}} \leq \frac{S_n}{\mathbb{E}S_n} \leq \frac{S_{n(k+1)}}{\mathbb{E}S_{n(k)}} = \frac{S_{n(k+1)}}{\mathbb{E}S_{n(k+1)}} \frac{\mathbb{E}S_{n(k+1)}}{\mathbb{E}S_{n(k)}}.$$

So it suffices to show that $\mathbb{E}S_{n(k+1)} / \mathbb{E}S_{n(k)} \to 1$ as $k \to \infty$. But this is obvious from the bound $k^2 \leq \mathbb{E}S_{n(k)} \leq \mathbb{E}S_{n(k+1)} \leq (k+1)^2 + 1$. □

Example 1.6.11 (Record values)**.**

Example 1.6.12 (Head runs)**.**

1.7 Strong Law of Large Numbers

In this section we will give a complete proof of the strong law of large numbers. As in the weak law, we assume a sequence of i.i.d. random variables $(X_i)_{i=1}^{\infty}$ with $\mathbb{E}|X_i| < \infty$. Denote $\bar{X}_n = X_n \mathbb{1}_{\{|X_n| \leq n\}}$, $T_n = \sum_{i=1}^{n} \bar{X}_i$ and $\mu = \mathbb{E}X_1$.

Theorem 1.7.1 (Strong law of large numbers)**.** *If $(X_i)_{i=1}^{\infty}$ is a sequence of i.i.d. random variables with $\mathbb{E}|X_1| < \infty$, then $\frac{S_n}{n} \to \mu$ almost surely.*

Proof. We start by showing that it is sufficient to prove the strong law for the truncated random variables.

Lemma 1.7.2. *If $T_n / n \to \mu$ a.s., then $S_n / n \to \mu$ a.s.*

Proof. WLOG, we may assume $\mu = 0$. For each $k \in \mathbb{Z}_+$ consider the events $A_k = \{|S_k/k| > \varepsilon\}$, $B_k = \{|(S_k - T_k)/k| > \varepsilon/2\}$, and $C_k = \{|T_k/k| > \varepsilon/2\}$. By the triangle inequality we have the implication: if $\omega \in B_k^c \cap C_k^c$, then $\omega \in A_k^c$. Taking complements, we have $A_k \subset B_k \cup C_k$. Thus,

$$\limsup A_n = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k \subset \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} (B_k \cup C_k) = \limsup B_n \cup \limsup C_n.$$

Then the union bound $\mathbb{P}(A_n \text{ i.o.}) \leq \mathbb{P}(B_n \text{ i.o.}) + \mathbb{P}(C_n \text{ i.o.})$ and the assumption $T_n/n \to 0$ a.s. imply that we only need to show $\mathbb{P}(B_n \text{ i.o.}) = 0$. Note that $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n) \leq \int_0^{\infty} \mathbb{P}(|X_1| > x)dx = \mathbb{E}|X_1| < \infty$, so by Borel-Cantelli lemma, $\mathbb{P}(X_n \neq \bar{X}_n \text{ i.o.}) = 0$. In particular, it follows that almost surely $|S_n(\omega) - T_n(\omega)| = R(\omega) < \infty$. Because $R(\omega)/n \to 0$, we have $\mathbb{P}(B_n \text{ i.o.}) = 0$. □

The next pieces are the following two lemmas. The proof of the second lemma is based on the first one.

Lemma 1.7.3. *If $y \geq 0$, then $2y \sum_{k>y} k^{-2} \leq 4$.*

Proof. First we observe $m \geq 2$, we compare

$$\sum_{k \geq m} k^{-2} \leq \int_{m-1} t^{-2} dt = \frac{1}{m-1}.$$

So when $y \geq 1$, $\lfloor y \rfloor + 1 \geq 2$, and

$$2y \sum_{k>y} \frac{1}{k^2} = 2y \sum_{k=\lfloor y \rfloor + 1}^{\infty} \frac{1}{k^2} \leq 2 \frac{y}{\lfloor y \rfloor} \leq 4,$$

because $y/\lfloor y \rfloor \leq 2$. When $y < 1$, we have

$$2y \sum_{k>y} \frac{1}{k^2} \leq 2 \frac{\pi^2}{6} \leq 4.$$

□

Lemma 1.7.4. $\sum_{k=1}^{\infty} \mathrm{Var}(\bar{X}_k)/k^2 \leq 4\mathbb{E}|X_1|$.

Proof. We start from the trivial bound, $\mathrm{Var}(\bar{X}_k) \leq \mathbb{E}\bar{X}_k^2$ for each $k \in \mathbb{Z}_+$. So the sum can be bounded,

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{\mathrm{Var}(\bar{X}_k)}{k^2} &\leq \sum_{k=1}^{\infty} \frac{\mathbb{E}\bar{X}_k^2}{k^2} \\
&= \sum_{k=1}^{\infty} \frac{1}{k^2} \int_0^{\infty} 2t\mathbb{P}(|\bar{X}_k| > t)dt \\
&\leq \sum_{k=1}^{\infty} \frac{1}{k^2} \int_0^{\infty} \mathbb{1}_{\{t<k\}} 2t\mathbb{P}(|X_1| > t)dt \\
&\leq \int_0^{\infty} \left(\sum_{k=1}^{\infty} \frac{1}{k^2} \mathbb{1}_{\{t<k\}} \right) 2t\mathbb{P}(|X_1| > t)dt \\
&\leq 4 \int_0^{\infty} \mathbb{P}(|X_1| > t)dt = 4\mathbb{E}|X_1|.
\end{aligned}$$

□

Now we go back to prove the main assertion. Note that in general we may write $X_n = X_n^+ - X_n^-$ where X_n^+, X_n^- denote the positive and negative parts of X_n respectively. So it is sufficient to prove strong laws assume $X_n \geq 0$ for $n \in \mathbb{Z}_+$. We shall seek a convergent subsequence then use monotonicity control to get the full convergence. Let $\alpha > 1$ and $n(k) = \lfloor \alpha^k \rfloor$. Given any $\varepsilon > 0$, by Chebyshev's inequality,

$$\begin{aligned}\sum_{k=1}^{\infty} \mathbb{P}(|T_{n(k)} - \mathbb{E}T_{n(k)}| > \varepsilon n(k)) &\leq \sum_{k=1}^{\infty} \frac{\mathrm{Var}(T_{n(k)})}{\varepsilon^2 n(k)^2} \\ &\leq \frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} \sum_{m=1}^{n(k)} \frac{\mathrm{Var}(\bar{X}_m)}{n(k)^2} \\ &= \frac{1}{\varepsilon^2} \sum_{m=1}^{\infty} \mathrm{Var}(\bar{X}_m) \sum_{k:n(k)\geq m} \frac{1}{n(k)^2}.\end{aligned}$$

Clearly we note $n(k) \geq \alpha^k/2$ for $k \geq 1$. So bounding this geometric series yields

$$\sum_{k:n(k)\geq m} \frac{1}{n(k)^2} \leq 4 \sum_{k:n(k)\geq m} \frac{1}{\alpha^{2k}} \leq \frac{4}{m^2(1-\alpha^{-2})}.$$

Hence,

$$\sum_{k=1}^{\infty} \mathbb{P}(|T_{n(k)} - \mathbb{E}T_{n(k)}| > \varepsilon n(k)) \leq \frac{4}{\varepsilon^2(1-\alpha^{-2})} \sum_{m=1}^{\infty} \frac{\mathrm{Var}(\bar{X}_m)}{m^2} \leq \frac{16\mathbb{E}|X_1|}{\varepsilon^2(1-\alpha^{-2})} < \infty,$$

by Lemma 1.7.4. This implies, by Borel-Cantelli lemma, $\frac{T_{n(k)} - \mathbb{E}T_{n(k)}}{n(k)} \to 0$ almost surely. Since dominated convergence theorem shows $\mathbb{E}\bar{X}_n \to \mathbb{E}X_1$, the quotient $\mathbb{E}T_{n(k)}/n(k) \to EX_1$ as well. Thus $T_{n(k)}/n(k) \to \mathbb{E}X_1$ almost surely. Now observe that for $n(k) \leq m < n(k+1)$, we have

$$\frac{n(k)}{n(k+1)} \frac{T_{n(k)}}{n(k)} = \frac{T_{n(k)}}{n(k+1)} \leq \frac{T_m}{m} \leq \frac{T_{n(k+1)}}{n(k)} = \frac{T_{n(k+1)}}{n(k+1)} \frac{n(k+1)}{n(k)}.$$

Note that $n(k+1)/n(k) \to \alpha$, so

$$\frac{1}{\alpha}\mathbb{E}X_1 \leq \liminf_{m\to\infty} \frac{T_m}{m} \leq \limsup_{m\to\infty} \frac{T_m}{m} \leq \alpha \mathbb{E}X_1.$$

Let $\alpha \to 1$ and the proof is complete. □

A trivial extension of the strong law is the following.

Theorem 1.7.5. *Let $(X_n)_{n=1}^{\infty}$ be a sequence of i.i.d. random variables. If $\mathbb{E}X_1^+ = \infty$ and $\mathbb{E}X_1^- < \infty$, then $S_n/n \to \infty$ almost surely.*

Proof. First we truncate X_n by M. Let $M > 0$ and define $X_n^M = \min\{X_n, M\}$. Clearly $(X_n^M)_{n=1}^{\infty}$ is a sequence of i.i.d. r.v.'s with $\mathbb{E}|X_1^M| < \infty$. So by the strong law, we see that $S_n^M/n \to \mathbb{E}X_1^M$ almost surely. Note because $X_n \geq X_n^M$, it follows that

$$\liminf_{n\to\infty} \frac{S_n}{n} \geq \lim_{n\to\infty} \frac{S_n^M}{n} = \mathbb{E}X_1^M.$$

Since this holds for all $M > 0$, letting $M \uparrow \infty$ yields $\liminf_{n\to\infty} S_n/n \geq \infty$ a.s. □

Several applications of the strong law are in order.

Example 1.7.6 (Renewal theory)**.** Let $(X_n)_{n=1}^\infty$ be a sequence of i.i.d. r.v.'s with $0 < X_i < \infty$. Consider a janitor who replaces a light bulb the instant it burns out. Suppose the first bulb is put at time 0 and X_i is the lifetime of the ith lightbulb. Denote $T_n = X_1 + \cdots + X_n$. Then T_n is the time the nth bulb burns out. Consider $N_t = \sup\{n : T_n \le t\}$ which is the number of burnt out lightbulbs before time t. We shall establish a limit theorem for the time average N_t/t.

Theorem 1.7.7. *If* $\mathbb{E}X_1 = \mu \le \infty$*, then* $N_t/t \to 1/\mu$ *as* $t \to \infty$ *almost surely.*

Proof. We shall show that $t/N_t \to \mu$ a.s. By the strong law and its variant Theorem x.7.5, $T_n/n \to \mu$ a.s. Observe that $T(N_t) \le t < T(N_t+1)$. Thus we get a squeezing bound

$$\frac{T(N_t)}{N_t} \le \frac{t}{N_t} < \frac{T(N_t+1)}{N_t+1}\frac{N_t+1}{N_t}.$$

Now note that for all $\omega \in \Omega$ and $n \in \mathbb{Z}_+$, $T_n(\omega) < \infty$. Thus for all $\omega \in \Omega$, $N_t(\omega) = \sup\{n : T_n(\omega) \le t\} \uparrow \infty$ as $t \to \infty$. Now note that by the strong law we can find a subset $\Omega_0 \subset \Omega$ of full measure, such that $T_n(\omega)/n \to \mu$ for all $\omega \in \Omega$. Thus for such ω's, we have

$$\frac{T_{N_t(\omega)}(\omega)}{N_t(\omega)} \to \mu \text{ and } \frac{N_t(\omega)+1}{N_t(\omega)} \to 1 \text{ as } t \to \infty.$$

Then it follows that $\frac{t}{N_t(\omega)} \to \mu$ for all $\omega \in \Omega_0$. □

Example 1.7.8 (Empirical distribution function)**.** Let $X_1, X_2, \ldots$ be i.i.d. r.v.'s with distribution function F. Define

$$F_n(x) = \frac{\sum_{m=1}^n \mathbb{1}_{\{X_m \le x\}}}{n}.$$

It is clear that $F_n(x)$ is the observed frequency of values that are $\le x$. The next result shows that F_n converges uniformly to F as $n \to \infty$.

Theorem 1.7.9 (Glivenko-Cantelli)**.** *As* $n \to \infty$*,*

$$\sup_x |F_n(x) - F(x)| \to 0 \text{ as } n \to \infty.$$

Example 1.7.10 (Shannon's theorem)**.**

1.8 Convergence of Random Series

To begin with, we introduce the notion of **tail σ-algebra**. Given a sequence $(X_n)_{n=1}^\infty$ of r.v.'s, we define the future after time n, $\mathcal{F}_{m\ge n} = \sigma(X_n, X_{n+1}, \ldots)$, i.e., the smallest σ-algebra w.r.t. which all the $X_m, m \ge n$ are measurable. The tail σ-algebra is defined by $\mathcal{T} = \cap_n \mathcal{F}_{m\ge n}$. Intuitively, $A \in \mathcal{T}$ if and only if changing a finite number of values does not affect the occurrence of the event. We will prove that all events in the tail σ-algebra have probability 0 or 1 under some circumstances. Before that, we illustrate the concepts with some examples.

Example 1.8.1. Let $(X_n)_{n=1}^\infty$ be a sequence of r.v.'s and let B_n be a Borel set for each n. It is easy to see that $\{X_n \in B_n \text{ i.o.}\} \in \mathcal{T}$.

Example 1.8.2. Let $S_n = \sum_{i=1}^n X_i$. We claim that
(i) $\{\lim_{n\to\infty} S_n \text{ exists}\} \in \mathcal{T}$;
(ii) $\{\limsup_{n\to\infty} S_n > 0\} \notin \mathcal{T}$;
(iii) $\{\limsup_{n\to\infty} S_n/c_n > x\} \in \mathcal{T}$ if $c_n \to \infty$ as $n \to \infty$.

Theorem 1.8.3 (Kolmogorov's 0-1 law)**.** *If* $X_1, X_2, \ldots$ *are independent and* $A \in \mathcal{T}$*, then* $\mathbb{P}(A) = 0$ *or* 1.

Proof. We shall prove that A is independent of itself, so that $\mathbb{P}(A \cap A) = \mathbb{P}(A)^2$. It follows that $\mathbb{P}(A) = 0$ or 1.

First, we claim that for each $k \geq 1$, $\sigma(X_1, X_2, ..., X_k)$ is independent of $\sigma(X_{k+1}, X_{k+2}, ...)$. Suppose $A \in \sigma(X_1, X_2, ..., X_k)$ and $B \in \sigma(X_{k+1}, X_{k+2}, ...)$. If for some $j \geq 1$, $B \in \sigma(X_{k+1}, ..., X_{k+j})$, then by Corollary 1.4.5, A and B are independent. The general case follows from $\cup_j \sigma(X_{k+1}, ..., X_{k+j}) \subset \sigma(X_{k+1}, X_{k+2}, ...)$ and $\cup_j \sigma(X_{k+1}, ..., X_{k+j}), \sigma(X_1, ..., X_k)$ are π-systems. Thus the claim is established.

Next, we show that $\sigma(X_1, X_2, ...)$ and $\mathcal{T}$ are independent. Note for each $k \geq 1$, by previous claim, $\sigma(X_1, ..., X_k)$ and $\mathcal{T}$ are independent because $\mathcal{T} \subset \sigma(X_{k+1}, X_{k+2}, ...)$. So it is clear that $\cup_k \sigma(X_1, ..., X_k)$ and $\mathcal{T}$ are independent. Moreover, since they form π-systems and contain Ω, the claim follows.

