MAP 6472 / MAP 6473 - Probability and Potential Theory 1 & 2

Jay Pantone
University of Florida

Last Edited: April 14, 2013

Contents

This packet consists mainly of notes, homework assignments, and exams, from MAP6472 / MAP6473 Probability and Potential Theory 1 & 2 taught during the Fall 2012 and Spring 2013 semesters at the University of Florida. The course was taught by Prof. S. McKinley. The notes for the course follow *A Probability Path*, by Sidney Resnick. Numbering in these notes corresponds to the numbering in the text. Additionally, the course page (with lecture notes, homework problems, simulation code, and the course syllabus) can be found at http://uflprob.wordpress.com/.

If you find any errors or you have any suggestions, please contact me at jay.pantone@gmail.com.

Introduction

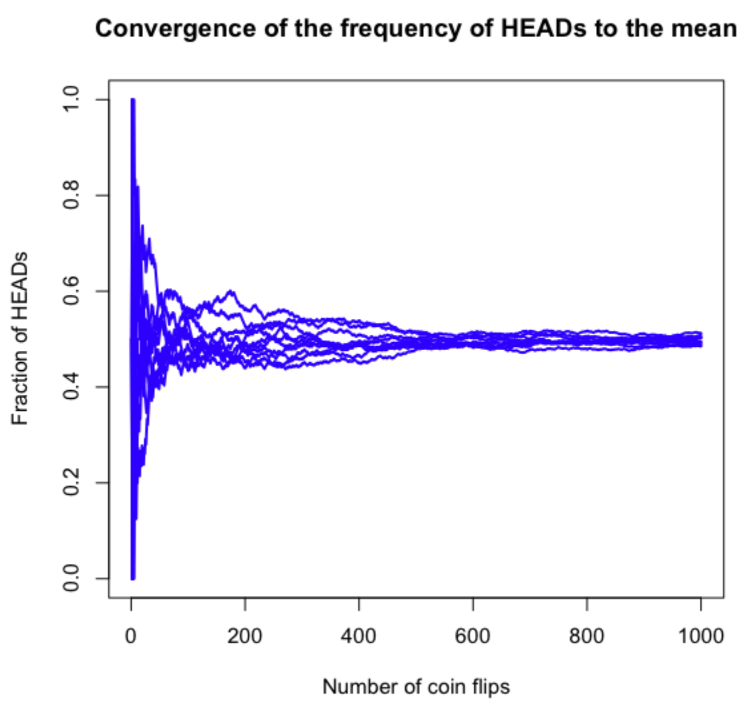
Example: Consider the idea of flipping a coin, i.e., consider $\{X_i\}_{i=1}^{\infty} \in \{0, 1\}$. Define

$$S_n := \sum_{i=1}^{n} X_i.$$

"Everyone knows" that

$$\lim_{n \to \infty} \frac{S_n}{n} = 0.5.$$

This is exhibited by the simulation:



This is an intuition we have about the Laws of Large Numbers, which we need to make rigorous.

Similarly, the intuition we have about the Central Limit Theorem is that

$$\frac{1}{\sqrt{n}} \left(\frac{S_n}{n} - \mu \right) \xrightarrow{\text{dist}} N(0, \sigma),$$

which says that the distribution approaches a normal curve.

Note: In this course, we'll consider some other types of convergence, such as the Martingale Convergence Theorem and Ergodicity.

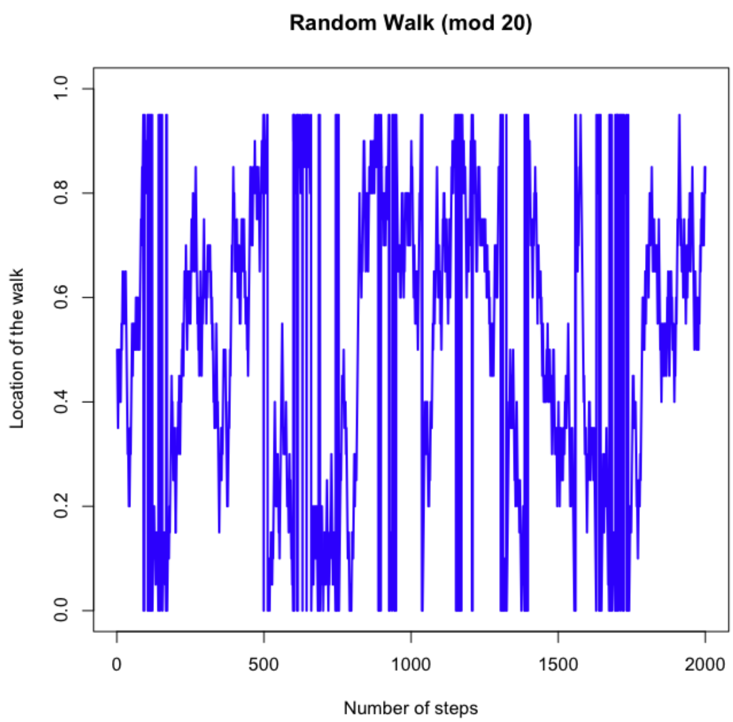
Example: (random walks modulo N) Let $X_i \in \{-1, 1\}$. Define

$$S_n := \sum_{i=1}^{n} X_i$$

and set

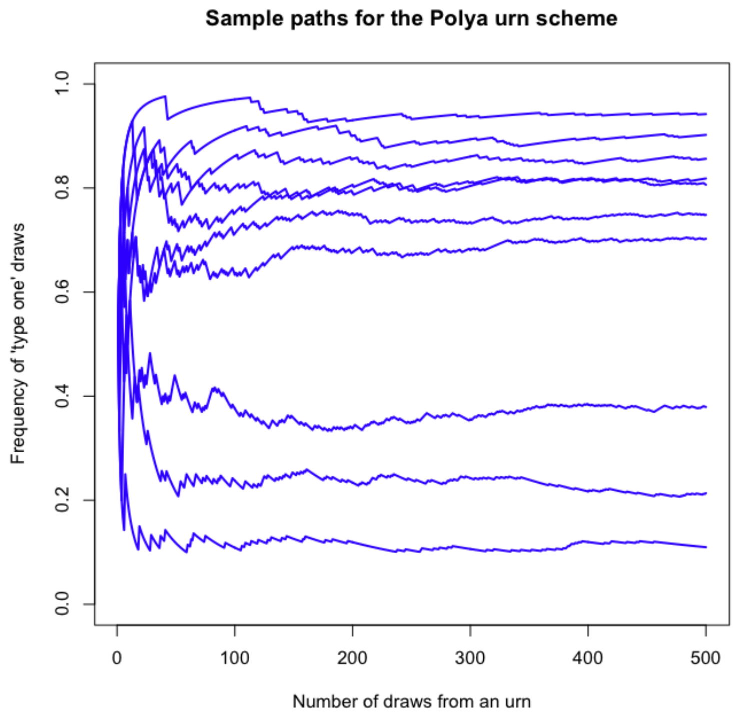
$$Z_n := S_n \pmod{N}.$$

This is exhibited by the simulation:



We're now interested in the quantity $\dfrac{Z_n}{N}$. These paths do not converge to a certain point, and so this does not have the same kind of Law of Large Numbers as the previous example. In fact, $\dfrac{Z_n}{N}$ is "uniformly distributed" in the intuitive sense. "Ergodicity" is the idea that a time-average is the same as a population-average (looking at a walk far into the future is equivalent to looking at a lot of walks at an instant in time).

Example: (Polya urn scheme) Start with an orange ball and a blue ball in an urn. Pull one out at random. Whichever color you pick, take an outside ball of the same color and put them both in. Now there are three balls, two of one color and one of the other. Repeat this process. We look at the proportion of balls which are a certain color. As you can see, this simulations has a very different appearance:



It turns out that this can converge to any point in $[0, 1]$ (and does converge "almost surely"). These are also called "reinforced random walks", and it is an example of the Martingale Convergence Theorem.

Chapter 1

Chapter 1 - Sets and Events

Section 1.5 - Set Operations and Closure (The Rigorous Representation of Information)

Example: Let $X \in \Omega := \{1,2,3,4,5,6\}$ be the outcome of the roll of a fair die. There are two observers. Observer A sees if the roll is even. Observer B sees if the roll has value ≤ 2. Let $\mathscr{A}$ and $\mathscr{B}$ denote the sets of all possible inferences based on the observations of A and B. In other words, $\mathscr{A}$ is the set of subsets of Ω for which we can know based on Observer A whether the roll is in the subset.

For starters, $\{1,2,3,4,5,6\} \in \mathscr{A}$ and $\emptyset \in \mathscr{A}$, trivially. Also, we see that $\{2,4,6\} \in \mathscr{A}$ and $\{1,3,5\} \in \mathscr{A}$. By contrast, $\{1\} \notin \mathscr{A}$, because you can never tell whether the outcome is 1 just by asking Observer A. So

$$\mathscr{A} = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \Omega\}.$$

Similarly,

$$\mathscr{B} = \{\emptyset, \{1,2\}, \{3,4,5,6\}, \Omega\}.$$

We observe two general properties of such an "inference set". Let Ω denote a set of possible outcomes and let $\mathcal{E}$ be a set of inferences. Then,

(1) $\emptyset, \Omega \in \mathcal{E}$

(2) If $E \in \mathcal{E}$, then $E^C \in \mathcal{E}$.

Now suppose that we have reports from A and B. For example, if A says TRUE and B says FALSE then we can infer $X \in \{4,6\}$. This shows us the properties:

(3) If $E_1 \in \mathcal{E}$ and $E_2 \in \mathcal{E}$, then $E_1 \cap E_2 \in \mathcal{E}$.

(3') If $E_1 \in \mathcal{E}$ and $E_2 \in \mathcal{E}$, then $E_1 \cup E_2 \in \mathcal{E}$.

Property (3') comes from applying DeMorgan's Law to Property (3).

Definition 1.5.2: A field is a non-empty class of subsets of Ω which satisfies (1), (2), and (3') above.

Definition 1.5.3: If a non-empty class of subsets of Ω satisfies (1), (2), and:

$$(3\sigma): \text{ If } E_i \in \mathcal{E} \text{ for } i \geq 1, \text{ then } \bigcup_{i=1}^{\infty} E_i \in \mathcal{E}$$

then we say it is a σ-field.

Examples of σ-fields:

(-) The power set: $\mathcal{E} = \mathcal{P}(\Omega)$.

(-) The trivial σ-field: $\mathcal{E} = \{\emptyset, \Omega\}$.

(-) The countable / co-countable σ-field: Let $\Omega := \mathbb{R}$ and define

$$\mathcal{E} := \{E \subseteq \mathbb{R} : E \text{ is countable or } E^C \text{ is countable}\}.$$

Example of a field which is not a σ-field: Let $\Omega = (0, 1]$ and let $\mathscr{A}$ contain the empty set and all finite unions of disjoint intervals of the form $(a, b]$. To prove this, we first need to verify that $\mathscr{A}$ is a field:

(i) By definition, $\emptyset \in \mathscr{A}$ and $\Omega = (0, 1] \in \mathscr{A}$.

(ii) Complementation within Ω:

$$(a, b]^C = (0, a] \cup (b, 1]$$

This also needs to be verifies for sets of finite unions, but we skip that here.

(iii) We will check that $\mathscr{A}$ is closed under finite intersections. Suppose $0 < a_1 < a_2 \leq 1$ and $a_1 < b_1 < b_2 \leq 1$. Well, if $a_2 \leq b_1$, then the intersection is empty, so this case is trivial. If $a_1 \leq b_1 \leq a_2 \leq b_2$, then the intersection is $(b_1, a_2]$, which is in $\mathscr{A}$. If $a_1 \leq b_1 \leq b_2 \leq a_2$, then the intersection is $(b_1, b_2]$, which is also in $\mathscr{A}$. Hence $\mathscr{A}$ is closed under pairwise intersections, and therefore under finite intersections.

Lastly we show that $\mathscr{A}$ is not a σ-field. Define

$$S_n := \sum_{i=1}^{n} \left(\frac{1}{2}\right)^i$$

and consider the set

$$A := (0, S_1] \cup (S_2, S_3] \cup (S_4, S_5] \cup \cdots.$$

Then, A is a countable union of sets in $\mathscr{A}$, but is not a member of $\mathscr{A}$.

Section 1.6 - The σ-field Generated by a Given Class $\mathcal{C}$

Example: In the leading example, we had $\Omega = \{1, 2, 3, 4, 5, 6\}$ with two observers A and B that know the truth of $X \in A := \{2, 4, 6\}$ and $X \in B := \{1, 2\}$, respectively. Then we get the σ-field:

$$\begin{aligned}\mathcal{E} &= \{\emptyset, \Omega, A, A^C, B, B^C, A \cap B, A \cap B^C, A^C \cap B, A^C \cap B^C, (A \cap B)^C, (A \cap B^C)^C, (A^C \cap B)^C, (A^C \cap B^C)^C\} \\ &= \{\emptyset, \Omega, \{2, 4, 6\}, \{1, 3, 5\}, \{1, 2\}, \{3, 4, 5, 6\}, \{2\}, \{4, 6\}, \{1\}, \{3, 5\}, \{1, 2, 4, 6\}, \{2, 3, 4, 5, 6\}, \{1, 2, 3, 5\}, \{1, 3, 4, 5, 6\}\}\end{aligned}$$

We call $\mathcal{E}$ the σ-field generated by A and B. Notationally,

$$\mathcal{E} = \sigma(A, B).$$

Remark: If Ω is finite, then $\sigma(\mathcal{C})$ can always be constructed by enumeration of intersections and complements. However, this method is not possible when Ω is infinite. (In fact, it turns out the σ-fields are either finite or uncountable, and we certainly can't enumerate an uncountable set!)

Corollary 1.6.1: The intersection of σ-fields is a σ-field.

Proof: Just check conditions.

Definition 1.6.1: Let $\mathcal{C}$ be a collection of subsets of Ω. The σ-field generated by $\mathcal{C}$, denoted $\sigma(\mathcal{C})$ is a σ-field satisfying:

(a) $\mathcal{C} \subset \sigma(\mathcal{C})$

(b) If $\mathcal{E}'$ is some σ-field containing $\mathcal{C}$, then $\sigma(\mathcal{C}) \subset \mathcal{E}'$.

Remark: $\sigma(\mathcal{C})$ is sometimes called the minimal σ-field containing $\mathcal{C}$.

Proposition 1.6.1: Given a class $\mathcal{C}$ of subsets of Ω, there is a unique minimal σ-field which contains $\mathcal{C}$.

Proof: Let

$$\aleph := \{\mathcal{E} \mid \mathcal{E} \text{ is a } \sigma\text{-field and } \mathcal{C} \subset \mathcal{E}\}.$$

Since $\mathbb{P}(\Omega)$ is a σ-field containing $\mathcal{C}$, $\aleph$ is nonempty. Now define

$$\aleph^{\#} := \bigcap_{\mathcal{E} \in \aleph} \mathcal{E}.$$

We claim that $\aleph^{\#} = \sigma(\mathcal{C})$.

By **Corollary 1.6.1**, $\aleph^{\#}$ is a σ-field itself. Furthermore, $\mathcal{C} \subset \aleph^{\#}$. To see minimality and uniqueness, suppose $\mathcal{E}'$ is a σ-field containing $\mathcal{C}$. Since $\mathcal{E}' \in \aleph$, it follows that $\aleph^{\#} \subseteq \mathcal{E}'$. □

Section 1.7 - Borel Sets on the Real Line

Definition: Suppose $\Omega := \mathbb{R}$ and let $\mathcal{C} := \{(a, b] \mid -\infty \leq a \leq b \leq \infty\}$. The Borel subsets of $\mathbb{R}$ are defined to be

$$\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{C}).$$

It is worth noting that:

$$\begin{aligned}
\mathcal{B}(\mathbb{R}) &= \sigma(\{(a, b) \mid -\infty \leq a \leq b \leq \infty\}) \\
&= \sigma(\{[a, b) \mid -\infty \leq a \leq b \leq \infty\}) \\
&= \sigma(\{[a, b] \mid -\infty \leq a \leq b \leq \infty\}) \\
&= \sigma(\{(-\infty, x] \mid x \in \mathbb{R}\}) \\
&= \sigma(\text{open subsets}) \\
&= \sigma(\text{closed subsets}).
\end{aligned}$$

In general, if we consider a metric space $\mathbb{S}$, then the Borel subsets of $\mathbb{S}$ are the collection

$$\mathcal{B}(\mathbb{S}) = \sigma(\text{open subsets of } \mathbb{S}).$$

Important examples: $\mathbb{S} = \mathbb{R}^d,\ \mathbb{R}^\infty,\ C((0, 1])$.

Chapter 2

Chapter 2 - Probability Spaces

Section 2.1 - Basic Definitions and Properties

Definition: A probability space is a triple $(\Omega, \mathcal{B}, \mathbb{P})$ where:

- Ω is the sample space,
- $\mathcal{B}$ is a σ-field of subsets of Ω,
- $\mathbb{P}$ is a probabiliy measure,

where a probability measure is a function $\mathbb{P} : \mathcal{B} \to [0, 1]$ such that

(i) $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{B}$,

(ii) $\mathbb{P}$ is σ-additive, meaning if $\{A_n\}_{n\in\mathbb{N}}$ are disjoint events in $\mathcal{B}$, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n),$$

(iii) $\mathbb{P}(\Omega) = 1$.

Properties: We should verify the following properties:

(1) $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$

Proof: $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^C) = \mathbb{P}(A) + \mathbb{P}(A^C)$

(2) $\mathbb{P}(\emptyset) = 0$

Proof: $\mathbb{P}(\emptyset) = \mathbb{P}(\Omega^C) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0$

(3) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$

Proof: $\mathbb{P}(A) = \mathbb{P}(AB) + \mathbb{P}(AB^C)$ and $\mathbb{P}(B) = \mathbb{P}(AB) + P(A^C B)$. So,

$$\mathbb{P}(A \cup B) = \mathbb{P}(AB \cup (AB^C) \cup (A^C B)) = \mathbb{P}(AB) + \mathbb{P}(AB^C) + \mathbb{P}(A^C B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB).$$

(4) Inclusion-Exclusion:

$$\mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{i} \mathbb{P}(A_i) - \sum_{i,j} \mathbb{P}(A_i A_j) + \sum_{i,j,k} \mathbb{P}(A_i A_j A_k) - \cdots + (-1)^{n+1} \mathbb{P}(A_1 A_2 \cdots A_n).$$

(5) Monotonicity: For events $A, B \in \mathcal{B}$, if $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Proof: Observe $\mathbb{P}(B) = \mathbb{P}(BA) + \mathbb{P}(BA^C) = \mathbb{P}(A) + \mathbb{P}(BA^C)$, since $BA = A$. Since all probabilities are non-negative, $\mathbb{P}(B) \geq \mathbb{P}(A)$.

(6) Subadditivity:

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

Proof: Follows from monotonicity.

(7) Continuity: If $\{A_n\}_{n\in\mathbb{N}}$ is a monotone increasing sequence of events in $\mathcal{B}$, i.e.,

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$$

and there exists A such that $A_n \nearrow A$, then

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A).$$

Proof: Define $B_1 := A_1$, $B_2 := A_2 \setminus A_1$, …, $B_k := A_k \setminus A_{k-1}$, etc. These are disjoint events whose union is A. So,

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(B_n) \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \mathbb{P}(B_n) \\ &= \lim_{N\to\infty} \mathbb{P}\left(\bigcup_{n=1}^{N} B_n\right) \\ &= \lim_{N\to\infty} \mathbb{P}(A_N)\end{aligned}$$

(8) Fatou's Lemma:

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) \leq \liminf_{n\to\infty} \mathbb{P}(A_n) \leq \limsup_{n\to\infty} \mathbb{P}(A_n) \leq \mathbb{P}\left(\limsup_{n\to\infty} A_n\right)$$

Each inequality can be strict. For example, pick $\Omega := (-1, 1)$ and pick X_k to be drawn "uniformly" between -1 and 1. Define A_{2k} be the event that $X_{2k} \in (0, 1)$ and let A_{2k+1} be the event that $X_{2k+1} \in (-1, 0)$. It follows that

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \emptyset.$$

So,

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) = \mathbb{P}(\emptyset) = 0.$$

On the other hand, $\mathbb{P}(A_n) = \dfrac{1}{2}$ for all n. So,

$$0 = \mathbb{P}\left(\liminf_{n\to\infty} A_n\right) < \liminf_{n\to\infty} \mathbb{P}(A_n) = \frac{1}{2}$$

and the inequality is strict.

Now we prove Fatou's Lemma:

Proof:

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k\right) = \mathbb{P}\left(\lim_{n\to\infty}\bigcap_{k=n}^{\infty} A_k\right).$$

Observe that

$$B_n := \bigcap_{k=n}^{\infty} A_k$$

is a monotone increasing sequene of sets (since we are intersecting over fewer sets). So, by continuity (7), we have

$$\mathbb{P}\left(\lim_{n\to\infty}\bigcap_{k=n}^{\infty} A_k\right) = \lim_{n\to\infty}\mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k\right) \le \liminf_{n\to\infty}\mathbb{P}(A_n)$$

where in the last step we use the fact that

$$\bigcap_{k=n}^{\infty} A_k \subset A_n. \ \square$$

(9) If $A_n \to A$, then $\lim_{n\to\infty}\mathbb{P}(A_n) = \mathbb{P}(A)$. (This is like (7) without the requirement of monotonicity.)

Proof: If $A_n \to A$ then by the definition of "limit"

$$\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = A.$$

By **Fatou's Lemma**,

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}\left(\liminf_{n\to\infty} A_n\right)\\ &\le \liminf_{n\to\infty}\mathbb{P}(A_n)\\ &\le \limsup_{n\to\infty}\mathbb{P}(A_n)\\ &\le \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \mathbb{P}(A).\end{aligned}$$

Hence there is equality throughout, and so

$$\lim_{n\to\infty}\mathbb{P}(A_n) = \mathbb{P}(A). \ \square$$

Example 2.1.2: (The Graduates and their Caps) A bunch of students are graduating and they all throw their caps in the air. An instantaneous tornado mixes them all up and each student catches a cap. What is the probability that every student ends up with a cap which is not their own?

Label the students (and their caps) from 1 through n. When each student picks up a cap, they look at the label. This induces a permutation on the set $[n]$. So,

$$\Omega = \{(x_1, x_2, \ldots, x_n) \mid x_i \in \{1, \ldots, n\} \text{ and } x_i \neq x_j, \text{ if } i \neq j\} = \{\text{permutations of } [n]\}.$$

The σ-field generated by Ω is $\mathcal{B} := \mathcal{P}(\Omega)$. For any $(x_1, x_2, \ldots, x_n) \in \Omega$, set

$$\mathbb{P}((x_1, x_2, \ldots, x_n)) := \frac{1}{n!}$$

and corresponding for a set $B \in \mathcal{B}$ in the σ-field, we set

$$\mathbb{P}(B) := \frac{1}{n!}\cdot(\text{the number of elements of } B).$$

We have now defined our probability space $(\Omega, \mathcal{B}, \mathbb{P})$.

Define $A_i \subset \Omega$ to be all events where $x_i = i$. The event that at least one student picks up their own cap is

$$A := \bigcup_{i=1}^{n} A_i.$$

We use inclusion-exclusion to see that

$$\mathbb{P}(A) = \sum_{i=1}^{n} A_i - \sum_{1 \le i < j \le n} A_i A_j + \sum_{1 \le i < j < k \le n} A_i A_j A_k - \cdots + (-1)^{n+1} \mathbb{P}(A_1 A_2 \cdots A_n).$$

It is clear that for all i,

$$\mathbb{P}(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

So,

$$\begin{aligned} \mathbb{P}(A) &= \binom{n}{1}\frac{(n-1)!}{n!} - \binom{n}{2}\frac{(n-2)!}{n!} + \binom{n}{3}\frac{(n-3)!}{n!} - \cdots + (-1)^{n+1}\binom{n}{n}\frac{1}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + \frac{(-1)^n}{n!} \end{aligned}$$

Recall that

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

and so

$$\mathbb{P}(A) \approx 1 - e^{-1} \approx 0.632\ldots$$

for n large. So, the probability that no one catches their cap is approximately

$$1 - (1 - e^{-1}) = e^{-1}.$$

Definition: Let $\Omega := \mathbb{R}$. Suppose $\mathbb{P}$ is a probability measure on $\mathcal{B}(\mathbb{R})$. Define $F(x)$ by

$$F(x) = \mathbb{P}((-\infty, x])$$

for all $x \in \mathbb{R}$. This is a distribution function.

Lemma: Let F be as above. Then,

(i) F is right-continuous.

(ii) F is monotone increasing.

(iii) F has limits at $\pm\infty$:

$$F(\infty) := \lim_{x \to \infty} F(x) = 1,$$
$$F(-\infty) := 0.$$

Proof of (i): Let $x \in \mathbb{R}$ and suppose that $\{x_n\} \subset \mathbb{R}$ such that $x_n \searrow x$. Then,

$$\begin{aligned} F(x) &= \mathbb{P}((-\infty, x]) \\ &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} (-\infty, x_n]\right) \\ &= \lim_{n \to \infty} \mathbb{P}((-\infty, x_n]) \\ &= \lim_{n \to \infty} F(x_n). \ \square \end{aligned}$$

Section 2.2 - More on Closure

We want to develop the mathematical tools to prove things such as the following corollary.

Corollary 2.2.2: Let $\Omega := \mathbb{R}$. Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be two probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such their cumulative distribution functions (cdfs) are equal, i.e.,

$$\forall x \in \mathbb{R} : F_1(x) := \mathbb{P}_1((-\infty, x]) = \mathbb{P}_2((-\infty, x]) =: F_2(x).$$

Then, $\mathbb{P}_1 \equiv \mathbb{P}_2$ on $\mathcal{B}(\mathbb{R})$.

Remark: To borrow language from Chapter 1, the probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ are "closures" of the information contained in their respective cdfs. We need to generalize the notion of closure for certain mathematical "structures" (defined to be a collection of subsets of Ω satisfying some closure axioms). Examples of some structures: σ-fields, σ-rings, monotone class, π-systems, λ-systems.

Definition 2.2.1: The minimal structure $\mathcal{S}$ generated by a class $\mathcal{C}$ is a nonempty structure satisfying:

(i) $\mathcal{C} \subseteq \mathcal{S}$,

(ii) If $\mathcal{S}'$ is another structure containing $\mathcal{C}$, then $\mathcal{S} \subseteq \mathcal{S}'$.

Proposition 2.2.1: The minimal structure $\mathcal{S}(\mathcal{C})$ generated by a class $\mathcal{C}$ in Ω exists and is unique.

Proof: Define $\aleph := \{G \mid G \text{ is a structure and } \mathcal{C} \subset G\}$. Then, define

$$\mathcal{S}(\mathcal{C}) := \bigcap_{G \in \aleph} G,$$

This requires our "minimal structure" to be closed under infinite intersections. □

Definition: A π-system is a class $\mathcal{P}$ which is closed under finite intersections.

Definition: A λ-system is a class $\mathscr{L}$ of subsets of Ω satisfying:

(λ_1) $\Omega \in \mathscr{L}$,

(λ_2) If $A \in \mathscr{L}$ then $A^C \in \mathscr{L}$,

(λ_3) If $\{A_n\}_{n=1}^{\infty}$ is a sequence of mutually disjoint sets in $\mathscr{L}$, then $\displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathscr{L}$.

Remark: Every σ-field is a λ-system, but not vice versa.

Theorem 2.2.2: (Dynkin's π-λ Theorem)

(a) If $\mathcal{P}$ is a π-system and $\mathscr{L}$ is a λ-system containing $\mathcal{P}$, then $\sigma(\mathcal{P}) \subseteq \mathscr{L}$.

(b) If $\mathcal{P}$ is a π-system, then $\sigma(\mathcal{P}) = \mathscr{L}(P)$.

We will prove Dynkin's Theorem later. For now, we will prove a proposition and some corollaries.

Proposition 2.2.3: Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be two probability measures on $(\Omega, \mathcal{B})$. The class

$$\mathscr{L} := \{A \in \mathcal{B} \mid \mathbb{P}_1(A) = \mathbb{P}_2(A)\}$$

is a λ-system.

Proof: We verify each of the properties:

(λ_1) $\mathbb{P}_1(\Omega) = \mathbb{P}_2(\Omega) = 1$, and so $\Omega \in \mathscr{L}$.