Therefore, A is independent of itself because $A \in \mathcal{T}$ implies $A \in \sigma(X_1, X_2, ...)$. ☐

Theorem 1.8.4 (Kolmogorov's maximal inequality)**.** *Suppose $X_1, ..., X_n$ are independent with $\mathbb{E}X_i = 0$ and $\operatorname{Var}(X_i) < \infty$. If $S_n = X_1 + \cdots + X_n$, then*

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| \geq x\right) \leq \frac{\operatorname{Var}(S_n)}{x^2}.$$

Proof. For $1 \leq k \leq n$, let us define $A_k = \{|S_1| < x, ..., |S_{k-1}| < x, |S_k| \geq x\}$. Clearly $A_i \cap A_j = \varnothing$ and $\cup_{k=1}^n A_k = \{\max_{1 \leq k \leq n} |S_k| \geq x\}$. We will bound the probability of each A_k. Observe that $|S_k|^2 \mathbb{1}_{A_k} \geq x^2 \mathbb{1}_{A_k}$. By integrating both sides, we obtain $\mathbb{P}(A_k) \leq \frac{\mathbb{E}S_k^2 \mathbb{1}_{A_k}}{x^2}$. Thus,

$$\begin{aligned}
\mathbb{P}(A_k) &\leq \frac{1}{x^2} \mathbb{E}S_k^2 \mathbb{1}_{A_k} \\
&\leq \frac{1}{x^2} \mathbb{E}[S_k^2 + (S_n - S_k)^2] \mathbb{1}_{A_k} \\
&= \frac{1}{x^2} \mathbb{E}[S_k^2 \mathbb{1}_{A_k} + (S_n - S_k)^2 \mathbb{1}_{A_k} + 2S_k(S_n - S_k)\mathbb{1}_{A_k}] = \frac{1}{x^2} \mathbb{E}S_n^2 \mathbb{1}_{A_k},
\end{aligned}$$

where we used the fact that $S_k, S_n - S_k$ are independent and $S_n - S_k$ has expectation 0. Summing over $1 \leq k \leq n$, we get $\mathbb{P}(\max_{1 \leq k \leq n} |S_k| \geq x) = \sum_{k=1}^n \mathbb{P}(A_k) \leq \frac{1}{x^2} \mathbb{E}S_n^2$. ☐

Next we shall prove Kolmogorov's results on convergence of random series. Before that we introduce the notion of almost surely Cauchy sequence. A sequence $(X_n)_{n=1}^\infty$ is Cauchy a.s. if for each $\varepsilon > 0$, $\lim_{n \to \infty} \mathbb{P}(\sup_{k, \ell > n} |X_k - X_\ell| > \varepsilon) = 0$, or $\lim_{n,m \to \infty} \mathbb{P}(\sup_{m < k \leq n} |X_k - X_m| > \varepsilon) = 0$. Several related results on this can be found in this article[1].

Theorem 1.8.5 (One-series theorem)**.** *If $X_1, X_2, ...$ are independent with $\mathbb{E}X_i = 0$ for each $i \in \mathbb{Z}_+$ and $\sum_{i=1}^\infty \operatorname{Var}(X_i) < \infty$, then $\sum_{i=1}^\infty X_i$ converges almost surely.*

Proof. Let $\varepsilon > 0$ and let $S_n = \sum_{i=1}^n X_i$. We shall prove that S_n is a Cauchy sequence with probability 1. Use maximal's inequality to estimate

$$\mathbb{P}(\sup_{m < k \leq n} |S_k - S_n| > \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{i=m+1}^n \mathbb{E}X_i^2 = \frac{1}{\varepsilon^2} \sum_{i=m+1}^n \operatorname{Var}(X_i).$$

Sending $n, m \to \infty$ implies that $(S_n)_{n=1}^\infty$ is a Cauchy sequence. ☐

Theorem 1.8.6 (Two-series theorem)**.** *If $X_1, X_2, ...$ are independent with $\sum_i \mathbb{E}X_i < \infty$ and $\sum_i \operatorname{Var}(X_i) < \infty$, then $\sum_{i=1}^\infty X_i$ converges almost surely.*

Proof. Define $Y_i = X_i - \mathbb{E}X_i$. Then $(Y_i)_{i=1}^\infty$ is a sequence of r.v.'s with $\mathbb{E}Y_i = 0$ and $\sum_i \operatorname{Var}(Y_i) = \sum \mathbb{E}Y_i = \sum \operatorname{Var}(X_i) < \infty$. By one-series theorem, $\sum_{i=1}^\infty Y_i$ converges almost surely. Thus $\sum_{i=1}^\infty X_i = \sum_{i=1}^\infty Y_i - \sum_{i=1}^\infty \mathbb{E}Y_i$ converges almost surely. ☐

[1] https://tzhouyi.wordpress.com/2016/05/18/cauchys-criterion-on-almost-sure-convergence/

Theorem 1.8.7 (Three-series theorem)**.** *Let $X_1, X_2, \ldots$ be independent r.v.'s. For the convergence of the series $\sum_i X_i$, it is necessary and sufficient that all the following three series converge:*

(i) for some cutoff value $C > 0$, $\sum_i \mathbb{P}(|X_i| > C)$ converge;
(ii) if $Y_i = X_i \mathbb{1}_{\{|X_i| \leq C\}}$, then $\sum_i \mathbb{E}Y_i$ converges;
(iii) $\sum_i \mathrm{Var}(Y_i)$ converges.

Proof. (Sufficiency). Let $Y_i = X_i \mathbb{1}_{\{|X_i| \leq C\}}$ as in (i). By Two-series theorem, the infinite sum $\sum_i Y_i$ converges almost surely. Then we note that

$$\mathbb{P}(X_i \neq Y_i \text{ i.o.}) = \mathbb{P}(|X_i| > C \text{ i.o.}).$$

By assumption (i), $\sum_i \mathbb{P}(|X_i| > C) < \infty$, so using Borel-Cantelli lemma, $\mathbb{P}(X_i \neq Y_i \text{ i.o.}) = 0$. This establishes the almost sure convergence of the sum $\sum_i X_i$.

(Necessity) If the first condition is not true for every $C > 0$, then by the second Borel-Cantelli lemma, for some $C > 0$, $\mathbb{P}(|X_i| > C \text{ i.o.}) = 1$. This clearly contradicts to the convergence of $\sum_i X_i$. So we can assume (i) holds for $A = 1$. Note that condition (iii) and the second-series theorem imply that $\sum_i (Y_i - \mathbb{E}Y_i)$ converges. Since $\sum_i Y_i$ is convergent, the sum $\sum_i \mathbb{E}Y_i$ converges as well. It remains to show (iii). By taking an i.i.d. copy Y_i' of Y_i and forming $Z_i = Y_i - Y_i'$, we can assume each Y_i has mean 0. To see this, note that Z_i is bounded by 2 if $|Y_i| \leq 1$ and $\mathrm{Var}(Z_i) = 2\mathrm{Var}(X_i)$, so the convergence of $\sum_i \mathrm{Var}(Z_i)$ and the convergence of $\sum_i \mathrm{Var}(X_i)$ are equivalent. We state the last portion of the proof in the following:

Lemma 1.8.8. *Let $X_1, X_2, \ldots$ be a sequence of independent r.v.'s with mean 0 and variances $\mathrm{Var}(X_i)$. If $\sum_i X_i$ converges, then $\sum_i \mathrm{Var}(X_i)$ converges.*

Proof of Lemma. Set $F_0 = \Omega$ and $S_0(\omega) = 0$ for all $\omega \in \Omega$. For $n \geq 1$, let $F_n = \{\omega : |S_1(\omega)| \leq \ell, \ldots, |S_n(\omega)| \leq \ell\}$. Since $\sum_i X_i$ converges, there is some $\ell > 0, \delta > 0$ such that $\mathbb{P}(F_n) \geq \delta$ for all n. Then we estimate, for each $n \geq 1$,

$$\begin{aligned}
\int_{F_{n-1}} S_n^2 d\mathbb{P} &= \int_{F_{n-1}} (S_{n-1} + X_n)^2 d\mathbb{P} \\
&= \int_{F_{n-1}} S_{n-1}^2 + 2S_{n-1}X_n + X_n^2 d\mathbb{P} \\
&= \int_{F_{n-1}} S_{n-1}^2 d\mathbb{P} + \sigma_n^2 \mathbb{P}(F_{n-1}) \geq \int_{F_{n-1}} S_{n-1}^2 d\mathbb{P} + \sigma_n^2 \delta.
\end{aligned}$$

On the other hand,

$$\int_{F_{n-1}} S_n^2 d\mathbb{P} = \int_{F_{n-1} \cap F_n^c} S_n^2 d\mathbb{P} + \int_{F_n} S_n^2 d\mathbb{P} \leq \int_{F_n} S_n^2 d\mathbb{P} + \mathbb{P}(F_{n-1} \cap F_n^c)(\ell + C)^2.$$

Combining these together, we have

$$\delta \sigma_n^2 \leq \int_{F_{n-1}} S_n^2 d\mathbb{P} - \int_{F_{n-1}} S_{n-1}^2 d\mathbb{P} \leq \int_{F_n} S_n^2 d\mathbb{P} - \int_{F_{n-1}} S_{n-1}^2 d\mathbb{P} + \mathbb{P}(F_{n-1} \cap F_n^c)(\ell + C)^2.$$

Thus,

$$\begin{aligned}
\delta\sum_{n=1}^{m}\sigma_n^2 &\le \sum_{n=1}^{m}\left(\int_{F_n} S_n^2 d\mathbb{P} - \int_{F_{n-1}} S_{n-1}^2 d\mathbb{P} + \mathbb{P}(F_{n-1}\cap F_n^c)(\ell+C)^2\right)\\
&= \int_{F_m} S_m^2 d\mathbb{P} + \sum_{n=1}^{m}\mathbb{P}(F_{n-1}\cap F_n^c)(\ell+C)^2\\
&= \int_{F_m} S_m^2 d\mathbb{P} + \mathbb{P}(\cup_{n=1}^{m}(F_{n-1}\cap F_n^c))(\ell+C)^2\\
&= \int_{F_m} S_m^2 d\mathbb{P} + \mathbb{P}(F_0\cap F_m^c)(\ell+C)^2\\
&\le \mathbb{P}(F_0)(\ell^2+(\ell+C)^2).
\end{aligned}$$

This concludes the proof of the lemma. □

The tool connecting the convergence of S_n/n and the convergence of random series is Kronecker's lemma.

Lemma 1.8.9 (Kronecker)**.** *If $a_n \uparrow \infty$ and $\sum_{n=1}^{\infty} x_n/a_n$ converges, then*

$$\frac{\sum_{m=1}^{n} x_m}{a_n} \to 0 \ \textit{as}\ n\to\infty.$$

Proof. The argument is similar to the argument of the usual summability theorems. Let $b_n = \sum_{m=1}^{n} x_m/a_m, b_0 = 0, a_0 = 0$. Note in this way, $x_n = a_n(b_n - b_{n-1})$. Then observe,

$$\begin{aligned}
\frac{\sum_{m=1}^{n} x_m}{a_n} &= \frac{1}{a_n}\sum_{m=1}^{n} a_m(b_m - b_{m-1})\\
&= b_n + \frac{1}{a_n}\sum_{m=1}^{n}(a_{m-1}-a_m)b_{m-1}\\
&= b_n - \sum_{m=1}^{n}\frac{a_m - a_{m-1}}{a_n}b_{m-1}.
\end{aligned}$$

We should observe the telescoping nature in the last term which is simply an weighted average of $b_0, ..., b_{n-1}$. It suffices to show this term goes to $b_\infty := \lim_{n\to\infty} b_n$ as $n\to\infty$. Clearly the error is

$$\left|\sum_{m=1}^{n}\frac{a_m - a_{m-1}}{a_n}b_{m-1} - b_\infty\right| \le \sum_{m=1}^{n}\frac{a_m - a_{m-1}}{a_n}|b_{m-1} - b_\infty| \tag{e1}$$

We shall divide the sum into two parts. The contribution coming from the first part will be controlled by $1/a_n$ and the contribution from the second part will be controlled by $|b_{m-1} - b_\infty|$.

Since $(b_n)_{n=1}^{\infty}$ is convergent, we can find a constant $C = \sup_n |b_n|$. Let us pick $M > 0$ such that when $n \ge M$, $|b_m - b_\infty| < \varepsilon/2$. Then we pick $N > M$ so that when $n > N$, $a_M/a_n < \varepsilon/4C$. Hence the sum of the first M terms

$$\sum_{m=1}^{M}\frac{a_m - a_{m-1}}{a_n}|b_{m-1} - b_\infty| \le 2C\sum_{m=1}^{M}\frac{a_m - a_{m-1}}{a_n} = 2C\frac{a_M}{a_n} < \frac{\varepsilon}{2}.$$

The rest of the sum

$$\sum_{m=M+1}^{n}\frac{a_m - a_{m-1}}{a_n}|b_{m-1} - b_\infty| \le \frac{a_n - a_M}{a_n}\cdot\frac{\varepsilon}{2} < \frac{\varepsilon}{2}.$$

This shows the error (e1) is bounded by ε and the proof is complete. □

Now we can give another proof of the strong law of large numbers based on tools given in this section.

Second Proof of Theorem 1.7.1. We start as usual by truncation. Let $Y_n = X_n \mathbb{1}_{\{|X_n| \le n\}}$ and $T_n = \sum_{i=1}^n Y_n$. By Lemma 1.7.2, it suffices to show $T_n/n \to \mu =: \mathbb{E}X_1$ almost surely. Denote $Z_n = (Y_n - \mathbb{E}Y_n)/n$. Note that $\mathbb{E}Z_n = 0$ and $\mathrm{Var}(Z_n) = \mathrm{Var}(Y_n)/n^2$. Lemma 1.7.4 implies $\sum_{k=1}^\infty \mathrm{Var}(Z_k) \le 4\mathbb{E}|X_1| < \infty$. So by the second-series theory, the sum $\sum_n Z_n$ converges almost surely. At these sample points, by Kronecker's lemma,

$$\frac{\sum_{k=1}^n Y_k - \mathbb{E}Y_k}{n} \to 0 \text{ as } n \to \infty.$$

That is, $\lim_{n\to\infty}(\sum_{k=1}^n Y_k)/n = \lim_{n\to\infty}(\sum_{k=1}^n \mathbb{E}Y_k)/n = \mu$ where the last equality follows from $\mathbb{E}Y_k \to \mu$ as $k \to \infty$. □

Using Kronecker's lemma and Kolmogorov's theorems on convergence of random series, we are able to investigate the convergence of S_n/a_n with $a_n \uparrow \infty$ as $n \to \infty$. Clearly if $\mathbb{E}|X_1| = \infty$, we should expect a_n growing faster than n.

Theorem 1.8.10. *Let $X_1, X_2, \ldots$ be i.i.d. r.v.'s with $\mathbb{E}X_1 = 0$ and $\mathrm{Var}(X_1) = \sigma^2 < \infty$. If $\varepsilon > 0$, then*

$$\frac{S_n}{n^{1/2}(\log n)^{1/2+\varepsilon}} \to 0 \ a.s.$$

Remark 1.8.11. The optimal scaling will be given by the law of iterated logarithm, which will be covered later.

Proof. Denote $a_n = n^{1/2}(\log n)^{1/2+\varepsilon}$. By Kronecker's lemma, we only need to show $\sum_{m=1}^n X_m/a_m$ converges almost surely. Note that for each $m \ge 1$, $\mathbb{E}X_m/a_m = 0$. So we compute

$$\sum_{m=2}^\infty \mathrm{Var}(X_m/(m^{1/2}(\log m)^{1/2+\varepsilon}) = \sigma^2 \sum_{m=2}^\infty \frac{1}{m(\log m)^{1+2\varepsilon}} < \infty.$$

Thus by two-series theorem, $\sum_{m=1}^n X_m/a_m$ converges almost surely. □

Theorem 1.8.12 (Marcinkiewicz-Zygmund)**.** *Let $X_1, X_2, \ldots$ be centered i.i.d. r.v.'s and $S_n = X_1 + \cdots + X_n$. If $\mathbb{E}|X_1|^p < \infty$ where $1 < p < 2$, then $S_n/n^{1/p} \to 0$ a.s.*

Proof. As in the proof of the strong law (Theorem 1.7.1), we truncate the r.v.'s by defining $Y_n = X_n \mathbb{1}_{\{|X_n| \le |n|^{1/p}\}}$. Then note

$$\sum_{n=1}^\infty \mathbb{P}(Y_n \ne X_n) \le \sum_{n=1}^\infty \mathbb{P}(|X_n|^p > n) = \mathbb{E}|X_1|^p < \infty,$$

and by Borel-Cantelli lemma, $X_n = Y_n$ eventually with probability 1. So it suffices to show $T_n/n^{1/p} \to 0$ a.s. where $T_n = Y_1 + \cdots + Y_n$. To do this, by Kronecker's lemma, we will show a.s. convergence of the series $\sum_{n=1}^\infty T_n/n^{1/p}$. We start by computing

$$\begin{aligned}
\sum_{n=1}^\infty \mathrm{Var}(T_n/n^{1/p}) &\le \sum_{n=1}^\infty \frac{\mathbb{E}T_n^2}{n^{2/p}} \\
&= \sum_{n=1}^\infty \sum_{m=1}^n \int_{(m-1)^{1/p}}^{m^{1/p}} \frac{1}{n^{2/p}} 2y\mathbb{P}(|X_1| > y)dy \\
&= \sum_{m=1}^\infty \int_{(m-1)^{1/p}}^{m^{1/p}} \sum_{n=m}^\infty \frac{1}{n^{2/p}} 2y\mathbb{P}(|X_1| > y)dy.
\end{aligned}$$

Then there is a constant $C > 0$ so that the sum $\sum_{n=m}^{\infty} 1/n^{2/p} \le \int_{m-1}^{\infty} 1/x^{2/p}dx = \frac{p}{2-p}(m-1)^{(p-2)/p} \le Cy^{p-2}$ whenever $y \in [(m-1)^{1/p}, m^{1/p}]$. Thus $\sum_{n=1}^{\infty} \mathrm{Var}(T_n/n^{1/p}) \le \int_0^{\infty} 2Cy^{p-1}\mathbb{P}(|X_1| > y)dy = 2C\mathbb{E}|X_1|^p/p < \infty$. Denote $\mu_n = \mathbb{E}Y_n$. It remains to show $\sum_n \mu_n/n^{1/p}$ converges. The rest follows from the three-series theorem. Note that because X_i's are centered, $\mu_n = -\mathbb{E}X_n\mathbb{1}_{\{|X_n|>n^{1/p}\}}$. Thus,

$$\begin{aligned}
|\mu_n| &\le \int |X_n|\mathbb{1}_{\{|X_n|>n^{1/p}\}}d\mathbb{P} \\
&= n^{1/p}\int \frac{|X_n|\mathbb{1}_{\{|X_n|>n^{1/p}\}}}{n^{1/p}}d\mathbb{P} \\
&\le n^{1/p}\int \frac{|X_n|^p\mathbb{1}_{\{|X_n|>n^{1/p}\}}}{n}d\mathbb{P} \\
&= \frac{\mathbb{E}|X_n|^p\mathbb{1}_{\{|X_n|>n^{1/p}\}}}{n^{1-1/p}}
\end{aligned}$$

Then we can bound the partial sum

$$\frac{1}{n^{1/p}}\sum_{m=1}^{n}\frac{1}{m^{1-1/p}} \le \frac{1+\int_1^n x^{-1+1/p}dx}{n^{1/p}} < C$$

by some $C > 0$ independent of n. Since $\mathbb{E}|X_n|^p\mathbb{1}_{\{|X_n|>n^{1/p}\}} \to 0$ as $n \to \infty$, by Dirichlet test, $\sum_{n=1}^{\infty} \mu_n/n^{1/p} < \infty$. This completes the proof. □

Recall in section 1.6, we proved if $\mathbb{E}|X_1| = \infty$, the probability that S_n/n has a finite limit is zero. We now can extend this result furthermore.

Theorem 1.8.13 (Feller)**.** *Let $X_1, X_2, \dots$ be i.i.d. r.v.'s with $\mathbb{E}|X_1| = \infty$. If a_n/n is positive and increasing, then $\lim_{n\to\infty} |S_n|/a_n = 0$ or ∞ according as $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > a_n) < \infty$ or $= \infty$.*

Proof. Suppose $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > a_n) < \infty$. Let $Y_n = X_n\mathbb{1}_{\{|X_n|\le a_n\}}$ and let $T_n = Y_1 + \cdots + Y_n$. The assumption implies $\mathbb{P}(X_n \ne Y_n \text{ i.o.}) = 0$. So we shall check conditions in the two-series theorem. To begin with, denote $a_0 = 0$ and we compute

$$\begin{aligned}
\sum_{n=1}^{\infty}\mathrm{Var}(Y_n/a_n) &\le \sum_{n=1}^{\infty}\frac{\mathbb{E}Y_n^2}{a_n^2} \\
&\le \sum_{n=1}^{\infty} a_n^{-2}\sum_{m=1}^{n}\int_{a_{m-1}}^{a_m} y^2 dF(y) \\
&= \sum_{m=1}^{\infty}\int_{a_{m-1}}^{a_m} y^2 dF(y)\sum_{n=m}^{\infty} a_n^{-2}.
\end{aligned}$$

Note that $a_n/n \ge a_m/m$, so $\sum_{n=m}^{\infty} a_n^{-2} \le (m^2/a_m^2)\sum_{n=m}^{\infty} 1/n^2 \le Cm/a_m^2$. Thus,

$$\sum_{n=1}^{\infty}\mathrm{Var}(Y_n/a_n) \le C\sum_{m=1}^{\infty}\int_{a_{m-1}}^{a_m} m\frac{y^2}{a_m^2}dF(y) \le C\sum_{m=1}^{\infty} m\mathbb{P}(a_{m-1} \le |X_1| < a_m).$$

By Fubini's theorem,

$$\sum_{m=1}^{\infty} m\mathbb{P}(a_{m-1} \le |X_1| < a_m) = \sum_{k=1}^{\infty}\sum_{m=k}^{\infty}\mathbb{P}(a_{m-1} \le |X_1| < a_m) = \sum_{k=1}^{\infty}\mathbb{P}(|X_1| \ge a_{k-1}) < \infty.$$

This implies $\sum_{n=1}^{\infty}\mathrm{Var}(Y_n/a_n) < \infty$. By two-series theorem, $\sum_{n=1}^{\infty}(Y_n - \mathbb{E}Y_n)/a_n$ converges almost surely. Thus Kronecker's lemma implies $\sum_{m=1}^{n}(Y_n - \mathbb{E}Y_n)/a_n \to 0$. It is then sufficient to show $(\sum_{i=1}^n \mathbb{E}Y_i)/a_n \to 0$ as $n \to \infty$. Note that $a_n/n \uparrow \infty$ for otherwise $a_n/n < C$ will contradict to the fact that $\sum_{n=1}^\infty \mathbb{P}(|X_1| > n) = \infty$ but $\sum_{n=1}^\infty \mathbb{P}(|X_1| > a_n) < \infty$ We observe

$$\begin{aligned}
\left|\frac{\sum_{i=1}^n \mathbb{E}Y_i}{a_n}\right| &\le \frac{\sum_{i=1}^n \mathbb{E}|X_1|\mathbb{1}_{\{|X_1|\le a_i\}}}{a_n} \\
&\le \frac{n\mathbb{E}|X_1|\mathbb{1}_{\{|X_1|\le a_N\}} - (n-N)\mathbb{E}|X_1|\mathbb{1}_{\{|X_1|\le a_N\}} + (n-N)\mathbb{E}|X_1|\mathbb{1}_{\{|X|\le a_n\}}}{a_n} \\
&= \frac{na_N + (n-N)\mathbb{E}|X_1|\mathbb{1}_{\{a_N<|X|\le a_n\}}}{a_n} \le \frac{na_N}{a_n} + \frac{n\mathbb{E}|X_1|\mathbb{1}_{\{a_N<|X|\le a_n\}}}{a_n}
\end{aligned}$$

holds for any integer $N > 0$. Note that $n/a_n \downarrow 0$, so $na_N/a_n \to 0$. The second term is

$$\begin{aligned}
\frac{n\mathbb{E}|X_1|\mathbb{1}_{\{a_N<|X|\le a_n\}}}{a_n} &\le \sum_{m=N+1}^{n} \frac{m}{a_m}\mathbb{E}|X_1|\mathbb{1}_{\{a_{m-1}<|X_1|\le_m\}} \\
&\le \sum_{m=N+1}^{\infty} m\mathbb{P}(a_{m-1} < |X_1| \le a_m) < \infty.
\end{aligned}$$

Letting $N \to \infty$, we see that $(\sum_{i=1}^n \mathbb{E}Y_i)/a_n \to 0 \to 0$ as $n \to \infty$. □

1.9 Large Deviations

Let $X_1, X_2, ...$ be i.i.d. and let $S_n = X_1 + \cdots + X_n$. We will investigate, in this section, the rate of convergence of $\mathbb{P}(S_n \ge na) \to 0$ for $a \ge \mu := \mathbb{E}X_1$. Assume the **moment-generating function** $\mathbb{E}e^{\theta X_1} < \infty$ for some $\theta > 0$. First let us show the existence of the following object,

$$\gamma(a) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \ge na).$$

Lemma 1.9.1. *If $\gamma_{m+n} \ge \gamma_m + \gamma_n$, then $\gamma_n/n \to \sup_m \gamma_m/m$ as $n \to \infty$.*

Thus we obtain the following bound

$$\mathbb{P}(S_n \ge na) \le e^{n\gamma(a)}.$$

Denote $\varphi(\theta) = \mathbb{E}e^{\theta X_1}$. Then Chebyshev's inequality implies

$$e^{\theta na}\mathbb{P}(S_n \ge na) \le \mathbb{E}e^{\theta S_n} = \varphi(\theta)^n.$$

Setting $\kappa(\theta) = \ln \varphi(\theta)$, we see that

$$\mathbb{P}(S_n \ge na) \le e^{-n(a\theta - \kappa(\theta))}.$$

Lemma 1.9.2. *If $a > \mu$ and $\theta > 0$ is small, then $a\theta - \kappa(\theta) > 0$.*

Now we attempt to optimize the upper bound, $\theta a - \kappa(\theta)$. We first formally differentiate this object,

$$\frac{d}{d\theta}\theta a - \kappa(\theta) = a - \frac{\varphi'(\theta)}{\varphi(\theta)}.$$

It can be checked that the maximum occurs when $a = \varphi'(\theta)/\varphi(\theta)$.

Example 1.9.3 (Normal distribution)**.** Suppose X has standard normal distribution.

$$\mathbb{E}e^{\theta X} = \int e^{\theta x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = e^{\frac{\theta^2}{2}} \int \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\theta)^2}{2}}dx = e^{\frac{\theta^2}{2}}.$$

In this case $\varphi'(\theta)/\varphi(\theta) = \theta$.

Example 1.9.4 (Exponential distribution)**.** Suppose $X \sim \exp(\lambda)$.

2 Central Limit Theorems

2.1 The De Moivre-Laplace Theorem

Recall the **Stirling's formula** $n! \sim n^n e^{-n}\sqrt{2\pi n}$ where $a_n \sim b_n$ means $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. Our main goal is to prove

Theorem 2.1.1 (De Moivre-Laplace)**.** *Let $X_1, X_2, ...$ be i.i.d. r.v.'s with $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$ and let $S_n = \sum_{i=1}^n X_i$. Then as $n \to \infty$*

$$\mathbb{P}(a \le S_n/\sqrt{n} \le b) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx.$$

We start with a few lemmas.

Lemma 2.1.2. *If $c_n \to 0$, $a_n \to \infty$ and $a_n c_n \to \lambda$, then $(1 + c_n)^{a_n} \to e^{\lambda}$.*

Proof. Recall that $x/(1 + x) \le \log(1 + x) \le x$ for $x > -1$. So for sufficiently large n,

$$\frac{a_n c_n}{1 + c_n} \le a_n \log(1 + c_n) \le a_n c_n.$$

Since both upper and lower bounds converge to λ, the lemma follows. □

Remark 2.1.3. This result can be generalized: if $\max_{1\le j\le n} |c_j, n| \to 0$, $\sum_{j=1}^n c_{j,n} \to \lambda$ and $\sup_n \sum_{j=1}^n |c_{j,n}| < \infty$, then $\prod_{j=1}^n (1 + c_{j,n}) \to e^{\lambda}$.

Clearly S_n can be visualized as the trajectory of a random walk with i.i.d. increments of size 1. After $2n$ steps, we first compute the probability

$$\mathbb{P}(S_{2n} = 2k) = \binom{2n}{n+k} 2^{-n} = \frac{n!}{(n-k)!(n+k)!} 2^{-n}.$$

Note that

$$\begin{aligned}
\frac{(2n)!}{(n-k)!(n+k)!} 2^{-2n} &\sim \frac{2n^{2n} e^{-2n}\sqrt{2\pi 2n}}{(n-k)^{n-k} e^{-n+k}\sqrt{2\pi(n-k)}(n+k)^{n+k} e^{-n-k}\sqrt{2\pi(n+k)}} 2^{-2n} \\
&= \frac{n^{2n} e^{-2n}\sqrt{2n}}{(n^2-k^2)^n e^{-2n}\sqrt{2\pi(n^2-k^2)}(n-k)^{-k}(n+k)^k} \\
&= \frac{n^{2n} e^{-2n}\sqrt{2n}}{n^{2n}(1-\frac{k^2}{n^2})^n e^{-2n}\sqrt{2\pi(n^2-k^2)}(1-\frac{k}{n})^{-k}(1+\frac{k}{n})^k} \\
&= \frac{1}{(1-\frac{k^2}{n^2})^n (1-\frac{k}{n})^{-k}(1+\frac{k}{n})^k \sqrt{\pi n(1-\frac{k^2}{n^2})}}
\end{aligned}$$

If $k = x\sqrt{n/2}$, then as $n \to \infty$, $(1-\frac{k^2}{n^2})^n \to e^{x^2/2}$, $(1-\frac{k}{n})^{-k} \to e^{-x^2/2}$, $(1+\frac{k}{n})^k \to e^{-x^2/2}$ and $(1-\frac{k^2}{n^2}) \to 1$. Hence we obtain

Lemma 2.1.4. *If $2k/\sqrt{2n} \to x$, then $\mathbb{P}(S_{2n} = 2k) \sim e^{-x^2/2}/\sqrt{\pi n}$.*

2.2 Weak Convergence

Definition 2.2.1. A sequence $(F_n)_{n=1}^{\infty}$ of distribution functions is said to **converge weakly to** a limit F (written $F_n \Rightarrow F$) if $F_n(y) \to F(y)$ for all y that are continuous points of F. A sequence $(X_n)_{n=1}^{\infty}$ is said to **converge weakly** or **converge in distribution** to a limit X_∞ if their distribution functions $F_n(x) = \mathbb{P}(X_n \le x)$ converge weakly.

It should be noted that convergence at points of continuity is enough to identify the limit. To see this, suppose we have distribution functions F, G which agree at points of continuity. It is clear that they have the same set of points of continuity. Let D denote the set of discontinuities. For $x \in D$, by the right continuity we cannot have a $x_n \in D$ converging to x. This implies that there is $\delta_x > 0$ such that $(x, x+\delta_x)$ on which F, G are continuous, so they agree on this set. Passing to the limit, we see that $F(x) = G(x)$ as well.

We have seen in previous section that if $X_1, X_2, ...$ are i.i.d. signed Bernoulli r.v.'s then $S_n/\sqrt{n} \Rightarrow \mathcal{N}(0,1)$ where $\mathcal{N}(\mu, \sigma^2)$ stands for the normal distribution with mean μ and variance σ^2. Also the Glivenko-Cantelli theorem tells us that for almost every ω,

$$F_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i(\omega)\le y\}} \to F(y) \text{ for all } y.$$

The next example shows why we restrict our attention to continuity points.

Example 2.2.2. Consider $X_n := X + 1/n$. Clearly $F_n(x) = \mathbb{P}(X \le x - 1/n) \to \lim_{y\uparrow x} F(x)$ as $n \to \infty$. The convergence only occurs at continuity points, but still we like to include this situation into consideration.

Example 2.2.3 (Waiting for rare events)**.**

Example 2.2.4 (Birthday problem)**.**

Theorem 2.2.5 (Scheffé)**.** *If a sequence of probability densities $f_n \to f$ pointwise, as $n \to \infty$, then the convergence also holds under* **total variation norm**

$$||\mu_n - \mu|| = \sup_{B\in\mathcal{B}} |\mu_n(B) - \mu(B)| \to 0$$

Theorem 2.2.6 (Skorokhod)**.** *If $F_n \Rightarrow F$, then there is r.v.'s Y_n with distributions F_n for $1 \le n \le \infty$, so that $Y_n \to Y_\infty$ a.s.*

Proof. Let us take the probability space $((0,1), \mathcal{B}(0,1), \mu)$ where μ is the Lebesgue measure on $(0,1)$. For $\omega \in (0,1)$, define $Y_n(\omega) = \sup\{x : F_n(x) < \omega\}$. As we saw in Theorem 1.1.13, $Y_n(\omega) \le x$ if and only if $\omega \le F_n(x)$ so that Y_n has distribution F_n. We shall show that $Y_n(\omega) \to Y(\omega)$ whenever Y is continuous at ω. Let $\varepsilon > 0$, and choose $x \in (Y(\omega) - \varepsilon, Y(\omega))$ such that $F(x) - F(x-) = 0$, i.e. F is continuous at x. Clearly $F(x) < \omega$. Since $F_n(x) \to F(x)$, for sufficiently large n, $F_n(x) < \omega$. This implies $Y_n(\omega) > x > Y_(\omega) - \varepsilon$. It follows that $\liminf_{n\to\infty} Y_n(\omega) \ge Y(\omega)$. Suppose now that $\omega' > \omega$. Pick $y \in (Y(\omega'), Y(\omega') + \varepsilon)$ such that F is continuous at y. Then observe $\omega < \omega' \le F(Y(\omega')) \le F(y)$. Thus for sufficiently large n, $F_n(y) \ge \omega$ and it follows that $Y_n(\omega) \le y < Y(\omega') + \varepsilon$. Thus $\limsup_{n\to\infty} Y_n(\omega) \le Y(\omega')$ whenever $\omega' > \omega$.