(λ_2) Suppose $A \in \mathscr{L}$. Then,

$$\mathbb{P}(A^C) = 1 - \mathbb{P}_1(A) = 1 - \mathbb{P}_2(A) = \mathbb{P}_2(A^C).$$

(λ_3) Suppose $\{A_j\}$ is a mutually disjoint sequence in $\mathscr{L}$. Then,

$$\mathbb{P}_1\left(\bigcup A_j\right) = \sum_j \mathbb{P}_1(A_j) = \sum_j \mathbb{P}_2(A_j) = \mathbb{P}_2\left(\bigcup A_j\right).$$

Hence $\mathscr{L}$ is a λ-system. □

Corollary 2.2.1: If $\mathbb{P}_1$ and $\mathbb{P}_2$ are probability measures on $(\Omega, \mathcal{B})$ and if $\mathcal{P} \subseteq \mathcal{B}$ is a π-system such that for all $A \in \mathcal{P}$ we have $\mathbb{P}_1(A) = \mathbb{P}_2(A)$, then

$$\forall B \in \sigma(\mathcal{P}) : \quad \mathbb{P}_1(B) = \mathbb{P}_2(B).$$

Proof: From **Proposition 2.2.3**, $\mathscr{L} := \{A \in \mathcal{B} \mid \mathbb{P}_1(A) = \mathbb{P}_2(A)\}$ is a λ-system. Furthermore, $P \subset \mathscr{L}$. By **Dynkin's Theorem**, $\sigma(P) \subset \mathscr{L}$, which proves the statement. □

Corollary 2.2.2: If $F_1(x) = F_2(x)$ for all $x \in \mathbb{R}$, then $\mathbb{P}_1 = \mathbb{P}_2$ on $\mathcal{B}(\mathbb{R})$. $F_1(x)$ and $F_2(x)$ are defined as in the lemma preceding this section.

Proof: Define $\mathcal{P} := \{(-\infty, x] \mid x \in \mathbb{R}\}$. Then $\mathcal{P}$ is a π-system since

$$(-\infty, x] \cup (-\infty, y] = (-\infty, \min(x, y)] \in \mathcal{P}.$$

Recall also that $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$. So, using **Corollary 2.2.1**, $\mathcal{B}(\mathbb{R}) \subseteq \mathscr{L} \subseteq \mathcal{B}(\mathbb{R})$ (where $\mathscr{L}$ is defined as in the previous corollary), and so $\mathscr{L} = \mathcal{B}(\mathbb{R})$, i.e.,

$$\mathbb{P}_1(A) = \mathbb{P}_2(A)$$

for all $A \in \mathcal{B}(\mathbb{R})$. □

Proposition 2.2.4: A class $\mathscr{L}$ that is both a π-system and a λ-system is a σ-field.

Proof: By (λ_1), $\Omega \in \mathscr{L}$. Suppose that $A \in L$. By (λ_2), $A^C \in \mathscr{L}$. Suppose that $\{A_i\}_{i=1}^{i=\infty} \subseteq \mathscr{L}$. We want to show that the infinite union of the A_i is an element of $\mathscr{L}$. Define

$$B_1 := A_1, \quad B_2 := A_2 \cap A_1^C, \quad B_3 := A_3 \cap A_1^C \cap A_2^C, \quad \ldots, \quad B_n := A_n \cap A_1^C \cap \cdots \cap A_{n-1}^C, \quad \ldots.$$

In general, $B_n \in \mathscr{L}$ only if $\mathscr{L}$ is a π-system, which we assumed it is. So, the union of the A_i equals the union of the B_i, which is now a disjoint union. So, $\mathscr{L}$ is closed under arbitrary unions. Therefore, $\mathscr{L}$ is a σ-field. □

Now we will prove **Dynkin's π-λ Theorem**. We restate it first.

Theorem 2.2.2: (Dynkin's π-λ Theorem)

(a) If $\mathcal{P}$ is a π-system and $\mathscr{L}$ is a λ-system containing $\mathcal{P}$, then $\sigma(\mathcal{P}) \subseteq \mathscr{L}$.

(b) If $\mathcal{P}$ is a π-system, then $\sigma(\mathcal{P}) = \mathscr{L}(P)$.

Proof of (a): Let $\mathscr{L}(\mathcal{P})$ be the λ-system generated by $\mathcal{P}$. By minimality,

$$\mathcal{P} \subseteq \mathscr{L}(\mathcal{P}) \subseteq \mathscr{L}.$$

Now suppose for a moment that $\mathscr{L}(\mathcal{P})$ is also a π-system. Then, by **Proposition 2.2.4**, $\mathscr{L}(\mathcal{P})$ is a σ-field. By minimality of $\sigma(\mathcal{P})$, we have the chain

$$\mathcal{P} \subseteq \sigma(\mathcal{P}) \subseteq \mathscr{L}(\mathcal{P}) \subseteq \mathscr{L}.$$

This would give the result. So it remains to show that $\mathscr{L}(\mathcal{P})$ is also a π-system.

Given $A \subseteq \Omega$, define

$$\mathscr{L}_A := \{B \subseteq \Omega \mid A \cap B \in \mathscr{L}(\mathcal{P})\}.$$

Claim 1: If $A \in \mathscr{L}(\mathcal{P})$, then $\mathscr{L}_A$ is a λ-system.

Proof: Since $A \in \mathscr{L}(\mathcal{P})$, by hypothesis it follows that $\Omega \cap A = A \in \mathscr{L}(\mathcal{P})$. Therefore, $\Omega \in \mathscr{L}_A$ and so (λ_1) is verified. Next, suppose $B \in \mathscr{L}_A$. We need to show that $B^C \in \mathscr{L}_A$, i.e., $B^C \cap A \in \mathscr{L}(\mathcal{P})$. So, we need to rewrite this as the complement of a disjoint union of sets in $\mathscr{L}(P)$. Clearly

$$B^C \cap A = (A^C \cup (A \cap B))^C \in \mathscr{L}(\mathcal{P}).$$

So, (λ_2) is verified. To show (λ_3), assume that $\{B_i\}_{i\in\mathbb{N}}$ is a disjoint sequence of sets. Then,

$$A \cap \left(\bigcup B_i\right) = \bigcup (A \cap B_i) \in \mathscr{L}(\mathcal{P})$$

because $A \cap B^C$ is also the union of disjoint sets of $\mathscr{L}(\mathcal{P})$. So, this claim is proved and $\mathscr{L}_A$ is a λ-system in the case where $A \in \mathscr{L}(\mathcal{P})$. □

Claim 2: If $A \in \mathcal{P}$, then $\mathscr{L}(\mathcal{P}) \subseteq \mathscr{L}_A$.

Proof: Since $A \in \mathcal{P} \subseteq \mathscr{L}(\mathcal{P})$, then $\mathscr{L}_A$ is a λ-system. By minimality (since $\mathscr{L}_A \supseteq \mathcal{P}$ because $\mathcal{P}$ is a π-system), $\mathscr{L}(\mathcal{P}) \subseteq \mathscr{L}_A$. □

Remark: Suppose $A \in \mathcal{P}$ and $B \in \mathscr{L}(\mathcal{P})$. Then since $B \in \mathscr{L}(P) \subseteq \mathscr{L}_A$, we have that $A \cap B \in \mathscr{L}(\mathcal{P})$.

Claim 3: If $A \in \mathscr{L}(\mathcal{P})$, then $\mathscr{L}(\mathcal{P}) \subseteq \mathscr{L}_A$.

Proof: First, suppose $B \in \mathcal{P}$. Then by the previous remark, $A \cap B \in \mathscr{L}(\mathcal{P})$. Therefore $P \subseteq \mathscr{L}_A$. But from **Claim 1**, $\mathscr{L}_A$ is a λ-system. By minimality, $\mathscr{L}(\mathcal{P}) \subseteq \mathscr{L}_A$.

If $A \in \mathscr{L}(\mathcal{P})$ and $B \in \mathscr{L}(\mathcal{P})$, then $B \in \mathscr{L}_A$. By definition of $\mathscr{L}_A$, $A \cap B \in \mathscr{L}(\mathcal{P})$. Hence, $\mathscr{L}(\mathcal{P})$ is a π-system. This completes **part (a)** of the theorem. □

Proof of (b): Since $\mathcal{P}$ is a π-system contained in the λ-system $\mathscr{L}(\mathcal{P})$, **part (a)** implies

$$\sigma(\mathcal{P}) \subseteq \mathscr{L}(\mathcal{P}).$$

On the other hand, every σ-field is a λ-system. Therefore $\sigma(\mathcal{P})$ is a λ-system by minimality. So,

$$\mathscr{L}(\mathcal{P}) \subseteq \sigma(\mathcal{P}).$$

Hence,

$$\mathscr{L}(\mathcal{P}) = \sigma(\mathcal{P}). \ \square$$

Section 2.4 - Construction of Probability Spaces

Definition 2.4.1: A class $\mathcal{S}$ of subsets of Ω is a semialgebra if the following hold:

(i) $\emptyset, \Omega \in S$

(ii) $\mathcal{S}$ is a π-system

(iii) If $A \in \mathcal{S}$, then there exists some finite collection of disjoint $C_1, C_2, \dots, C_n \in \mathcal{S}$ such that

$$A^C = \bigsqcup_{i=1}^{n} C_i.$$

Remark: For the disjoint union, rather than the conventional notation $\bigsqcup_{i=1}^{n}$, Resnick uses the notation $\sum_{i=1}^{n}$.

Note: A^C is not necessarily in $\mathcal{S}$, since $\mathcal{S}$ is not assumed to be a λ-system.

Example: Let $\Omega := \mathbb{R}$ and define $\mathcal{S}_1 := \{(a, b] \mid -\infty \leq a \leq b \leq \infty\}$. Then,

(i) $\emptyset, \Omega \in \mathcal{S}_1$

(ii) Suppose $I_1, I_2 \in \mathcal{S}_1$. Then it's clear that $I_1 \cap I_2 \in \mathcal{S}_1$.

(iii) Suppose $I = (a, b] \in \mathcal{S}_1$. Then,

$$I^C = (-\infty, a] \sqcup (b, \infty].$$

Remark: Note that $\mathcal{S}_1$ is not closed under complements, since $I^C \notin \mathcal{S}_1$.

Example: Let $\Omega := \mathbb{R}^k = \{(x_1, x_2, \dots, x_k) \mid x_i \in \mathbb{R}\}$. Define $\mathcal{S}_k$ to be the set of all rectangles, i.e., sets of the form

$$A = I_1 \times I_2 \times \cdots \times I_k.$$

Then, $\mathcal{S}_k$ is a semialgebra.

Lemma 2.4.1: (the field generated by a semialgebra) Suppose $\mathcal{S}$ is a semialgebra of subsets of Ω. Then, the smallest field containing $\mathcal{S}$, denoted $\mathscr{A}(\mathcal{S})$ can also be written

$$\mathscr{A}(\mathcal{S}) = \left\{ \bigsqcup_{i \in I} S_i \,\middle|\, |I| < \infty,\ S_i \in \mathcal{S} \right\} \tag{$*$}$$

Proof: Define

$$\Lambda := \left\{ \bigsqcup_{i \in I} S_i \,\middle|\, |I| < \infty,\ S_i \in \mathcal{S} \right\}.$$

We first show that $\mathcal{S} \subseteq \Lambda$. Note that for any $S \in \mathcal{S}$, we have that $S \in \Lambda$ by taking $I := \{1\}$ and $S_1 := \{S\}$. Now show that Λ is a field.

(i) Since $\Omega \in \mathcal{S}$ by definition of semialgebra, we have $\Omega \in \Lambda$.

(ii) Suppose $\bigsqcup_i S_i \in \Lambda$ and $\bigsqcup_j S'_j \in \Lambda$. We need to show that their intersection is also contained in Λ. Note

$$\left(\bigsqcup_i S_i \right) \bigcap \left(\bigsqcup_j S'_j \right) = \bigsqcup_{i,j} (S_i \cap S'_j).$$

(You may need to think about it a bit to see that this is indeed a disjoint union.) Now, each $S_i \cap S'_j \in \mathcal{S}$ since semialgebras are π-systems.

(iii) Suppose $\bigsqcup_i S_i \in \Lambda$, then

$$\left(\bigsqcup_i S_i\right)^C = \bigcap_i S_i^C.$$

Each S_i^C can be written as

$$S_i^C = \bigsqcup_{i_k} S_{i_k}$$

for some collection S_{i_k}. Since Λ is closed under finite intersections, we have

$$\left(\bigsqcup_i S_i\right)^C \in \Lambda.$$

This shows that Λ is a field containing $\mathcal{S}$, and so $\mathscr{A}(\mathcal{S}) \subseteq \Lambda$. To see the reverse, note that each element is a (disjoint) finite union of elements of $\mathcal{S}$, and $\mathscr{A}(\mathcal{S})$ contains these elements of $\mathcal{S}$ and is closed under finite unions, and so $\Lambda \subseteq \mathscr{A}(\mathcal{S})$. So, we have shown equality. $\square$

Theorem 2.4.1: (First Extension Theorem) Suppose $\mathcal{S}$ is a semialgebra of subsets of Ω and $P : \mathcal{S} \to [0, 1]$ is σ-additive on $\mathcal{S}$ and satisfies $P(\Omega) = 1$. Then, there exists a unique extension P' of P to $\mathscr{A}(\mathcal{S})$ defined by

$$P'\left(\bigsqcup_i S_i\right) := \sum_i P(S_i)$$

which is a probability measure on $\mathscr{A}(\mathcal{S})$.

Proof: First we show uniqueness of the extension. Suppose that $A \in \mathscr{A}(\mathcal{S})$ has two distinct representations,

$$A = \bigsqcup_i S_i = \bigsqcup_j S'_j.$$

We must confirm that

$$\sum_i P(S_i) = \sum_j P(S'_j).$$

Indeed,

$$\begin{aligned}
\sum_i P(S_i) &= \sum_i P(S_i \cap A) \\
&= \sum_i P\left(S_i \cap \left(\bigsqcup_i S'_j\right)\right) \\
&= \sum_i P\left(\bigsqcup_j (S_i \cap S'_j)\right).
\end{aligned}$$

Now, since $\mathcal{S}$ is a π-system, $S_i \cap S'_j \in \mathcal{S}$. So, by the additivity of P,

$$\sum_i P\left(\bigsqcup_j (S_i \cap S'_j)\right) = \sum_i \sum_j P(S_i \cap S'_j) = \sum_j P(S'_j).$$

The last equality is shown by reversing the logic.

Next, we need to show that P' is σ-additive on $\mathscr{A}(\mathbb{S})$. Suppose

$$A = \bigsqcup_{i=1}^{\infty} A_i \in \mathscr{A}(\mathcal{S})$$

and that for each A_i,

$$A_i = \bigsqcup_{j \in J_i} S_{i,j} \in \mathscr{A}(\mathcal{S})$$

where $|J_i| < \infty$ and $S_{i,j} \in \mathcal{S}$. Since $A \in \mathscr{A}(\mathcal{S})$, there must be some finite representation

$$A = \bigsqcup_{k \in K} S_k$$

where $|K| < \infty$ and $S_k \in \mathcal{S}$. By the definition of P',

$$P'(A) = \sum_{k \in K} P(S_k).$$

But,

$$S_k = S_k \cap A = \bigsqcup_{i=1}^{\infty} [S_k \cap A_i] = \bigsqcup_{i=1}^{\infty} \bigsqcup_{j \in J_i} [S_k \cap S_{i,j}].$$

Since $\mathcal{S}$ is a π-system, $S_k \cap S_{i,j} \in \mathcal{S}$. Separately, we know that $S_k \in \mathcal{S}$. Since P is σ-additive,

$$P(S_k) = \sum_{i=1}^{\infty} \sum_{j \in J_i} P(S_k \cap S_{i,j})$$

and so

$$\begin{aligned} P'(A) &= \sum_{k \in K} P(S_k) \\ &= \sum_{k \in K} \sum_{i=1}^{\infty} \sum_{j \in J_i} P(S_k \cap S_{i,j}) \\ &= \sum_{i=1}^{\infty} \sum_{j \in J_i} \sum_{k \in K} P(S_k \cap S_{i,j}) \\ &= \sum_{i=1}^{\infty} \sum_{j \in J_i} P(S_{i,j}) \\ &= \sum_{i=1}^{\infty} P'\left(\bigsqcup_{j \in J_i} S_{i,j} \right) \\ &= \sum_{i=1}^{\infty} P'(A_i). \end{aligned}$$

Thus, P' is σ-additive. □

Theorem 2.4.2: (Second Extension Theorem) A probability measure $\mathbb{P}$ defined on a field $\mathscr{A}$ of subsets of Ω has a unique extension to a probability measure on $\sigma(\mathscr{A})$.

Remark: See book for proof.

Theorem 2.4.3: (Combo Extension Theorem) Suppose $\mathcal{S}$ is a semialgebra of subsets of Ω and that $\mathbb{P}$ is a σ-additive set function mapping $\mathcal{S}$ into $[0, 1]$ such that $\mathbb{P}(\Omega) = 1$. Then, there exists a unique probability measure on $\sigma(S)$ that extends $\mathbb{P}$.

Remark: This combines the first and second extension theorems. We are responsible for being able to state this theorem on a test, while defining each of the underlined terms in the process.

Section 2.5 - Measure Constructions

Construction: To define the Lebesgue Measure on $(0,1]$, take $\Omega := (0,1]$. Let $\mathcal{B}$ be the Borel sets on $(0,1]$ and $\mathcal{S} = \{(a,b] \mid 0 \le a \le b \le 1\}$ (we showed already that $\mathcal{S}$ is a semialgebra and that $\sigma(\mathcal{S}) = \mathcal{B}$). Define $\lambda : S \to [0,1]$ by $\lambda(\emptyset) = 0$ and $\lambda((a,b]) = b - a$ for $b > a$.

The only thing left to be shown is σ-additivity on $\mathcal{S}$. For finite additivity, let $(a,b] \in \mathcal{S}$ be written by a finite disjoint union of S, i.e.,

$$(a,b] = \bigsqcup_{i=1}^{k} (a_i, b_i].$$

Then (with some possible renumbering), $a_i = b_{i-1}$ for all $2 \le i \le k$. Well,

$$\mathbb{P}((a,b]) = b - a$$

by definition, but now we have

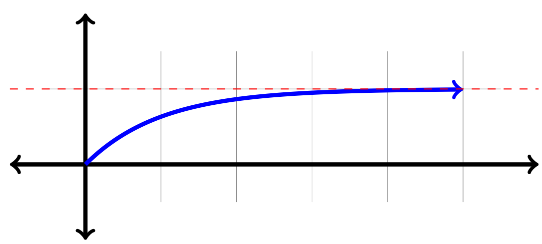
$$\mathbb{P}\left(\bigsqcup_{i=1}^{k} (a_i, b_i]\right) = \sum_{i=1}^{k} b_i - a_i = b_k - a_1 = b - a.$$

So we have shown finite additivity. See the book for the σ-additive part.

Remark: How do we draw from arbitrary probability distributions from uniform random variables?

Example: We now discuss generating exponential random variables: $\mathbb{P}(X > x) = e^{-\lambda x}$, where $\lambda > 0$ and $x > 0$. Then, the cdf is

$$F(x) := \mathbb{P}(X < x) = 1 - e^{-\lambda x}.$$



We claim that if $u \sim \text{Unif}(0,1)$ then $F^{-1}(u) \sim \text{Exp}(\lambda)$. Indeed,

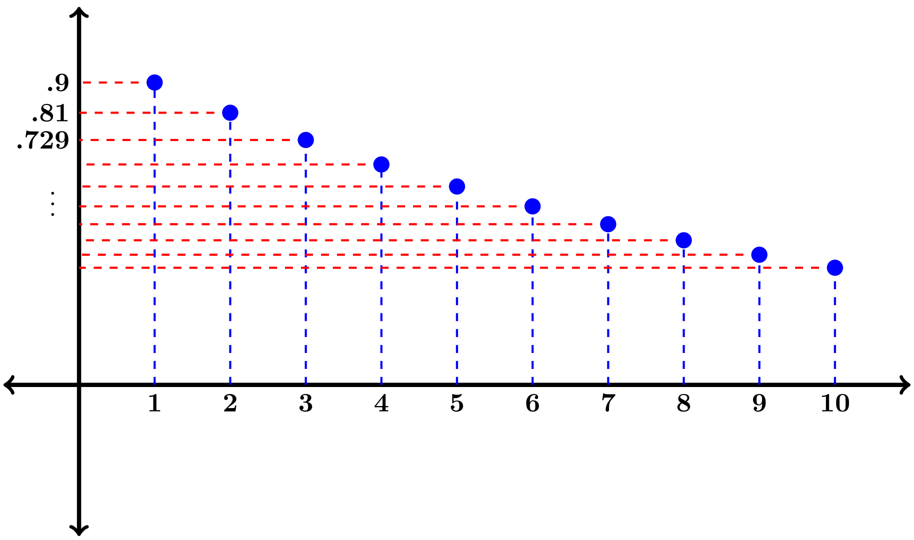
$$\begin{aligned}\mathbb{P}(X > x) &= \mathbb{P}(F^{-1}(u) > x) \\ &= \mathbb{P}(U > F(x)) \\ &= 1 - F(x).\end{aligned}$$

Example: Now we look at discrete random variables. Let X be the time of the first successful trial and let $p \in [0,1]$ be the probability of success. Then,

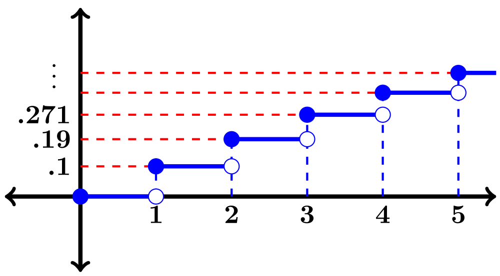
$$\mathbb{P}(X = k) = (1-p)^{k-1} p$$

for $k \ge 1$.

We get the density function:



The corresponding cdf:



Construction: Now we construct a probability measure from a <u>cumulative distribution function</u> . Let $F(x)$ be a cdf. Define

$$\mathbb{P}_F((-\infty, x]) := F(x).$$

Define the left-continuous inverse of F by

$$F^{-1}(y) := \inf\{x \mid F(x) \geq y\}$$

and define

$$A(y) := \{x \mid F(x) \geq y\}.$$

Suppose $A \in \mathcal{B}(\mathbb{R})$. Then, let

$$\xi_F(A) := \{u \in (0, 1] \mid F^{-1}(u) \in A\}.$$

By **Lemma 2.5.1**, if $A \in \mathcal{B}(\mathbb{R})$, then $\xi_F(A) \in \mathcal{B}((0, 1])$.

The complete definition of $\mathbb{P}_F$ is given by $\mathbb{P}_F := \lambda \circ \xi_F$ with

$$A \mapsto \lambda(\xi_F(A)).$$

Chapter 3

Chapter 3 - Random Variables

Section 3.0 - Motivation for Random Variables

Colloquially, what is a random variable? It's a characterization of a random outcome (as opposed to a cold census).

Example: (Yahtzee!) A 5-of-a-kind when rolling fair dice is called a "Yahtzee!". We will construct a random variable for this event. Our universe of rolls is

$$\Omega := \{1, 2, 3, 4, 5, 6\}^5.$$

We write a roll $w \in \Omega$ as $w = (w_1, w_2, w_3, w_4, w_5)$. We define

$$\mathcal{B} := \mathcal{P}(\Omega)$$

and define $\mathbb{P}$ on Ω by $\mathbb{P}(w) = \frac{1}{6^5}$ for all $w \in \Omega$. Now we define the random variable:

$$Y := \begin{cases} 1, & \text{if } w_1 = w_2 = w_3 = w_4 = w_5 \\ 0, & \text{otherwise} \end{cases}.$$

Let E be the event $\{Y = 1\}$. Then,

$$\mathbb{P}(E) = (\text{\# of ways to Yahtzee!}) \cdot (\text{probability each happens}) = 6 \cdot \frac{1}{6^5} = \frac{1}{6^4}.$$

Now suppose that there are two observers with restricted points of view:

- Observer A can only see how many of the dice are odd.
- Observer B can only check to see if pairs of dice have equal value.

Observer B can determine the value of Y, but A will not necessarily be able to (A may be able to rule out a Yahtzee in some circumstances).

In the context of Chapter 2, $\sigma(B) \subsetneqq \mathcal{B}$, $\sigma(A) \subsetneqq \mathcal{B}$ and both $\sigma(B) \subsetneqq \sigma(A)$ and $\sigma(A) \subsetneqq \sigma(B)$.

These statements motivate a desire to want to write something like

$$\rho(Y) \subseteq \sigma(B), \quad \text{but} \quad \rho(Y) \not\subseteq \sigma(A)$$

where ρ is some structure generated by knowledge of Y.

Section 3.1 - Inverse Maps

Definition: Suppose Ω, R are two sets. (R will almost always be $\mathbb{R}$ and is always a "range" of a function.) Let $X : \Omega \to R$. Then, X determines an inverse set function

$$X^{-1} : \mathcal{P}(R) \to \mathcal{P}(\Omega)$$

defined by

$$X^{-1}(A) = \{w \in \Omega \mid X(w) \in A\}.$$

Proposition 3.1.1: If $\mathcal{B}$ is a σ-field of subsets of R, then $X^{-1}(\mathcal{B})$ is also a σ-field.

Proof: We verify the three properties.

(i) Note that $R \in \mathcal{B}$ and also $X^{-1}(R) = \Omega$. Hence, $\Omega \in X^{-1}(\mathcal{B})$. Also, $\emptyset \in \mathcal{B}$ and $X^{-1}(\emptyset) = \emptyset$ by definition. Hence, $\emptyset \in X^{-1}(\mathcal{B})$.

(ii) Suppose $A \in \mathcal{B}$. This means that $A^C \in \mathcal{B}$ as well. This implies that both

$$X^{-1}(A) \in X^{-1}(\mathcal{B}) \qquad \text{and} \qquad X^{-1}(A^C) \in X^{-1}(\mathcal{B}).$$

Now,

$$\begin{aligned} (X^{-1}(A))^C &= \{\omega \in \Omega \mid X(\omega) \in A\}^C \\ &= \{\omega \in \Omega \mid X(\omega) \in A^C\} \\ &= X^{-1}(A^C) \in X^{-1}(\mathcal{B}). \end{aligned}$$

(iii) Suppose that $\{A_n\}_{n\in\mathbb{N}}$ is a sequence in $\mathcal{B}$. Then, $\bigcup A_n \in \mathcal{B}$ since $\mathcal{B}$ is a σ-field. Now observe

$$\begin{aligned} \textbf{Definition: } \bigcup_{n\in\mathbb{N}} X^{-1}(A_n) &= \{\omega \in \Omega \mid X(\omega) \in A_n \text{ for some } n\} \\ &= \left\{\omega \in \Omega \;\middle|\; X(\omega) \in \bigcup_{n\in\mathbb{N}} A_n \right\} \\ &= X^{-1}\left(\bigcup_{n\in\mathbb{N}} A_n\right) \in \mathcal{B}. \end{aligned}$$

Hence $X^{-1}(\mathcal{B})$ is a σ-field. $\square$

Proposition 3.1.2: If $\mathcal{C}$ is a class of subsets of R, then

$$X^{-1}(\sigma(\mathcal{C})) = \sigma(X^{-1}(\mathcal{C})).$$

Proof: From **Proposition 3.1.1** we have that $X^{-1}(\sigma(\mathcal{C}))$ is a σ-field. Note that since $\mathcal{C} \subseteq \sigma(\mathcal{C})$, we have

$$X^{-1}(\mathcal{C}) \subseteq X^{-1}(\sigma(\mathcal{C})).$$

Therefore, by minimality,

$$\sigma(X^{-1}(\mathcal{C})) \subseteq X^{-1}(\sigma(\mathcal{C}))\}.$$

To show the reverse inclusion, define

$$\mathcal{F} := \{A \in \mathcal{P}(R) \mid X^{-1}(A) \in \sigma(X^{-1}(\mathcal{C})).$$

We claim that $\mathcal{F}$ is a σ-field. See the textbook or lecture notes for the proof of this claim.

By definition, $X^{-1}(\mathcal{F}) \subseteq \sigma(X^{-1}(\mathcal{C}))$. Furthermore $\mathcal{C} \subseteq \mathcal{F}$ and so $X^{-1}(\mathcal{C}) \subseteq X^{-1}(\mathcal{F})$. Because $\mathcal{F}$ is a σ-field, $\sigma(\mathcal{C}) \subseteq \mathcal{F}$. Hence by minimality,

$$X^{-1}(\sigma(\mathcal{C})) \subseteq X^{-1}(\mathcal{F}) \subseteq \sigma(X^{-1}(\mathcal{C})).$$

This completes the proof. □

Section 3.2 - Measurable Maps and Random Elements

Definition: If $(\Omega, \mathcal{F})$ and $(\mathcal{R}, \mathcal{B})$ are measure spaces and $X : \Omega \to \mathcal{R}$, we say that X is $\mathcal{F}$-measurable if

$$X^{-1}(\mathcal{B}) \subseteq \mathcal{F}.$$

In the special case where $(\mathcal{R}, \mathcal{B})$ is $\mathbb{R}$ with the Borel sets, then we say that X is a random variable.

Definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and suppose X is $\mathcal{F}$-measurable. For a set $A \in \mathcal{R}$, we define

$$[X \in A] := X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}.$$

Define the set function

$$(\mathbb{P} \circ X^{-1})(A) := \mathbb{P}(X^{-1}(A)).$$

Then, $\mathbb{P} \circ X^{-1}$ is a probability measure on $(\mathcal{R}, \mathcal{B})$ and is called the probability measure induced by X, or the distribution of X

Remark: To verify that this is a probability measure, we check:

(a) $(\mathbb{P} \circ X^{-1})(\mathcal{R}) = \mathbb{P}(X^{-1}(\mathcal{R})) = \mathbb{P}(\Omega) = 1$.

(b) $(\mathbb{P} \circ X^{-1})(A) \geq 0$ for all $A \in \mathcal{B}$ since $\mathbb{P}(E) \geq 0$ for all $E \in \mathcal{F}$.

(c) Suppose that $\{A_n\}_{n\in\mathbb{N}}$ is disjoint and in $\mathcal{B}$. Then,

$$\begin{aligned}(\mathbb{P} \circ X^{-1})\left(\bigsqcup_n A_n\right) &= \mathbb{P}\left(X^{-1}\left(\bigsqcup_n A_n\right)\right) \\ &= \mathbb{P}\left(\bigsqcup_n X^{-1}(A_n)\right) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(X^{-1}(A_n)) \\ &= \sum_{n=1}^{\infty} (\mathbb{P} \circ X^{-1})(A_n).\end{aligned}$$

Notation: We often write $(\mathbb{P} \circ X^{-1})(A) = \mathbb{P}[X \in A]$ or $\mathbb{P}\{X \in A\}$.

Example: (Yahtzee!) In the context of our earlier example using the game of Yahtzee!, we can write:

$$\begin{aligned}\mathbb{P}[Y = 1] &= \mathbb{P}(Y^{-1}(\{1\})) \\ &= \mathbb{P}\left(\bigsqcup_{i=1}^{6} (i, i, i, i, i)\right) \\ &= \sum_{i=1}^{6} \mathbb{P}((i, i, i, i, i)) \\ &= 6 \cdot \frac{1}{6^5} = \frac{1}{6^4}.\end{aligned}$$

Meanwhile, $\mathbb{P}[Y = 0] = 1 - \frac{1}{6^4}$.

The range measure space "$(\mathcal{R}, \mathcal{B}, \mathbb{P})$" is

$$(\{0,1\}, \{\emptyset, \{1\}, \{0\}, \{0,1\}\}, \mathbb{P} \circ Y^{-1}).$$

Proposition 3.2.1: (Test for Measurability) Suppose $X : \Omega \to \mathcal{R}$ where $(\Omega, \mathcal{F})$ and $(\mathcal{R}, \mathcal{B})$ are measure spaces. Let $\mathcal{C}$ be a class such that $\sigma(\mathcal{C}) = \mathcal{B}$. Then,

$$[X \text{ is } \mathcal{F}\text{-measurable}] \iff [X^{-1}(\mathcal{C}) \subseteq \mathcal{F}].$$

Proof:

($\Longleftarrow$) Suppose $X^{-1}(\mathcal{C}) \subseteq \mathcal{F}$. Then by minimality, $\sigma(X^{-1}(\mathcal{C})) \subseteq \mathcal{F}$. However, $X^{-1}(\sigma(\mathcal{C})) = X^{-1}(\mathcal{B})$ since we assume that $\sigma(\mathcal{C}) = \mathcal{B}$. Also, we proved that $X^{-1}(\sigma(\mathcal{C})) = \sigma(X^{-1}(\mathcal{C}))$ in **Proposition 3.1.2**. Hence

$$X^{-1}(\mathcal{B}) = \sigma(X^{-1}(\mathcal{C})) \subseteq \mathcal{F}. \ \square$$

($\Longrightarrow$) Suppose X is $\mathcal{F}$-measurable, i.e., $X^{-1}(\mathcal{B}) \subseteq \mathcal{F}$. Since $\mathcal{C} \subseteq \mathcal{B}$, we have

$$X^{-1}(\mathcal{C}) \subseteq X^{-1}(\mathcal{B}) \subseteq \mathcal{F}. \ \square$$

Corollary 3.2.1: $X : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable if $X^{-1}((-\infty, x]) = [X \leq x] \in \mathcal{F}$.

Example: (Yahtzee! with two dice) We have

$$\Omega = \{\omega = (\omega_1, \omega_2) \mid \omega_i \in \{1,2,3,4,5,6\}\}.$$

Define the random variable

$$Y := \begin{cases} 0, & \text{if } \omega_1 = \omega_2 \\ 1, & \text{otherwise} \end{cases}.$$

Observer E knows the number of even dice, so $E \in \{0,1,2\}$. E itself is a random variable. Then, we can define the σ-field generated by E as

$$\begin{aligned} \sigma(E) &= \sigma(E^{-1}(\mathcal{P}(\{0,1,2\}))) \\ &= \sigma(\{E^{-1}(\{0\}), E^{-1}(\{1\}), E^{-1}(\{2\})\}) \end{aligned}$$

Now, we calculate these sets:

$$\begin{aligned} E^{-1}(\{0\}) &= \{(1,1),(1,3),(1,5),(3,1),(3,3),(3,5),(5,1),(5,3),(5,5)\}, \\ E^{-1}(\{1\}) &= \{(2,1),(1,2),(2,3),(3,2),(2,5),(5,2),(4,1),(1,4),(4,3), \\ &\qquad (3,4),(4,5),(5,4),(6,1),(1,6),(6,3),(3,6),(6,5),(5,6)\}, \\ E^{-1}(\{2\}) &= \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}. \end{aligned}$$

The intersections of each pair of these is empty. Hence,

$$\sigma(E) = \sigma(E^{-1}(\mathcal{P}(\{0,1,2\})) = \{\emptyset, E^{-1}(\{0\}), E^{-1}(\{1\}), E^{-1}(\{2\}), E^{-1}(\{0\})^C, E^{-1}(\{1\})^C, E^{-1}(\{2\})^C, \Omega).$$

Proposition 3.2.6: Let $X_1, X_2, \dots$ be random variables defined in $(\Omega, \mathcal{F})$. Then,

(i) $\inf_{n\in\mathbb{N}} X_n$ and $\sup_{n\in\mathbb{N}} X_n$ are random variables.

Proof: First note that

$$\begin{aligned}[\sup X_n \le x] &= \{\omega \in \Omega \mid \sup X_n(\omega) \le x\} \\ &= \bigcap_n \{\omega \in \Omega \mid X_n(\omega) \le x\}.\end{aligned}$$

This is a countable intersection of sets in the σ-field $\mathcal{F}$, and hence the intersection itself is in $\mathcal{F}$. Therefore,

$$[\sup X_n \le x] \in \mathcal{F}.$$

The proof for the infimum is analogous. □

(ii) $\liminf_{n\to\infty} X_n$ and $\limsup_{n\to\infty} X_n$ are random variables.

Proof: $\liminf_{n\to\infty} X_n = \sup_{n\in\mathbb{N}} \left(\inf_{k\ge n} X_n \right)$. □

(iii) If $\lim_{n\to\infty} X_n(\omega)$ exists for all $\omega \in \Omega$, then $\lim_{n\to\infty} X_n$ is a random variable.

Proof: If $\lim_{n\to\infty} X_n(\omega)$ exists, then $\lim_{n\to\infty} X_n(\omega) = \liminf_{n\to\infty} X_n(\omega)$. □

(iv) The set on which $\{X_n\}$ has a limit is $\mathcal{F}$-measurable, i.e.,

$$\{\omega \mid \lim_{n\to\infty} X_n(\omega) \text{ exists}\} \in \mathcal{F}.$$

Proof: Let $\mathbb{Q}$ be the set of all rationals. Then,

$$\begin{aligned}\{\omega \mid \lim_{n\to\infty} X_n(\omega) \text{ exists}\}^C &= \{\omega \mid \liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n\} \\ &= \bigcup_{r\in\mathbb{Q}} \left[\liminf_{n\to\infty} X_n \le r < \limsup_{n\to\infty} X_n\right] \\ &= \bigcup_{r\in\mathbb{Q}} \left(\left[\liminf_{n\to\infty} X_n \le r\right] \cap \left[\limsup_{n\to\infty} X_n > r\right] \right). \ \square\end{aligned}$$

Chapter 4

Chapter 4 - Independence

Section 4.0 - Motivation for Independence

The motivation for the definition of independence comes from conditional probabilities, i.e., the formula

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

The intuitive notion of independence says that if A is independent of B, then the probability of A happening in the B-universe should be equal to the probability of A happening in the Ω-universe, i.e.,

$$\mathbb{P}(A) = \frac{\mathbb{P}(A)}{\mathbb{P}(\Omega)} = \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Cross-multiplying, we see that if this occurs, then

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

Section 4.1 - Basic Definitions

Definition 4.1.2: Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. The events $A_1, \ldots, A_n \in \mathcal{F}$ are defined to be independent if

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i)$$

for all finite $I \subseteq \{1, \ldots, n\}$. Note that this is actually

$$\sum_{k=2}^{n} \binom{n}{k} = 2^n - n - 1$$

equations to check, all collections of 2 or more elements of $\mathcal{F}$.

Theorem 4.1.1: (Basic Criterion) If for each $i \in [n]$, $\mathcal{C}_i$ is a class of events satisfying

(i) $\mathcal{C}_i$ is a π-system,

(ii) $\{\mathcal{C}_i\}_{i=1}^n$ is independent,

then $\sigma(\mathcal{C}_1), \ldots, \sigma(\mathcal{C}_n)$ are independent.

Remark: A set of classes is considered to be independent if for any choices $\{A_i \in \mathcal{C}_i\}$, we have that the events $\{A_i\}$ are independent.

Proof: We show the $n = 2$ case. Fix an event $A_2 \in \mathcal{C}_2$ and define

$$\mathscr{L} := \{A \in \mathcal{F} \mid \mathbb{P}(A \cap A_2) = \mathbb{P}(A)\mathbb{P}(A_2)\}.$$

We now show that $\mathscr{L}$ is a λ-system.

(λ_1) $\mathbb{P}(\Omega \cap A_2) = \mathbb{P}(A_2) = \mathbb{P}(\Omega)\mathbb{P}(A_2)$. So, $\Omega \in \mathscr{L}$.

(λ_2) Suppose $A \in \mathscr{L}$. Then,

$$\begin{aligned}
\mathbb{P}(A^C \cap A_2) &= 1 - \mathbb{P}((A \cap A_2) \sqcup A_2^C) \\
&= 1 - (\mathbb{P}(A \cap A_2) + \mathbb{P}(A_2^C)) \\
&= 1 - (\mathbb{P}(A)\mathbb{P}(A_2) + 1 - \mathbb{P}(A_2)) && \text{(since } A \in \mathscr{L}\text{)} \\
&= \mathbb{P}(A_2) - \mathbb{P}(A)\mathbb{P}(A_2) \\
&= \mathbb{P}(A_2)(1 - \mathbb{P}(A)) \\
&= \mathbb{P}(A_2)\mathbb{P}(A^C).
\end{aligned}$$

Hence $A^C \in \mathscr{L}$.

(λ_3) Suppose $B_1, \ldots, B_n \in \mathcal{F}$ are disjoint. Then.

$$\begin{aligned}
\mathbb{P}\left(\left(\bigsqcup_i B_i\right) \bigcap A\right) &= \mathbb{P}\left(\bigsqcup_i (B_i \cap A)\right) \\
&= \sum_i \mathbb{P}(B_i \cap A) \\
&= \sum_i \mathbb{P}(B_i)\mathbb{P}(A) \\
&= \mathbb{P}\left(\bigsqcup_i B_i\right)\mathbb{P}(A).
\end{aligned}$$

Hence $\mathscr{L}$ is a λ-system. Note that $\mathcal{C}_1 \subseteq \mathscr{L}$. Therefore, $\sigma(\mathcal{C}_1) \subseteq \mathscr{L}$, i.e., $\sigma(\mathcal{C}_1), A_2$ are independent. Since A_2 was arbitrary, $\sigma(\mathcal{C}_1), \mathcal{C}_2$ are independent. Reversing the entire argument, $A_1, \sigma(\mathcal{C}_2)$ are independent for all $A_1 \in \sigma(\mathcal{C}_1)$. Combining, we get $\sigma(\mathcal{C}_1), \sigma(\mathcal{C}_2)$ are independent.

Proceed by induction to prove cases where $n > 2$. □

Definition 4.1.4: Let T be an arbitrary index set. The classes $\{\mathcal{C}_t\}_{t \in T}$ are independent families if for each finite $I \subseteq T$, the set $\{\mathcal{C}_t\}_{t \in I}$ is independent.

Corollary 4.1.1: If $\{\mathcal{C}_t\}_{t \in T}$ are nonempty π-systems that are independent, then $\{\sigma(\mathcal{C}_t)\}_{t \in T}$ are independent.

Proof: Proceed as in the above theorem. □

Section 4.2 - Independent Random Variables

Definition 4.2.1: $\{X_t\}_{t\in T}$ is an independent family of random variables if $\{\sigma(X_t)\}_{t\in T}$ are independent σ-fields. Recall that $\sigma(X_t) = \sigma(X^{-1}(\mathcal{B}(\mathbb{R})))$ where $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Remark: An important concept related to this section is finite dimensional distributions:

$$F_J\left(\{x_t\}_{t\in T}\right) = \mathbb{P}\{X_t \le x_t \ \forall t \in J\}$$

for all finite subsets $J \subseteq T$.

Theorem 4.2.1: (Factorization Condition) A family of random variables $\{X_t\}_{t\in T}$ is independent if and only if for all finite $J \subseteq T$

$$F_J\left(\{x_t\}_{t\in T}\right) = \prod_{t\in J} \mathbb{P}\{X_t \le x_t\} \qquad (\bigstar)$$

for all $x_t \in \mathbb{R}$.

Proof: By **Definition 4.1.4**, it suffices to show for finite index sets J that $\{X_t\}_{t\in J}$ is independent if and only if ($\bigstar$) holds. Define

$$\mathcal{C}_t := \big\{\{\omega \in \Omega \mid X_t(\omega) \le x\} \mid x \in \mathbb{R}\big\} = \{[X_t \le x] \mid x \in \mathbb{R}\}.$$

(i) We claim that $\mathcal{C}_t$ is a π-system:

$$[X_t \le x] \cap [X_t \le y] = [X_t \le \min(\{x, y\})] \in \mathcal{C}_t.$$

(ii) Additionally, $\sigma(\mathcal{C}_t) = \sigma(X_t)$.

($\bigstar$) is sufficient to imply that $\{\mathcal{C}_t\}_{t\in J}$ is an independent family. By the "Basic Criterion" in **Theorem 4.1.1**, $\{\sigma(\mathcal{C}_t)\}_{t\in J}$ is independent. □

Corollary: A finite collection of random variables $\{X_i\}_{i=1}^k$ is independent if and only if

$$\mathbb{P}\{X_1 \le x_1,\ X_2 \le x_2,\ \cdots,\ X_k \le x_k\} = \prod_{i=1}^k \mathbb{P}\{X_i \le x_i\}.$$

Section 4.3 - Ranks, Records, Renyi's Theorem

Let $\{X_n\}_{n\ge1}$ be iid with a common continuous cdf $F(x)$.

Lemma: There can be no ties (almost surely), where

$$\{\texttt{ties}\} = \bigcup_{i\neq j}\{X_i = X_j\}.$$

Proof: By subadditivity:

$$\mathbb{P}\{X_i = X_j \text{ for some } i, j\} \le \sum_{i\neq j} \mathbb{P}\{X_i = X_j\} = 0.$$

We will show the equality with zero above momentarily, which will then complete the theorem. □

Definition: Call X_n a <u>record</u> if
$$X_n > \max_{i \in [n-1]} X_i.$$

Warnings: As always, there are hidden ω's. The correct interpretation is
$$A_n(\omega) := \begin{cases} 1, & X_n(\omega) > X_i(\omega)\ \forall i \in [n-1] \\ 0, & \text{otherwise} \end{cases}.$$

This is the official definition associated with Resnick's "$A_n = \{X_n \text{ is a record}\}$".

Remark: Renyi's first result says that:
$$\mathbb{P}\{A_n = 1\} = \frac{1}{n}.$$
Furthermore, the family $\{A_n\}_{n \geq 1}$ is independent. This is a slightly surprising result. For example, the fact that the 10^{th} draw was a record has no bearing on whether or not the 11^{th} draw will be a record. We will prove this shortly.

Lemma: Suppose that X_1 and X_2 are independent with a common continuous cumulative distribution function F. Then,
$$\mathbb{P}\{X_1 = X_2\} = 0.$$

Proof: We start by relaxing the exact equality and bounding the probability above by the probability of landing in the same partition of width $\frac{1}{2^n}$.

$$\begin{aligned}
\mathbb{P}\{X_1 = X_2\} &\leq \sum_{k=-\infty}^{\infty} \mathbb{P}\left\{\frac{k-1}{2^n} < X_1 \leq \frac{k}{2^n},\ \ \frac{k-1}{2^n} < X_2 \leq \frac{k}{2^n}\right\} \\
&= \sum_{k=-\infty}^{\infty} \left[\left[F\left(\frac{k}{2^n}\right) - F\left(\frac{k-1}{2^n}\right)\right]\left[F\left(\frac{k}{2^n}\right) - F\left(\frac{k-1}{2^n}\right)\right]\right] \\
&\leq \sup_{k \in \mathbb{Z}} \left[F\left(\frac{k}{2^n}\right) - F\left(\frac{k-1}{2^n}\right)\right] \cdot \sum_{k=-\infty}^{\infty} \left[F\left(\frac{k}{2^n}\right) - F\left(\frac{k-1}{2^n}\right)\right] \\
&= \sup_{k \in \mathbb{Z}} \left[F\left(\frac{k}{2^n}\right) - F\left(\frac{k-1}{2^n}\right)\right] \qquad \text{(telescoping sum)}
\end{aligned}$$

Demanding that $\sup_{k \in \mathbb{Z}} \left[F\left(\frac{k}{2^n}\right) - F\left(\frac{k-1}{2^n}\right)\right]$ is bounded for all n is equivalent to the condition that F is uniformly continuous on $\mathbb{R}$. Recall that a function which is continuous on a compact set is uniformly continuous.

Claim: Suppose $f \in \mathcal{C}([0,\infty))$ and $\lim_{x \to \infty} f(x) = L < \infty$ exists. Then, $f \in \mathcal{UC}([0,\infty)$.

Proof: Say that $|x - y| < 1$, for convenience. Because $\lim_{x \to \infty} f(x) = L$, there exists M such that for all $x \geq M$, $|f(x) - L| < \epsilon/2$. So, for all $x, y \geq M$, $|f(x) - f(y)| \leq |f(x) - L| + |L - f(y)| < \epsilon$. Otherwise, $x, y \in [0, M+1]$, and since this is compact the function is uniformly continuous here. $\square$

Since F is continuous on $(-\infty, \infty)$ and $\lim_{x \to \pm\infty} F(x)$ each exist, we have that $F \in \mathcal{UC}((-\infty, \infty))$, i.e., given $\epsilon > 0$, there exists n such that for all $|x - y| < \frac{1}{2^n}$, $|F(x) - F(y)| < \epsilon$.

So, for any $\epsilon > 0$, $\mathbb{P}\{X_1 = X_2\} < \epsilon$, i.e., $\mathbb{P}\{X_1 = X_2\} = 0$. $\square$

Notation: Suppose $X_1, X_2, \ldots$ are iid random variables with common continuous cdf. For each n, let R_n be the relative rank of X_n among $\{X_1, \ldots, X_n\}$, i.e.,

$$R_n = \sum_{i=1}^{n} 1_{\{X_i \geq X_n\}}$$

so that $R_n = 1$ if X_n is a record.

Theorem 4.3.1: (Renyi's Theorem) The sequence of random variables $\{R_n\}_{n\geq 1}$ is independent, and

$$\mathbb{P}\{R_n = k\} = \frac{1}{n}$$

for all $k \in [n]$.

Proof: For (almost) all $\omega \in \Omega$, note that there are $n!$ orderings of $\{X_i(\omega)\}_{i=1}^n$. Because the values are iid, all orderings occur with the same probability: $\frac{1}{n!}$. (Remember, there are no ties, almost surely!)

Each realization $R_1, \ldots, R_n$ uniquely determines an ordering. For example, if for ω_1 we have

$$R_1(\omega_1) = 1 \quad R_2(\omega_1) = 1 \quad R_3(\omega_1) = 1$$

then we have

$$X_1(\omega_1) < X_2(\omega_1) < X_3(\omega_1).$$

If for ω_2 we have

$$R_1(\omega_2) = 1 \quad R_2(\omega_2) = 2 \quad R_3(\omega_2) = 3$$

then we have

$$X_1(\omega_2) > X_2(\omega_2) > X_3(\omega_2).$$

Of course, each ordering has probability $\frac{1}{n!}$. So, for any sequence $(r_1, \ldots, r_n)$, where $r_i \in [i]$, the probability that

$$\mathbb{P}\{R_1 = r_1, R_2 = r_2, \cdots, R_n = r_n\} = \frac{1}{n!}.$$

Now fix n. We will compute $\mathbb{P}\{R_n = r_n\}$.

$$\begin{aligned}
\mathbb{P}\{R_n = r_n\} &= \sum_{r_1, r_2, \cdots, r_{n-1}} \mathbb{P}\{R_1 = r_1, R_2 = r_2, \cdots, R_n = r_n\} \\
&= \sum_{r_1, r_2, \cdots, r_{n-1}} \frac{1}{n!} \\
&= (n-1)! \cdot \frac{1}{n!} \\
&= \frac{1}{n}.
\end{aligned}$$

Show the general argument with induction. □

Example: (Independence / Dyadic Expansions of Uniform Real Numbers) Define $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1], \mathcal{B}((0,1]), \lambda)$ where λ is the Lebesgue measure. For a given $\omega \in \Omega$, write

$$\omega = \sum_{n=1}^{\infty} \frac{1}{2^n} d_n(\omega) = 0.d_1(\omega) d_2(\omega) d_3(\omega) \cdots$$

where each $d_i(\omega) \in \{0, 1\}$. Note that with this representation

$$1 = \sum_{n=1}^{\infty} \frac{1}{2^n} = 0.1111\cdots,$$

and

$$\frac{1}{2} = 0.1 = 0.01111\cdots.$$

To overcome this, we only consider nonterminating sequences.

Fact 1: Each d_n is a random variable.

We need to show that the sets $\{d_n = 0\}$ and the set $\{d_n = 1\}$ are both in $\mathcal{B}((0, 1])$. (Recall, $\{d_n = 0\} = \{\omega \in \Omega \mid d_n(\omega) = 0\}$.) Since they are complements, we only need to show this for one of them. Observe that

$$\{d_1 = 1\} = \{\omega \in (0, 1] \mid d_1(\omega) = 1\}$$

i.e., ω has the form $\omega = 0.1d_2d_3d_4\cdots > \frac{1}{2}$. (The inequality is strict because we're ignoring terminating sequences.) So

$$\{d_1 = 1\} = \left(\frac{1}{2}, 1\right] \in \mathcal{B}((0, 1]).$$

Similarly, $\{d_2 = 1\}$ consists of ω of the form $0.d_1 1 d_4 d_4\cdots$, i.e.,

$$\{d_2 = 1\} = \left(\frac{1}{4}, \frac{1}{2}\right] \cup \left(\frac{3}{4}, 1\right] \in \mathcal{B}((0, 1]).$$

We see now that for arbitrary n, the set $\{d_n = 1\}$ is measurable. This proves the fact.

Fact 2: For all n, $\mathbb{P}\{d_n = 1\} = \lambda(\{\omega \in \Omega \mid d_n(\omega) = 1\}) = \frac{1}{2}$. So, the d_n are identically distributed.

Fact 3: $\{d_n\}_{n\geq 1}$ is an independent sequence.

It suffices to show independence for $\{d_i\}_{i=1}^n$. We need to calculate

$$\mathbb{P}\left\{\bigcap_{i=1}^{n} \{d_i = u_i\}\right\}$$

for a given vector $u = (u_1, u_2, \ldots, u_n) \in \{0, 1\}^n$. Note that saying $\omega \in \bigcap_{i=1}^{n} \{d_i = u_i\}$ means that ω has the form

$$\omega = 0.u_1u_2\cdots u_n d_{n+1} d_{n+2}\cdots \in (0.u_1u_2\cdots u_n 0000\cdots, \ 0.u_1u_2\cdots u_n 1111\cdots)$$

and this interval has size $\frac{1}{2^n}$.

$$\mathbb{P}\left\{\bigcap_{i=1}^{n} \{d_i = u_i\}\right\} = \frac{1}{2^n} = \prod_{i=1}^{n} \mathbb{P}\{d_i = u_i\}.$$

Section 4.5 - Independence, Zero-One Laws, Borel-Cantelli Lemma

Borel-Cantelli Lemma: Let $\{A_n\}_{n\geq 1}$ be a sequence of events in $\mathcal{F}$. If $\sum_{n=1}^{\infty} \mathbb{P}\{A_n\} < \infty$, then

$$\mathbb{P}\{A_n \text{ i.o.}\} = \mathbb{P}\{\limsup_{n\to\infty} A_n\} = 0.$$

Proof: Recall that using ωs, we are showing

$$\mathbb{P}\{\omega \in A_n \text{ i.o.}\} = \mathbb{P}\{\omega \in \limsup_{n\to\infty} A_n\} = 0.$$

We know that

$$\begin{aligned}
\mathbb{P}\{A_n \text{ i.o.}\} &= \mathbb{P}\big\{\omega \in \lim_{n\to\infty} \bigcup_{j\geq n} A_j\big\} && \text{(by definition)}\\
&= \lim_{n\to\infty} \mathbb{P}\big\{\bigcup_{j\geq n} A_j\big\} && \text{(by continuity of } \mathbb{P})\\
&\leq \lim_{n\to\infty} \sum_{j=n}^{\infty} \mathbb{P}(A_j).
\end{aligned}$$

We claim this sum is zero. Since $\sum_{j=1}^{\infty} \mathbb{P}\{A_j\} < \infty$, for all $\epsilon > 0$ there exists N such that for all $n > N$, $\sum_{j=n}^{\infty} \mathbb{P}(A_j) < \epsilon$. □

Example: Suppose $\{X_n\}_{n\geq 1}$ is a sequence of Bernoulli random variables where $p_n := \mathbb{P}\{X_n = 1\}$ satisfies $\sum_{n=1}^{\infty} p_n < \infty$. Then, $\mathbb{P}\{\lim_{n\to\infty} X_n = 0\} = 1$.

Proof:

$$\begin{aligned}
\liminf_{n\to\infty}\{X_n = 0\} &= \liminf_{n\to\infty}\{\omega \in \Omega \mid X_n(\omega) = 0\}\\
&= \{\omega \in \Omega \mid X_n(\omega) = 1 \text{ i.o.}\}^C\\
&= \{\omega \in \Omega \mid \exists N_\omega \text{ such that } \forall n > N_\omega,\ X_n(\omega) = 0\}\\
&= \{\omega \mid \lim_{n\to\infty} X_n(\omega) = 0\}.
\end{aligned}$$

Borel Zero-One Laws: If $\{A_n\}_{n\geq 1}$ is a sequence of independent events, then

$$\mathbb{P}\{A_n \text{ i.o.}\} = \begin{cases} 0, & \sum_{n=1}^{\infty} \mathbb{P}\{A_n\} < \infty \\ 1, & \sum_{n=1}^{\infty} \mathbb{P}\{A_n\} = \infty \end{cases}.$$

Proof: By the **Borel-Cantelli Lemma**, if $\sum_{n=1}^{\infty} \mathbb{P}\{A_n\} < \infty$, then $\mathbb{P}\{A_n \text{ i.o.}\} = 0$.