Since Y is increasing, the set D of discontinuities is at most countable. So $\mu(D) = 0$ and $Y_n \to Y$ on $(0,1) \setminus D$. This completes the proof. □

Theorem 2.2.7. *$X_n \Rightarrow X$ if and only if for any bounded continuous function f, $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$.*

Proof. $\Rightarrow$ Using previous theorem, we may assume that $Y_n \to Y$ a.s. and $Y_n =_d X_n, Y =_d X$. Then bounded convergence theorem implies

$$\mathbb{E}f(X_n) = \mathbb{E}f(Y_n) \to \mathbb{E}f(Y) = \mathbb{E}f(X).$$

$\Leftarrow$ Let $\varepsilon > 0$. Define

$$f_{x,\varepsilon}(t) = \begin{cases} 1 : t \le x \\ \text{linear} : x \le t \le x+\varepsilon \\ 0 : t \ge x+\varepsilon \end{cases}$$

Clearly f is bounded and continuous, so $\mathbb{E}f_{x,\varepsilon}(X_n) \to \mathbb{E}f_{x,\varepsilon}(X)$ as $n \to \infty$. Suppose x is a continuous point of the distribution of X. Note that

$$\limsup_{n\to\infty} \mathbb{P}(X_n \le x) \le \limsup_{n\to\infty} \mathbb{E}f_{x,\varepsilon}(X_n) = \mathbb{E}f_{x,\varepsilon}(X) \le \mathbb{P}(X \le x + \varepsilon),$$

and

$$\liminf_{n\to\infty} \mathbb{P}(X_n \le x) \ge \liminf_{n\to\infty} \mathbb{E}f_{x-\varepsilon,\varepsilon}(X_n) = \mathbb{E}f_{x-\varepsilon,\varepsilon}(X) \ge \mathbb{P}(X \le x - \varepsilon).$$

Letting $\varepsilon \to 0$, we have

$$\limsup_{n\to\infty} \mathbb{P}(X_n \le x) \le \mathbb{P}(X \le x) = \mathbb{P}(X < x) \le \liminf_{n\to\infty} \mathbb{P}(X_n \le x).$$

Thus, $\lim_{n\to\infty} \mathbb{P}(X_n \le x) = \mathbb{P}(X \le x)$ whenever x is a continuous point of the distribution of X, i.e., $X_n \Rightarrow X$. □

A trivial but useful generalization of this characterization of weak convergence is

Theorem 2.2.8 (Continuous mapping)**.** *Let g be a measurable and let D_g be the set of discontinuities of g. If $X_n \Rightarrow X$ and $\mathbb{P}(X \in D_g) = 0$, then $g(X_n) \Rightarrow g(X)$. In addition, if g is bounded, then $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$.*

Proof. Let f be a bounded continuous function. We will show that $\mathbb{E}f(g(X_n)) \to \mathbb{E}f(g(X))$. For simplicity, denote $h = f \circ g$. Clearly $D_h \subset D_g$ because wherever g is continuous h is continuous there as well. Pick $Y_n =_d X_n$ and $Y =_d X$ such that $Y_n \to Y$ a.s. So $\mathbb{P}(Y \in D_h) = 0$ and it follows that $h(Y_n) \to h(Y)$ a.s. Hence by the bounded convergence theorem $\mathbb{E}h(X_n) \to \mathbb{E}h(X)$ which justifies $g(X_n) \Rightarrow g(X)$. The rest is trivial. □

We now state a set of equivalent notions of weak convergence.

Theorem 2.2.9 (Portmanteau)**.** *The following statements are equivalent:*

(i) $X_n \Rightarrow X$.
(ii) For all open sets G, $\liminf_{n\to\infty} \mathbb{P}(X_n \in G) \ge \mathbb{P}(X \in G)$.
(ii) For all closed sets C, $\limsup_{n\to\infty} \mathbb{P}(X_n \in C) \le \mathbb{P}(X \in C)$.
(iv) For all sets A with $\mathbb{P}(X \in \partial A) = 0$, $\lim_{n\to\infty} \mathbb{P}(X_n \in G) = \mathbb{P}(X \in G)$.

Theorem 2.2.10 (Helly's selection)**.** *If $(F_n)_{n=1}^\infty$ is a sequence of distribution functions, then there is a right continuous increasing function F such that $F_{n(k)} \Rightarrow F$ along some subsequence $(F_{n(k)})_{k=1}^\infty$.*

Proof. Let $(q_i)_{i=1}^\infty$ be an enumeration of the rationals. Since F_n's are bounded, by Cantor's diagonal process we can find a subsequence $(F_{n(k)})_{k=1}^\infty$ such that $F_{n(k)}(q_i)$ converges at all q_i's as $k \to \infty$. Let $F(q_i) = \lim_{k\to\infty} F_{n(k)}(q_i)$. We will modify F to make it right-continuous. Define $G(x) = \inf\{F(q) : q \in \mathbb{Q}, q > x\}$. Clearly G is increasing and $F(q) \le G(q)$ whenever $q \in \mathbb{Q}$. We observe

$$\begin{aligned}
\lim_{y_n \downarrow x} G(y_n) &= \lim_{y_n \downarrow x} \inf\{G(q) : q \in \mathbb{Q}, q > y_n\} \\
&= \inf\{G(q) : q \in \mathbb{Q}, q > y_n \text{ for some } n\} \\
&= \inf\{G(q) : q \in \mathbb{Q}, q > x\} = G(x).
\end{aligned}$$

Now we claim that $F_{n(k)} \Rightarrow G$. Let x be a continuity point of G. Let $\varepsilon > 0$. We find rational numbers s_1, s_2, t such that $s < x < t$ and

$$G(x) - \varepsilon < G(s_1) \le G(s_2) \le G(x) \le G(t) < G(x) + \varepsilon.$$

Since $F_{n(k)}(t) \to F(t) \le G(t)$, we have $F_{n(k)}(t) \le G(t)$ for large enough k. Similarly $F_{n(k)}(s_2) \ge F(s_2) \ge G(s_1)$ for sufficiently large k. Hence we obtain

$$G(x) - \varepsilon < F_{n(k)}(s_2) \le F_{n(k)}(x) \le F_{n(k)}(t) < G(x) + \varepsilon$$

for sufficiently large k. This concludes the proof. □

Remark 2.2.11. It is possible, however, that the limit is not a distribution function. For instance, suppose G is a distribution function and let $F_n(x) = a\mathbb{1}_{\{x\geq n\}} + b\mathbb{1}_{\{x\geq -n\}} + cG(x)$ where $a+b+c=1$. Then $F_n(x) \to F(x) := b + cG(x)$, $\lim_{x\to\infty} F(x) = b + c = 1 - a$ and $\lim_{x\to-\infty} F(x) = c$. Intuitively, an amound of mass a escapes to $+\infty$ and mass b escapes to $-\infty$.

To avoid this situation, we impose an additional condition called **tightness**.

Theorem 2.2.12. *Let $(F_n)_{n=1}^\infty$ be a sequence of distribution functions. Every subsequential limit is the distribution function of a probability measure if and only if $(F_n)_{n=1}^\infty$ is* **tight**, *i.e., for all $\varepsilon > 0$, there is an M_ε so that*

$$\limsup_{n\to\infty} 1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon) < \varepsilon.$$

Proof. $\Leftarrow$ Let $\varepsilon > 0$. Suppose that F is a weak limit of a subsequence $(F_{n(k)})_{k=1}^\infty$. Pick continuity points s,t of F such that $F(t) > F(M_\varepsilon)$ and $F(s) < F(-M_\varepsilon)$. Then we obtain

$$1 - F(t) = \lim_{k\to\infty} 1 - F_{n(k)}(t) \leq \limsup_{n\to\infty} 1 - F_n(M_\varepsilon) < \varepsilon$$

and

$$F(s) = \lim_{k\to\infty} F_{n(k)}(s) \leq \limsup_{n\to\infty} F_n(-M_\varepsilon) < \varepsilon.$$

It follows that $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$ and thus F is the distribution function of a probability measure.

$\Rightarrow$ Now we suppose $(F_n)_{n=1}^\infty$ is not tight. By definition, there is some $\delta > 0$ such that for all M, $\limsup_{n\to\infty} 1 - F_n(M) + F_n(-M) > \delta$. This implies the existence of a subsequence $(F_{n(k)})_{k=1}^\infty$ such that

$$1 - F_{n(k)}(M) + F_{n(k)}(-M) > \delta \tag{2.1}$$

for all k. By Helly's selection theorem, we can find a subsequence $(F_{n(k_j)})_{j=1}^\infty$ whose weak limit F. Clearly (2.1) implies that F cannot be a distribution function. □

The following sufficient condition for tightness is useful.

Theorem 2.2.13. *Suppose $\varphi \geq 0$ and $\varphi(x) \to \infty$ as $|x| \to \infty$. If*

$$C = \sup_n \int \varphi(x) dF_n(x) < \infty,$$

then $(F_n)_{n=1}^\infty$ is tight.

Proof. We just note that $[1 - F_n(M) - F_n(-M)] \inf_{|x|\geq M} \varphi(x) \leq C$. Dividing both sides by $\inf_{|x|\geq M} \varphi(x)$ yields the result. □

2.3 Characteristic Functions

Definition 2.3.1. Suppose X is a random variable. Its **characteristic function (ch.f.)** is defined to be $\varphi(t) = \mathbb{E}e^{itX} = \mathbb{E}\cos tX + i\mathbb{E}\sin tX$.

Immediately we have the following
Properties:

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

(b) $\overline{\varphi(t)} = \varphi(-t)$.

(c) $|\varphi(t)| \leq 1$.

(d) $t \mapsto \varphi(t)$ is uniformly continuous.

(e) $\mathbb{E}e^{it(aX+b)} = e^{itb}\varphi(at)$.

(f) If X_1, X_2 are independent r.v.'s and have ch.f.'s φ_1, φ_2, then $X_1 + X_2$ has ch.f. $\varphi_1(t)\varphi_2(t)$.

Example 2.3.2 (Coin flips)**.** If $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$, then X has ch.f.

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

Example 2.3.3 (Poisson)**.** If $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k = 0, 1, 2, ...$, then

$$\mathbb{E}e^{itX} = \sum_{k=0}^{\infty} e^{-\lambda}\frac{(e^{it}\lambda)^k}{k!} = e^{\lambda(e^{it}-1)}.$$

Example 2.3.4 (Normal)**.** Suppose $X \sim \mathcal{N}(0, 1)$. We claim that $\mathbb{E}e^{itX} = e^{-t^2/2}$. To see this, first note that since $\sin t$ is an odd function, we only need to compute

$$\mathbb{E}e^{itX} = \mathbb{E}\cos tX = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\cos(tx)dx.$$

Differentiating w.r.t. t under the integral, we obtain

$$[\mathbb{E}e^{itX}]' = \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-x^2/2} x(-\sin tx)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}(-t\cos tx)dx = -t\mathbb{E}e^{itX}.$$

The unique solution of this ODE, is then given by $\mathbb{E}e^{itX} = e^{-t^2/2}$. Combing this and Property (e) yield that for a r.v. $X \sim \mathcal{N}(\mu, \sigma^2)$, its ch.f. is $\varphi(t) = e^{it\mu-(\sigma t)^2/2}$.

Example 2.3.5 (Uniform(a,b))**.** Suppose $X \sim \text{Uniform}(a, b)$. Its ch.f. is

$$\mathbb{E}e^{itX} = \frac{1}{b-a}\int_a^b e^{itx}dx = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

In particular if $a = -c, b = c$, then $\mathbb{E}e^{itX} = (\sin ct)/ct$.

Example 2.3.6 (Triangular)**.**

Example 2.3.7 (Exponential)**.** Suppose $X \sim \exp(\lambda)$. Then

$$\mathbb{E}e^{itX} = \int_0^{\infty} \lambda e^{itx}e^{-\lambda x}dx = \frac{\lambda e^{(it-\lambda)x}}{it-\lambda}\Big|_0^{\infty} = \frac{\lambda}{\lambda - it}.$$

Example 2.3.8 (Bilateral exponential)**.** Suppose X has density $e^{-|x|}/2$. Its ch.f. is $1/(1+t^2)$. To prove this, we need

Lemma 2.3.9. *If the probability distributions $F_1, ..., F_n$ have ch.f. $\varphi_1, ..., \varphi_n$ and $\lambda_i \geq 0$ have $\sum_{i=1}^n \lambda_i = 1$, then $\sum_{i=1}^n \lambda_i F_i$ has ch.f. $\sum_{i=1}^n \lambda_i\varphi_i$.*

The first deep result is that the ch.f. determines the distribution uniquely.

Theorem 2.3.10 (The inversion formula)**.** *If $\varphi(t) = \int e^{itx}\mu(dx)$ where μ is a probability measure, and $a < b$, then*

$$\lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\varphi(t)dt = \mu(a, b) + \frac{1}{2}\mu\{a, b\}.$$

Proof. Denote

$$I_T = \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = \int_{-T}^{T} \int \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \mu(dx) dt.$$

Observe the integrand has

$$\left| \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \right| \leq \int_a^b |e^{ity}| dy \leq b - a.$$

So we can apply Fubini's theorem to obtain

$$I_T = \int \int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt \mu(dx) = \int \int_{-T}^{T} \frac{\sin(t(x-a)) - \sin(t(x-b))}{t} dt \mu(dx),$$

where the second equality follows from Euler's formula and symmetries of $\cos, \sin$. Denote $R(\theta, T) = \int_{-T}^{T} (\sin \theta t)/t dt$. We may rewrite the last equation as

$$I_T = \int R(x-a, T) - R(x-b, T) \mu(dx).$$

Setting $S(T) = \int_0^T (\sin x)/x dx$ and changing variables $t = x/\theta$, we relate the two quantities by

$$R(\theta, T) = \int_{-T}^{T} \frac{\sin \theta t}{t} dt = 2 \int_0^{T\theta} \frac{\sin x}{x} dx = 2S(T\theta), \theta > 0;$$

and for $\theta < 0$, $R(\theta, T) = -R(|\theta|, T) = -S(T|\theta|)$. Define the usual signum function $\operatorname{sgn} x := -\mathbb{1}_{(-\infty,0)}(x) + \mathbb{1}_{(0,\infty)}(x)$. The above expression can be written as

$$R(\theta, T) = 2(\operatorname{sgn} \theta) S(T|\theta|).$$

Recall the classical result that $S(T) \to \pi/2$ as $T \to \infty$ which can be proved using the residue theory. As $T \to \infty$, $R(\theta, T) \to (\operatorname{sgn} \theta)\pi$ and hence

$$R(x-a, T) - R(x-b, T) \to 2\pi \mathbb{1}_{\{(a,b)\}} + \pi \mathbb{1}_{\{\{a,b\}\}} \text{ as } T \to \infty.$$

It then follows from the bounded convergence theorem that

$$\lim_{T \to \infty} \frac{1}{2\pi} I_T = \mu(a, b) + \frac{1}{2} \mu(\{a, b\}).$$

□

Theorem 2.3.11. *If $\int |\varphi(t)| dt < \infty$, then μ has bounded continuous density*

$$f(y) = \frac{1}{2\pi} \int e^{-ity} \varphi(t) dt.$$

Proof. Recall the observation we made in the proof of inversion formula 2.3.10 that

$$\left| \frac{e^{-ita} - e^{-itb}}{it} \right| \leq b - a.$$

Thus the integrand in the inversion formula is absolutely integrable, and we obtain

$$\mu((a, b)) + \frac{1}{2} \mu(\{a, b\}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt \leq \frac{b-a}{2\pi} \int_{-\infty}^{\infty} |\varphi(t)| dt.$$

Taking $a_n \uparrow b$, we see that $\frac{1}{2}\mu\{b\} \leq \lim_{n\to\infty} \frac{1}{2}\mu(\{a_n, b\}) \leq 0$. This shows that μ has no point masses. Moreover, by the Fubini's theorem,

$$\begin{aligned}
\mu((x, x+h)) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-it(x+h)}}{it} \varphi(t)dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{x}^{x+h} e^{-ity} dy \right) \varphi(t)dt \\
&= \frac{1}{2\pi} \int_{x}^{x+h} \left(\int_{-\infty}^{\infty} e^{-ity} \varphi(t)dt \right) dy.
\end{aligned}$$

Hence μ has density function $f(y) = \frac{1}{2\pi} \int e^{-ity}\varphi(t)dt$. □

Theorem 2.3.12. *Suppose $(\mu_n)_{n=1}^{\infty}$ is a sequence of probability measures with ch.f. φ_n.*

(i) If $\mu_n \Rightarrow \mu_\infty$, then $\varphi_n(t) \to \varphi_\infty(t)$.

(ii) If $\varphi_n(t) \to \varphi(t)$ pointwise and $\varphi(t)$ is continuous at 0, then the associated sequence $(\mu_n)_{n=1}^{\infty}$ of distributions is tight and $\mu_n \Rightarrow \mu$ where μ has ch.f. φ.

Proof. (i) follows from an equivalent formulation of the weak convergence. To prove (ii), we first compute a seemingly irrelevant integral.

$$\begin{aligned}
\frac{1}{u} \int_{-u}^{u} 1 - \varphi_n(t)dt &= \frac{1}{u} \int_{-u}^{u} \int 1 - e^{itx} \mu_n(dx)dt \\
&= \frac{1}{u} \int \int_{-u}^{u} 1 - e^{itx} dt \mu_n(dx) \\
&= \frac{1}{u} \int 2u - \frac{2\sin ux}{x} \mu_n(dx) \\
&= 2 \int 1 - \frac{\sin ux}{ux} \mu_n(dx) \\
&\geq 2 \int_{|x|>2/u} 1 - \frac{\sin ux}{ux} \mu_n(dx) \\
&\geq 2 \int_{|x|>2/u} 1 - \frac{1}{ux} \mu_n(dx) \geq \mu_n(|x| > 2/u),
\end{aligned}$$

where the last inequality follows from $1 - 1/ux \geq 1/2$ whenever $|x| > 2/u$. Let $\varepsilon > 0$. Since $\varphi(t)$ continuous at 0, we can find u so small that $\frac{1}{u} \int_{-u}^{u} 1 - \varphi(t)dt < \varepsilon$. Then the dominated convergence theorem implies

$$\overline{\lim_{n\to\infty}}\, \mu_n(|x| > 2/u) \leq \overline{\lim_{n\to\infty}}\, \frac{1}{u} \int_{-u}^{u} 1 - \varphi_n(t)dt < \varepsilon.$$

Hence the sequence $(\mu_n)_{n=1}^{\infty}$ is tight. Suppose $(\mu_{n(k)})_{k=1}^{\infty}$ is a subsequence. Then by Helly's theorem there is a μ such that $\mu_{n(k)} \Rightarrow \mu$. Moreover, the tightness of $(\mu_n)_{n=1}^{\infty}$ shows that μ is a probability measure and part (i) implies μ has ch.f. φ. Thus we proved that every subsequence of $(\mu_n)_{n=1}^{\infty}$ has a further subsequence converging weakly to μ with ch.f. φ. Applying Lemma 1.6.3 to the sequence $(\int f d\mu_n)_{n=1}^{\infty}$ where f is an arbitrary bounded continuous function, we see that $\mu_n \Rightarrow \mu$. □

In the proof of the continuity theorem, we related the tail behavior of a distribution to its ch.f. via the following bound

$$\mu(|x| > 2/u) \leq \int_{-u}^{u} 1 - \varphi(t)dt.$$

Exercise 2.3.14 shows that finite n-th moment implies the n-time differentiability of $\varphi(t)$ and $\varphi^{(n)}(0) = \mathbb{E}(iX)^n$. Expanding $\varphi(t)$ about $t = 0$ yields

$$\varphi(t) = \sum_{k=0}^{n} \frac{t^k \varphi^{(k)}(0)}{k!} + o(t^n) = \sum_{k=0}^{n} \frac{\mathbb{E}(itX)^k}{k!} + o(t^n).$$

We can derive quantitative bound on the error term.

Lemma 2.3.13. $|e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!}| \le \min\{\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\}$.

Proof. First we claim that

$$e^{ix} - \sum_{k=0}^{k} \frac{(ix)^k}{k!} = \frac{i^{n+1}}{n!} \int_0^x e^{it}(x-t)^n dt. \tag{2.2}$$

The argument is similar to the proof of Taylor's formula with an integral remainder. When $n = 0$, the RHS is simply

$$i \int_0^x e^{it} dt = e^{it}\Big|_0^x = e^{ix} - 1 = \text{LHS}.$$

Assume the inductive hypothesis. Integration by parts implies

$$\begin{aligned}
\text{RHS} &= \frac{i^n}{n!} e^{it}(x-t)^n\Big|_0^x + \frac{i^n}{n!}\int_0^x n e^{it}(x-t)^{n-1} dt \\
&= -\frac{i^n}{n!} x^n + \frac{i^n}{(n-1)!}\int_0^x e^{it}(x-t)^{n-1} dt \\
&= -\frac{i^n}{n!} x^n + e^{ix} - \sum_{k=0}^{n-1} \frac{(ix)^k}{k!} \\
&= e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!} = \text{LHS}.
\end{aligned}$$

This completes the verification of (2.2). Now we attempt to bound the desired quantity. Clearly, via (2.2) we have

$$\left| e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!} \right| \le \frac{1}{n!} \left| \int_0^x (x-t)^n dt \right| = \frac{|x|^{n+1}}{(n+1)!}.$$

To obtain the other one, we note in the second line of the above verification that

$$\begin{aligned}
\left| \frac{i^{n+1}}{n!} \int_0^x e^{ix}(x-t)^n dt \right| &= \left| -\frac{i^n}{n!} x^n + \frac{i^n}{(n-1)!} \int_0^x e^{it}(x-t)^{n-1} dt \right| \\
&= \left| -\frac{i^n}{(n-1)!} \int_0^x (x-t)^{n-1} dt + \frac{i^n}{(n-1)!} \int_0^x e^{it}(x-t)^{n-1} dt \right| \\
&= \left| \frac{i^n}{(n-1)!} \int_0^x (x-t)^{n-1}(e^{it} - 1) dt \right| \\
&\le \left| \frac{2}{(n-1)!} \int_0^x (x-t)^{n-1} dt \right| = 2\frac{|x|^n}{n!},
\end{aligned}$$

where in the second step we used $x^n/n = \int_0^x (x-t)^{n-1} dt$. $\square$

Theorem 2.3.14. $|\mathbb{E}e^{itX} - \sum_{k=0}^{n} \mathbb{E}\frac{(itX)^k}{k!}| \le \mathbb{E}\min\{\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\}$.

Proof. Applying previous lemma to x=tX, we see that

$$\left| e^{tX} - \sum_{k=0}^{n} \frac{(itX)^k}{k!} \right| \le \min\left\{ \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right\}.$$

Taking expectation on both sides then applying Jensen's inequality yields

$$\left| \mathbb{E}e^{tX} - \sum_{k=0}^{n} \mathbb{E}\frac{(itX)^k}{k!} \right| \le \mathbb{E}\min\left\{ \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right\}.$$

□

Corollary 2.3.15. *If $\mathbb{E}|X|^2 < \infty$, then*

$$\varphi(t) = 1 + it\mathbb{E}X - t^2\mathbb{E}X^2/2 + o(t^2).$$

Theorem 2.3.16. *If*

$$\overline{\lim_{h\downarrow 0}} \frac{\varphi(h) + \varphi(-h) - 2\varphi(0)}{h^2} > -\infty,$$

then $\mathbb{E}|X|^2 < \infty$.

Proof. We observe that

$$\frac{e^{ihx} + e^{-ihx} - 2}{h^2} = \frac{2(\cos(hx) - 1)}{h^2} \le 0$$

and

$$\frac{2(1 - \cos(hx))}{h^2} \to x^2 \text{ as } h \to 0.$$

Thus it follows from the Fatou's lemma that

$$\mathbb{E}|X|^2 \le 2 \varliminf_{h\to 0} \int \frac{1 - \cos hx}{h^2} dF(x) = -2 \varlimsup_{h\to 0} \int \frac{e^{ihx} + e^{-ihx} - 2}{h^2} dF(x) < \infty.$$

□

2.4 Central Limit Theorems

Theorem 2.4.1. *Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}X_1 = \mu$, $\mathrm{Var}(X_1) = \sigma^2 < \infty$. Then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow Z \sim \mathcal{N}(0, 1).$$

Proof. WLOG, assume $\mu = 0$. We know from Corollary 2.3.15 that

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2),$$

so the ch.f. of $S_n/\sigma\sqrt{n}$ is

$$\varphi_1(t) = \mathbb{E}e^{itS_n/\sigma\sqrt{n}} = \left(1 - \frac{t^2}{2n} + o(1/n)\right)^n \to e^{-t^2/2}.$$

To fully justify the last convergence, we need to extend Lemma 2.1.2 to complex numbers. □

Theorem 2.4.2. *If $c_n \to c \in \mathbb{C}$, then $(1 + c_n/n)^n \to e^c$.*

Lemma 2.4.3. *Let $z_i, w_i, 1 \le i \le n$ be complex numbers with modulus $\le \theta$, then*

$$\left|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i\right| \le \theta^{n-1} \sum_{i=1}^n |z_i - w_i|.$$

Proof. The result is true for $n = 1$. Assume its validity for $n = k$. We directly compute

$$\begin{aligned}\left|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i\right| &= \left|z_1 \prod_{i=2}^n z_i - z_1 \prod_{i=2}^n w_i\right| + \left|z_1 \prod_{i=2}^n w_i - w_1 \prod_{i=2}^n w_i\right| \\ &\le \theta \left|\prod_{i=2}^n z_i - \prod_{i=2}^n w_i\right| + \theta^{n-1}|z_1 - w_1|.\end{aligned}$$

The induction hypothesis concludes the proof. □

Lemma 2.4.4. *If $b \in \mathbb{C}$ and $|b| \le 1$, then $|e^b - (1 + b)| \le |b|^2$.*

Proof. Use Taylor series for e^x to proceed. □

Proof of Theorem 2.4.2. For $1 \le i \le n$, let $z_i = (1 + c_n/n)$, $w_i = e^{c_n/n}$ and $\gamma > |c|$. For sufficiently large n, we have $|c_n/n| \le 1$, $|c_n| < \gamma$. Thus by the previous two results, we obtain

$$|(1 + c_n/n)^n - e^{c_n}| \le (e^{\gamma/n})^{n-1} \sum_{i=1}^n \frac{|c_n|^2}{n^2} \to 0.$$

□

Theorem 2.4.5 (Lindeberg-Feller)**.** *For each n, let $X_{n,m}, 1 \le m \le n$, be independent r.v.'s with $\mathbb{E}X_{n,m} = 0$. If*
(i)$\sum_{m=1}^n \mathbb{E}X_{n,m}^2 \to \sigma^2 > 0$;
(ii) for all $\varepsilon > 0$, $\sum_{m=1}^n \mathbb{E}X_{n,m}^2 \mathbb{1}_{\{|X_{n,m}| > \varepsilon\}} \to 0$,
then $S_n = \sum_{m=1}^n X_{n,m} \Rightarrow \sigma Z$.

Proof. Denote $\varphi_{n,m}(t) = \mathbb{E}e^{itX_{n,m}}$ and $\sigma_{n,m} = \mathbb{E}|X_{n,m}|^2$. Our task is to show

$$\prod_{m=1}^n \varphi_{n,m}(t) \to e^{-\frac{\sigma^2 t^2}{2}}.$$

In view of Remark 2.1.3 and assumption (i) & (ii), it suffices to show

$$\left|\prod_{m=1}^n \varphi_{m,n}(t) - \prod_{m=1}^n (1 - \frac{\sigma_{n,m}^2 t^2}{2})\right| \to 0. \tag{$*$}$$

To see that $\sup_m \sigma_{n,m}^2 \to 0$ as $n \to \infty$, note that by (ii),

$$\begin{aligned}\sup_m \sigma_{n,m}^2 &= \sup_m \{\mathbb{E}|X_{n,m}|^2 \mathbb{1}_{\{|X_{n,m}| \le \varepsilon\}} + \mathbb{E}|X_{n,m}|^2 \mathbb{1}_{\{|X_{n,m}| > \varepsilon\}}\} \\ &\le \sup_m \{\varepsilon + o(1)\} \to 0.\end{aligned}$$

We now apply the bound 2.4.3, using $|\varphi_{n,m}(t)| \le 1$ and $|1-\sigma_{n,m}^2 t^2/2| \le 1$ for sufficiently large n, to the LHS of $(*)$,

$$\begin{aligned}\left|\prod_{m=1}^n \varphi_{n,m}(t) - \prod_{m=1}^n (1-\frac{\sigma_{n,m}^2 t^2}{2})\right| &\le \sum_{m=1}^n \left|\varphi_{n,m}(t) - (1-\sigma_{n,m}^2 t^2/2)\right| \\ &= \sum_{m=1}^n \mathbb{E}\left[\frac{|tX_{n,m}|^3}{6} \wedge \frac{2|tX_{n,m}|^2}{2}\right] \\ &= \sum_{m=1}^n \mathbb{E}\frac{|tX_{n,m}|^3}{6}\mathbb{1}_{\{|X_{n,m}|\le\varepsilon\}} + \mathbb{E}|tX_{n,m}|^2\mathbb{1}_{\{|X_{n,m}|>\varepsilon\}} \\ &\le \sum_{m=1}^n \frac{\varepsilon t^3}{6}\mathbb{E}|X_{n,m}|^2\mathbb{1}_{\{|X_{n,m}|\le\varepsilon\}} + \mathbb{E}|tX_{n,m}|^2\mathbb{1}_{\{|X_{n,m}|>\varepsilon\}}.\end{aligned}$$

As the sum of the second term converges to 0, we see that

$$\overline{\lim_{n\to\infty}}\left|\prod_{m=1}^n \varphi_{n,m}(t) - \prod_{m=1}^n (1-\frac{\sigma_{n,m}^2 t^2}{2})\right| \le \frac{\varepsilon\sigma^2 t^3}{6}.$$

Since $\varepsilon > 0$ is arbitrary, the proof is complete. □

Next we give a few applications of the Lindeberg-Feller theorem.

Example 2.4.6 (Cycles in random permutations)**.** Let $Y_1, Y_2, \ldots$ be independent with $\mathbb{P}(Y_m = 1) = 1/m$ and $\mathbb{P}(Y_m = 0) = 1 - 1/m$. Then $\mathbb{E}Y_m = 1/m$ and $\text{Var}(Y_m) = 1/m - 1/m^2$. Set $S_n = \sum_{m=1}^n Y_m$, so $\mathbb{E}S_n \sim \log n$ and $\text{Var}(S_n) \sim \log n$. Also let

$$X_{n,m} = \frac{Y_m - 1/m}{\sqrt{\log n}}.$$

We claim that

$$\frac{S_n - \sum_{m=1}^n 1/m}{\sqrt{\log n}} = \sum_{m=1}^n X_{n,m} \Rightarrow Z$$

where $Z \sim \mathcal{N}(0,1)$. It suffices to check that the $(X_{n,m})_{n=1}^\infty$ satisfies the conditions in the Lindeberg-Feller Theorem. Clearly,

$$\sum_{m=1}^n \mathbb{E}|X_{n,m}|^2 = \sum_{m=1}^n \frac{\text{Var}(Y_m)}{\log n} = \frac{\text{Var}(S_n)}{\log n} \to 1 \text{ as } n\to\infty.$$

Let $\varepsilon > 0$. Note that as long as n is so large that $\sqrt{\log n} < \varepsilon$, $\frac{|Y_m - 1/m|}{\sqrt{\log n}} \le \varepsilon$ a.s. because $Y_m - 1/m < 1$ a.s.. Hence it follows that $\mathbb{E}|X_{n,m}|^2\mathbb{1}_{\{|X_{n,m}|>\varepsilon\}} = 0$ for n large, and

$$\sum_{m=1}^n \mathbb{E}|X_{n,m}|^2\mathbb{1}_{|X_{n,m}|>\varepsilon} \to 0 \text{ as } n\to\infty.$$

Therefore, the claim is true. Indeed, we can replace $\sum_{m=1}^n 1/m$ by $\log n$ to obtain a better-looking weak convergence

$$\frac{S_n - \log n}{\sqrt{\log n}} \Rightarrow Z,$$

because $|\log n - \sum_{m=1}^n 1/m|/\sqrt{\log n} \to 0$.