Conversely, suppose $\sum_{n=1}^{\infty} \mathbb{P}\{A_n\} = \infty$. Well,

$$\begin{aligned}
\mathbb{P}\{A_n \text{ i.o.}\} &= \mathbb{P}\{\omega \mid \omega \in \limsup_{n\to\infty} A_n\} \\
&= 1 - \mathbb{P}\{\liminf_{n\to\infty} A_n^C\} && \text{(we are leaving out the } \omega\text{s now)} \\
&= 1 - \mathbb{P}\left\{ \lim_{n\to\infty} \bigcap_{k\geq n} A_k^C \right\} \\
&= 1 - \lim_{n\to\infty} \mathbb{P}\left\{ \bigcap_{k\geq n} A_k^C \right\} && \text{(since the measure is finite)} \\
&= 1 - \lim_{n\to\infty} \mathbb{P}\left\{ \lim_{m\to\infty} \bigcap_{k=n}^{m} A_k^C \right\} \\
&= 1 - \lim_{n\to\infty} \lim_{m\to\infty} \mathbb{P}\left\{ \bigcap_{k=n}^{m} A_k^C \right\} \\
&= 1 - \lim_{n\to\infty} \lim_{m\to\infty} \prod_{k=n}^{m} \mathbb{P}(A_k^C) && \text{(by independence)} \\
&= 1 - \lim_{n\to\infty} \lim_{m\to\infty} \prod_{k=n}^{m} \left[1 - \mathbb{P}(A_k)\right].
\end{aligned}$$

Note that for large k, $\mathbb{P}\{A_k\}$ is < 1, so we use the approximation $e^{-x} \approx 1 - x$. Recall that

$$e^{-x} = 1 - x + \frac{x^2}{2} - \frac{x^3}{3} + \cdots$$

and so in fact for $x \in (0, 1)$,

$$1 - x \leq e^{-x} \leq 1 - x + \frac{x^2}{2}.$$

Claim: $e^{-x} \geq 1 - x$ for all $x \in [0, 1]$

Proof: By the Mean Value Theorem, for any $x \in (0, 1)$, there exists $c \in (0, 1)$ such that

$$\frac{e^{-x} - e^{-0}}{x - 0} = -e^{-c}.$$

So, $e^{-}x = 1 - xe^{-c} > 1 - x$.

Now,

$$\begin{aligned}
\lim_{m\to\infty} \prod_{k=n}^{m} (1 - \mathbb{P}\{A_n\}) &\leq \lim_{m\to\infty} \prod_{k=n}^{m} e^{-\mathbb{P}\{A_k\}} \\
&= \lim_{m\to\infty} \exp\left(-\sum_{k=n}^{\infty} \mathbb{P}\{A_k\} \right) \\
&= \exp\left(-\sum_{k=n}^{\infty} \mathbb{P}\{A_k\} \right) \\
&= 0.
\end{aligned}$$

This completes the proof. □

Example: (Behavior of Exponential Random Variables) Let $\{E_n\}_{n\geq 1}$ be a sequence of independent and identically distributed random variables on exp(1), i.e., for all n, $\mathbb{P}\{E_n > t\} = e^{-t}$ for $t > 0$. Then,

$$\limsup_{n\to\infty} \frac{E_n}{\log(n)} = 1.$$

Proof: Translate the statement: For a given $\omega \in \Omega$, $\limsup_{n\to\infty} \frac{E_n(\omega)}{\log(n)} = 1$ means that

$$\forall \epsilon > 0,\ \exists N_\omega \text{ such that } \frac{E_n(\omega)}{\log(n)} < 1 + \epsilon,$$

and, because of the equality to 1,

$$\forall M,\ \exists n > M \text{ such that } \frac{E_n(\omega)}{\log(n)} > 1 - \epsilon.$$

So,

$$1 - \epsilon < \frac{E_n(\omega)}{\log(n)} < 1 + \epsilon$$

for the first inequality happens infinitely often and the second inequality happens eventually always. This is the statement that we will prove.

First we show the left-hand inequality, i.e., that

$$\mathbb{P}\left\{\frac{E_n}{\log(n)} > 1 - \epsilon \quad \text{i.o.}\right\} = 1.$$

To apply the Borel Zero-One Law, we compute

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{\frac{E_n}{\log(n)} > 1 - \epsilon\right\}$$

and show that this is infinite. Well,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}\left\{\frac{E_n}{\log(n)} > 1 - \epsilon\right\} &= \sum_{n=1}^{\infty} \mathbb{P}\left\{E_n > (1-\epsilon)\log(n)\right\} \\
&= \sum_{n=1}^{\infty} e^{-(1-\epsilon)\log(n)} \\
&= \sum_{n=1}^{\infty} \frac{1}{n^{1-\epsilon}} = \infty.
\end{aligned}$$

On the other hand,

$$\mathbb{P}\left\{\frac{E_n}{\log(n)} > 1 + \epsilon\right\} = 0$$

since by a similar argument

$$\sum_{n=1}^{\infty} \frac{1}{n^{1+\epsilon}} < \infty.$$

After some bookkeeping which can be found in the textbook, this proves the theorem. □

Definition: Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables and define

$$\mathcal{F}'_n := \sigma(X_{n+1}, X_{n+2}, \ldots)$$

where $n \in \mathbb{N}$. The tail σ-field $\mathcal{T}$ is defined by

$$\mathcal{T} := \bigcap_{n=0}^{\infty} \mathcal{F}'_n.$$

Examples:

(1) $\left\{\omega \;\middle|\; \sum_{n=1}^{\infty} X_n(\omega) \text{ converges}\right\} \in \mathcal{T}$. Why is this true? For all m,

$$\sum_{n=1}^{\infty} X(\omega) \text{ converges if and only if } \underbrace{\sum_{n=m}^{\infty} X_n(\omega) \text{ converges.}}_{\in \mathcal{F}'_m}$$

(2) $\limsup_{n\to\infty} X_n$, $\liminf_{n\to\infty} X_n$, $\left\{\omega \;\middle|\; \lim_{n\to\infty} X_n(\omega) \text{ exists}\right\} \in \mathcal{T}$.

(3) (Important!) Define $S_n := X_1 + X_2 + \cdots + X_n$. Then,

$$\left\{\omega \;\middle|\; \lim_{n\to\infty} \frac{S_n(\omega)}{n} = 0\right\} \in \mathcal{T}.$$

Definition: A σ-field whose events all have probability 0 or 1 is called almost trivial.

Lemma 4.5.1: Let $\mathcal{G}$ be an almost trivial σ-field and let X be a random variable such that $X \in \mathcal{G}$. Then, there exists c such that $\mathbb{P}\{X = c\} = 1$.

Proof: See book.

Theorem 4.5.3: (Kolmogorov Zero-One Law) Let $\{X_n\}_{n\in\mathbb{N}}$ be independent random variables with tail σ-field $\mathcal{T}$. Then, for any $A \in \mathcal{T}$,

$$\mathbb{P}\{A\} \in \{0, 1\}.$$

Proof: Suppose $A \in \mathcal{T}$. We claim that A is independent of itself. This immediately implies that $\mathbb{P}(A) \in \{0, 1\}$ and the theorem follows. We now show that A is independent of itself.

Define

$$\mathcal{F}_n := \sigma(X_1, X_2, \ldots, X_n) = \bigvee_{j=1}^{n} \sigma(X_j)$$

and

$$\mathcal{F}_\infty := \sigma(X_1, X_2, \ldots) = \bigvee_{j=1}^{\infty} \sigma(X_j).$$

Note that $A \in \mathcal{T} \subseteq \mathcal{F}'_n$ for any n. In particular,

$$A \subseteq \mathcal{T} \subseteq \mathcal{F}'_1 = \mathcal{F}_\infty.$$

Since the X_n's are independent,

$$\mathcal{F}_n \perp \mathcal{F}'_n.$$

Therefore, $A \perp \mathcal{F}_n$ (since $A \in \mathcal{F}'_n$). This is true for all n. Hence,

$$A \perp \bigcup_{n=1}^{\infty} \mathcal{F}_n.$$

Even though $\bigcup_{n=1}^{\infty} \mathcal{F}_n \neq \mathcal{F}_\infty$, we do have that $\sigma\left(\bigcup_{n=1}^{\infty} \mathcal{F}_n\right) = \mathcal{F}_\infty$.

Let $\mathcal{C}_1 := \{A\}$ and let $\mathcal{C}_2 := \bigcup_{n=1}^{\infty} \mathcal{F}_n$. Note that $\mathcal{C}_1 \perp \mathcal{C}_2$. Recall a basic criterion for independence:

$$\sigma(\mathcal{C}_1) \perp \sigma(\mathcal{C}_2) \text{ if } \mathcal{C}_1 \text{ and } \mathcal{C}_2 \text{ are } \pi\text{-systems.}$$

But, $A \in \sigma(\{A\})$ and $A \in \mathcal{F}_\infty = \sigma(\mathcal{C}_2)$. So, $A \perp A$, which completes the theorem. $\square$

Chapter 5

Chapter 5 - Integration and Expectation

Section 5.1 - Preparation for Integration

Definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We say that $X : \Omega \to \mathbb{R}$ is a simple function (or a simple random variable) if it has finite range. We can write

$$X(\omega) = \sum_{i=1}^{k} a_i 1_{\{A_i\}}$$

where $a_i \in \mathbb{R}$, $A_i \in \mathcal{F}$, $\{A_i\}$ are disjoint, and $\bigsqcup_{i=1}^{k} A_i = \Omega$.

Remark: Let $\mathcal{E}$ be the space of all simple random variables. We claim that $\mathcal{E}$ is a vector space. We check the conditions.

- Closure under scalar multiplication: If

  $$X = \sum_{i=1}^{k} a_i 1_{\{A_i\}}$$

  then

  $$\alpha X = \sum_{i=1}^{k} \alpha a_i 1_{\{A_i\}} \in \mathcal{E}.$$

- Closure under addition: If

  $$Y = \sum_{j=1}^{\ell} b_j 1_{\{B_j\}}$$

  then

  $$(X + Y)(\omega) = \sum_{i=1}^{k} \sum_{j=1}^{\ell} (a_i + b_j) 1_{\{A_i \cap B)j\}}(\omega) \in \mathcal{E}.$$

Additionally, note that if $X, Y \in \mathcal{E}$, then $XY \in \mathcal{E}$. Also, $X \vee Y \in \mathcal{E}$ and $X \wedge Y \in \mathcal{E}$.

Theorem 5.1.1: (Measurability) Suppose $X(\omega) \geq 0$ for all $\omega \in \Omega$. Then, X is $\mathcal{F}$-measurable if and only if there exist simple random variables $\{X_n\} \subseteq \mathcal{E}$ such that

$$0 \leq \lim_{n\to\infty} X_n \nearrow X.$$

Proof: Let $X_n \in \mathcal{E}$ such that $0 \leq \lim_{n\to\infty} X_n \nearrow X$. Then, $X_n \in \mathcal{F}$ (since $\sigma(X_n) = \sigma(A_1^n, \ldots, A_k^n)$, where each $A_i^n \in \mathcal{F}$). If $X_n \nearrow X$, then X is $\mathcal{F}$-measurable as well, by the results in **Section 3.2**.

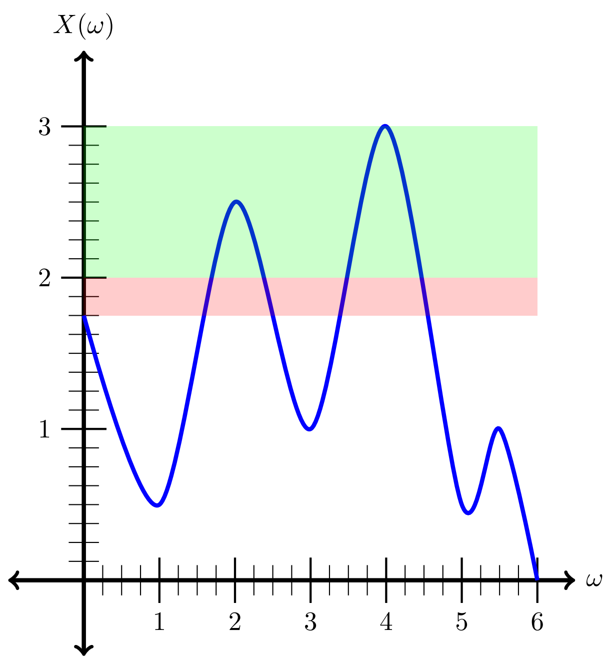
Conversely, suppose X is $\mathcal{F}$-measurable. Define,

$$X_n := \sum_{k=1}^{n2^n} \left(\frac{k-1}{2^n}\right) 1_{\left\{\frac{k-1}{2^n} \leq X < \frac{k}{2^n}\right\}} + n1_{\{X\geq n\}}.$$

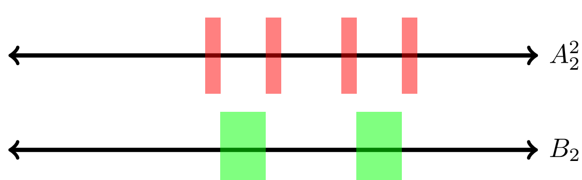
X_n is $\mathcal{F}$-measurable by construction since both

$$\left\{\frac{k-1}{2^n} \leq X < \frac{k}{2^n}\right\} \qquad \text{and} \qquad \{X \geq n\}$$

are. Consider the following picture, with $\Omega = [0, \infty)$.



Let $A_k^n = \left\{\frac{k-1}{2^n} \leq X < \frac{k}{2^n}\right\}$ and $B_n = \{X \geq n\}$. Then,using our picture:



For any given $\omega \in \Omega$,

$$X_n(\omega) \leq X_{n+1}(\omega) \leq X(\omega).$$

If $X(\omega) < \infty$, then $\lim_{n\to\infty} X_n(\omega)$ exists. Furthermore,

$$|X(\omega) - X_n(\omega)| < \frac{1}{2^n}$$

for sufficiently large n. So,

$$\lim_{n\to\infty} X_n(\omega) = X(\omega).$$

If $X(\omega) = \infty$, then

$$\lim_{n\to\infty} X_n(\omega) = \infty.$$

This completes the proof. □

Remark: Note further that if $\sup_{\omega\in\Omega} X_(\omega) < \infty$, then $\sup_{\omega\in\Omega} |X_n(\omega) - X(\omega)| \to 0$.

Section 5.2 - Expectation and Integration

The undergraduate version of expectation is the formula:

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \, dP(x).$$

For us:

$$\mathbb{E}[X] := \sum_{i=1}^{k} a_i \mathbb{P}\{A_i\}.$$

Example: $\mathbb{E}[c] = c\mathbb{P}\{\Omega\} = 1$ and

$$\mathbb{E}[1_A] = 1 \cdot \mathbb{P}\{A\} + 0 \cdot \mathbb{P}\{A^C\} = \mathbb{P}\{A\}.$$

Remark: Expectation is linear. Recall that

$$\alpha X + \beta Y = \sum_{i,j} (\alpha a_i + \beta b_j) 1_{\{A_i \cap B_j\}}.$$

So,

$$\begin{aligned}
\mathbb{E}[\alpha X + \beta Y] &= \sum_{i,j} (\alpha a_i + \beta b_j) \mathbb{P}\{A_i \cap B_j\} \\
&= \sum_{i,j} \alpha a_i \mathbb{P}\{A_i \cap B_j\} + \sum_{i,j} \beta b_j \mathbb{P}\{A_i \cap B_j\} \\
&= \alpha \sum_i a_i \sum_j \mathbb{P}\{A_i \cap B_j\} + \beta \sum_j b_j \sum_i \mathbb{P}\{A_i \cap B_j\} \\
&= \alpha \sum_i a_i \mathbb{P}\{A_i\} + \beta \sum_j b_j \mathbb{P}\{B_j\} \\
&= \alpha \mathbb{E}[X] + \beta \mathbb{E}[X].
\end{aligned}$$

Remark: A random variable can have $\mathbb{P}\{X < \infty\} < 1$. For example, consider the return time of a Brownian motion in 3 dimensions to a finite sphere. Note that this allows that the cdf of a probability distribution not approach 1 as $x \to \infty$.

Remark: A random variable can have $\mathbb{P}\{X < \infty\} = 1$ yet $\mathbb{E}(X) = \infty$. An example of this is the Cauchy distribution. The density function of a Cauchy random variable is

$$f(x) = \frac{c}{1 + x^2}.$$

Then,

$$\mathbb{E}[X] = 2\int_0^\infty \frac{cx}{1+x^2}\,dx = 2\int_1^\infty \frac{c}{u}\,du = 2[\ln(u)]_{u=1}^\infty = \infty,$$

where $u = 1 + x^2$ and $du = 2x\,dx$.

Notation: Let $\mathcal{E}_+$ be the set of nonnegative simple random variables and define

$$X_+ := \{X : \Omega \to \overline{\mathbb{R}} \mid X \geq 0,\ \sigma(X) \in \mathcal{F}\}$$

where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Construction: Let $X \in X_+$ and suppose $\mathbb{P}\{X = \infty\} > 0$. Then define $\mathbb{E}[X] = \infty$.

Now suppose $\mathbb{P}\{X = \infty\} = 0$. By the measurability theorem above, there exists a sequence of simple nonnegative nondecreasing random variables $\{X_n\}_{n\geq 1} \subset \mathcal{E}_+$ such that

$$\lim_{n\to\infty} X_n = X.$$

By monotonicity of expectation, $\{\mathbb{E}[X_n]\}_{n\geq 1}$ is a nondecreasing sequence of reals, and so $\lim_{n\to\infty} \mathbb{E}[X_n]$ exists in $\overline{\mathbb{R}}_+$. So, it is unambiguous to define

$$\mathbb{E}[X] := \lim_{n\to\infty} \mathbb{E}[X_n]$$

where $\mathbb{E}[X] \in \overline{\mathbb{R}}_+$.

Example: For the Cauchy distribution, construct an approximating sequence X_n as in the above measurability theorem and compute $\mathbb{E}[X_n]$.

Proposition 5.2.1: The operator $\mathbb{E}$ is well defined on X_+ in the sense that if $X_n \nearrow X$ and $Y_n \nearrow X$, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \lim_{n\to\infty} \mathbb{E}[Y_n].$$

Proof: It suffices to show that if

$$\lim_{n\to\infty} X_n \leq \lim_{n\to\infty} Y_n$$

then

$$\lim_{n\to\infty} \mathbb{E}[X_n] \leq \lim_{n\to\infty} \mathbb{E}[Y_n].$$

Fix $n \in \mathbb{N}$ for the moment. For any $m \in \mathbb{N}$, we have $X_n \wedge Y_m \in \mathbb{E}_+$. Additionally, note that

$$\lim_{m\to\infty} X_n \wedge Y_m = X_n$$

(since $\lim_{m\to\infty} Y_m \geq \lim_{n\to\infty} X_n \geq X_n$). By monotonicity,

$$\mathbb{E}[X_n] = \lim_{m\to\infty} \mathbb{E}[X_n \wedge Y_m] \leq \lim_{m\to\infty} \mathbb{E}[Y_m].$$

This is true for all n¿ So,

$$\lim_{n\to\infty} E[X_n] \leq \lim_{m\to\infty} \mathbb{E}[Y_m].$$

Symmetry gives the reverse inequality. □

Basic Properties: Suppose $X \in X_+$. Then we have,

(1) $0 \leq \mathbb{E}[X] \leq \infty$,

(2) linearity,

(3) monotone convergence theorem.

Proof: Suppose $\{X_n\}_{n\geq 1}$ is an ω-by-ω increasing sequence such that

$$\lim_{n\to\infty} X_n = X.$$

Then,

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X],$$

or, as Resnick says,

$$\mathbb{E}\left[\lim_{n\to\infty} \nearrow X_n\right] = \lim_{n\to\infty} \nearrow \mathbb{E}[X_n].$$

For each n, there exists a sequence $\{Y_m^{(n)}\}_{m\geq 1} \subseteq \mathcal{E}_+$ such that

$$Y_m^{(n)} \nearrow X_n.$$

We need to construct a sequence $\{Z_m\}_{m\geq 1} \subset \mathcal{E}_+$ such that

$$Z_m \nearrow X.$$

To this end, define

$$X_m := \bigvee_{n\leq m} Y_m^{(n)}.$$

Observe that $\{Z_n\}$ is indeed increasing. For $n \leq m$:

$$\begin{array}{ccccccc}
Y_m^{(n)} & \leq & Z_m & \leq & X_m & & \\
\downarrow & & \downarrow & & \downarrow & & m\to\infty \\
X_n & \leq & \lim_{m\to\infty} Z_m & \leq & \lim_{m\to\infty} X_m & & \\
\downarrow & & \downarrow & & \downarrow & & n\to\infty \\
X = \lim_{n\to\infty} X_n & \leq & \lim_{n\to\infty} X_n & \leq & \lim_{n\to\infty} X_n = X & &
\end{array}$$

Also,

$$\begin{aligned}
\mathbb{E}[X_n] &= \lim_{m\to\infty} \mathbb{E}[Y_m^{(n)}] \\
&\leq \lim_{m\to\infty} \mathbb{E}[Z_m] \\
&\leq \lim_{m\to\infty} \mathbb{E}[X_m].
\end{aligned}$$

Taking $n\to\infty$,

$$\lim_{n\to\infty} \mathbb{E}[X_n] \leq \lim_{m\to\infty} \mathbb{E}[Z_m] \leq \lim_{m\to\infty} E[X_m].$$

Hence,

$$\mathbb{E}[X] = \mathbb{E}\left[\lim_{m\to\infty} Z_m\right] = \lim_{m\to\infty} \mathbb{E}[Z_m] = \lim_{m\to\infty} \mathbb{E}[X_m]. \ \square$$

Definition: Given an $\mathcal{F}$-measurable random variable X, define

$$X_+ := X \vee 0$$

and

$$X_- := -(X \wedge 0).$$

Note that $|X| = X_+ + X_-$. If $\mathbb{E}[X_+] < \infty$ and $\mathbb{E}[X_-] < \infty$ then we say X is integrable ($X \in L_1(\mathbb{P})$) and

$$\mathbb{E}[X] := \mathbb{E}[X_+] - \mathbb{E}[X_-].$$

Remark: If only one of $\mathbb{E}[X_+]$ and $\mathbb{E}[X_-]$ is finite, then we say that X is quasi-integrable, and $\mathbb{E}[X] = \pm\infty$. $\mathbb{E}[X]$ is considered undefined if both are infinite.

Properties:

(1) If $X \in L_1(P)$, then $\mathbb{P}\{X = \pm\infty\} = 0$.

(2) (Linearity) If $\mathbb{E}[X]$ exists, then $\mathbb{E}[cX] = x\mathbb{E}[X]$. Futhermore, if $X, Y \in L_1(\mathbb{P})$, then $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

(3) (Monotonicity) If $X \geq 0$, then $\mathbb{E}[X] \geq 0$.

(4) (Continuity) If $\{X_n\}_{n\geq 1} \subseteq L_1$ such that $X_n \nearrow X$, then $\mathbb{E}[X_n] \nearrow \mathbb{E}[X]$.

(5) (Modulus Inequality) If $X \in L_1$, then $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

(5) (Variance, Covariance) Recall $\text{Var}(X) := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$. Well,

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &= \mathbb{E}\left[X^2 + 2X\mathbb{E}[X] + (\mathbb{E}[X])^2\right] \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2.\end{aligned}$$

Recall $\text{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$. By a similar calculation

$$\text{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Later in the course, we will show that independence implies $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, i.e., $X \perp Y$ implies $\text{Cov}(X,Y) = 0$. However, the converse is not true.

(6) Independence implies that the variance of a sum is equal to the sum of the variances.

$$\begin{aligned}\text{Var}\left(\sum_{i=1}^n X_i\right) &= \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i\right) \\ &= \sum_i \text{Cov}(X_i, X_i) + \sum_{i\neq j} \text{Cov}(X_i, X_j) && \text{(by bilinearity)} \\ &= \sum_i \text{Cov}(X_i, X_i) \qquad = \sum_{i=1}^n \text{Var}(X_i). && \text{(by independence)}\end{aligned}$$

(7) (Markov Inequality) If $X \in L_1$, then

$$\mathbb{P}\{|X| \geq \lambda\} \leq \frac{\mathbb{E}[|X|]}{\lambda}.$$

(8) (Chebyshev Inequality) If $X \in L_2$, then

$$\mathbb{P}\{|X - \mathbb{E}[X]| \geq \lambda\} \leq \frac{\text{Var}(X)}{\lambda^2}.$$

(9) (Weak Law of Large Numbers) Let $\{X_n\}_{n\geq 1}$ be iid with finite mean μ and finite variance σ. Then, for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| > \epsilon\right\} = 0.$$

Proof: Calculating,

$$\begin{aligned}
\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right\} &\le \frac{1}{\epsilon^2}\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) && \text{(Chebyshev Inequality)} \\
&= \frac{1}{\epsilon^2}\frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2\epsilon^2} n \operatorname{Var}(X_1) \\
&= \frac{1}{n\epsilon^2}\operatorname{Var}(X_1).
\end{aligned}$$

Clearly,

$$\lim_{n\to\infty}\frac{\operatorname{Var}(X_1)}{n\epsilon^2} = 0. \ \square$$

Section 5.3 - Limits and Integrals

Recall the Monotone Convergence Theorem: If $0 \le X_n \nearrow X$, then $0 \le \mathbb{E}[X_n] \nearrow \mathbb{E}[X]$.

Series Version of Monotone Convergence Theorem: Suppose $\{X_n\}_{n\ge1} \ge 0$. Then,

$$\mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right] = \sum_{n=1}^{\infty}\mathbb{E}[X_n].$$

Proof: We see that

$$\begin{aligned}
\mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right] &= \mathbb{E}\left[\lim_{N\to\infty}\sum_{n=1}^{N} X_n\right] \\
&= \lim_{N\to\infty}\mathbb{E}\left[\sum_{n=1}^{N} X_n\right] && \text{(MCT)} \\
&= \lim_{N\to\infty}\sum_{n=1}^{N}\mathbb{E}[X_n] && \text{(linearity)} \\
&= \sum_{n=1}^{\infty}\mathbb{E}[X_n]. \ \square
\end{aligned}$$

Fatou's Lemma: If $\{X_n\}_{n\ge1} \ge 0$, then

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \le \liminf_{n\to\infty}\mathbb{E}[X_n]. \qquad (\bigstar)$$

In general, if there exists $Z \in L_1$ such that $X_n \ge Z$ for all n, then $(\bigstar)$ still holds.

Proof: Observe

$$\begin{aligned}
\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] &= \mathbb{E}\left[\lim_{n\to\infty}\bigwedge_{k=n}^{\infty} X_k\right] \\
&= \lim_{n\to\infty}\mathbb{E}\left[\bigwedge_{k=n}^{\infty} X_k\right] && \text{(MCT)} \\
&\le \liminf_{n\to\infty}\mathbb{E}[X_k].
\end{aligned}$$

The justification for the last step is that for almost all $\omega \in \Omega$,

$$\mathbb{E}\left(\bigwedge_{k=n}^{\infty} X_k(\omega)\right) \leq \mathbb{E}(X_{k_*}) < \left(\inf_{k>n} \mathbb{E}(X_K)\right) + \epsilon,$$

where we define $k_*(\epsilon)$ so that $\mathbb{E}(X_{k_*}) < \left(\inf_{k>n} \mathbb{E}(X_n)\right) + \epsilon$. □

Corollary: $\mathbb{E}\left[\limsup_{n\to\infty} X_n\right] \geq \limsup_{n\to\infty} \mathbb{E}[X_n]$.

Dominated Convergence Theorem: If $X_n \to X$ and if there exists a dominating random variable $Z \in L_1$ in the sense that $|X_n| \leq Z$ for all n, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$ and $\mathbb{E}[|X_n - X|] \to 0$ (convergence in L_1).

Proof: We have $-Z \leq X_n \leq Z$. So, apply **Fatou's Lemma** to both sides. Then,

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \\
&\leq \liminf_{n\to\infty} \mathbb{E}[X_n] \\
&\leq \limsup_{n\to\infty} \mathbb{E}[X_n] \\
&\leq \mathbb{E}\left[\limsup_{n\to\infty} X_n\right] \\
&\leq \mathbb{E}[X].
\end{aligned}$$

So, we have equality throughout. □

Example: Take $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$ and define

$$X_n := n^2 1_{[0,1/n]}.$$

Note that

$$\lim_{n\to\infty} X_n(\omega) = 0$$

for all $\omega \in (0,1]$. However,

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \lim_{n\to\infty} n^2 \mathbb{P}\{\omega < 1/n\} = \lim_{n\to\infty} n = \infty.$$

So in this case

$$\lim_{n\to\infty} \mathbb{E}[X_n] > \mathbb{E}\left[\lim_{n\to\infty} X_n\right].$$

Section 5.4 - Indefinite Integrals

Definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If $X \in L_1(\mathbb{P})$, then we define

$$\int_A X d\mathbb{P} := \mathbb{E}\left[X 1_A\right].$$

This is called the integral of X restricted to the event A.

Section 5.5 - The Transformation Theorem and Densities

Let there be two spaces $(\Omega, \mathcal{F})$, $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$. Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$. Suppose

$$T : (\Omega, \mathcal{F}) \to (\widetilde{\Omega}, \widetilde{\mathcal{F}}).$$

Define

$$\widetilde{\mathbb{P}} := \mathbb{P} \circ T^{-1}$$

so that

$$\widetilde{\mathbb{P}}(\widetilde{A}) = \mathbb{P}(T^{-1}(\widetilde{A})).$$

Transformation Theorem: Suppose $\widetilde{X} : (\widetilde{\Omega}, \widetilde{\mathcal{F}}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

(i) If $\widetilde{X} \geq 0$, then

$$\int_\Omega \widetilde{X}(T(\omega))\mathbb{P}(d\omega) = \int_{\widetilde{\Omega}} \widetilde{X}(\widetilde{\omega})\widetilde{\mathbb{P}}(d\widetilde{\omega}),$$

i.e.,

$$\mathbb{E}\left[\widetilde{X} \circ T\right] = \widetilde{\mathbb{E}}\left[\widetilde{X}\right].$$

(ii) $\widetilde{X} \in L_1(\widetilde{\mathbb{P}})$ if and only if $X \circ T \in L_1(\mathbb{P})$, in which case

$$\int_{T^{-1}(\widetilde{A})} \widetilde{X}(T(\omega))\mathbb{P}(d\omega) = \int_{\widetilde{A}} \widetilde{X}(\widetilde{\omega})\widetilde{\mathbb{P}}(d\widetilde{\omega}).$$

Proof of (i): We prove the statement first for indicator random variables, then for simple random variables, then for nonnegative random varibles (by MCT), then for all random variables (by DCT).

Suppose $\widetilde{X} = 1_{\widetilde{A}}$ where $\widetilde{A} \in \widetilde{\mathcal{F}}$, i.e.,

$$\widetilde{X}(T(\omega)) = 1_{\widetilde{A}}(T(\omega)) = 1_{T^{-1}(\widetilde{A})}(\omega).$$

Then,

$$\begin{aligned}\int_\Omega \widetilde{X}(T(\omega))\mathbb{P}(d\omega) &= \int_\Omega 1_{T^{-1}(\widetilde{A})}(\omega)\mathbb{P}(d\omega) \\ &= \mathbb{P}(T^{-1}(\widetilde{A})) \\ &= \widetilde{\mathbb{P}}(\widetilde{A}) \\ &= \int_{\widetilde{\Omega}} \widetilde{X}(\widetilde{\omega})\widetilde{P}(d\widetilde{\omega}).\end{aligned}$$

Now suppose $\widetilde{X}$ is simple, i.e.,

$$\widetilde{X} = \sum_{i=1}^{k} \widetilde{a}_i 1_{\widetilde{A_i}}.$$

Then,

$$\begin{aligned}\int_\Omega \widetilde{X}(T(\omega))\mathbb{P}(d\omega) &= \sum_{i=1}^{k} \widetilde{a}_i \mathbb{P}(T^{-1}(\widetilde{A_i}) \\ &= \sum_{i=1}^{k} \widetilde{a}_i \widetilde{\mathbb{P}}(\widetilde{A_i}) \\ &= \int_\Omega \underbrace{\sum_{i=1}^{k} \widetilde{a}_i 1_{\widetilde{A_i}}(\widetilde{\omega})}_{\widetilde{X}(\widetilde{\omega})} \widetilde{\mathbb{P}}(\widetilde{\omega}).\end{aligned}$$

Next, let $\widetilde{X}$ be nonnegative (i.e., $\widetilde{X} \geq 0$) and $\mathcal{F}$-measurable. Then, there exists a sequence of random variables $\widetilde{X_n}$ such that $\widetilde{X_n} \uparrow \widetilde{X}$. This implies

$$\widetilde{X_n} \circ T \uparrow \widetilde{X} \circ T.$$

This should be checked by the reader.

Next, we calculate, using the Monotone Convergence Theorem:

$$\begin{aligned}
\int_\Omega \widetilde{X}(T(\omega))\mathbb{P}(d\omega) &= \int_\Omega \lim_{n\to\infty} \widetilde{X_n}(T(\omega))\mathbb{P}(d\omega) \\
&= \lim_{n\to\infty} \int_\Omega \widetilde{X_n}(T(\omega))\mathbb{P}(d\omega) && \text{(MCT)} \\
&= \lim_{n\to\infty} \int_{\widetilde{\Omega}} \widetilde{X_n}(\widetilde{\omega})\widetilde{\mathbb{P}}(d\widetilde{\omega}) \\
&= \int_{\widetilde{\Omega}} \lim_{n\to\infty} \widetilde{X_n}(\widetilde{\omega})\widetilde{\mathbb{P}}(d\widetilde{\omega}) && \text{(MCT)} \\
&= \int_{\widetilde{\Omega}} \widetilde{X}(\widetilde{\omega})\widetilde{\mathbb{P}}(d\widetilde{\omega}).
\end{aligned}$$

Lastly, by the Dominated Convergence Theorem, we can apply the result to all random variables (not just the nonnegative ones). □

Proof of (ii): Follows the same ideas. □

Remark: We have said that the cdf F can be thought of as $F(x) = \mathbb{P}\{X \leq x\}$. In this case, we are treating F like a function. However, we can also view F as a measure

$$F : \mathcal{B}(\mathbb{R}) \to [0, 1]$$

by

$$F(A) := (\mathbb{P} \circ X^{-1})(A)$$

for all $A \in \mathcal{B}(\mathbb{R})$.

Corollary to the Transformation Theorem:

(i) Suppose X is a random variable with cdf F. Then,

$$\mathbb{E}[X] = \int xF(dx).$$

(ii) Suppose $X : (\Omega, \mathcal{F}) \to (\mathbb{S}, \mathcal{S})$ with cdf $F = \mathbb{P} \circ X^{-1}$. Furthermore, suppose

$$g : (\mathbb{S}, \mathcal{S}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

and that f is $\mathcal{S}$-measurable. Then, the expected value of $g(X)$ is

$$\mathbb{E}[g(X)] = \int_\Omega g(X(\omega))\mathbb{P}(d\omega) = \int_{x\in\mathcal{S}} g(x)F(dx).$$

Proof of (i): Let $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $\widetilde{X} : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\widetilde{X}(x) = x$. Recall the equation

$$\int_\Omega \widetilde{X}(T(\omega))\mathbb{P}(d\omega) = \int_{\widetilde{\Omega}} \widetilde{X}(\widetilde{\omega})\widetilde{\mathbb{P}}(d\widetilde{\omega}).$$

Here, we have that the right hand side is

$$\int_{\mathbb{R}} xF(dx).$$

Set $T(\omega) = X(\omega)$. Then the the left-hand side becomes

$$\int_{\Omega} X(\omega)\mathbb{P}(d\omega).$$

Since these are equal, we get the result. $\square$

Remark: We say that X or F is absolutely continuous if there exists a nonnegative function $f : \mathbb{S} \to \mathbb{R}$ such that

$$F(A) = \int_A f(x)dx$$

for all $A \in \mathcal{S}$. In particular,

$$\mathbb{E}[g(X)] = \int_{\mathbb{S}} g(x)f(x)dx.$$

We say that $F(dx)$ has density $f(x)dx$.

Chapter 6

Chapter 6 - Convergence Concepts

Section 6.1 - Almost Sure Convergence

Definition: (Almost Sure Convergence)

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then this is defined by: there exists $N \in \mathcal{F}$ with $\mathbb{P}(N) = 0$ such that the statement of interest holds if $\omega \in N^C$. N is often called the exception set.

Remark: A typical statement may be $X = X'$ a.s. or $X \geq X'$ a.s. or $\lim_{n\to\infty} X_n = X$ a.s..

Example: Consider the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$. Define

$$X_n(\omega) = \begin{cases} n, & \text{if } \omega \in [0, 1/n] \\ 0, & \text{if } \omega \in (1/n, 1] \end{cases} .$$

Then, $X_n \to X$ almost surely. The exception set is $N = \{0\}$.

Proposition: Let $\{X_n\}_{n\geq 1}$ be iid random variables with a common cdf $F(x)$. Suppose $F(x) < 1$ for all $x \in \mathbb{R}$. Define

$$M_n = \bigvee_{i=1}^{n} X_i.$$

For all ω, $M_n(\omega) = \max_{i\in[n]} X_i(\omega)$. Then, $M_n \nearrow \infty$ almost surely.

Proof:

$$\begin{aligned} \mathbb{P}\{M_n \leq x\} &= \mathbb{P}\{X_1 \leq x, \ldots, X_n \leq x\} \\ &= \prod_{i=1}^{n} \mathbb{P}\{X_i \leq x\} && \text{(by independence)} \\ &= (F(x))^n. && \text{(common cdf)} \end{aligned}$$

Note that,

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}\{M_n \leq j\} &= \sum_{n=1}^{\infty} (F(j))^n \\ &= \frac{F(j)}{1 - F(j)} \\ &< \infty. \end{aligned}$$

So, by the **Borel-Cantelli Lemma** (note that the M_ns are dependent), we have that

$$\mathbb{P}\{M_n \leq j \text{ i.o.}\} = 0,$$

i.e.,

$$\mathbb{P}\left\{\limsup_{n\to\infty}\{\omega \mid M_n(\omega) \leq j\}\right\} = 0.$$

Define

$$N_j := \limsup_{n\to\infty}\{M_n \leq j\}$$

and then note that

$$N_j^C = \liminf_{n\to\infty}\{M_n > j\}.$$

So, for all $\omega \in N_j^C$ and for all n_*, there exists $n > n_*$ such that $M_n(\omega) > j$.

Define

$$N := \bigcup N_j.$$

Now,

$$\mathbb{P}\{N\} \leq \sum_{j=1}^{\infty} \mathbb{P}\{N_j\} = \sum_{j=1}^{\infty} 0.$$

Hence,

$$\mathbb{P}\left\{N^C\right\} = 1. \ \square$$

Section 6.2 - Convergence in Probability

Definition: (Convergence in Probability) $\{X_n\}_{n\geq 1}$ converges in probability to X, denoted $X_n \xrightarrow{\mathbb{P}} X$ if, for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\{|X_n - X| > \epsilon\} = 0.$$

Remark: This is usually demonstrated by **Chebyshev's Inequality**.

Remark: We can write the definition of almost sure convergence alternatively:

$$\mathbb{P}\left\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = 1.$$

Theorem: [Almost Sure Convergence] $\Longrightarrow$ [Convergence in Probability].

Proof: Suppose $X_n \xrightarrow{\text{a.s.}} X$. Then, for all $\epsilon > 0$,

$$\begin{aligned} 0 &= \mathbb{P}\{|X_n - X| > \epsilon \text{ i.o.}\} \\ &= \mathbb{P}\left\{\limsup_{n\to\infty}\{|X_n - X| > \epsilon\}\right\} \\ &= \lim_{N\to\infty} \mathbb{P}\left\{\bigcup_{n\geq N}\{|X_n - X| > \epsilon\}\right\} \\ &\geq \lim_{n\to\infty} \mathbb{P}\{|X_n - X| > \epsilon\}. \end{aligned}$$

Hence,

$$\lim_{n\to\infty} \mathbb{P}\{|X_n - X| > \epsilon\} = 0$$

which gives convergence in probability. □

Section 6.3 - Connections Between Almost Sure and In Probability Convergence

Theorem: Suppose $\{X_n\}_{n\geq 1}$ and X are random variables.

(a) The Cauchy criterion for random variables is: for all $\epsilon, \delta > 0$, there exists N such that for all $n, m > N$, we have

$$\mathbb{P}\{|X_n - X_m| > \epsilon\} < \delta.$$

This is true if and only if $X_n \xrightarrow{P} X$.

Proof: Assume $X_n \to X$ in probability. Let $\epsilon > 0$ be given. By the triangle inequality,

$$\{|X_n - X_m| > \epsilon\} \subseteq \{|X_n - X| > \epsilon/2\} \cup \{|X_m - X| > \epsilon/2\}.$$

Taking probabilities and using subadditivity,

$$\mathbb{P}\{|X_n - X_m| > \epsilon\} < \mathbb{P}\{\{|X_n - X| > \epsilon/2\} \cup \{|X_m - X| > \epsilon/2\}\} \leq \mathbb{P}\{|X_n - X| > \epsilon/2\} + \mathbb{P}\{|X_m - X| > \epsilon/2\}.$$

Now, for $\epsilon/2$, and $\delta/2$, there exists N such that for all $n, m > N$,

$$\mathbb{P}\{|X_n - X| > \epsilon/2\} < \delta/2.$$

Conversely, suppose $\{X_n\}_{n\geq 1}$ is Cauchy in probability. We claim that there exists a subsequence $\{X_{n_j}\}$ which converges almost surely. To see this, define $n_0 = 1$ and

$$n_j = \inf\left\{N > n_{j-1} \mid \forall n, m \geq N\ \mathbb{P}\left\{|X_n - X_m| > 2^{-j}\right\} < 2^{-j}\right\}.$$

It follows that

$$\mathbb{P}\left\{|X_{n_{j+1}} - X_{n_J}| > 2^{-j}\right\} < 2^{-j}.$$

Hence,

$$\sum_{j=1}^{\infty} \mathbb{P}\left\{|X_{n_{j+1}} - X_{n_j}| > 2^{-j}\right\} < \infty.$$

Define

$$N := \limsup_{j\to\infty} \{|X_{n_{j+1}} - X_{n_j}| > 2^{-j}\},$$

i.e.,

$$N = \{\omega \in \Omega \mid |X_{n_{j+1}}(\omega) - X_{n_j}(\omega)| > 2^{-j} \text{ i.o.}\}.$$

By **Borel-Cantelli**,

$$\mathbb{P}\{N\} = 0.$$

Well, observe that for all $\omega \in N^C$, the sequence $\{X_{n_j}(\omega)\}_{j\geq 0}\}$ is a Cauchy sequence of real numbers. So, there exists $X(\omega)$ such that $X_{n_j}(\omega) \to X(\omega)$, for all $\omega \in N^C$.

To show that having a convergence subsequence and being Cauchy in probability implies convergence in probability note that

$$\mathbb{P}\{|X_n - X| > \epsilon\} \leq \mathbb{P}\left\{|X_n - X_{n_j}| > \epsilon/2\right\} + \mathbb{P}\left\{|X_{n_j} - X| > \epsilon/2\right\}.$$

The second summand on the right-hand side is $< \delta/2$ by the almost sure convergence of the subsequence. The first summand on the right-hand side is $< \delta/2$ by the Cauchy subsequence, as long as we pick n_j to be the smallest index in the convergence subsequence which is $> n$. □

(b) $X_n \xrightarrow{P} X$ if and only if for each subsequence $\{X_{n_k}\}$ there exists a further subsequence $X_{n_{k(i)}}$ such that convergence is almost sure.

Proof: Suppose $X_n \to X$ in probability. Then, any subsequence $\{X_{n_j}\}$ converges in probability (to X) as well. By part (a), $\{X_{n_j}\}$ is Cauchy in probability as well, and we already showed that this implies that there exists a further subsequence that converges almost surely.

Conversely, suppose that for all subsequences $\{X_{n_j}\}$ there exists a further subsequence that converges to X almost surely. We proceed by contradiction. Suppose $X_n \not\to X$ in probability. Then, there exists a subsequence $\{X_{n_k}\}$ and an $\epsilon > 0$ and $\delta > 0$ such that for all n_k,

$$\mathbb{P}\{|X_{n_k} - X| > \epsilon\} \geq \delta. \qquad (\bigstar)$$

However, by hypothesis, X_{n_k} has a convergent subsequence. This is a contradiction to $(\bigstar)$, because

$$\mathbb{P}\left\{X_{n_{k(i)}} \text{ converges to } X\right\} = 1 - \mathbb{P}\left\{|X_{n_{k(i)}} - X| > \epsilon \text{ i.o.}\right\}$$

By a Borel-Cantelli Argument, using the fact that

$$\sum_{i=1}^{\infty} \mathbb{P}\left\{|X_{n_{k(i)}} - X| > \epsilon\right\} \geq \sum_{i=1}^{\infty} \delta = \infty,$$

we see that we have a contradiction. $\square$

Section 6.5 - L_p Convergence

Definition: For $p \in (0, \infty)$, define L_p by

$$X \in L_p \text{ if } \mathbb{E}\left[|X|^p\right] < \infty.$$

The distance on this space is

$$d(X,Y) = \left(\mathbb{E}\left[|X-Y|^p\right]\right)^{1/p}.$$

The L_p norm is

$$\|X\|_p := \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p}.$$

Definition: We say that $X_n \xrightarrow{L_p} X$ if

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^p\right] = 0.$$

Example: Suppose $\{X_n\}_{n\geq 1}$ are iid with $\mathbb{E}[X_n] = \mu$ and $\mathrm{Var}(X_n) = \sigma^2$. Then,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{L_2} \mu.$$

Proof: Define

$$S_n := \sum_{i=1}^{n} X_i.$$

Then,

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{1}{n}S_n - \mu\right|^2\right] &= \frac{1}{n^2}\mathbb{E}\left[|S_n - n\mu|^2\right] \\
&= \frac{1}{n^2}\,\mathrm{Var}(S_n) \\
&= \frac{n\sigma^2}{n^2} \to 0. \quad \square
\end{aligned}$$

Remark: Observe that
$$[X_n \xrightarrow{L_p} X] \quad \Longrightarrow \quad [X_n \xrightarrow{P} x].$$
To see this, use the generalized Chebyshev inequality:
$$\mathbb{P}\left\{|X_n - X| > \epsilon\right\} \leq \frac{\mathbb{E}\left[|X_n - X|^p\right]}{\epsilon^p} \to 0.$$
The converse is false. Consider the space $([0,1], \mathcal{B}([0,1]), \lambda)$. Define
$$X_n(\omega) := 2^n \cdot 1_{(0,1/n]}(\omega).$$
This is because
$$\mathbb{P}\left\{X_n(\omega) > \frac{1}{n}\right\} = \frac{1}{n} \to 0,$$
yet
$$\mathbb{E}\left[|X_n|^p\right] = \frac{2^{np}}{n} \to \infty.$$

Remark: Hoeffding's Inequality provides exponential estimates, but without calculating $\mathbb{E}[e^{aX}]$. This opents the way for what are called "Concentration Inequalities", which will appear in Homework 5.

Lemma: If $\mathbb{E}[X] = 0$ and $a \leq X \leq b$, then
$$\mathbb{E}[e^{tX}] \leq e^{t^2(b-a)^2/8}.$$

Proof: First, by convexity,
$$e^{tX} \leq \alpha e^{tb} + (1-\alpha)^{ta}$$
where $\alpha = \dfrac{X-a}{b-a}$ and $1 - \alpha = \dfrac{b-X}{b-a}$. Take expextations of each side:
$$\begin{aligned} \mathbb{E}[e^{tx}] &\leq -\frac{ae^{tb}}{b-a} + \frac{be^{ta}}{b-a} \\ &= e^{g(t(b-a))}, \end{aligned}$$
where
$$g(u) = -\gamma u + \log(1 - \gamma + \gamma e^u,$$
where
$$\gamma = -\frac{a}{b-a}.$$
Note that $g(0) = 0$ and $g'(0) = 0$. Furthermore, $g''(u) \leq 1/4$ for all $u > 0$. By Taylor's Theorem, there exists $\xi \in (0, u)$ such that
$$\begin{aligned} g(u) &= g(0) + ug'(0) + \frac{1}{2}u^2 g''(\xi) \\ &\leq \frac{t^2(b-a)^2}{8}. \ \square \end{aligned}$$

Lemma: (Chernoff's Method) Let X be a random variables. Then,
$$\mathbb{P}\left\{X > \epsilon\right\} \leq \inf_{t \geq 0} e^{-t\epsilon}\mathbb{E}[e^{tX}].$$

Proof: Generalized Chebychev Inequality together with optimization over t. $\square$

Theorem: (Hoeffding's Inequality) Let $\{X_n\}_{n\geq 1}$ be iid such that $\mathbb{E}[X_n] = \mu$ and $a \leq X_n \leq b$. Then, for all $\epsilon > 0$,

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right\} \leq 2e^{\dfrac{-2n\epsilon^2}{(b-a)^2}}.$$

Corollary: If $\{X_1, \ldots, X_n\}$ are iid with $a \leq X_i \leq b$, then with probability at least $1-\delta$,

$$\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \leq \sqrt{\frac{(b-a)^2}{2n}\log\left(\frac{2}{\delta}\right)}.$$

Proof: Without loss of generality, $\mu = 0$. Now, define

$$\overline{X_n} := \frac{1}{n}\sum_{i=1}^{n} X_n.$$

So,

$$\mathbb{P}\left\{\left|\overline{X_n}\right| \geq \epsilon\right\} = \mathbb{P}\left\{\overline{X_n} \geq \epsilon\right\} + \mathbb{P}\left\{-\overline{X_n} \geq \epsilon\right\}.$$

Observe that

$$\begin{aligned}
\mathbb{P}\left\{\overline{X_n} \geq \epsilon\right\} &= \mathbb{P}\left\{t\sum_{i=1}^{n} X_i \geq tn\epsilon\right\} \\
&= \mathbb{P}\left\{e^{t\sum_{i=1}^{n} X_i} \geq e^{tn\epsilon}\right\} \\
&\leq e^{-tn\epsilon\mathbb{E}\left[e^{t\sum_{i=1}^{n} X_i}\right]} && \text{(Gen. Cheb. and Chern. Ineq.)} \\
&= e^{-tn\epsilon}\prod_{i=1}^{n}\mathbb{E}\left[e^{tX_i}\right] && \text{(independence)} \\
&= e^{-tn\epsilon}\mathbb{E}\left[e^{tX_1}\right]^n && \text{(independence)} \\
&\leq e^{-tn\epsilon}\left(e^{t^2(b-a)^2/8}\right). && \text{(Lemma)}
\end{aligned}$$

So,

$$\mathbb{P}\left\{\overline{X_n} \geq \epsilon\right\} \leq e^{-nt^2(b-a)^2/8 - t\epsilon}.$$

By calculus, the exponent is minimized when $t = 4\epsilon/(b-a)^2$. Apply also to $-\overline{X_n}$ and then plug in. $\square$

Inequalities:

(1) Cauchy-Schwartz Inequality:

$$|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|] \leq \mathbb{E}[X^2]^{1/2}\mathbb{E}[Y^2]^{1/2}.$$

(2) Hölder's Inequalitky:

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p}\mathbb{E}[|Y|^q]^{1/q}$$

where $\dfrac{1}{p} + \dfrac{1}{q} = 1$.

(3) Minkowski's Inequality:

$$\|X+Y\|_p \leq \|X\|_p + \|Y\|_p.$$

(4) Jensen's Inequality: If $f(x)$ is convex, then

$$\mathbb{E}[f(x)] \geq f(\mathbb{E}[x]).$$

Chapter 7

Chapter 7 - Laws of Large Numbers and Sums of Independent Random Variables

Section 7.1 - Truncation and Equivalence

Definition: The truncation of a random variable is

$$X_n \cdot 1_{\{|X_n| \leq n\}}.$$

Example: Suppose $\mathrm{Var}(X_n) = \infty$, but nevertheless, $\mathrm{Var}(X_n \cdot 1_{\{|X_n| \leq n\}}) < \infty$ for all n. The truncations will help.

Definition: Two sequences of random variables $\{X_n\}$ and $\{\widetilde{X_n}\}$ are <u>tail equivalent</u> if

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{X_n \neq \widetilde{X_n}\right\} < \infty.$$

Proposition: (Equivalence) Suppose $\{X_n\}$ and $\{\widetilde{X_n}\}$ are tail equivalent. Then,

(1) $\displaystyle\sum_{n=1}^{\infty}(X_n - \widetilde{X_n})$ converges almost surely,

(2) $\displaystyle\sum_{n=1}^{\infty} X_n$ converges if and only if $\displaystyle\sum_{n=1}^{\infty} \widetilde{X_n}$ converges,

(3) If there exists a sequence $\{a_n\} \subseteq \mathbb{R}$ such that $a_n \nearrow \infty$ and if there exists a random variable X such that

$$\frac{1}{a_n}\sum_{j=1}^{n} X_j \to X \text{ almost surely,}$$

then also

$$\frac{1}{a_n}\sum_{j=1}^{n} \widetilde{X_j} \to X \text{ almost surely,}$$

Proof: By **Borel-Cantelli**, $\mathbb{P}\left\{X_n \neq \widetilde{X_n} \text{ i.o.}\right\} = 0$ implies that

$$\mathbb{P}\left\{\liminf_{n\to\infty}(X_n = \widetilde{X_n})\right\} = 1.$$

Restrict to the set $\omega \in \left\{\liminf_{n\to\infty}(X_n = \widetilde{X_n})\right\}$. We have that there exists $N(\omega)$ such that for all $n > N(\omega)$, $X_n(\omega) = \widetilde{X_n}(\omega)$. This immediately yields (1).

Furthermore, for each such ω, $\sum_{n=N(\omega)}^{\infty} X_n(\omega) = \sum_{n=N(\omega)}^{\infty} \widetilde{X_n}(\omega)$, which implies (2).

To prove (3), observe that

$$\frac{1}{a_n}\sum_{j=1}^{n}\widetilde{X_j} = \frac{1}{a_n}\sum_{j=1}^{n}\widetilde{X_j} - X_j + \frac{1}{a_n}\sum_{j=1}^{n} X_j.$$

The first term on the left-hand side goes to 0 almost surely as $n \to \infty$, and the second term on the right-hand side goes to X almost surely as $n \to \infty$. This completes the theorem. □

Section 7.2 - A General Weak Law of Large Numbers

Theorem: (General WLLN) Suppose $\{X_n\}_{n\geq 1}$ are independent random variables and define $S_n := \sum_{j=1}^{n} X_j$.

Then, if

(i) $\sum_{j=1}^{n} \mathbb{P}\{|X_j| > n\} \to 0$, and

(ii) $\frac{1}{n^2}\sum_{j=1}^{n} \mathbb{E}\left[X_j^2 \cdot 1_{\{|X_j|\leq n\}}\right] \to 0$,

then we have that

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

where

$$a_n = \sum_{j=1}^{n} \mathbb{E}\left[X_j \cdot 1_{\{|X_j|\leq n\}}\right].$$

Remark: In the special case where the random variables are iid, with $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$, then

$$\sum_{j=1}^{n} \mathbb{P}\{|X| > n\} \to 0 \text{ (in probability)} \qquad \text{iff} \qquad n\mathbb{P}\{|X| > n\} \to 0 \text{ (in probability)}.$$

(To prove this special case, use Chebyshev.)

Proof: Define $\widetilde{X_{nj}} := X_j 1_{\{|X_j|\leq n\}}$, and $\widetilde{S_n} := \sum_{j=1}^{n} \widetilde{X_{nj}}$. Then,

$$\sum_{j=1}^{n} \mathbb{P}\left\{\widetilde{X_{nj}} \neq X_j\right\} = \sum_{j=1}^{n} \mathbb{P}\{|X_j| > n\} \to 0.$$

We claim that $S_n - \widetilde{S_n} \xrightarrow{P} 0$. To see this, note that

$$\begin{aligned}\mathbb{P}\left\{|S_n - \widetilde{S_n}| > \epsilon\right\} &\le \mathbb{P}\left\{S_n \neq \widetilde{S_n}\right\} \\ &\le \mathbb{P}\left\{\bigcup_{j=1}^{n} \widetilde{X_{nj}} \neq X_j\right\} \\ &\le \sum_{j=1}^{n} \mathbb{P}\left\{\widetilde{X_{nj}} \neq X_J\right\} \\ &\xrightarrow{P} 0.\end{aligned}$$

Apply Chebyshev's inequality to $\mathbb{P}\left\{\frac{1}{n}\left|S_n - \mathbb{E}[\widetilde{S_n}]\right| > \epsilon\right\}$. So,

$$\begin{aligned}\mathbb{P}\left\{\frac{1}{n}\left|S_n - \mathbb{E}[\widetilde{S_n}]\right| > \epsilon\right\} &\le \frac{\mathrm{Var}(\widetilde{S_n})}{n^2\epsilon^2} \\ &\le \frac{1}{n^2\epsilon^2}\sum_{j=1}^{n} \mathbb{E}\left[\widetilde{X_{nj}}^2\right] \\ &= \frac{1}{\epsilon^2}\frac{1}{n^2}\sum_{j=1}^{n} \mathbb{E}\left[X_j^2 1_{\{|X_j|\le n\}}\right] \\ &\to 0,\end{aligned}$$

by hypothesis. Above, we used the fact that

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 \le \mathbb{E}(Y^2).$$

Therefore,

$$\frac{\widetilde{S_n} - a_n}{n} \xrightarrow{P} 0.$$

Hence,

$$\frac{S_n - a_n}{n} = \underbrace{\frac{S_n - \widetilde{S_n}}{n}}_{\searrow^P 0} + \underbrace{\frac{\widetilde{S_n} - a_n}{n}}_{\searrow^P 0} \xrightarrow{P} 0. \quad \square$$

Theorem: (Khintchin's WLLN, with only first moment) Let $\{X_n\}$ be iid and $\mathbb{E}[X_1] = \mu$. Assume $X_1 \in L_1$, i.e., $\mathbb{E}[|X_1|] < \infty$. Then,

$$\frac{S_n}{n} \xrightarrow{P} \mu.$$

Proof: We need to check the two conditions.