Example 2.4.7. (Second proof of the Three-series theorem). The proof of the first two necessities is the same. Let us prove the last one. Let $X_1, X_2, ...$ be i.i.d. and let $Y_m = X_m \mathbb{1}_{\{|X_m| \le A\}}$. Suppose to the contrary that $\lim_{n\to\infty} c_n := \lim_{n\to\infty} \sum_{m=1}^n \mathrm{Var}(Y_m) = \infty$. Consider

$$X_{n,m} = \frac{Y_m - \mathbb{E}Y_m}{\sqrt{c_n}}$$

We claim that $\sum_{m=1}^n X_{n,m} \Rightarrow Z$, where $Z \sim \mathcal{N}(0,1)$. Clearly $\sum_{m=1}^n \mathbb{E}|X_{n,m}|^2 = 1$ for all n. Also note that as long as $2A/\sqrt{c}_n < \varepsilon$, by a similar reasoning as in the previous example, $\mathbb{E}|X_{n,m}|^2 \mathbb{1}_{\{|X_{n,m}|>\varepsilon\}} = 0$ for large n. Thus the claim is true. Now note that the convergence of $\sum_{m=1}^\infty X_n$ implies the convergence of $\sum_{m=1}^\infty Y_n$. Moreover, $\sum_{m=1}^n Y_n/\sqrt{c}_n \to 0$ a.s. This implies that

$$\sum_{m=1}^n X_{n,m} - \sum_{m=1}^n \frac{Y_m}{\sqrt{c}_n} \Rightarrow Z.$$

But indeed the LHS

$$\sum_{m=1}^n X_{n,m} - \sum_{m=1}^n \frac{Y_m}{\sqrt{c}_n} = -\frac{\sum_{m=1}^n \mathbb{E}Y_m}{\sqrt{c}_n},$$

which is a deterministic object. So the weak convergence is weird.

Example 2.4.8 (Infinite variance)**.** Suppose that X is a symmetric r.v. supported on $(-\infty, 1] \cup [1, \infty)$ with $\mathbb{P}(|X| > x) = 1/x^2$ for $x \ge 1$. Let $X_1, X_2, ...$ be i.i.d. copies of X. Note that

$$\mathbb{E}|X_1|^2 = \int 2x\mathbb{P}(|X_1| > x)dx = \infty.$$

But still we can find some suitable normalizing constants to ensure the weak convergence of S_n/c_n to the standard normal. Let $a_n = \sqrt{n}\log\log n$ and let $Y_{n,m} = X_m \mathbb{1}_{\{|X_m| \le a_n\}}$. If we can prove the result for $S_n' = \sum_{m=1}^n Y_{n,m}$, then because

$$\mathbb{P}(S_n \ne S_n') \le \sum_{m=1}^n \mathbb{P}(Y_{n,m} \ne X_m) \le n\mathbb{P}(|X_m| > a_n) \to 0,$$

the result holds true for S_n. Set $c_n = \sqrt{n \log n}$. We shall apply the Lindeberg-Feller theorem to $Z_{n,m} = Y_{n,m}/c_n$. First we show $\mathbb{E}Y_{n,m}^2 \sim \log n$. Clearly note that $\mathbb{P}(|Y_{n,m}| > x) \le \mathbb{P}(|X_1| > x)$ and in particular $\mathbb{P}(|Y_{n,m}| > x)$ vanishes if $x > c_n$. Thus

$$\mathbb{E}Y_{n,m}^2 = \int_1^{c_n} 2x\mathbb{P}(|Y_{n,m}| > x)dx \le \int_1^{c_n} 2x\mathbb{P}(|X_1| > x)dx \sim \log n.$$

Also since $\mathbb{P}(|Y_{n,m}| > x) = \mathbb{P}(|X_1| > x) - \mathbb{P}(|X_1| > c_n) \ge (1 - (\log\log n)^{-2})x^{-2}$ when $x \le \sqrt{n}$, we have

$$\mathbb{E}Y_{n,m}^2 \ge \int_1^{\sqrt{n}} 2(1 - (\log\log n)^{-2})x^{-1}dx \sim \log n.$$

Hence it follows that $\sum_{m=1}^n \mathbb{E}|Z_{n,m}|^2 = n\mathbb{E}|Y_{n,m}|^2/(n\log n) \to 1$ as $n \to \infty$. Moreover, for any $\varepsilon > 0$, since $|Y_{n,m}| < \sqrt{n}\log\log n$, the summand of $\sum_{m=1}^n \mathbb{E}|Z_{n,m}|^2 \mathbb{1}_{\{|Z_{n,m}|>\varepsilon\}}$ is zero for n sufficiently large. Therefore, the Lindeberg-Feller theorem and our argument imply

$$\frac{X_1 + \cdots + X_n}{\sqrt{n \log n}} \Rightarrow Z.$$

Theorem 2.4.9.

2.5 Poisson Convergence

In this section, we shall investigate the approximating distribution of a sum of independent Bernoulli r.v.'s with parameter rapidly decreasing.

Theorem 2.5.1. *For each n, let $(X_{n,m})_{m=1}^n$ be independent r.v.'s with $\mathbb{P}(X_{n,m}=1)=p_{n,m}$, and $\mathbb{P}(X_{n,m}=0)=1-p_{n,m}$. If*

(i) $\sum_{m=1}^n p_{n,m} \to \lambda \in (0,\infty)$, *and*

(ii) $\max_{1\le m\le n} p_{n,m} \to 0$,

then $S_n=\sum_{m=1}^n X_{n,m} \Rightarrow N$ where $N\sim \mathrm{Poi}(\lambda)$.

Proof. Let $\varphi_{n,m}(t)=\mathbb{E}e^{itX_{n,m}}=(1-p_{n,n})+e^{it}p_{n,m}=1+p_{n,m}(e^{it}-1)$. Then $\varphi_n(t):=\mathbb{E}e^{itS_n}=\prod_{m=1}^n(1+p_{n,m}(e^{it}-1))$. Instead of proving the convergence of this product to the ch.f. of a Poisson distribution, we show the convergence for a close relative $\psi_n := \prod_{m=1}^n \psi_{n,m}(t) = \prod_{m=1}^n \psi_{n,m}(t)$ where $\psi_{n,m}(t)=e^{\varphi_{n,m}(t)-1}$. Note that the ch.f. of N is $\mu(t)=e^{\lambda(e^{it}-1)}$, so we shall compute

$$\begin{aligned}
|\log\psi_n(t)-\log\mu(t)| &= \left|\sum_{m=1}^n(\varphi_{n,m}(t)-1)-\lambda(e^{it}-1)\right| \\
&= \left|\sum_{m=1}^n p_{n,m}(e^{it}-1)-\lambda(e^{it}-1)\right| \\
&\le \left|\sum_{m=1}^n p_{n,m}-\lambda\right| \to 0.
\end{aligned}$$

It suffices to show $\varphi_n(t)$ and $\psi_n(t)$ are identical in the limit. To this end, we estimate

$$\begin{aligned}
|\log\varphi_n(t)-\log\psi_n(t)| &= \left|\sum_{m=1}^n \log(1+p_{n,m}(e^{it}-1))-p_{n,m}(e^{it}-1)\right| \\
&\le C\sum_{m=1}^n |p_{n,m}(e^{it}-1)|^2 \\
&\le C\max_{1\le m\le n}|p_{n,m}|\sum_{m=1}^n |p_{n,m}| \to 0
\end{aligned}$$

where we used $\log(1+x)=x+O(x^2)$. □

Example 2.5.2.

Example 2.5.3.

Example 2.5.4.

There is a different way to prove the Poisson convergence theorem. Define the **total variance distance** on the space of measures on a countable set S, via

$$||\mu-\nu|| := \sum_{s\in S}|\mu(s)-\nu(s)| = 2\sup_{A\subset S}|\mu(A)-\nu(A)|.$$

Note the a for any $A\subset S$ by triangle inequality, we always have

$$2|\mu(A)-\nu(A)| = |\mu(A)-\nu(A)|+|\mu(A^c)-\nu(A^c)| \le \sum_{s\in S}|\mu(s)-\nu(s)|.$$

The equality is actually attained on $A=\{s:\mu(s)\ge\nu(s)\}$. Given two measures μ_1,μ_2 on $\mathbb{Z}$, we define $\mu_1\times\mu_2(x,y):=\mu_1(x)\mu_2(y)$.

Lemma 2.5.5. $||\mu_1 \times \mu_2 - \nu_1 \times \nu_2|| \leq ||\mu_1 - \nu_1|| + ||\mu_2 - \nu_2||$.

Recall that the discrete convolution of two measures results in a new measure

$$\mu_1 * \mu_2(x) = \sum_y \mu_1(x-y)\mu_2(y).$$

Lemma 2.5.6. $||\mu_1 * \mu_2 - \nu_1 * \nu_2|| \leq ||\mu_1 \times \mu_2 - \nu_1 \times \nu_2||$.

Next we give an upper bound on the total variation distance between a Bernoulli(p) distribution and a Poi(λ) distribution.

Lemma 2.5.7. *Let μ be a measure with $\mu(1) = p$ and $\mu(0) = 1 - p$. Let ν be a Poisson distribution with parameter p. Then $||\mu - \nu|| \leq 2p^2$.*

Second proof of Theorem 2.5.1. Let $\mu_{n,m}$ denote the distribution of $X_{n,m}$ and $\mu_n = \mu_{n,1} * \cdots * \mu_{n,n}$, which have the distribution of S_n. Let $\nu_{n,m}$ be a Poi($p_{n,m}$) distribution and ν_n be a Poi($\sum_{m=1}^n p_{n,m}$) distribution. Then applying previous three lemmas all together, we have

$$||\mu_n - \nu_n|| \leq \sum_{m=1}^n ||\mu_{n,m} - \nu_{n,m}|| \leq 2\sum_{m=1}^n p_{n,m}^2.$$

Then

$$\sup_{A \subset S} |\mu_n(A) - \nu_n(A)| \leq \sum_{m=1}^n p_{n,m}^2 \to 0.$$

where the last convergence is justified by (i) and (ii). □

3 Martingales

3.1 Conditional Expectations

One of the main object throughout this section will be the conditional expectation w.r.t. a σ-algebra. Let us fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X with $\mathbb{E}|X| < \infty$.

Definition 3.1.1. Given a sub σ-algebra $\mathcal{G} \subset \mathcal{F}$, we define **conditional expectation** $\mathbb{E}(X|\mathcal{G})$ **of** X **given** $\mathcal{G}$ to be any random variable Y such that
(i) $Y \in \mathcal{G}$, i.e., Y is $\mathcal{G}$-measurable;
(ii) for all $A \in \mathcal{G}$, $\int_A Y d\mathbb{P} = \int_A X d\mathbb{P}$.

The existence of the conditional expectation $\mathbb{E}(X|\mathcal{G})$ can be shown in the following way. First assume $X \geq 0$. The set function $\mu : \mathcal{G} \to [0, +\infty)$ given by $\mu(A) = \int_A X d\mathbb{P}$ is readily seen to be a measure on $\mathcal{G}$ and moreover μ is absolutely continuous w.r.t. $\mathbb{P}$. So by the Radon-Nikodym theorem, there is a $\mathcal{G}$-measurable function f such that $\mu(A) = \int_A f d\mathbb{P}$ for all $A \in \mathcal{G}$. Thus the function f satisfies our requirement. For a general r.v. X, we apply Radon-Nikodym theorem to X^+, X^- respectively and to obtain the conditional expectation in this case.

Remark 3.1.2. It is possible to drop the assumption $\mathbb{E}|X| < \infty$.

Now we collect a number of basic properties of the conditional expectation.

Theorem 3.1.3. *(a) If $X \in \mathcal{G}$, then $\mathbb{E}(X|\mathcal{G}) = X$.*
(b) If X is independent of $\mathcal{G}$, then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}X$.
(c) $\mathbb{E}(aX + Y|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + \mathbb{E}(Y|\mathcal{G})$.
(d) If $X \leq Y$, then $\mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(Y|\mathcal{G})$.
(e) If $X_n \geq 0$ and $X_n \uparrow X$ with $\mathbb{E}X < \infty$, then $\mathbb{E}(X_n|\mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G})$.
(f) If φ is convex and $\mathbb{E}|X|, \mathbb{E}|\varphi(X)| < \infty$, then $\varphi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(\varphi(X)|\mathcal{G})$.
(g) If $p \geq 1$, then $\mathbb{E}|\mathbb{E}(X|\mathcal{G})|^p \leq \mathbb{E}|X|^p$.
(h) $\mathbb{E}(\mathbb{E}(Y|\mathcal{G})) = \mathbb{E}Y$.
(i) If $\mathcal{G}_1 \subset \mathcal{G}_2$, then $\mathbb{E}(\mathbb{E}(X|\mathcal{G}_1)|\mathcal{G}_2) = \mathbb{E}(X|\mathcal{G}_1)$ and $\mathbb{E}(\mathbb{E}(X|\mathcal{G}_2)|\mathcal{G}_1) = \mathbb{E}(X|\mathcal{G}_1)$.
(j) If $X \in \mathcal{F}$, and $\mathbb{E}|Y| < \infty, \mathbb{E}|XY| < \infty$, then $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$.

3.2 Martingales

Definition 3.2.1. A **filtration** is an increasing family of σ-algebras. A sequence of r.v.'s $(X_n)_{n=1}^{\infty}$ is said to be **adapted** to a filtration $(\mathcal{F}_n)_{n=1}^{\infty}$ if $X_n \in \mathcal{F}_n$ for each n.

Definition 3.2.2. Let $(\mathcal{F}_n)_{n=1}^{\infty}$ be a filtration. Suppose $(X_n)_{n=1}^{\infty}$ is adapted to $(\mathcal{F}_n)_{n=1}^{\infty}$ and $\mathbb{E}|X_n| < \infty$ for each n. We say $(X_n)_{n=1}^{\infty}$ is a **submartingale/martingale/supermartingale** if $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq / = / \leq X_n$ repsectively.

Immediately we have

Theorem 3.2.3. *If X_n is a supermartingale then $\mathbb{E}(X_n|\mathcal{F}_m) \leq X_m$ whenever $n \geq m$.*

Proof. The trick is to insert conditional expectation w.r.t. a richer σ-algebra one at a time. We obtain

$$\begin{aligned}\mathbb{E}(X_n|\mathcal{F}_m) &= \mathbb{E}(\mathbb{E}(\cdots \mathbb{E}(X_n|\mathcal{F}_{n-1})|\cdots)|\mathcal{F}_m) \\ &\leq \mathbb{E}(\mathbb{E}(\cdots X_{n-1}|\cdots)|\mathcal{F}_m) \\ &\leq \cdots \leq \mathbb{E}(X_{m+1}|\mathcal{F}_m) \leq X_m.\end{aligned}$$

□

Corollary 3.2.4. *If X_n is a submartingale/martingale, $\mathbb{E}(X_n|\mathcal{F}_m) \geq / = X_m$ whenever $n \geq m$.*

Proof. Clearly $-X_n$ is a supermartingale, so the previous result immediately imply the conclusion. □

From now on, we only prove theorems for one of supermartingae/martingale/submartingale only. The statements for other cases can be easily translated.

Theorem 3.2.5. *If X_n is a martingale w.r.t. $\mathcal{F}_n$ and φ is a convex function with $\mathbb{E}|\varphi(X_n)| < \infty$ for all n, then $\varphi(X_n)$ is a submartingale.*

Proof. Applying conditional Jensen's inequality yields

$$\mathbb{E}(\varphi(X_{n+1})|\mathcal{F}_n) \geq \varphi(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) = \varphi(X_n).$$

□

Corollary 3.2.6. *Let $p \geq 1$. If X_n is a martingale w.r.t. $\mathcal{F}_n$ and $\mathbb{E}|X_n|^p < \infty$ for all n, then $|X_n|^p$ is a submartingale.*

Theorem 3.2.7. *If X_n is a submartingale and φ is an increasing convex function with $\mathbb{E}|\varphi(X_n)| < \infty$ for all n, then $(\varphi(X_n))_{n=1}^{\infty}$ is a submartingale. Similarly if X_n supermartingale and $\varphi(x)$ is a increasing concave function then $(\varphi(X_n))_{n=1}^{\infty}$ is a supermartingale.*

Proof. The last identity in the previous equation can be changed to $\geq$ in our case. □

Corollary 3.2.8. *If X_n is a submartingale then $(X_n - a)^+$ is a submartingale. If X_n is a supermartingale then $X_n \wedge a$ is a supermartingale.*

Definition 3.2.9. We say (H_n) is **predictable** if $H_n \in \mathcal{F}_{n-1}$ for all $n \geq 1$. The **discrete martingale transform** $(H \cdot X)$ is the sequence

$$(H \cdot X)_n = \sum_{k=1}^{n} H_k(X_k - X_{k-1}).$$

Theorem 3.2.10. *Let $X_n, n \geq 0$ be a supermartingale. If $H_n \geq 0$ is predictable and bounded, then $(H \cdot X)_n$ is a supermartingale.*

Proof. This follows from a direct calculation

$$\mathbb{E}((H \cdot X)_{n+1} - (H \cdot X)_n|\mathcal{F}_n) = \mathbb{E}(H_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n) = H_{n+1}\mathbb{E}(X_{n+1} - X_n|\mathcal{F}_n) \leq 0.$$

□

Remark 3.2.11. The key point is that the martingale transform preserves the martingale property of a sequence.