(1)

$$\begin{aligned}\sum_{j=1}^{n} \mathbb{P}\left\{|X_j| > n\right\} &= n\mathbb{P}\left\{|X_1| > n\right\} \\ &= n\mathbb{E}\left[1_{\{|X_1|>n\}}\right] \\ &= \mathbb{E}\left[n1_{\{|X_1|>n\}}\right] \\ &= \mathbb{E}\left[|X_1|1_{\{|X_1|>n\}}\right] \\ &=\to 0,\end{aligned}$$

because $|X_1|1_{\{|X_1|>n\}}(\omega) \xrightarrow{\text{a.s.}} 0$ and since $|X_1| \in L_1(\mathbb{P})$ and $|X_1|$ is dominating.

(2) For iid random variables, we're actually showing (by linearity) that,

$$\frac{1}{n}\mathbb{E}\left[X^2 1_{\{|X|\leq n\}}\right] \to 0.$$

Well,

$$\begin{aligned}\frac{1}{n}\mathbb{E}\left[X^2 1_{\{|X|\leq n\}}\right] &\leq \frac{1}{n}\left(\mathbb{E}\left[X^2 1_{\{|X|<\epsilon\sqrt{n}\}}\right] + \mathbb{E}\left[X^2 1_{\{\epsilon\sqrt{n}\leq|X|\leq n\}}\right]\right)\\ &\leq \frac{1}{n}\left(\epsilon^2 n + \mathbb{E}\left[n|X|1_{\{\epsilon\sqrt{n}\leq|X|\leq n\}}\right]\right)\\ &\leq \epsilon^2 + \mathbb{E}\left[|X|1_{\{\epsilon\sqrt{n}\leq|X|\leq n\}}\right] \qquad \to \epsilon^2. \qquad \text{(DCT)}\end{aligned}$$

Also,

$$\frac{1}{n}a_n = \frac{n\mathbb{E}\left[X_1 1_{\{|X_1|\leq n\}}\right]}{n} \to \mathbb{E}[X_1] = \mu$$

by DCT. □

Theorem: (Feller's WLLN, without first moment) Let $\{X_n\}$ be iid random variables, with

$$\lim_{x\to\infty} x\mathbb{P}\{|X_n| > x\} \to 0.$$

Then,

$$\frac{S_n}{n} - \mathbb{E}\left[X_1 1_{\{|X_1|\leq n\}}\right] \xrightarrow{P} 0.$$

Section 7.3 - Almost Sure Convergence of Sums of Independent Random Variables

Theorem: (Lévy's Theorem) If $\{X_n\}$ is an independent sequence of random variables, then

$$\sum_n X_n \text{ converges a.s.} \qquad \text{iff} \qquad \sum_n X_n \text{ converges in probability.}$$

This means that the following are equivalent:

(1) S_n is Cauchy in probability.

(2) S_n converges in probability.

(3) S_n converges a.s.

(4) S_n is a.s. Cauchy.

Remark: Formally, (1) says that

$$\forall\epsilon,\delta\exists\mathbb{N}_{\epsilon,\delta} \text{ such that } \forall n,m > N_{\epsilon,\delta} : \mathbb{P}\{|S_n - S_m| > \epsilon\} < \delta.$$

(4) says that

$$\forall\epsilon\exists N_\epsilon(\omega) \text{ such that (for } \omega\in\widetilde{\Omega} \text{ where } \mathbb{P}\left\{\widetilde{\Omega}\right\} = 1)\forall n,m > N_\epsilon(\omega) : |S_n(\omega) - S_m(\omega)| < \epsilon.$$

Proof: (1) $\iff$ (2) by **Theorem 6.3.1(a)**. Also, (3) $\iff$ (4) by the corresponding equivalence for sequences of real numbers shown in analysis. We've already shown that (3) $\implies$ (2) by **Theorem 6.2.1**. The complete the proof, we will shoe (1) $\implies$ (4).

Define

$$\xi_N := \sup_{m,n \geq N} |S_m - S_n| .$$

We need to show that (1) implies that

$$\xi_N \xrightarrow{\text{a.s.}} 0.$$

Note that $\{\xi_N\}_{N\in\mathbb{N}}$ is a positive non-increasing sequence of random variables. By **Exercise 6.1**,

$$[\xi_N \xrightarrow{P} 0] \text{ iff } [\xi_N \xrightarrow{\text{a.s.}} 0].$$

So, we will show that $\xi_N \xrightarrow{P} 0$. Now,

$$\begin{aligned}
\xi_N &:= \sup_{m,n \geq N} |S_m - S_N + S_N - S_n| \\
&\leq 2 \sup_{n \geq N} |S_n - S_N| \\
&= 2 \sup_{j \geq 0} |S_{N+j} - S_N| .
\end{aligned}$$

We will now show that this goes to zero in probability. To this end, let $\epsilon > 0$ and $S \in (0, 1/2)$. The fact that $\{S_n\}$ is Cauchy in probability implies that there exists $N_{\epsilon,\delta}$ such that for all $n, m > N_{\epsilon,\delta}$, we have

$$\mathbb{P}\{|S_n - S_m| > \epsilon/2\} \leq \delta.$$

So, for any $N > N_{\epsilon,\delta}$, we have that

$$\mathbb{P}\{|S_{N+j} - S_N| > \epsilon/2\} \leq \delta$$

for all j.

Define

$$\begin{aligned}
\widetilde{X_i} &:= X_{N+i} \\
\widetilde{S_j} &:= \sum_{i=1}^{j} \widetilde{X_i} = S_{N+j} - S_N.
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{P}\left\{\sup_{j\geq 0} |S_{N+j} - S_N| > \epsilon/2\right\} &= \mathbb{P}\left\{\lim_{N'\to\infty} \sup_{j\in[N']} |S_{N+j} - S_N| > \epsilon/2\right\} \\
&= \lim_{N'\to\infty} \mathbb{P}\left\{\sup_{j\in[N']} |S_{N+j} - S_N| > \epsilon/2\right\}. \\
&= \mathbb{P}\left\{\sup_{0\leq N'} \left|\widetilde{S_j}\right| > \epsilon/2\right\}.
\end{aligned}$$

For the next step, we will use **Skorohod's Inequality**, which tells us that there exists $c \in (0, 1)$ such that

$$\mathbb{P}\left\{\sup_{0\leq N'} \left|\widetilde{S_j}\right| > \epsilon/2\right\} \leq \frac{1}{1-c}\mathbb{P}\left\{\left|\widetilde{S_n}\right| > \epsilon/4\right\},$$

where

$$c = \sup_{j\leq N'} \mathbb{P}\left\{\left|\widetilde{S_{N'}} - \widetilde{S_j}\right| > \epsilon/4\right\}.$$

Since we have $c < \delta$ by the Cauchy property, the whole thing goes to zero, i.e., $\xi_N \xrightarrow{P} 0$. $\square$

Section 7.4 - Strong Law of Large Numbers

We will discuss two motivating examples.

(1) Logarithm Growth of New Records.

(2) Explosions in Arrival Processes.

Example (1): Let $\{X_n\}$ be an iid sequence of random variables with common cdf F. We defined X_n to be a record if

$$\max_{i\in[n]} X_i = X_n.$$

Define $\mu_n = \sum_{i=1}^{n} \mathbf{1}_{\{X_i \text{ is a record}\}}$. The Strong Law of Large Numbers will allow us to prove the following result characterizing the rate at which records are broken:

$$\lim_{n\to\infty} \frac{\mu_n}{\log(n)} = 1 \text{ a.s.}.$$

Example (2): Define $\{\tau_n\}_{n\geq 1}$ to be the independent sequence of random variables with $\mathbb{P}\{\tau_n > t\} = e^{-\lambda_n t}$ for all $t > 0$, i.e., $\tau_n \sim \text{Exp}(\lambda_n)$.

Think of τ_n as the times between events and let N_t be the number of events as of time t. The Strong Law of Large Numbers will help us characterize when such a process explodes to infinity.

Lemma 7.4.1: (Kronecker's Lemma) Suppose $\{a_n\}_{n\geq 1}$ and $\{x_n\}_{n\geq 1}$ satisfy $x_n \in \mathbb{R}$ and $0 < a_n \nearrow \infty$. If $\sum_{n=1}^{\infty} \frac{x_n}{a_n}$ converges, then

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} x_k = 0.$$

Proof: Define $r_n := \sum_{k=n+1}^{\infty} \frac{x_k}{a_k}$. Then, $r_n \to 0$, and given $\epsilon > 0$, there exists N_ϵ such that for all $n > N_\epsilon$, we have $|r_n| \leq \epsilon$. Now, note that

$$\frac{x_n}{a_n} = r_{n-1} - r_n.$$

So, $x_n = a_n(r_{n-1} - r_n)$ and thus

$$\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} (r_{k-1} - r_k) a_k.$$

The rest follows by some calculations which can be found on pg. 214 of Resnick. □

Recall: (Kolmogorov Convergence Criterion) If $\{X_n\}_{n\geq 1}$ is independent with $\sum_{n=1}^{\infty} \text{Var}(X_n) < \infty$, then

$$\sum_{n=1}^{\infty} (X_n - \mathbb{E}[X_n])$$

converges almost surely.

Corollary 7.4.1: (Strong Law of Large Numbers) Let $\{X_n\}_{n\geq 1}$ be independent with $\mathbb{E}\left[X_n^2\right] < \infty$. Suppose that $\{b_n\}$ is a sequence that increases to ∞ such that

$$\sum_{k=1}^{\infty} \mathrm{Var}\left(\frac{X_k}{b_k}\right) < \infty.$$

Then,

$$\frac{1}{b_n}\left(S_n - \mathbb{E}[S_n]\right) \xrightarrow{\text{a.s.}} 0.$$

Apply the Kolmogorov Convergence Criterion and Kronecker's Lemma. □

Example (1): (Record Counts) Define

$$\mathbf{1}_k := \mathbf{1}_{\{X_k \text{ is a record}\}}$$

and

$$\mu_n := \sum_{k=1}^{n} \mathbf{1}_k.$$

Proposition: $\lim_{n\to\infty} \frac{\mu_n}{\log(n)} = 1$ almost surely.

Proof: Assume $\mathbb{P}\{\mathbf{1}_k\} = \frac{1}{k}$ and recall that $\mathbb{E}[\mathbf{1}_k] = \frac{1}{k}$. Note that

$$\begin{aligned}\mathrm{Var}(\mathbf{1}_k) &= \mathbb{E}\left[\mathbf{1}_k^2\right] - \mathbb{E}[\mathbf{1}_k]^2 \\ &= \frac{1}{k} - \frac{1}{k^2} \\ &= \frac{k-1}{k^2}.\end{aligned}$$

We now check convergence.

$$\begin{aligned}\sum_{k=2}^{\infty} \mathrm{Var}\left(\frac{\mathbf{1}_k}{\log(k)}\right) &= \sum_{k=2}^{\infty} \frac{1}{(\log(k))^2}\mathrm{Var}(\mathbf{1}_k) \\ &= \sum_{k=2}^{\infty} \frac{k-1}{k^2(\log(k))^2}.\end{aligned}$$

By the integral test,

$$\int_2^{\infty} \frac{1}{x(\log(x))^2}\,dx = \int_{\log(2)}^{\infty} \frac{1}{u^2}\,du < \infty.$$

Therefore, we conclude that

$$\sum_{k=2}^{\infty} \frac{\mathbf{1}_k - \mathbb{E}[\mathbf{1}_k]}{\log(k)} < \infty.$$

By the **Strong Law of Large Numbers**,

$$\frac{1}{\log(n)}\left(\sum_{k=1}^{n}(\mathbf{1}_k - \mathbb{E}[\mathbf{1}_k])\right) \xrightarrow{\text{a.s.}} 0$$

and so

$$\frac{\mu_n}{\log(n)} \xrightarrow{\text{a.s.}} \frac{\sum_{k=1}^{n}\mathbb{E}[\mathbf{1}_k]}{\log(n)}.$$

We need to show that the right hand side goes to 1 as $n \to \infty$. Well,

$$\frac{1}{\log(n)} \sum_{k=1}^{n} \frac{1}{k} \xrightarrow{1} -\frac{\gamma}{\log(n)} \to 0,$$

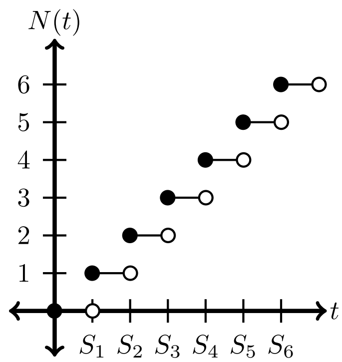
where γ is Euler's Constant.

Example (2): (Explosions in Arrival Times) Let $\{\tau_n\}_{n=1}^{\infty}$ be independent with $\tau_n \sim \text{Exp}(\lambda_n)$. Say we have a counting process

$$S_n = \sum_{k=1}^{n} \tau_k$$

with $S_0 = 0$. Define

$$N(t) := \sum_{n=0}^{\infty} \mathbf{1}_{\{S_n \le t\}}.$$



We get an explosion when

$$\sum_{k=1}^{n} \tau_k < \infty,$$

which implies that there exists T such that

$$\lim_{t \to T} N(t) = \infty.$$

Claim: $\mathbb{P}\{\text{explosion}\} = \begin{cases} 1, & \sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty \\ 0, & \sum_{n=1}^{\infty} \frac{1}{\lambda_n} = \infty \end{cases}$. (It can only be 0 or 1 because it is a tail event.)

Proof: Suppose that $\mathbb{P}\left\{\sum_{k=1}^{\infty} \tau_k < \infty\right\} = 1$. We aim to show that $\sum_{k=1}^{\infty} \frac{1}{\lambda_k} < \infty$.

By hypothesis, $\exp\left(-\sum_{k=1}^{\infty} \tau_k\right) > 0$, and so

$$\mathbb{E}\left[\exp\left(-\sum_{k=1}^{\infty} \tau_k\right)\right] > 0.$$

Thus,

$$\begin{aligned}
0 &< \mathbb{E}\left[\exp\left(-\sum_{k=1}^{\infty}\tau_k\right)\right] \\
&= \mathbb{E}\left[\prod_{k=1}^{\infty} e^{-\tau_k}\right] \\
&= \mathbb{E}\left[\lim_{n\to\infty}\prod_{k=1}^{n} e^{-\tau_k}\right] \\
&= \lim_{n\to\infty}\mathbb{E}\left[\prod_{k=1}^{n} e^{-\tau_k}\right] && \text{(by MCT)} \\
&= \lim_{n\to\infty}\prod_{k=1}^{n}\mathbb{E}\left[e^{-\tau_k}\right] && \text{(by independence)} \\
&= \lim_{n\to\infty}\prod_{k=1}^{n}\int_0^{\infty} e^{-t}\lambda_k e^{-\lambda_k t}\,dt \\
&= \lim_{n\to\infty}\prod_{k=1}^{n}\frac{\lambda_k}{1+\lambda_k}.
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}\left[\exp\left(-\sum_{k=1}^{\infty}\tau_k\right)\right] > 0 &\iff -\log\left(\mathbb{E}\left[\exp\left(-\sum_{k=1}^{\infty}\tau_k\right)\right]\right) < \infty \\
&\iff -\sum_{k=1}^{\infty}\log\left(\frac{\lambda_k}{1+\lambda_k}\right) < \infty \\
&\iff \sum_{k=1}^{\infty}\log\left(1+\frac{1}{\lambda_k}\right) < \infty \\
&\iff \sum_{k=1}^{\infty}\frac{1}{\lambda_k} < \infty.
\end{aligned}$$

The last step is by the Limit Comparison Test and the fact that

$$\lim_{x\to 0}\frac{\log(1+x)}{x} = 1. \quad \square$$

Section 7.5 - Strong Law of Large Numbers for IID Sequences

Theorem: (Kolmogorov's SLLN) Let $\{X_n\}_{n\geq 1}$ be iid random variables, and define $S_n := \sum_{k=1}^{n} X_k$. Then, there exists $\mu \in \mathbb{R}$ such that

$$\overline{X_n} := \frac{S_n}{n} \xrightarrow{\text{a.s.}} \mu$$

if and only if

$$\mathbb{E}\left[|X_1|\right] < \infty.$$

When convergence does occur, $\mu = \mathbb{E}\left[X_1\right]$.

Proof:

($\Longrightarrow$) Suppose that $\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mu$.

Lemma 7.5.1: Let $\{X_n\}$ be iid. Then, the following are equivalent.

(a) $\mathbb{E}[|X_1|] < \infty$.

(b) $\lim_{n\to\infty} \left|\frac{X_n}{n}\right| = 0$ almost surely.

(c) $\sum_{n=1}^{\infty} \mathbb{P}\{|X_1| > \epsilon n\} < \infty$, for all $\epsilon > 0$.

Proof: (sketch) [(b) $\Longleftrightarrow$ (c)] can be shown using the Borel 0-1 Law.

[(a) $\Longleftrightarrow$ (c)] can be shown by using the tail formula for expectation

$$\mathbb{E}[|X_1|] = \int_0^{\infty} \mathbb{P}\{|X_1| > x\}\, dx$$

and comparing to the sum. □

So, by **Lemma 7.5.1**, $\mathbb{E}[|X_1|] < \infty$ if and only if

$$\lim_{n\to\infty} \left|\frac{X_n}{n}\right| = 0 \text{ almost surely.}$$

But, note that

$$\begin{aligned}\frac{X_n}{n} &= \frac{S_n - S_{n-1}}{n} \\ &= \frac{S_n}{n} - \frac{S_n}{n-1} \cdot \frac{n-1}{n} \\ &\xrightarrow{\text{a.s.}} \mu - \mu \cdot 1 \\ &= 0. \ \square\end{aligned}$$

($\Longleftarrow$) Suppose $\mathbb{E}[|X_1|] < \infty$.

Step 1: (Truncation.)

Define $\widetilde{X}_n := X_n \mathbf{1}_{\{|X_n| \le n\}}$. We need to show that X_n and $\widetilde{X}_n$ are tail equivalent, i.e., that

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{X_n \neq \widetilde{X}_n\right\} = \sum_{n=1}^{\infty} \mathbb{P}\{|X_n| > n\} < \infty$$

by **Lemma 7.5.1** [(a) $\Longrightarrow$ (c)]. Therefore,

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mu \qquad \Longleftrightarrow \frac{\widetilde{S}_n}{n} \xrightarrow{\text{a.s.}} \mu.$$

Step 2: If the truncated series converges, then it converges to the right value.

Claim: $\left(\frac{\widetilde{S}_n}{n} - \mathbb{E}\left[X_1\right]\right) \xrightarrow{\text{a.s.}} 0 \iff \frac{1}{n}\sum_{k=1}^{n}(\widetilde{X}_k - \mathbb{E}\left[\widetilde{X}_k\right]) \xrightarrow{0}.$

Proof: Note that

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=1}^{n}\left(\widetilde{X}_k - \mathbb{E}\left[\widetilde{X}_k\right]\right) - \frac{1}{n}\sum_{k=1}^{n}(\widetilde{X}_k - \mathbb{E}\left[X_k\right])\right| \\ &= \left|\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[\widetilde{X}_k\right] - \mathbb{E}[X_1]\right| \\ &\to 0.\end{aligned}$$

The last step is by DCT. □

Step 3: The truncated series converges.

Confirm that $\frac{1}{n}\sum_{k=1}^{n}\left(\widetilde{X}_k - \mathbb{E}\left[\widetilde{X}_k\right]\right) \to 0$ by showing that

$$\sum_{k=1}^{\infty}\text{Var}\left(\frac{\widetilde{X}_k}{k}\right) < \infty.$$

See **Corollary 7.4.1**. □

Applications of the Strong Law of Large Numbers

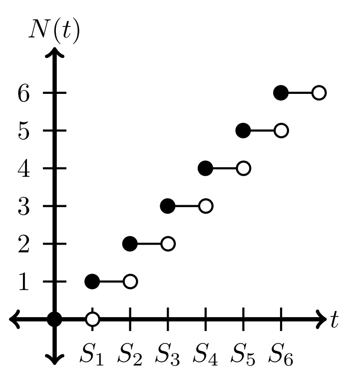
Renewal Theorem: Let $\{X_n\}_{n\geq 1}$ be iid with $X_n \geq 0$. Let $\mathbb{E}[X_1] =: \mu \in (0,\infty)$ and let $\mathbb{E}[|X_1|] < \infty$. By the Strong Law,

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mu.$$

Thus, $S_n \to \infty$ almost surely.

Define a counting process

$$N(t) = \sum_{n=1}^{\infty}\mathbf{1}_{\{S_n \leq t\}}.$$



Observe that

$$\{N(t) \le n\} = \{S_{n+1} > t\}$$

and thus

$$S_{N(t)} \le t \le S_{N(t)+1}.$$

We claim that $\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}$ almost surely.

Proof: Define

$$\Omega_1 := \{\omega \in \Omega \mid \frac{S_n(\omega)}{n} \to \mu\},$$
$$\Omega_2 := \{\omega \in \Omega \mid (N(t))(\omega) \to \infty\}.$$

Then, $\Omega_0 := \Omega_1 \cap \Omega_2$ has $\mathbb{P}\{\Omega_0\} = 1$. (We assumed that $\mathbb{P}\{\Omega_1\} = 1$, but we need to show that $\mathbb{P}\{\Omega_2\} = 1$ for this to be true. But,

$$\lim_{t\to\infty} \mathbb{P}\{N(t) \le m\} = \lim_{t\to\infty} \mathbb{P}\{S_{m+1} > t\} \to 0.$$

We have $N(t) \xrightarrow{P} \infty$ and by monotonicity of $N(t)$, we see $N(t) \to$ a.s.∞.)

Well,

$$\frac{S_{N(t)}}{N(t)} \xrightarrow{t\to\infty} \mu \text{ a.s..}$$

Now,

$$\frac{t}{N(t)} \le \frac{S_{N(t)}}{N(t)} \to \mu$$

and

$$\frac{t}{N(t)} \ge \frac{S_{N(t)-1}}{N(t)} = \frac{S_{N(t)-1}}{N(t)-1} \cdot \frac{N(t)-1}{N(t)} \to \mu \cdot 1 = \mu.$$

Thus $\frac{t}{N(t)} \to \mu$ and $\frac{N(t)}{t} \to \mu^{-1}$. So, we have shown the claim. □

Theorem 7.5.2: (Glivenko-Cantelli Theorem)

Question: Let $\{X_n\}_{n\ge1}$ be iid random variables with common cdf $F(x)$. How many samples do we need to take to estimate the cdf well? (This is called the Kolmogorov-Smirnoff Test.)

Empirical Distribution Formula: Define

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \le x\}}.$$

Precise Statement:

$$\sup_{x\in\mathbb{R}} \left|\widehat{F}_n(x) - F(x)\right| \xrightarrow{\text{a.s.}} 0.$$

In other words, the convergence of $\widehat{F}_n(x)$ to $F(x)$ is uniform.

Proof: See the online notes here:

`http://uflprob.files.wordpress.com/2013/01/glivenko-cantelli-shaikh.pdf`.

Section 7.6 - The Kolmogorov Three Series Theorem

Theorem: (Kolmogorov's Three-Series Theorem) Let $\{X_n\}_{n\geq 1}$ be independent. Then, $\sum_{n=1}^{\infty} X_n$ converges almost surely if and only if there exists $c > 0$ such that:

(i) $\sum_{n=1}^{\infty} \mathbb{P}\{|X_n| > c\} < \infty$,

(ii) $\sum_{n=1}^{\infty} \operatorname{Var}\left(X_n \mathbf{1}_{\{|X_n|\leq c\}}\right) < \infty$,

(iii) $\sum_{n=1}^{\infty} \mathbb{E}\left[X_n \mathbf{1}_{\{|X_n|\leq c\}}\right]$ converges.

Remark: If $X_i \geq 0$ for all i, then (i) and (iii) suffice. This is **Exercise 7.15**.

Proof: The necessity is hard to prove. There is a version in the book. A more natural proof uses the Central Limit Theorem. Once we learn this, we'll come back.

The proof of sufficiency is much easier. Define

$$\widetilde{X}_n := X_n \mathbf{1}_{\{|X_n|\leq c\}}.$$

The first step is to check the tail equivalence of $\{X_n\}$ and $\{\widetilde{X}_n\}$, i.e, that they converge/diverge together i.e., that

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{\widetilde{X}_n \neq X_n\right\} < \infty.$$

But, by construction

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{\widetilde{X}_n \neq X_n\right\} = \sum_{n=1}^{\infty} \mathbb{P}\{X_n > c\} < \infty.$$

It remains to show now that the truncated series $\sum_{n=1}^{\infty} \widetilde{X}_n$ converges. Note that

$$\sum_{n=1}^{\infty} \operatorname{Var}\left(\widetilde{X}_n\right) = \sum_{n=1}^{\infty} \operatorname{Var}\left(X_n \mathbf{1}_{\{X_n\leq c\}}\right) < \infty$$

by assumption (ii). So, by **Kolmogorov's Convergence Criterion** implies that

$$\sum_{n=1}^{\infty} \left(\widetilde{X}_n - \mathbb{E}\left[\widetilde{X}_n\right]\right)$$

converges almost surely.

Noting that by (iii) that

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left[\widetilde{X}_k\right]$$

exists, so too does $\lim_{n\to\infty} \sum_{k=1}^{n} \widetilde{X}_k$. By the tail equivalence that we showed, the proof is complete. □

Remark: As mentioned in the remark before the proof, if $X_i \geq 0$ for all i, then we don't need the hypothesis (ii).

Proof: Suppose $X_i \geq 0$. Then,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathrm{Var}\left(X_n \mathbf{1}_{\{|X_n| \leq c|\}}\right) &= \sum_{n=1}^{\infty} \mathbb{E}\left[X_n^2 \mathbf{1}_{\{X_n \leq c\}}\right] - \mathbb{E}\left[X_n \mathbf{1}_{\{X_n \leq c\}}^2\right] \\
&\leq \sum_{n=1}^{\infty} \mathbb{E}\left[X_n^2 \mathbf{1}_{\{X_n \leq c\}}\right] \\
&\leq \sum_{n=1}^{\infty} c\mathbb{E}\left[X_n \mathbf{1}_{\{X_n \leq c\}}\right] < \infty
\end{aligned}$$

So, (ii) follows from (iii). We still need (i) to show tail-equivalent. □

Application: This applies to the convergence of series of "heavy-tailed random variables", meaning that $\mathbb{E}\left[|X_i|\right] = \infty$. Examples are the Cauchy distribution and the Pareto distribution.

Chapter 8

Chapter 8 - Convergence in Distribution

Section 8.1 - Basic Definitions

Example: The following is an example of the type of problems that we are interested in.

Define U_n such that

$$\mathbb{P}\left\{U_n = \frac{k}{n}\right\} = \frac{1}{n}$$

for $k \in [n]$. In what sense does $U_n \to U := \text{Unif}([0,1])$?

We may attempt the following failed reasoning: Suppose we ask that for every $A \in \mathcal{B}((0,1])$, we have

$$\mathbb{P}\{U_n \in A\} \to \mathbb{P}\{U \in A\}.$$

Why does this fail? Choose $A = \mathbb{Q} \cap [0,1]$. Then, for all n,

$$\mathbb{P}\{U_n \in A\} = 1$$

but

$$\mathbb{P}\{U \in A\} = 0.$$

So, this attempt at a definition of convergence in distribution fails.

Four Notations Of Convergence in Distribution:

Reminder: F is a distribution function if:

(i) F is right-continuous,

(ii) F is non-decreasing,

(iii) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

We use the notation $F((a,b]) = F(b) - F(a)$.

(1) **Vague Convergence:** (Chung '68, Feller '71) We say that

$$F_n \xrightarrow{v} F$$

if for every interval I with endpoints at which F is continuous

$$F_n(I) \to F(I).$$

(2) **Proper Convergence:** (Feller '71) F_n converges properly to F if $F_n \xrightarrow{v} F$ (as above) and additionally $F(\mathbb{R}) = 1$.

(3) **Weak Convergence:** (Billingsly '68) We say that

$$F_n \xrightarrow{w} F$$

if, for all $x \in \mathcal{C}(F) := \{x \in \mathbb{R} \mid F \text{ is continuous}\}$,

$$F_n(x) \to F(x).$$

(4) **Complete Convergence:** (Lōene, 77) We say that F_n converges completely to F if $F_n \xrightarrow{w} F$ (as above) and $F(\mathbb{R}) = 1$.

Theorem: If $F(\mathbb{R}) = 1$, then (1)–(4) are equivalent.

Proof:

(4) $\implies$ (2): Let $a, b \in \mathcal{C}(F)$. Then,

$$\begin{aligned} F_n((a,b]) &= F_n(b) - F_n(a) \\ &\to F(b) - F(a) \\ &= F((a,b]). \end{aligned}$$

To extend this to all intervals (not just those with endpoints in the continuity set), we need to following lemma.

Lemma 8.1: A distribution function is uniquely determined by its values on a dense set.

Proof: Suppose that F_D is defined on a dense set D, such that F_D satisfies the definition of a distribution on the set D. Define

$$F(x) := \inf_{\substack{y \in D \\ y > x}} F_D(y).$$

Now check that $F(x)$ has all of the right properties. See the textbook for the full proof. □

This completes this portion of the proof.

(2) $\implies$ (4) Assume that $F_n(I) \to F(I)$ for all I such that F is continuous on I. Suppose that $a, b \in \mathcal{C}(F)$. Then,

$$F_n(b) \geq F_n((a,b]) \xrightarrow{F} ((a,b]).$$

Thus,

$$\liminf_{n \to \infty} F_n(b) \geq F((a,b])$$

and hence, taking the limit as $a \to -\infty$

$$\liminf_{n \to \infty} F_n(b) \geq F(b).$$

Now suppose that $\ell < b < r$ for $\ell, r \in \mathcal{C}(F)$, and furthermore suppose that

$$F\left((\ell, r]^C\right) < \epsilon.$$

Now,

$$F_n\left((\ell, r]^C\right) \to F\left((\ell, r]^C\right).$$

So, given $\epsilon > 0$, there exists N such that for all $n > N$, we have

$$F_n\left((\ell, r]^C\right) < 2\epsilon.$$

Thus

$$\begin{aligned} F_n(b) &= F_n(b) - F_n(\ell) + F_n(\ell) \\ &= F_n((\ell, b]) + F_n(\ell) \\ &\leq F((\ell, b]) + 2\epsilon. \end{aligned}$$

Hence

$$\limsup_{n\to\infty} F_n(b) \leq F((\ell, b]) + \epsilon.$$

Taking $\epsilon \to 0$ and $\ell \to -\infty$, we get

$$\limsup_{n\to\infty} F_n(b) \leq F(b).$$

So, we have equality.

The other equivalences are trivial. $\square$

Section 8.2 - Scheffé's Lemma

Scheffé's Lemma: Suppose F and F_n are cdfs. Then,

$$\sup_{B\in\mathcal{B}(\mathbb{R})} |F_n(B) - F(B)| = \frac{1}{2}\int_{\mathbb{R}} |f_n(x) - f(x)|\, dx.$$

If $f_n(x) \to f(x)$ almost everywhere, then $F_n \xrightarrow{w} F$.

Proof: Let $B \in \mathcal{B}\mathbb{R}$. Then, since

$$\int_{\mathbb{R}} (f_n(x) - f(x))dx = \int_{\mathbb{R}} f_n(x)dx - \int_{\mathbb{R}} f(x)dx = 1 - 1 = 0,$$

we have

$$0 = \int_{\mathcal{B}} (f_n(x) - f(x))dx + \int_{\mathcal{B}^C} (f_n(x) - f(x))dx,$$

which implies that

$$\left|\int_{\mathcal{B}^C} (f_n(x) - f(x))dx\right| = \left|\int_{\mathcal{B}} (f_n(x) - f(x))dx\right|.$$

Thus,

$$\begin{aligned} 2|F_n(B) - F(B)| &= \left|\int_{\mathcal{B}} (f_n(x) - f(x))dx\right| + \left|\int_{\mathcal{B}^C} (f_n(x) - f(x))dx\right| \\ &\leq \int_{\mathcal{B}} |f_n(x) - f(x)|\, dx + \int_{\mathcal{B}^C} |f_n(x) - f(x)|\, dx \\ &= \int_{\mathbb{R}} |f_n(x) - f(x)|\, dx. \end{aligned}$$

Hence,

$$|F_n(B) - F(B)| \leq \frac{1}{2}\int_{\mathbb{R}} |f_n(x) - f(x)|\, dx.$$

In order to equality under the supremum, we need only find one set B where equality holds. Define

$$B = \{x : f_n(x) \geq f(x)\}.$$

Verify that this works.

Now suppose $f_n(x) \to f(x)$ almost everywhere. Then, $(f - f_n)^+ \to 0$ almost everywhere. Also,

$$(f - f_n)^+ \le f$$

since f_n is a density and so is always positive. Additionally, f is integrable on $\mathbb{R}$ with respect to Lebesgue measure.

Since

$$\begin{aligned} 0 &= \int_{\mathbb{R}} (f(x) - f_n(x))dx \\ &= \int_{\mathbb{R}} (f(x) - f_n(x))^+ dx - \int_{\mathbb{R}} (f(x) - f_n(x))^- dx, \end{aligned}$$

it follows that

$$\begin{aligned} \int_{\mathbb{R}} |f(x) - f_n(x)|\, dx &= \int_{\mathbb{R}} (f(x) - f_n(x))^+ dx + \int_{\mathbb{R}} (f(x) - f_n(x))^- dx \\ &= 2\int_{\mathbb{R}} (f(x) - f_n(x))dx. \end{aligned}$$

Thus, $(f - f_n)^+ \in L_1$ and $(f - f_n)^+ \to 0$ almost everywhere, so by dominated convergence,

$$\int_{\mathbb{R}} |f(x) - f_n(x)|\, dx \to 0. \ \square$$

Example: (Order Statistics) Let $\{U_n\}_{n\ge1}$ be a sequence of iid Unif$([0,1])$ random variables. Among the first n, rewrite these in ascending order

$$\begin{aligned} &U_{(1,1)} \\ &U_{(1,2)} \le U_{(2,2)} \\ &U_{(1,3)} \le U_{(2,3)} \le U_{(3,3)} \\ &\quad \cdots, \end{aligned}$$

etc. Let k_n be a sequence of integers such that $\dfrac{k_n}{n} \to 0$. Define

$$\xi_n := \frac{U_{(k_n,n)} - \dfrac{k_n}{n}}{\sqrt{\dfrac{k_n}{n}\left(1 - \dfrac{k_n}{n}\right)\dfrac{1}{n}}}.$$

Then, the density of ξ_n converges to a standard normal density

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Hence, by **Scheffé's Lemma**,

$$\sup_{B\in\mathcal{B}(\mathbb{R})} \left| \mathbb{P}\left\{\xi_n \in B\right\} - \int_B \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right| \to 0.$$

Section 8.3 - The Baby Skorohod Theorem

Remark: This section explores the relationship between almost sure convergence and weak convergence.

Notation: If random variables X_n have cdfs F_n which converge weakly to some proper F, then we write $X_n \Longrightarrow X$ where F is the cdf of X.

Proposition 8.3.1: Suppose X and $\{X_n\}_{n\geq 1}$ are random variables. Then, if $X_n \xrightarrow{\text{a.s.}} X$, we have $X_n \Longrightarrow X$.

Proof: Suppose $X_n \xrightarrow{\text{a.s.}} X$, i.e., $\exists N$ with $\mathbb{P}\{N\} = 0$ such that $\forall \omega \in N^C$, we have $X_n(\omega) \to X(\omega)$. We want to show that for all $x \in \mathcal{C}(F)$, we have

$$F_n(x) \to F(x).$$

Observe that

$$\begin{aligned} N^C \cap \{\omega : X(\omega) < x - h\} &\subseteq N^C \cap \liminf_{n\to\infty} \{\omega : X_n(\omega) \leq x\} \\ &\subseteq N^C \cap \limsup_{n\to\infty} \{\omega : X_n(\omega) \leq x\} \\ &\subseteq N^C \cap \{\omega : X(\omega) \leq x\}. \end{aligned}$$

Taking probabilities,

$$\begin{aligned} F(x-h) &\leq \mathbb{P}\left\{\liminf_{n\to\infty}\{X_n \leq x\}\right\} \\ &\leq \liminf_{n\to\infty} \mathbb{P}\{X_n \leq x\} && \text{(by Fatou's Lemma)} \\ &\leq \limsup_{n\to\infty} \mathbb{P}\{X_n \leq x\} \\ &\leq \mathbb{P}\left\{\limsup_{n\to\infty}\{X_n \leq x\}\right\} && \text{(by Fatou's Lemma)} \\ &\leq F(x). \end{aligned}$$

Letting $h \to 0$, we see that $F(x-h) \to F(x)$ because we assume $x \in \mathcal{C}(F)$. So,

$$F(x) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x).$$

Thus, equality holds, and so $F_n(x) \to F(x)$. $\square$

Lemma: Suppose F_n is the cdf of X_n so that $F_n \xrightarrow{c} F$. If $t \in (0,1) \cup \mathcal{C}\left(F_0^{\leftarrow}\right)$, then

$$F_n^{\leftarrow} \to F_0^{\leftarrow}.$$

Proof: Let t be given. Given $\epsilon > 0$, there exists $x \in \mathcal{C}(F_0)$ such that

$$F_0^{\leftarrow}(t) - \epsilon < x < F_0^{\leftarrow}(t)$$

because $\mathcal{C}(F_0)^C$ is countable. By definition, if $x < F_0^{\leftarrow}(t)$, then

$$F_0(x) < t.$$

Also, $x \in \mathcal{C}(F_0)$, and so $F_n(x) \to F_0(x)$. So, for sufficiently large n, we have that

$$F_n(x) < t.$$

Thus,

$$F_0^{\leftarrow}(t) - \epsilon < x \leq F_m^{\leftarrow}(t)$$

which implies that

$$F_0^{\leftarrow}(t) \leq \liminf_{n\to\infty} F_n^{\leftarrow}(t).$$

To see the reverse inequality, let $t' > t$. Then, there exists $y \in \mathcal{C}(F_0)$ such that

$$F_0^{\leftarrow}(t') < y < F_0^{\leftarrow}(t') + \epsilon.$$

This implies that

$$F_0(y) \geq t' > t.$$

Since $y \in \mathcal{C}(F_0)$, we have that $F_n(y) \to F(y)$. So, for sufficiently large n, $F_n(y) \geq y$, which implies that $y \leq F_n^{\leftarrow}(t)$. So,

$$F_0^{\leftarrow}(t') + \epsilon > y \geq F_n^{\leftarrow}(t).$$

Thus,

$$\limsup_{n\to\infty} F_n^{\leftarrow}(t) \leq F_0^{\leftarrow}(t')$$

and so

$$\limsup_{n\to\infty} F_n^{\leftarrow}(t) \leq F_0^{\leftarrow}(t). \quad \square$$

Baby Skorohod Theorem: Suppose $\{X_n\}_{n\geq 1}$ are random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n \Rightarrow X_0$. Then, there exist random variables $\{X_n^{\#}\}_{n\geq 0}$ defined on $([0,1], \mathcal{B}([0,1]), \lambda)$ such that for each $n \geq 0$,

$$X_n \stackrel{d}{=} X_n^{\#}$$

and $X_n^{\#} \xrightarrow{\text{a.s.}} X_0^{\#}$.

Proof: Define $U \sim \text{Unif}([0,1])$ and $X_n^{\#} = F_n^{\leftarrow}(U)$. Note that $X_n^{\#} : [0,1] \to \mathbb{R}$. Then,

$$\begin{aligned}
\mathbb{P}_\lambda \left\{X_n^{\#} \leq y\right\} &= \lambda \left\{t \in [0,1] : F_n^{\leftarrow}(t) \leq y\right\} \\
&= \lambda \left\{t \in [0,1] : t \leq F_n(y)\right\} \\
&= F_n(y),
\end{aligned}$$

i.e, $X_n^{\#} \stackrel{d}{=} X_0$.

Now,

$$\begin{aligned}
\lambda \left\{t \in [0,1] : X_n^{\#}(t) \not\to X_0^{\#}(t)\right\} &= \lambda \left\{t \in [0,1] : F_n^{\leftarrow}(t) \not\to F_0^{\#}(t)\right\} \\
&= \lambda \left\{t \in [0,1] : F_0^{\leftarrow}(t) \text{ is not continuous at } t\right\} \\
&= 0,
\end{aligned}$$

since the set of discontinuities is either empty or countable. $\square$

Continuous Mapping Theorem: Suppose $X_n \Rightarrow X_0$. Let $g : \mathbb{R} \to \mathbb{R}$ be such that

$$\mathbb{P}\left\{X_0 \in \text{Disc}(g)\right\} = 0$$

(recall that $\text{Disc}(g)$ is the discontinuity set of the function g). Then, $g(X_n) \Rightarrow g(X_0)$, and furthermore, if g is bounded, then

$$\mathbb{E}\left[g(X_n)\right] \to \mathbb{E}\left[g(X_0)\right].$$

Proof: By Skorohod, there exists a sequence $\{X_n^{\#}\}$ such that $X_n^{\#} \stackrel{d}{=} X_n$, defined on $[0,1]$ such that $X_n^{\#} \xrightarrow{\text{a.s.}} X_0^{\#}$.

If $X_0^{\#} \in \mathcal{C}(g)$, then $X_n^{\#} \to X_0^{\#}$ implies $g\left(X_n^{\#}\right) \xrightarrow{\text{a.s.}} g\left(X_0^{\#}\right)$. So,

$$\begin{array}{ccc} g\left(X_n^{\$}\right) & \Longrightarrow & g\left(X_0^{\#}\right) \\ \Big\| d & & \Big\| d \\ g(X_n) & \Longrightarrow & g(X_0) \end{array} . \square$$

Delta Method: Suppose

$$\sqrt{n}\left(\frac{\frac{1}{n}S_n - \mu}{\sigma}\right) \Rightarrow Z.$$

Then, if g has a nonzero derivative at μ, we have

$$\sqrt{n}\left(\frac{g\left(\frac{1}{n}S_n\right) - g(\mu)}{\sigma g'(\mu)}\right) \Rightarrow Z.$$

Proof: Define

$$Z_n^{\#} \stackrel{d}{=} \underbrace{\sqrt{n}\left(\frac{\frac{1}{n}S_n - \mu}{\sigma}\right)}_{= Z_n},$$

where $Z_n^{\#}$ is defined on $([0,1], \mathcal{B}([0,1]), \lambda)$. Then, $Z_n^{\#} \xrightarrow{\text{a.s.}} Z^{\#}$. Well,

$$\begin{aligned} \sqrt{n}\left(\frac{g\left(\frac{1}{n}S_n\right) - g(\mu)}{\sigma}\right) &\stackrel{d}{=} \sqrt{n}\left(\frac{g\left(\mu + \frac{\sigma Z_n^{\#}}{\sqrt{n}}\right) - g(\mu)}{\sigma g'(\mu)}\right) \\ &= \sqrt{n}\left(\frac{g\left(\mu + \frac{\sigma Z_n^{\#}}{\sqrt{n}}\right) - g(\mu)}{\frac{\sigma Z_n^{\#}}{\sqrt{n}}}\right)\left(\frac{\frac{Z_n^{\#}}{\sqrt{n}}}{g'(\mu)}\right) \\ &\xrightarrow{\text{a.s.}} Z^{\#} \\ &\stackrel{d}{=} Z. \end{aligned}$$

Section 8.4 - Portmanteau Theorem

Portmanteau Theorem: Let $\{F_n\}$ be a sequence of proper cdfs. Then, the following are equivalent.

(1) $F_n \Rightarrow F_0$

(2) $\int f dF_n \to \int f dF_0$. Equivalently, if X_n is a random variable with distribution F_n, then for f bounded and continuous,

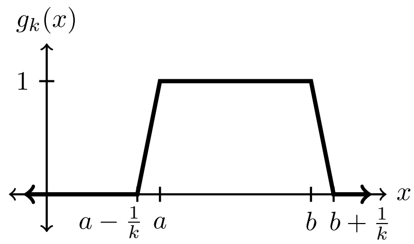
$$\mathbb{E}\left[f(X_n)\right] \to \mathbb{E}\left[f(X_0)\right].$$

(3) If $A \in \mathcal{B}(\mathbb{R})$, and $F(\partial(A)) = 0$, then $F_n(A) \to F_0(A)$.

Proof:

(1) $\Longrightarrow$ (2): Apply the **Continuous Mapping Theorem**. $\square$

(2) $\Longrightarrow$ (1): Let $a, b \in \mathcal{C}(F_0)$. We want to show that $F_n((a,b]) \to F_0((a,b])$. Consider $\{g_k(x)\}$ defined by the picture below.



Note that the g_k' are continuous, bounded, and that

$$\lim_{k\to\infty} g_k(x) = \mathbf{1}_{[a,b]}.$$

Well,

$$\begin{aligned} F_n((a,b]) &= \int_{\mathbb{R}} \mathbf{1}_{(a,b]}(x) dF_n(x) \\ &\le \int_{\mathbb{R}} g_k(x) dF_n(x) \\ &\xrightarrow{n\to\infty} \int_{\mathbb{R}} g_n dF_0(x). \qquad \text{(by assumption)} \end{aligned}$$

Note that $g_k(x) \le 1$ and $g_k(x) \searrow \mathbf{1}_{[a,b]}$. Thus,

$$\begin{aligned} \int_{\mathbb{R}} g_n dF_0(x) &\xrightarrow{k\to\infty} \int_{\mathbb{R}} \mathbf{1}_{[a,b]} dF_0(x) \\ &= F_0([a,b]). \end{aligned}$$

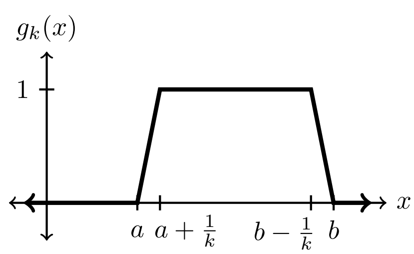
Also,

$$\limsup_{n\to\infty} F_n((a,b]) \le F_0([a,b]) = F_0[(a,b]]$$

since $a \in \mathcal{C}(F_0)$. To show that

$$\liminf_{n\to\infty} F_n((a,b]) \ge F_0(a,b])$$

we need to pick $h_k(x)$ as follows.



The rest of the argument is analogous. $\square$

Example: Define $\{X_n\}_{n=1}^{\infty}$ such that $X_n \in \left\{\frac{i}{n}\right\}_{i=i}^{n}$ and for each i,

$$\mathbb{P}\left\{X_n = \frac{i}{n}\right\} = \frac{1}{n}.$$

Show that $X_n \Rightarrow X$, where $X \sim \text{Unif}([0,1])$.

Proof: Note that $F(x) = x$ for all $x \in [0,1]$. Use the **Portmanteau Theorem**. Let f be bounded and continuous on $[0,1]$. Then,

$$\mathbb{E}\left[f(X_n)\right] = \sum_{i=1}^{n} f\left(\frac{1}{n}\right)\frac{1}{n}.$$

By Riemann Approximation, this converges to $\int_0^1 f(x)dx$ (note $dx = dF(x)$) which equals $\mathbb{E}\left[f(x)\right]$. So,

$$\begin{aligned}
\left|\mathbb{E}\left[f(X_n)\right] - \mathbb{E}\left[f(X)\right]\right| &= \left|\sum_{i=1}^{n}\int_{\frac{i-1}{n}}^{\frac{i}{n}}\left(f(x) - f\left(\frac{i}{n}\right)\right)dx\right| \\
&\leq \sum_{i=1}^{n}\int_{\frac{i-1}{n}}^{\frac{i}{n}}\left|f(x) - f\left(\frac{i}{n}\right)\right|dx.
\end{aligned}$$

Since f is continuous on $[0,1]$, it must also be uniformly continuous on this interval. So, let $\epsilon > 0$ be give Then, there exists $\delta_f > 0$ such that for all x and y with $|x-y| < \delta_f$, we must have $|f(x)-f(y)| < \epsilon$. Now, for $\epsilon > 0$, choose n_f large enough such that

$$\frac{1}{n_f} < \delta_f.$$

Then for $n > n_f$,

$$\begin{aligned}
\sum_{i=1}^{n}\int_{\frac{i-1}{n}}^{\frac{i}{n}}\left|f(x) - f\left(\frac{i}{n}\right)\right|dx &\leq \sum_{i=1}^{n}\int_{\frac{i-1}{n}}^{\frac{i}{n}}\epsilon dx \\
&= \epsilon. \quad \square
\end{aligned}$$

Section 8.5 - More Relations Among Modes of Convergence

Proposition 8.5.1: Let $\{X_n\}_{n\geq 1}$ and X be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$.

(i) If $X_n \xrightarrow{\text{a.s.}} X$ then $X_n \xrightarrow{\mathbb{P}} X$.

(ii) If $X_n \xrightarrow{\mathbb{P}} X$ then $X_n \Rightarrow X$.

Both converses are false.

Section 8.6 - New Convergences From Old

Theorem 8.6.1:

(a) If $X_n \Rightarrow X$ and $X_n - Y_n \xrightarrow{\mathbb{P}} 0$, then $Y_n \Rightarrow X$.

(b) Equivalently, if $X_n \Rightarrow X$ and $\xi_n \xrightarrow{\mathbb{P}} 0$, then $X_n + \xi_n \Rightarrow X$.

Chapter 9

Chapter 9 - Characteristic Functions and the Central Limit Theorem

Section 9.1 - Review of Moment Generating Functions and the Central Limit Theorem

Remark: The fundamental technical issue in studying the distribution of sums of random variables is the following. Suppose X_1 and X_2 are independent with cdfs F_1 and F_2. Then,

$$\begin{aligned}
\mathbb{P}\left\{X_1 + X_2 \le t\right\} &= \iint_{x+y\le t} (F_1 \times F_2)(dx, dy) \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{t-y} F_1(dx) \right) F_2(dy) && \text{(by Fubini's Theorem)} \\
&= \int_{-\infty}^{\infty} F_1(t-y) F_2(dy) \\
&=: (F_1 * F_2)(t).
\end{aligned}$$

Definition: The <u>characteristic function</u> of X is:

$$\begin{aligned}
\phi_x(t) &= \mathbb{E}\left[e^{itX}\right] \\
&= \int_{\mathbb{R}} e^{itx} F(dx) \\
&= \int_{\mathbb{R}} e^{itx} f(x) dx \\
&=: \widehat{f}(t).
\end{aligned}$$

Recall: The <u>moment generating function</u> is

$$M_X(t) = \mathbb{E}\left[e^{tX}\right].$$

(For negative values of t, this it the Laplace Transform.) The problem with this is that the moment generating function doesn't always (or even typically) exist. On the other hand,

$$\begin{aligned}
|\phi_x(t)| &= \left|\mathbb{E}\left[e^{itX}\right]\right| \\
&= \mathbb{E}\left[\left|e^{itX}\right|\right] \\
&= 1.
\end{aligned}$$

Remark: Always remember that $e^{itX} = \cos(tX) + i\sin(tX)$.

Remark: Why is $\mathbb{E}\left[e^{tX}\right]$ called a moment generating function? The answer comes from the taylor expansion.

$$\begin{aligned}
e^{tX} &= 1 + tX + \frac{t^2X^2}{2} + \frac{t^3X^3}{3!} + \cdots \\
\mathbb{E}\left[e^{tX}\right] &= 1 + t\mathbb{E}\left[X\right] + \frac{t^2}{2}\mathbb{E}\left[X^2\right] + \frac{t^3}{3!}\mathbb{E}\left[X^3\right] + \cdots \\
M_X(0) &= 1 \\
M_X'(t) &= \mathbb{E}\left[X\right] + t\mathbb{E}\left[X^2\right] + \frac{t^2}{2}\mathbb{E}\left[X^2\right] + \cdots \\
M_X'(0) &= \mathbb{E}\left[X\right].
\end{aligned}$$

In general,

$$\frac{d^n}{dt^n}M_X(0) = \mathbb{E}\left[X^n\right]$$

and

$$\frac{d^n}{dt^n}\phi_X(0) = u^n\mathbb{E}\left[X^n\right].$$

Example: Let $Z \sim \text{Norm}(0,1)$. Then,

$$\begin{aligned}
\phi_z(t) &= \mathbb{E}\left[e^{itZ}\right] \\
&= \sum_{n=0}^{\infty} \frac{(it)^n}{n!}\mathbb{E}\left[Z^n\right] \\
&= \sum_{k=0}^{\infty} \left(\frac{(it)^{2k}}{(2k)!} \cdot \frac{(2k)!}{k! \cdot 2^k}\right) \\
&= \sum_{k=0}^{\infty} \left((-1)^k \left(\frac{t^2}{2}\right)^k \cdot \frac{1}{k!}\right) \\
&= e^{-t^2/2}.
\end{aligned}$$

(Baby) Central Limit Theorem: Let $\{X_n\}_{n=1}^{\infty}$ be iid random variables with finite moments. Without loss of generality, $\mathbb{E}\left[X_n\right] = 0$ and $\mathbb{E}\left[X_n^2\right] = 1$. Define

$$S_n = \sum_{i=1}^{n} X_i.$$

We claim that

$$\frac{S_n}{\sqrt{n}} \Rightarrow N$$

where $N \sim \text{Norm}(0,1)$.

Section 9.3 - Expansion of e^{ix}

Method: We compute that

$$\begin{aligned}
\phi_{n^{-\alpha}S_n}(t) &= \mathbb{E}\left[\exp\left(it\left(\frac{1}{n^\alpha}\right)\sum_{i=1}^n X_i\right)\right] \\
&= \mathbb{E}\left[\exp\left(i\left(\frac{t}{n^\alpha}\right)\sum_{i=1}^n X_i\right)\right] \\
&= \mathbb{E}\left[\prod_{i=1}^N \exp\left(i\left(\frac{t}{n^\alpha}\right)X_i\right)\right] \\
&= \prod_{i=1}^n \mathbb{E}\left[\exp\left(i\left(\frac{t}{n^\alpha}\right)X_i\right)\right] && \text{(by independence)} \\
&= \left(\mathbb{E}\left[\exp i\left(\frac{t}{n^\alpha}\right)X_1\right]\right)^n \\
&= \phi_{X_1}\left(\frac{t}{n^\alpha}\right)^n.
\end{aligned}$$

Expanding,

$$\begin{aligned}
\left(\mathbb{E}\left[\exp\left(i\left(\frac{t}{n^\alpha}\right)X_1\right)\right]\right)^n &= \left(1 + \frac{it}{n^\alpha}\mathbb{E}\left[X_1\right] - \frac{t^2}{2n^{2\alpha}}\mathbb{E}\left[X_1^2\right] - \frac{it^3}{3!n^{3\alpha}}\mathbb{E}\left[X_1^3\right] + \cdots\right)^n \\
&= \left(1 - \frac{t^2}{2n^{2\alpha}} + o\left(\frac{1}{n^{3\alpha}}\right)\right)^n \\
&= \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n^{3/2}}\right)\right)^n && (\text{if } \alpha = \frac{1}{2}) \\
&\xrightarrow{n\to\infty} e^{-t^2/2}.
\end{aligned}$$

Section 9.5 - Two Big Theorems: Uniqueness and Continuity

Remark: We now illustrate the above ideas as presented in *Probability: Theory & Examples* by Rick Durrett.

Notation: $F(x) = \mathbb{P}\{X \leq x\}$. The associated probability measure will be denoted $\mu(dx)$, so that

$$F(x) = \int_{-\infty}^x \mu(dx).$$

If F has a density, then

$$F(x) = \int_{-\infty}^x f(x)dx.$$

Uniqueness of chfs: We look at the Inversion Formula: Let

$$\phi(x) = \mathbb{E}\left[e^{itX}\right] = \int_{\mathbb{R}} e^{itx}\mu(dx).$$

If $a < b$, then

$$\lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it}\phi(t)dt \text{ exists}$$

and equals $\mu((a,b)) + \frac{1}{2}\mu(\{a,b\})$.