Definition 3.2.12. We say a r.v. N is a **stopping time** if $\{N = n\} \in \mathcal{F}_n$ for each n. Given a stopping time N, we define the **stopped process** $X_n^N := X_{N\wedge n} = X_n \mathbb{1}_{\{n<N\}} + X_N \mathbb{1}_{\{n \geq N\}}$.

Theorem 3.2.13. *If N is a stopping time and X_n is a supermartingale, then $X_{N\wedge n}$ is a supermartingale.*

Proof. Define a predictable sequence $H_n = \mathbb{1}_{\{N \geq n\}}$ whose predictability can be seen by $\{N \geq n\} = \{N \leq n-1\}^c \in \mathcal{F}_{n-1}$. WLOG, assume $N \geq 1$ almost surely. Set $X_{-1} = 0$. Then note that

$$
\begin{aligned}
(H \cdot X)_n &= \sum_{k=1}^{n} \mathbb{1}_{\{N \geq k\}}(X_k - X_{k-1}) \\
&= \sum_{k=1}^{n-1} X_k(\mathbb{1}_{\{N \geq k\}} - \mathbb{1}_{\{N \geq k+1\}}) + X_n \mathbb{1}_{\{N \geq n\}} - X_0 \mathbb{1}_{\{N \geq 1\}} \\
&= \sum_{k=1}^{n-1} X_k \mathbb{1}_{\{N = k\}} + X_n \mathbb{1}_{\{N \geq n\}} - X_0 \mathbb{1}_{\{N \geq 1\}} \\
&= \sum_{k=1}^{n-1} X_k \mathbb{1}_{\{N = k\}} + X_n \mathbb{1}_{\{N \geq n\}} - X_0 \mathbb{1}_{\{N \geq 1\}} \\
&= \sum_{k=0}^{n-1} X_k \mathbb{1}_{\{N = k\}} + X_n \mathbb{1}_{\{N \geq n\}} - X_0 \\
&= X_{N \wedge n} - X_0.
\end{aligned}
$$

The LHS is a supermartingale by Theorem 4.2.10, so the result follows as X_0 is also a supermartingale. □

Remark 3.2.14. Let H be as in the previous proof. So far $(H \cdot X)_n$ is only defined for $n \geq 1$. We claim that if we set $(H \cdot X)_0 = 0$, the sequence $(H \cdot X)_{n=0}^{\infty}$ is still a supermartingale, provided that $(X_n)_{n=0}^{\infty}$ is a supermartingale. It is enough to compute

$$\mathbb{E}((H \cdot X)_1 | \mathcal{F}_0) = \mathbb{E}(X_{N \wedge 1} - X_0 | \mathcal{F}_0) = \mathbb{E}((X_0 - X_0)\mathbb{1}_{\{N=0\}} + (X_1 - X_0)\mathbb{1}_{\{N>0\}} | \mathcal{F}_0) \leq 0.$$

Now suppose $X_n, n \geq 0$ is a submartingale. Let $a < b$, $N_0 = -1$ and for $k \geq 1$, define

$$N_{2k-1} = \inf\{m > N_{2k} : X_m \leq a\},$$

and

$$N_{2k} = \inf\{m > N_{2k-1} : X_m \geq b\}.$$

Note $\{N_1 = k\} = \{X_1 > a, ..., X_{k-1} > a, X_k \leq a\} \in \mathcal{F}_k$, so N_1 is a stopping time. By induction, it is easy to show that N_i are stopping times. Thus $\{N_{2k-1} < m \leq N_{2k}\} = \{N_{2k-1} \leq m - 1\} \cap \{N_{2k} \leq m - 1\}^c \in \mathcal{F}_{m-1}$ and it follows

$$H_n = \begin{cases} 1 : N_{2k-1} < n \leq N_{2k} \text{ for some } k \\ 0 : \text{otherwise} \end{cases}$$

defines a predictable sequence. In particular, $X_{N_{2k-1}} \leq a$, $X_{N_{2k}} \geq b$ and X_m evolves from below a to above b between time $N_{2k} < m \leq N_{2k-1}$. Let $U_n = \sup\{k : N_{2k} \leq n\}$.

Lemma 3.2.15 (The upcrossing inequality)**.** *If $X_m, m \geq 0$ is a submartingale, then*

$$(b - a)\mathbb{E}U_n \leq \mathbb{E}(X_n - a)^+ - \mathbb{E}(X_0 - a)^+.$$

Proof. Let $Y_m = a + (X_m - a)^+$. Clearly Y_m upcrosses $[a, b]$ the same number of times that X_m does. Also Y_m is a submartingale. Since the final possible incomplete upcrossing only increases the total profit, we see that $(b - a)U_n \leq (H \cdot Y)_n$. Now let $K_m = 1 - H_m$ and note that

$$(H \cdot Y)_n + (K \cdot Y)_n = \sum_{k=1}^{n}(Y_k - Y_{k-1}) = Y_n - Y_0.$$

Realizing that $(K \cdot Y)$ is also a submartingale and taking expectation on both sides get

$$\mathbb{E}(X_n - a)^+ - \mathbb{E}(X_0 - a)^+ \geq (b - a)\mathbb{E}U_n + \mathbb{E}(K \cdot Y)_0 \geq (b - a)\mathbb{E}U_n.$$

□

Theorem 3.2.16. *[Martingale convergence theorem]If X_n is a submartingale with $\sup \mathbb{E}X_n^+ < \infty$ then X_n converges a.s. to a limit X with $\mathbb{E}|X| < \infty$.*

Proof. First note that we may bound $(X-a)^+ \le X^+|a|$, so the upcrossing inequality implies

$$\mathbb{E}U_n \le \frac{\mathbb{E}X_n^+ + |a|}{b-a}.$$

As $n \to \infty$, $U_n \uparrow U$, the total number of upcrossings of $(X_n)_{n=1}^\infty$. It follows that $\mathbb{E}U \le \sup(\mathbb{E}X_n^+ + |a|)/(b-a) < \infty$ and $U < \infty$ a.s. As this holds for all $a, b \in \mathbb{Q}$, the event

$$\cup_{a,b\in\mathbb{Q}}\{ \varliminf_{n\to\infty} X_n < a < b < \varlimsup_{n\to\infty} X_n\}$$

has probability 0. Hence $\lim_{n\to\infty} X_n$ exists a.s. Now by Fatou's lemma, $\mathbb{E}X^+ \le \varliminf_{n\to\infty} \mathbb{E}X_n^+ < \infty$ and $X < \infty$ a.s. To see $X > -\infty$, we observe, using submartingale property of X_n, that

$$\mathbb{E}X_n^- = \mathbb{E}X_n^+ - \mathbb{E}X_n \le \mathbb{E}X_n^+ - \mathbb{E}X_0.$$

Another application of Fatou's lemma yields the result. □

Corollary 3.2.17. *If $X_n \ge 0$ is a supermartingale then as $n \to \infty$, $X_n \to X$ a.s. and $\mathbb{E}X_n \le X_0$.*

Proof. Clearly $-X_n$ is a submartingale and $\mathbb{E}(-X_n)^+ = 0$. Thus the existence of X follows from the martingale convergence theorem. Applying Fatou's lemma we get $\mathbb{E}X \le \varliminf_{n\to\infty} \mathbb{E}X_n \le \mathbb{E}X_0$. □

Theorem 3.2.18 (Doob's decomposition)**.** *Any submartingale $X_n, n \ge 0$ can be written uniquely as $X_n = A_n + M_n$ where M_n is a martingale and A_n is a predictable increasing sequence with $A_0 = 0$.*

3.3 Examples

3.4 Doob's Inequality, Convergence in L^p

Theorem 3.4.1. *If X_n is a martingale and N is a stopping time with $\mathbb{P}(N \le k) = 1$, then*

$$\mathbb{E}X_0 \le EX_N \le \mathbb{E}X_k.$$

Proof. Recall by Theorem 4.2.13, $(X_{N\wedge n})_{n=1}^\infty$ is a submartingale. This implies

$$\mathbb{E}X_0 \le \mathbb{E}X_{N\wedge k} = \mathbb{E}X_N.$$

On the other hand, we define similarly a predictable sequence $K_n = \mathbb{1}_{\{N<n\}}$. Then $(K\cdot X)_n = X_n - X_{N\wedge n}, n \ge 1$ is a submartingale. In particular, following the remark below Theorem 4.2.13, we may set $(K\cdot X)_0 = 0$ and $(K\cdot X)_{n=0}^\infty$ is still a submartingale. Thus

$$\mathbb{E}X_k - \mathbb{E}X_N = \mathbb{E}(K\cdot X)_k \ge \mathbb{E}(K\cdot X)_0 = 0.$$

This proves the second inequality. □

Theorem 3.4.2 (Doob's maximal inequality)**.** *Let $(X_n)_{n=1}^\infty$ be a submartingale and $\lambda > 0$. Then*

$$\mathbb{P}\left(\max_{0\le m\le n} X_m \ge \lambda\right) \le \frac{1}{\lambda}\mathbb{E}X_n \mathbb{1}_{\{\max_{0\le m\le n} X_m \ge \lambda\}} \le \frac{1}{\lambda}\mathbb{E}X_n^+.$$

Proof. Let $N = \inf\{m : X_m \geq \lambda\}$ and $A = \{\omega : \max_{0\leq m\leq n} X_m(\omega) \geq \lambda\}$. Clearly $\lambda\mathbb{P}(A) \leq \mathbb{E}X_{N\wedge n}\mathbb{1}_A$. Note that $X_{N\wedge n}\mathbb{1}_{A^c} = X_n\mathbb{1}_{A^c}$ and $\mathbb{E}X_{N\wedge n} \leq \mathbb{E}X_n$ by Theorem 4.4.1, so we obtain

$$\lambda\mathbb{P}(A) \leq \mathbb{E}X_{N\wedge n}\mathbb{1}_A \leq \mathbb{E}X_n\mathbb{1}_A.$$

The proves the first inequality. The other one is trivial. □

Theorem 3.4.3 (L^p maximal inequality)**.** *Let $M_n = \max_{1\leq m\leq n} X_n$ and $1 < p < \infty$. If X_n is a submartingale, then*

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

Consequently if Y_n is a martingale and $\tilde{M}_n = \max_{0\leq m\leq n}|Y_m|$, then

$$\mathbb{E}\tilde{M}_n^p \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}|Y_n|^p.$$

Proof. The second part follows directly from an application of the first part to the submartingale $(|Y_n|)_{n=0}^\infty$. To avoid integrability problems, we fix a positive integer $N > 0$. Note that we can apply Doob's maximal inequality on $\{M_n \wedge N \geq \lambda\}$ because either $\lambda \leq N$, in which case $\{M_n \wedge N \geq \lambda\} = \{M_n \geq \lambda\}$, or $\lambda > N$ in which case $\{M_n \wedge N \geq \lambda\} = \varnothing$. Thus we obtain

$$\begin{aligned}
\mathbb{E}(M_n \wedge N)^p &= \int_0^\infty p\lambda^{p-1}\mathbb{P}(M_n \wedge N \geq \lambda)d\lambda \\
&\leq \int_0^\infty p\lambda^{p-1}\int \lambda^{-1}X_n^+\mathbb{1}_{\{M_n\wedge N\geq\lambda\}}d\mathbb{P}d\lambda \\
&= \int X_n^+\int_0^{M_n\wedge N} p\lambda^{p-2}d\lambda d\mathbb{P} \\
&= \frac{p}{p-1}\int X_n^+(M_n\wedge N)^{p-1}d\mathbb{P} \\
&\leq \frac{p}{p-1}(\mathbb{E}(X_n^+)^p)^{1/p}(\mathbb{E}(M_n\wedge N)^p)^{1-1/p}.
\end{aligned}$$

Thus dividing on both sides by $(\mathbb{E}(M_n\wedge N)^p)^{1-1/p}$ and raising to the pth power yields

$$\mathbb{E}(M_n\wedge N)^p \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(X_n^+)^p.$$

Finally sending $N \to \infty$ and monotone convergence theorem imply the desired result. □

Theorem 3.4.4 (L^p convergence)**.** *If X_n is a martingale with $\sup_n \mathbb{E}|X_n|^p < \infty$ where $p > 1$, then $X_n \to X$ a.s. and in L^p.*

Proof. Note $\sup_n(\mathbb{E}X_n^+)^p \leq \sup_n \mathbb{E}|X_n|^p < \infty$, so by martingale convergence theorem $X_n \to X$ a.s. Using L^p maximal inequality we see

$$\varlimsup_{n\to\infty} \mathbb{E}\left(\max_{0\leq m\leq n}|X_m|\right)^p \leq \sup_n\left(\frac{p}{p-1}\right)^p \mathbb{E}|X_n|^p < \infty.$$

It follows that $\sup_n|X_n| \in L^p$ and since $|X_n - X| \leq (2\sup_n|X_n|)^p$, the dominated convergence theorem implies $X_n \to X$ in L^p as well. □

We shall now discuss the special case for $p = 2$.

Theorem 3.4.5. *Let X_n be a martingale with $\mathbb{E}X_n^2 < \infty$ for all n. If $m \leq n$ and $Y \in \mathcal{F}_m$ with $\mathbb{E}Y^2 < \infty$, then*

$$\mathbb{E}((X_n - X_m)Y|\mathcal{F}_m) = 0.$$

Proof. By Cauchy-Schwartz inequality, $\mathbb{E}|(X_n - X_m)Y| < \infty$. Thus we can apply properties of conditional expectation to obtain

$$\mathbb{E}((X_n - X_m)Y|\mathcal{F}_m) = Y\mathbb{E}(X_n - X_m|\mathcal{F}_m) = 0.$$

□

Theorem 3.4.6. *If X_n is a martingale with $\mathbb{E}X_n^2 < \infty$, then*

$$\mathbb{E}((X_n - X_m)^2|\mathcal{F}_m) = \mathbb{E}(X_m^2|\mathcal{F}_m) - X_m^2.$$

Proof. We simply expand the quantity,

$$\mathbb{E}((X_n - X_m)^2|\mathcal{F}_m) = \mathbb{E}(X_n^2|\mathcal{F}_m) - 2X_m\mathbb{E}(X_n|\mathcal{F}_m) + X_m^2 = \mathbb{E}(X_n^2|\mathcal{F}_m) - X_m^2.$$

□

3.5 Uniform Integrability, Convergence in L^1

Definition 3.5.1. A collection of random variables $\{X_i : i \in I\}$ is said to be **uniformly integrable** if $\lim_{N\to\infty} \sup_{i\in I} \mathbb{E}|X_i|\mathbb{1}_{\{|X_i|>N\}} = 0$. It is called L^1**-bounded** if $\sup_i \mathbb{E}|X_i| < \infty$ and is called **uniformly absolutely continuous** if for every $\varepsilon > 0$ there is a $\delta > 0$ such that whenever $\mathbb{P}(A) < \delta$, $\mathbb{E}|X_i|\mathbb{1}_A < \varepsilon$ for all $i \in I$.

Lemma 3.5.2. *$\{X_i : i \in I\}$ is uniformly integrable if and only if it is L^1 bounded and uniformly absolutely continuous.*

Proof. Suppose X_i is uniformly integrable. We may pick N so large that

$$\sup_{i\in I} \mathbb{E}|X_i|\mathbb{1}_{\{|X_i|>N\}} \leq 1.$$

It follows that $\sup_{i\in I} \mathbb{E}|X_i| \leq N + 1 < \infty$, so it is L^1 bounded. Next let $\varepsilon > 0$. We can find N so large that $\sup_{i\in I} \mathbb{E}|X_i|\mathbb{1}_{\{|X_i|>\varepsilon\}} < \varepsilon/2$. Pick $\delta > 0$ so that $\delta < \varepsilon/2N$ and let $A \in \mathcal{F}$ with $\mathbb{P}(A) < \delta$. Then

$$\mathbb{E}|X_i|\mathbb{1}_A \leq \mathbb{E}|X_i|\mathbb{1}_{\{|X_i|>N\}} + \mathbb{E}|X_i|\mathbb{1}_{A\cap\{|X_i|\leq N\}} < \varepsilon/2 + N\mathbb{P}(A) < \varepsilon$$

for all $i \in I$.

We now prove the converse. Let $\varepsilon > 0$ and let δ be in the definition of uniformly absolute continuity. It suffices to show that for $\varepsilon > 0$, there is a positive integer N such that $\sup_{i\in I} \mathbb{P}(|X_i| > N) < \delta$. We simply observe

$$\sup_{i\in I} \mathbb{P}(|X_i| > N) \leq \sup_{i\in I} \frac{\mathbb{E}|X_i|}{N} \to 0 \text{ as } N \to \infty,$$

where the convergence follows from $\sup_{i\in I} \mathbb{E}|X_i| < \infty$. □

Theorem 3.5.3. *If $X \in L^1$, then $\{\mathbb{E}(X|\mathcal{G}) : \mathcal{G} \subset \mathcal{F}$ is a σ-algebra$\}$ is uniformly integrable.*

Proof. Clearly $\sup_{\mathcal{G}} \mathbb{E}|\mathbb{E}(X|\mathcal{G})| = \mathbb{E}|X| < \infty$, so the collection is L^1 bounded. Let $\varepsilon > 0$. We observe the dominated convergence theorem implies that $\{X\}$ is uniformly integrable and so there is a $\delta > 0$ such that $\mathbb{E}|X|\mathbb{1}_A < \varepsilon$ whenever $\mathbb{P}(A) < \delta$. Now observe

$$\mathbb{E}|\mathbb{E}(X|\mathcal{G})|\mathbb{1}_A \leq \mathbb{E}\mathbb{E}(|X||\mathcal{G})\mathbb{1}_A = \mathbb{E}|X|\mathbb{1}_A < \varepsilon.$$

Thus the collection is uniformly absolutely continuous. □

Theorem 3.5.4. *Suppose $X_n \to X$ in probability. Then the following are equivalent:*
(i) $\{X_n : n \geq 0\}$ is uniformly integrable;
(ii) $X_n \to X$ in L^1;
(iii) $\mathbb{E}|X_n| \to \mathbb{E}|X| < \infty$.

Proof. $(i) \Rightarrow (ii)$ Define a truncation function

$$\varphi_N(x) := x\mathbb{1}_{[-N,N]}(x) + N\mathbb{1}_{(N,+\infty)}(x) - N\mathbb{1}_{(-\infty,-N)}(x).$$

Then we may estimate

$$|X_n - X| \leq |X_n - \varphi_N(X_n)| + |\varphi_N(X_n) - \varphi_N(X)| + |\varphi_N(X) - X|.$$

Note in general $|Y - \varphi_N(Y)| = (Y - N)^+\mathbb{1}_{\{|Y|>N\}} \leq |Y|\mathbb{1}_{\{|Y|>N\}}$. We denote $I_1 = \mathbb{E}|X_n - \varphi_N(X_n)|, I_2 = \mathbb{E}|\varphi_N(X_n) - \varphi_N(X)|$ and $I_3 = \mathbb{E}|\varphi_N(X) - X|$. By (i), we can find N that $I_1 \leq \mathbb{E}|X_n|\mathbb{1}_{\{|X_n|>N\}} < \varepsilon/3$ for all n. The bounded convergence theorem implies $I_2 < \varepsilon/3$ for sufficiently large n. To bound I_3, we observe by Fatou's lemma and L^1-boundedness of $\{X_n : n \geq 0\}$, $\mathbb{E}|X| < \infty$. Making N large enough we can ensure $I_3 < \varepsilon/3$. Thus we have

$$\mathbb{E}|X_n - X| \leq I_1 + I_2 + I_3 < \varepsilon \text{ for } n \text{ sufficiently large.}$$

$(ii) \Rightarrow (iii)$ Clearly we have

$$|\mathbb{E}|X_n| - \mathbb{E}|X|| \leq \mathbb{E}||X_n| - X| \leq \mathbb{E}|X_n - X| \to 0.$$

$(iii) \Rightarrow (i)$ Define $\psi_N(x) = x\mathbb{1}_{[0,N-1)}(x) + (\text{affine})\mathbb{1}_{[N-1,N]}(x)$. Observe

$$\mathbb{E}|X_n|\mathbb{1}_{\{|X_n|>N\}} \leq \mathbb{E}|X_n| - \mathbb{E}\psi_N(|X_n|)$$

Since $\mathbb{E}|X_n| \to \mathbb{E}|X|$ and $\mathbb{E}\psi_N(|X_n|) \to \mathbb{E}\psi_N(|X|)$ we can find n_0 such that $\mathbb{E}|X_n| \leq \mathbb{E}|X| + \varepsilon/2$ and $\mathbb{E}\psi_N(|X_n|) \geq \mathbb{E}\psi_N(|X|) - \varepsilon/2$ for $n > n_0$. Also the dominated convergence theorem implies we can find N large enough to make $\mathbb{E}|X| - \mathbb{E}\psi_N(|X|) < \varepsilon$. This yields

$$\mathbb{E}|X_n|\mathbb{1}_{\{|X_n|>N\}} \leq \mathbb{E}|X| - \mathbb{E}\psi_N(|X|) + \varepsilon < 2\varepsilon \text{ for } n > n_0$$

By adjusting M if necessary to deal with $\mathbb{E}|X_n|\mathbb{1}_{\{|X_n|>N\}}$ for $0 \leq n \leq n_0$, we can conclude that $\{X_n : n \geq 0\}$ is uniformly integrable. ☐

The results on uniformly integrability makes our work on proving main theorems easy. We now work towards this direction.

Theorem 3.5.5. *For a submartingale $(X_n)_{n=1}^{\infty}$, the following are equivalent:*
(i) it is uniformly integrable;
(ii) it converges a.s. and in L^1;
(iii) it converges in L^1.

Proof. $(i) \Rightarrow (ii)$ The uniform integrability of $\{X_n : n \in \mathbb{N}\}$ implies $\sup_n \mathbb{E}|X_n| < \infty$. Hence by martingale convergence theorem, $X_n \to X$ a.s. Thus by previous theorem $X_n \to X$ in L^1 as well.

$(ii) \Rightarrow (iii)$ Trivial.

$(iii) \Rightarrow (i)$ L^1 convergence implies convergence in probability thus the implication holds true by previous theorem. ☐

Before state the results on martingales, we prove two lemmas.

Lemma 3.5.6. *If $X_n \to X$ in L^1, then $X_n\mathbb{1}_A \to X\mathbb{1}_A$ for any $A \in \mathcal{F}$.*

Proof. This follows directly the computation $\mathbb{E}|X_n\mathbb{1}_A - X\mathbb{1}_A| \leq \mathbb{E}|X_n - X| \to 0$. ☐

Lemma 3.5.7. *If X_n is a martingale and $X_n \to X$ in L^1 then $X_n = \mathbb{E}(X|\mathcal{F}_n)$.*

Proof. Fix n. The martingale property implies $\mathbb{E}(X_m|\mathcal{F}_n) = X_n$ for any $m > n$. Thus

$$\int_A X_m d\mathbb{P} = \int_A X_n d\mathbb{P}, \forall A \in \mathcal{F}_n.$$

Taking limits on LHS, we see

$$\int_A X d\mathbb{P} = \int_A X_n d\mathbb{P}, \forall A \in \mathcal{F}_n.$$

This is indeed saying $\mathbb{E}(X|\mathcal{F}_n) = X_n$. □

Theorem 3.5.8. *For a martingale (X_n), the following are equivalent:*
(i) it is uniformly integrable;
(ii) it converges a.s. and in L^1;
(iii) it converges in L^1;
(iv) there is an integrable r.v. X such that $X_n = \mathbb{E}(X|\mathcal{F}_n)$.

Proof. Because a martingale is also a submartingae, we only need to show the last two implication.
$(iii) \Rightarrow (iv)$ This is proved by the previous lemma.
$(iv) \Rightarrow (i)$ This is proved by Theorem 4.5.3. □

From now on, let $\mathcal{F}_\infty = \sigma(\cup_{n=1}^\infty \mathcal{F}_n)$.

Theorem 3.5.9. *Suppose $\mathcal{F}_n \uparrow \mathcal{F}_\infty$. Then $\mathbb{E}(X|\mathcal{F}_n) \to \mathbb{E}(X|\mathcal{F}_\infty)$ a.s. and in L^1.*

Proof. First note that $(\mathbb{E}(X|\mathcal{F}_n))_{n=1}^\infty$ is a uniformly integrable martingale. Let X_∞ be its a.s. (and L^1) limit. Lemma 4.5.7 implies $\mathbb{E}(X|\mathcal{F}_n) = \mathbb{E}(X_\infty|\mathcal{F}_n)$ for all n, i.e.

$$\int_A X d\mathbb{P} = \int_A X_\infty d\mathbb{P}, \forall A \in \mathcal{F}_n \tag{$*$}$$

Now observe $\mathcal{C} = \{A \in \mathcal{F} : (*) \text{ holds}\}$ is a λ-system containing the π-system $\cup_{n=1}^\infty \mathcal{F}_n$. Thus by π-λ theorem, $\mathcal{F}_\infty = \sigma(\cup_{n=1}^\infty \mathcal{F}_n) \subset \mathcal{C}$. This shows $X_\infty = \mathbb{E}(X|\mathcal{F}_\infty)$. □

Immediately, we have

Theorem 3.5.10 (Levy's 0-1 law)**.** *If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and $A \in \mathcal{F}_\infty$, then $\mathbb{E}(\mathbb{1}_A|\mathcal{F}_n) \to \mathbb{1}_A$.*

Theorem 3.5.11. *Suppose $Y_n \to Y$ a.s. and $|Y_n| \le Z$ for all n where $\mathbb{E}Z < \infty$. If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, then*

$$\mathbb{E}(Y_n|\mathcal{F}_n) \to \mathbb{E}(Y|\mathcal{F}_\infty) a.s.$$

Proof. Let $W_N = \sup\{|Y_n - Y_m| : n, m \ge N\}$. We start from an immediate estimate

$$|\mathbb{E}(Y_n|\mathcal{F}_n) - \mathbb{E}(Y|\mathcal{F}_\infty)| \le |\mathbb{E}(Y_n|\mathcal{F}_n) - \mathbb{E}(Y|\mathcal{F}_n)| + |\mathbb{E}(Y|\mathcal{F}_n) - \mathbb{E}(Y|\mathcal{F}_\infty)|.$$

The second term converges to 0 by Theorem 4.5.9. To bound the first term, observe

$$\begin{aligned} \varlimsup_{n\to\infty} |\mathbb{E}(Y_n|\mathcal{F}_n) - \mathbb{E}(Y|\mathcal{F}_n)| &\le \varlimsup_{n\to\infty} \mathbb{E}(|Y_n - Y||\mathcal{F}_n) \\ &\le \varlimsup_{n\to\infty} \mathbb{E}(W_N|\mathcal{F}_n) = \mathbb{E}(W_N|\mathcal{F}_\infty) \end{aligned}$$

for all N. Since $W_N \downarrow 0$ as $N \to \infty$, sending N to ∞ concludes the proof. □

3.6 Backward Martingales

Definition 3.6.1. A **backwards martingale** is a martingale indexed by the negative integers, i.e. $X_n, n \leq 0$ adapted to an increasing sequence of σ-algebras $\mathcal{F}_n$ with

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n \text{ for } n \leq -1.$$

The convergence theory for backwards martingale is much easier.

Theorem 3.6.2. *$X_\infty = \lim_{n\to-\infty} X_n$ exists a.s. and in L^1.*

Proof. Let $a < b \in \mathbb{R}$ and let $U_n(a,b)$ be the number of upcrossings by $X_{-n}, ..., X_0$. Then by the upcrossing inequality, we have $\mathbb{E}U_n(a,b) \leq \mathbb{E}(X_0 - a)^+/(b-a)$. Note that $U_n(a,b) \uparrow U(a,b)$ which is the upcrossings by the whole sequence $(X_{-n})_{n=0}^{\infty}$, so the monotone convergence theorem implies $\mathbb{E}U(a,b) < \infty$. So it holds that $\mathbb{P}(\cup_{a,b\in\mathbb{Q}} \underline{\lim}_{n\to\infty} X_n < a < b < \overline{\lim}_{n\to\infty} X_n) = 0$ and hence X_∞ exists a.s. Now observe also that $X_n = \mathbb{E}(X_0|\mathcal{F}_n)$ for every n, so $\{X_n : -n \in \mathbb{N}\}$ is uniformly integrable. Thus $X_n \to X_\infty$ in L^1 as well. □

Theorem 3.6.3. *Suppose $X_\infty = \lim_{n\to-\infty} X_n$ and $\mathcal{F}_\infty := \cap_n \mathcal{F}_n$, then $X_\infty = \mathbb{E}(X_0|\mathcal{F}_\infty)$.*

Proof. Observe that for every $n \leq 0$, $X_n = \mathbb{E}(X_0|\mathcal{F}_n)$ which in particular is uniformly integrable. Let $A \in \mathcal{F}_\infty$. It follows that

$$\int_A X_0 d\mathbb{P} = \int_A X_n d\mathbb{P}$$

Uniform integrability of $\{X_n : -n \in \mathbb{N}\}$ implies we can pass to the limits on the RHS. Thus

$$\int_A X_0 d\mathbb{P} = \int_A \lim_{n\to\infty} X_n d\mathbb{P} = \int_A X_\infty d\mathbb{P}.$$

Since this holds true for any $A \in \mathcal{F}_\infty$, by definition $X_\infty = \mathbb{E}(X_0|\mathcal{F}_\infty)$. □

Theorem 3.6.4. *If $\mathcal{F}_n \downarrow \mathcal{F}_\infty$, then $\mathbb{E}(Y|\mathcal{F}_n) \to \mathbb{E}(Y|\mathcal{F}_\infty)$ a.s. and in L^1.*

Proof. The seuence $(E(Y|\mathcal{F}_n))_{n=0}$ is clearly a UI martingale, so by previous theorem

$$\lim_{n\to\infty} \mathbb{E}(Y|\mathcal{F}_n) = \mathbb{E}(\mathbb{E}(Y|\mathcal{F}_0)|\mathcal{F}_\infty) = \mathbb{E}(Y|\mathcal{F}_\infty).$$

□

3.7 Optional Sampling Theorem

Theorem 3.7.1. *If X_n is a uniformly integrable submartingale then for any stopping time N, $X_{N\wedge n}$ is uniformly integrable.*

Proof. First note that X_n^+ is a submartingale and Theorem 4.4.1 implies $\mathbb{E}X_{N\wedge n}^+ \leq \mathbb{E}X_n^+$. Obviously X_n^+ is uniformly integrable and thus $\sup_n \mathbb{E}X_{N\wedge n}^+ \leq \sup_n \mathbb{E}X_n^+ < \infty$. Then the martingale convergence theorem (4.2.16) implies $X_{N\wedge n} \to X_N$ a.s. and $\mathbb{E}|X_N| < \infty$. Now we check that $\{X_{N\wedge n} : n \in \mathbb{N}\}$ is uniformly integrable. We compute

$$\mathbb{E}|X_{N\wedge n}|\mathbb{1}_{\{|X_{N\wedge n}|>K\}} = \mathbb{E}|X_N|\mathbb{1}_{\{|X_N|>K, N\leq n\}} + \mathbb{E}|X_n|\mathbb{1}_{\{|X_n|>K, N>n\}}.$$

The RHS can be made arbitrarily small uniformly in n because $\mathbb{E}|X_N| < \infty$ and X_n is uniformly integrable. □

The proof provides a criterion to show the uniformly integrability of $X_{N\wedge n}$.

Corollary 3.7.2. *If $\mathbb{E}|X_N| < \infty$ and $X_n \mathbb{1}_{\{N>n\}}$ is uniformly integrable, then $X_{N\wedge n}$ is uniformly integrable.*

For a UI submartingale, we obtain the following generalization of Theorem 4.4.1.

Theorem 3.7.3. *If X_n is a uniformly integrable submartingale, then for any stopping time N, we have $\mathbb{E}X_0 \leq \mathbb{E}X_N \leq \mathbb{E}X_\infty$ where $X_\infty = \lim_n X_n$.*

Proof. The result directly follows from passing to the limit in $\mathbb{E}X_0 \leq \mathbb{E}X_{N\wedge n} \leq \mathbb{E}X_n$, which is legitimate due to the uniform integrability of $X_{N\wedge n}$ and X_n. □

Theorem 3.7.4 (The optional stopping theorem)**.** *If $L \leq M$ are stopping times and $Y_{M\wedge n}$ is a uniformly integrable submartingale, then $\mathbb{E}Y_L \leq \mathbb{E}Y_M$ and $Y_L \leq \mathbb{E}(Y_M|\mathcal{F}_L)$.*