Proof:

Derivation:

Define

$$\begin{aligned} I_T &:= \int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it}\phi(t)dt \\ &= \int_{-T}^{T}\int_{\mathbb{R}} \frac{e^{-ita}-e^{-itb}}{it}e^{itx}\mu(dx)dt \\ &= \int_{\mathbb{R}}\int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it}e^{itx}dt\mu(dx) \\ &= \int_{\mathbb{R}}\left[\int_{-T}^{T}\frac{\sin(t(x-a))}{t}dt - \int_{-T}^{T}\frac{\sin(t(x-b))}{t}dt\right]\mu(dx) \\ &= \int_{\mathbb{R}}(R(x-a,T)-R(x-b,T))\mu(dx) \\ &= \int_{\mathbb{R}}(2\operatorname{sgn}(x-a)S(T|x-a|) - 2\operatorname{sgn}(x-b)S(T|x-b|))\mu(dx). \end{aligned}$$

Thoughts / Inspiration:

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

Use $e^{ix} = \cos(x) + i\sin(x)$ and note that the $\cos(x)$ terms cancel.

Define

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

Note if $\theta > 0$, then

$$R(\theta,T) = 2S(T\theta)$$

and if $\theta < 0$, then

$$R(\theta,T) = -R(|\theta|,T).$$

Hence,

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

Note $\lim_{T\to\infty}\int_0^T \frac{\sin(x)}{x}dx = \frac{\pi}{2}$. So,

$$\lim_{T\to\infty}(R(x-a,T)-R(x-b,T)) = \begin{cases} 2\pi, & x\in(a,b) \\ \pi, & x\in\{a,b\} \\ 0, & x\in(-\infty,a)\cup(b,\infty) \end{cases}.$$

Since $R(\theta, T) \leq 2 \sup_{y \in \mathbb{R}} S(y) < \infty$, by BCLT,

$$\frac{I_T}{2\pi} \to \mu((a,b)) + \frac{1}{2}\mu(\{a,b\}). \quad \square$$

Theorem: If $\int |\phi(t)| dt < \infty$, then μ has a density

$$f(y) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ity} \phi(t) dt.$$

The key formula is

$$\begin{aligned}
\mu((x, x+h)) &= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-itx} - e^{-it(x+h)}}{it} \phi(t) dt \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \int_x^{x+h} e^{-ity} dy \phi(t) dt \\
&= \int_x^{x+h} \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ity} \phi(t) dt dy.
\end{aligned}$$

Continuity Theorem: Suppose $\{\mu_n\}_{n \geq 1}$ is a sequence of probability measures with chfs $\{\phi(n)\}_{n \geq 1}$.

(i) If $\mu_n \Rightarrow \mu$, then $\phi_n(t) \to \phi(t)$ for all t.

(ii) If $\phi_n(t) \to \phi(t)$ for all t and for some ϕ which is continuous at 0, then the sequence of measures $\{\mu_n\}_{n \geq 1}$ is "tight" (see below), and $\mu_n \Rightarrow \mu$.

Example: Note that

$$\phi(0) = \mathbb{E}\left[e^{i \cdot 0 \cdot x}\right] = 1.$$

However, for example, let μ_n be the probability measure for a sequence of Gaussian random variables with mean 0 and variance n. Now, the density is

$$\frac{1}{n\sqrt{2\pi}} e^{-x^2/(2n^2)}$$

and the chf is

$$\phi_n(t) = e^{-n^2 t/2}.$$

Well, pointwise,

$$\phi(t) := \lim_{n \to \infty} \phi_n(t) = \begin{cases} 1, & t = 0 \\ 0, & t \neq 0 \end{cases}.$$

So, in this example, $\phi(t)$ is not continuous at $t = 0$.

Theorem: (Helly's Selection Theorem) Suppose F_n is a sequence of distribution functions. Then, there exists a subsequence F_{n_k} and a right continuous function F such that

$$\lim_{k \to \infty} F_{n_k}(y) = F(y)$$

for all $y \in \mathcal{C}(F)$.

Remark: F may not be a distribution function. For example, suppose $a + b + c = 1$ for $a, b, c > 0$ and define

$$F_n(x) = a\mathbf{1}_{\{X \geq n\}} + b\mathbf{1}_{\{X \geq -n\}} + cG(x)$$

where $G(x)$ is some distribution function. We have

$$F_n(x) \to \frac{b + cG(x)}{F(x)}$$

but

$$\lim_{x\to-\infty} F(x) = b > 0$$

and

$$\lim_{x\to\infty} F(x) = b + c < 1.$$

Proof: Let $\{q_i\}$ be an enumeration of the rationals. Note that for any k, $\{F_m(q_k)\}_{m\geq 1} \subseteq [0,1]$. Therefore, there exists a subsequence $m_k \subseteq m_{k-1}$ such that $F_{m_k(i)}(q_k)$ converges to some limit point, which we denote $G(q_k)$.

Now define $F_{n_k} := F_{m_k(k)}$. We claim that $F_{n_k}(q) \to G(q)$ for all q. Note that

$$F(x) = \inf_{\substack{q>x\\ q\in\mathbb{Q}}} G(q).$$

It remains to show that $F_n(x) \to F(x)$ for all $x \in \mathcal{C}(F)$. Let

$$r_1 < r_2 < x < s$$

for $r_1, r_2, s \in \mathbb{Q}$. Then,

$$F(x) - \epsilon < F(r_1) \leq F(r_2) < F(x) \leq F(s) \leq F(x + \epsilon).$$

Also

$$F_{n_k}(r_2) \to G(r_2) > F(r_1)$$

and

$$F_{n_k}(s) \to G(s) \leq F(s).$$

Hence

$$F_n(x) \to F(x). \ \square$$

Definition: A sequence of probability measures $\{\mu_n\}_{n\geq 1}$ is <u>tight</u> if for all $\epsilon > 0$, there exists a finite interval (a, b) such that

$$\mu_n((a, b]) > 1 - \epsilon$$

for all n sufficiently large.

Theorem: Every subsequential limit is the distribution function of a probability measure if and only if the sequence $\{F_n\}_{n\geq 1}$ is tight, i.e., for all ϵ, there exists M such that

$$\limsup_{n\to\infty}(1 - F_n(M)) - F_n(-M) \leq \epsilon.$$

Proof:

($\Longleftarrow$) Assume $\{F_n\}$ is tight and $F_{n_k} \to F$ weakly. Let $r, s \in \mathcal{C}(F)$ such that

$$r < -M < M < s.$$

Since $F_{n_k}(r) \to F(r)$ and $F_{n_k}(s) \to F(s)$, we have that

$$\begin{aligned} 1 - F(s) + F(r) &= \lim_{k\to\infty}(1 - F_{n_k}(s) + F_{n_k}(r)) \\ &\leq \limsup_{k\to\infty}(1 - F_{n_k}(s) + F_{n_k}(r)) \\ &\leq \epsilon. \ \square \end{aligned}$$

($\Longrightarrow$) (by contradiction) Suppose F_n is not tight. Then, there exists n_k such that

$$1 - F_{n_k}(k) + F_{n_k}(-k) \geq \epsilon.$$

But, by **Helly's Selection Theorem**, there exists n_{k_j} such that $F_{n_j} \Rightarrow F$ for some F. So,

$$\begin{aligned} 1 - F(s) + F(r) &= \lim_{j\to\infty} (1 - F_{n_{k_j}}(s) + F_{n_{k_j}}(r)) \\ &\geq \liminf_{j\to\infty} (1 - F_{n_{k_j}}(k_j) + F_{n_{k_j}}(-k_j)) \\ &\geq \epsilon. \end{aligned}$$

Hence, F is not a distribution function. □

Remark: We now prove the **Continuity Theorem** that we stated above.

Continuity Theorem: Suppose $\{\mu_n\}_{n\geq 1}$ is a sequence of probability measures with chfs $\{\phi(n)\}_{n\geq 1}$.

(i) If $\mu_n \Rightarrow \mu$, then $\phi_n(t) \to \phi(t)$ for all t.

(ii) If $\phi_n(t) \to \phi(t)$ for all t and for some ϕ which is continuous at 0, then the sequence of measures $\{\mu_n\}_{n\geq 1}$ is "tight" (see below), and $\mu_n \Rightarrow \mu$.

Proof: Proving (i) is easy, using the **Continuous Mapping Theorem**. Note that $e^{i\theta}$ is bounded and continuous. Hence, for all t,

$$\begin{array}{ccc} \mathbb{E}\left[e^{itX_n}\right] & \longrightarrow & \mathbb{E}\left[e^{itX}\right] \\ \| & & \| \\ \phi_n(t) & \longrightarrow & \phi(t). \end{array}$$

To prove (ii), we'll first prove tightness by means of the estimate

$$\mu_n \left\{ x \mid |x| > \frac{2}{u} \right\} \leq \frac{1}{u} \int_{-u}^{u} (1 - \phi_n(t))dt \leq 2\epsilon$$

for sufficiently large n.

Note that

$$\begin{aligned} \frac{1}{u}\int_{-u}^{u}(1 - e^{itx})dt &= 2 - \frac{1}{u}\int_{-u}^{u}(\cos(tx) + i\sin(tx))dt \\ &= 2 - \frac{1}{u}\left[\frac{\sin(tx)}{x}\right]_{t=-u}^{t=u} \\ &= 2 - 2\frac{\sin(ux)}{x}. \end{aligned}$$

Now, integrate both sides with respect to $\mu_n(dx)$. Then,

$$\frac{1}{u}\int_{-u}^{u}(1 - \phi_n(t))dt = 2 - 2\int_{\mathbb{R}} \frac{\sin(ux)}{ux}\mu_n(dx).$$

We can consider two estimates for the integral, for two different regions. On the interval $\left(-\frac{2}{u}, \frac{2}{u}\right)$, we have that

$$1 - \frac{\sin(ux)}{ux} \geq 0.$$

On the remaining region $\left(-\infty, -\frac{2}{u}\right) \cup \left(\frac{2}{u}, \infty\right)$,

$$\left|\frac{\sin(ux)}{ux}\right| \leq \frac{1}{|ux|}.$$

Hence,

$$\begin{aligned}
\frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt &= 2 - 2\int_{\mathbb{R}} \frac{\sin(ux)}{ux}\mu_n(dx) \\
&= 2\int_{\mathbb{R}}\left(1-\frac{\sin(ux)}{ux}\right)\mu_n(dx) \\
&\geq \int_{\{x:|x|>\frac{2}{u}\}}\left(1-\frac{1}{|ux|}\right)\mu_n(dx) \\
&\geq 2\int_{\{x:|x|>\frac{2}{u}\}}\frac{1}{2}\mu_n(dx) \\
&= \mu_n\left\{x : |x| > \frac{2}{u}\right\}.
\end{aligned}$$

Recall the assumption that $\phi(t)$ is continuous at $t = 0$. Also note that

$$\phi(0) = \mathbb{E}\left[e^{i(0)X}\right] = 1.$$

So, there exists $\delta > 0$ such that for all $t \in [-\delta, \delta]$, we have

$$|1-\phi(t)| < \frac{\epsilon}{2}.$$

Pick u such that $0 < u < \min\left(\delta, \frac{\epsilon}{2}\right)$. Then,

$$\left|\frac{1}{u}\int_{-u}^{u}(1-\phi(t))dt\right| \leq \frac{1}{u}\int_{-u}^{u}\frac{\epsilon}{2}\,dt = \epsilon.$$

Now, since $\phi_n(t) \to \phi(t)$ for all t, this implies by the **Dominated Convergence Theorem** that

$$\frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt \to \frac{1}{u}\int_{-u}^{u}(1-\phi(t))dt.$$

Therefore, for sufficiently large n, we have that

$$\left|\frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))\,dt\right| \leq 2\epsilon.$$

Therefore, $\{\mu_n\}$ is tight.

Fact: by (i), if a subsequence $\{\mu_{n_k}\}$ is convergent, then it must converge to μ. Also, by tightness, every subsequence contains a further subsequence that converges (which, by (i), converges to μ).

We claim that this implies that $\mu_n \Rightarrow \mu$.

Lemma: Let y_n be a sequence of elements in a topological space. If every subsequence y_{n_k} has a further subsequence $y_{n_{k_j}}$ such that $y_{n_{k_j}} \to y$, then $y_n \to y$.

> **Proof:** Suppose not to the contrary that $y_n \not\to y$. Then, there exists an open set G containing y and a subsequence y_{n_m} such that $y_{n_m} \notin G$ for all n_m. But, by hypothesis, there exists a further subsequence $y_{n_{m_\ell}} \to y$. This is a contradiction, since this would require $y_{n_{m_\ell}}$ to eventually visit G for all but finitely many ℓ. □

Remark: The smoothness of ϕ at 0 is related to the decay of μ at $\pm\infty$.

Example: If $\int |x|^n \mu(dx) < \infty$, then the chf ϕ of μ has derivatives of order n given by

$$\phi^{(n)}(t) = \int (ix)^n e^{itx} \mu(dx).$$

This implies that if X has n finite moments, then

$$\phi(t) = \sum_{k=0}^{n} \frac{\mathbb{E}\left[(itX)^k\right]}{k!} + o(t^n)$$

where $f \sim o(t^n)$ if $\lim_{t\to 0} \frac{f(t)}{t^n} = 0$.

To be precise, one can show that

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

For small values of x, the first term in the minimum is small. For large values of x, the second term in the minimum is small. Hence,

$$\begin{aligned} \left| \mathbb{E}\left[e^{itX}\right] - \sum_{k=0}^{n} \frac{\mathbb{E}\left[(itX)^k\right]}{k!} \right| &\leq \mathbb{E}\left[\left| e^{itX} - \sum_{k=0}^{n} \frac{(itX)^k}{k!} \right| \right] \\ &\leq t^n \mathbb{E}\left[\min\left(\frac{t|X|^{n+1}}{(n+1)!}, \frac{2|X|^n}{n!} \right) \right]. \end{aligned}$$

Note that this is $o(t^n)$ because

$$\lim_{t\to 0} \mathbb{E}\left[\min\left(\frac{t|X|^{n+1}}{(n+1)!}, \frac{2|X|^n}{n!} \right) \right] = 0,$$

since for any fixed t,

$$\mathbb{E}\left[\min\left(\frac{t|X|^{n+1}}{(n+1)!}, \frac{2|X|^n}{n!} \right) \right] = \int_{|x|<\widehat{x}} \frac{t|X|^{n+1}}{(n+1)!} \mu(dx) + \int_{x>\widehat{x}}^{\infty} \frac{2|X|^n}{n!} \mu(dx),$$

where

$$\widehat{x} := \frac{2(n+1)}{t}.$$

So, as $t \to 0$, we see that $|\widehat{x}| \to \infty$. Hence, the integrands are converging to the first term of the minimum.

Corollary: Suppose $\mathbb{E}[X] = \mu$ and $\mathbb{E}[X^2] = \sigma^2 < \infty$. Then, $\phi(t) = 1 + it\mu - \frac{t^2\sigma^2}{2} + o(t^2)$.

Remark: We've shown that if we have n moments, then $\phi^{(n)}$ exists. However, $\phi'(0)$ may exist even if $\mathbb{E}[|X|] = \infty$. Despite this, we have the following proposition.

Proposition: Suppose that $\limsup_{h\searrow 0} \frac{\phi(h) - 2\phi(0) + \phi(-h)}{h^2} > -\infty$. Then, $\mathbb{E}\left[|X|^2\right] < \infty$.

Proof: Firstly,

$$\frac{e^{ihx} - 2 + e^{-ihx}}{h^2} = \frac{\left(e^{ihx/2} - e^{-ihx/2}\right)^2}{h^2} = \frac{-4\sin^2\left(\frac{hx}{2}\right)}{h^2}.$$

Hence,

$$\infty > -\limsup_{h\searrow 0} \frac{\phi(h)-2\phi(0)+\phi(-h)}{h^2} = \liminf_{h\to 0} 4\int \frac{\sin^2\left(\frac{hx}{2}\right)}{h^2}\mu(dx) \geq \int |x|^2\mu(dx) = \mathbb{E}\left[|X|^2\right]. \ \square$$

Central Limit Theorem for iid Sequences: Suppose $\{X_n\}_{n\geq 1}$ is iid with $\mathbb{E}[X_n] = \mu$ and $\text{Var}[X_n] = \sigma^2 \in (0,\infty)$. Then,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow Z$$

where $Z \sim N(0,1)$.

Proof: Without loss of generality, suppose $\mu = 0$. (Otherwise, define $\sim X_n = X_n - \mu$.) Then,

$$\phi(t) = \mathbb{E}\left[e^{itX_1}\right] = 1 - \frac{\sigma^2 t^2}{2} + o(t^2).$$

Also,

$$\begin{aligned}
\mathbb{E}\left[e^{it\frac{S_n}{\sigma\sqrt{n}}}\right] &= \mathbb{E}\left[e^{i\frac{t}{\sigma\sqrt{n}}X_1}e^{i\frac{t}{\sigma\sqrt{n}}X_2}\cdots e^{i\frac{t}{\sigma\sqrt{n}}X_n}\right] \\
&= \left(1 - \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + o\left(\frac{t^2}{\sigma^2 n}\right)\right)^n \\
&= \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n. \qquad \text{(since } t \text{ is fixed)}
\end{aligned}$$

We now claim that as $n \to \infty$, the quantity above converges to $e^{-t^2/2}$.

Lemma: Suppose that $c_n \to c$, where $c_n, c \in \mathbb{C}$. Then,

$$\left(1 + \frac{c_n}{n}\right)^n \to e^c.$$

Proof of Lemma: Let $\{z_1,\ldots,z_n\}, \{w_1,\ldots,w_n\} \subset \mathbb{C}$ with modulus $\leq \theta \in \mathbb{R}_+$. Then,

$$\left|\prod_{k=1}^n z_k - \prod_{k=1}^n w_k\right| \leq \theta^{n-1}\sum_{k=1}^n |z_k - w_k|.$$

(The proof of this fact follows by induction.) Let

$$z_k = \left(1 + \frac{c_n}{n}\right)$$

and let

$$w_k = e^{c_n/n}.$$

Pick $\gamma > |c|$ so that for sufficiently large n, we have $|c_n| < \gamma$ and $\left|\frac{c_n}{n}\right| < 1$. So,

$$\begin{aligned}
\left|\left(1 + \frac{c_n}{n}\right)^n - e^{c_n}\right| &\leq \left(1 + \frac{\gamma}{n}\right)^{n-1}\sum_{k=1}^n \left|1 + \frac{c_n}{n} - e^{c_n/n}\right| \\
&\leq \left(1 + \frac{\gamma}{n}\right)^{n-1} n\left(\frac{c_n}{n}\right)^2 \\
&\leq e^{\gamma}\left(\frac{\gamma^2}{2n}\right) \to 0. \ \square
\end{aligned}$$

Now, the theorem follows immediately if we set $c_n = -\frac{t^2}{2n}$. $\square$

Section 9.8 - The Lindeberg-Feller CLT

Lindeberg-Feller Theorem: For every n, let $\{X_{n,k}\}$ for $1 \le k \le n$ be independent random variables with $\mathbb{E}[X_{n,k}] = 0$. Suppose that

(i) $\sum_{k=1}^{n} \mathbb{E}\left[X_{n,k}^2\right] \to \sigma^2 > 0$, and

(ii) For all $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left[|X_{n,k}|^2 \mathbf{1}_{\{|X_{n,k}|\ge\epsilon\}}\right] = 0.$$

Then,

$$S_n := \sum_{k=1}^{n} X_{n,k} \Rightarrow N(0,\sigma^2).$$

Historically, this was phrased: "The sum of a large number of small independent erros is approximately normal."

Proof: By the continuity theorem, it suffices to show that

$$\prod_{k=1}^{n} \phi_{n,k}(t) = \exp\left(\frac{-t^2\sigma^2}{2}\right).$$

Observe that

$$\begin{aligned}
\left|\phi_{n,k}(t) - \left(1 - \frac{-t^2\sigma^2}{2}\right)\right| &= \left|\mathbb{E}\left[e^{itX_{n,k}}\right] - \left(1 - \frac{-t^2\mathbb{E}\left[X_{n,k}^2\right]}{1}\right)\right| \\
&\le \left|\mathbb{E}\left[e^{itX_{n,k}}\right] - \left(1 - \mathbb{E}\left[\frac{-t^2X_{n,k}^2}{2}\right]\right)\right| && \text{(Jensen's Inequality)} \\
&\le \mathbb{E}\left[\min\left(\frac{|tX_{n,k}|^3}{3!}, \frac{2|tX_{n,k}|^2}{2!}\right)\right] \\
&\le \mathbb{E}\left[\frac{|t|^3|X_{n,k}|^3}{6}\mathbf{1}_{\{|X_{n,k}|<\epsilon\}} + |t|^2|X_{n,k}|^2\mathbf{1}_{\{|X_{n,k}|\ge\epsilon\}}\right] \\
&\le \frac{|t|^3\epsilon}{6}\mathbb{E}\left[|X_{n,k}|^2\mathbf{1}_{\{|X_{n,k}|<\epsilon\}}\right] + t^2\mathbb{E}\left[|X_{n,k}|^2\mathbf{1}_{\{|X_{n,k}|\ge\epsilon\}}\right] \\
&\to 0.
\end{aligned}$$

Therefore,

$$\phi_{n,k}(t) \xrightarrow{n\to\infty} 1 - \frac{t^2\sigma_{n,k}^2}{2}.$$

Earlier, we showed that if $\{z_n\}$ and $\{w_n\}$ have modulus $\le \theta$, then

$$\left|\prod_{k=1}^{n} z_k - \prod_{k=1}^{n} w_k\right| \le \theta^{n-1}\sum_{k=1}^{n}|z_k - w_k|.$$

Choose $z_{n,k} := \phi_{n,k}(t)$ and $w_{n,k} := 1 - \dfrac{t^2\sigma_{n,k}^2}{2}$. We can then show (after some work) that

$$\left|\prod_{k=1}^{n}\phi_{n,k}(t) - \prod_{k=1}^{n}\left(1 - \frac{t^2\sigma_{n,k}^2}{2}\right)\right| \to 0.$$

Exercise: If $\max_{1\le k\le n} c_{n,k} \to 0$ and $\sum_{k=1}^{n} c_{n,k} \to \lambda$, then

$$\prod_{k=1}^{n}(1+c_{n,k}) \to e^{\lambda}.$$

So, using the exercise, take $c_{n,k} := -\frac{t^2\sigma_{n,k}^2}{2}$. This completes the theorem. □

Theorem: Without the hypothesis that the random variables are identically distributed, suppose $\mathbb{E}[X_k] = 0$ and $\mathbb{E}\left[X_k^2\right] = \sigma^2$, and suppose that there exists $\delta > 0$ such that $\mathbb{E}\left[X_k^{2+\delta}\right] \le c$. Then,

$$\frac{S_n}{\sqrt{n}} \Rightarrow N(0,\sigma^2).$$

Converse of Three Series Theorem: Let $\{X_k\}$ be independent and define

$$\widetilde{X_k} := X_k \mathbf{1}_{\{|X_k|\le c\}}.$$

Then,

(i) $\sum_{k=1}^{\infty} \mathbb{P}\{|X_k| > c\} < \infty$,

(ii) $\sum_{k=1}^{\infty} \mathbb{E}\left[\widetilde{X_k}\right]$ converges,

(iii) $\sum_{k=1}^{\infty} \mathrm{Var}\left(\widetilde{X_k}\right)$ converges.

Suppose (i) fails. Then, since $\sum_{k=1}^{\infty} \mathbb{P}\{|X_k| > c\} = \infty$, we have that

$$\mathbb{P}\{|X_k| > c \text{ infinitely often}\} = 1.$$

Hence, $\sum_{k=1}^{\infty} X_k$ cannot converge, a contradiction.

Suppose (iii) fails but (i) holds. [See book.]

Chapter 10

Chapter 10 - Martingales

Section 10.1 - The Radon-Nikodym Theorem

Definition: A Martingale prcoess is fair game, i.e., it has the property

$$\mathbb{E}\left[X_n \mid \mathcal{F}_m\right] = X_m,$$

where $\mathcal{F}_m$ is everything that has happened as of time $m \leq n$.

Example: Let X, Y be the outcomes of two independent fair dice. Define $S := X + Y$ and $M := \min(X, Y)$. Then,

$$\begin{aligned}
\mathbb{P}\left\{S = 5 \mid X = 2\right\} &= \frac{\mathbb{P}\left\{S = 5 \wedge X = 2\right\}}{\mathbb{P}\left\{X = 2\right\}} \\
&= \frac{\mathbb{P}\left\{Y = 3 \wedge X = 2\right\}}{\mathbb{P}\left\{X = 2\right\}} \\
&= \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}.
\end{aligned}$$

We can also calculate that, for example,

$$\mathbb{P}\left\{M = 5 \mid X = 2\right\} = 0.$$

Consider the conditional expectations $\mathbb{E}\left[S \mid X\right]$ and $\mathbb{E}\left[M \mid X\right]$. We compute the intermediate step:

$$\begin{aligned}
\mathbb{E}\left[S \mid X = 1\right] &= \sum_{x=0}^{\infty} k \cdot \mathbb{P}\left\{S = k \mid X = 1\right\} \\
&= \sum_{k=2}^{7} k \cdot \frac{\mathbb{P}\left\{Y = k - 1 \wedge X = 1\right\}}{\mathbb{P}\left\{X = 1\right\}} \\
&= \sum_{i=1}^{6} (i + 1) \cdot \frac{\mathbb{P}\left\{Y = i \wedge X = 1\right\}}{\mathbb{P}\left\{X = 1\right\}} \\
&= \mathbb{E}\left[Y\right] + 1.
\end{aligned}$$

In general, for $k \in [6]$,

$$\mathbb{E}\left[S \mid X = k\right] = \mathbb{E}[Y] + k$$

and so we say

$$\mathbb{E}\left[S \mid X\right] = \mathbb{E}[Y] + X.$$

Similarly, for $\mathbb{E}[M \mid X]$,

$$\mathbb{E}[M \mid X = 1] = 1$$

$$\mathbb{E}[M \mid X = 2] = 1 \cdot \mathbb{P}\{Y = 1\} + 2 \cdot \mathbb{P}\{Y \in \{2,3,4,5,6\}\} = \frac{11}{6}$$

$$\mathbb{E}[M \mid X = 3] = 1 \cdot \mathbb{P}\{Y = 1\} + 2 \cdot \mathbb{P}\{Y = 2\} + 3 \cdot \mathbb{P}\{Y \in \{3,4,5,6\}\} = \frac{15}{6}$$

$$\mathbb{E}[M \mid X = 4] = \frac{18}{6}$$

$$\mathbb{E}[M \mid X = 5] = \frac{20}{6}$$

$$\mathbb{E}[M \mid X = 6] = \frac{21}{6}.$$

So,

$$\mathbb{E}[M \mid X] = \begin{cases} 1, & \text{with probability } 1/6 \\ \frac{11}{6}, & \text{with probability } 1/6 \\ \frac{15}{6}, & \text{with probability } 1/6 \\ \frac{18}{6}, & \text{with probability } 1/6 \\ \frac{20}{6}, & \text{with probability } 1/6 \\ \frac{21}{6}, & \text{with probability } 1/6 \end{cases}.$$

Thus,

$$\mathbb{E}[\mathbb{E}[M \mid X]] = \int \mathbb{E}[M \mid X] \, d\mathbb{P}^x = \frac{91}{36}.$$

We can also calculate that

$$\mathbb{E}[M] = \frac{91}{36}.$$

In general, if $\mathcal{G} \subseteq \mathcal{F}$, then for all $A \in \mathcal{G}$,

$$\int_A M d\mathbb{P} = \int_A \mathbb{E}[M \mid \mathcal{G}] \, d\mathbb{P}.$$

Example: Let $T_1, T_2 \sim \text{Exp}(\lambda)$ be iid. Set $S = T_1 + T_2$ and $M = \min(T_1, T_2)$. Then, $\mathbb{P}\{T_1 \leq t\} = 1 - e^{-\lambda t}$ and the density is $f(t) = \lambda e^{-\lambda t}$. Then, we have the conditional probability "formula"

$$\mathbb{P}\{S \leq t \mid T_1 = t_1\} \text{ “} = \text{”} \frac{\mathbb{P}\{S \leq t \wedge T = t_1\}}{\mathbb{P}\{T = t_1\}}.$$

Remember that

$$\mathbb{P}\{T_1 \in [t, t + \Delta]\} = \int_t^{t+\Delta} f(s) ds.$$

So,

$$\begin{aligned} \mathbb{P}\{S \in I \mid T \in I_1\} &= \frac{\mathbb{P}\{S \in I \wedge T_1 \in I_1\}}{\mathbb{P}\{T_1 \in I_1\}} \\ &= \frac{\mathbb{P}\{T_2 \in I \setminus I_1 \wedge T_1 \in I_1\}}{\mathbb{P}\{T_1 \in I_1\}}. \end{aligned}$$

Remark: To construct conditional expectation, we need the notion of a Radon-Nikodym derivative.

Definition: Let $(\Omega, \mathcal{F})$ be a measure space and suppose μ and ν are two positive bounded measures on $(\Omega, \mathcal{F})$. We say that ν is absolutely continuous with respect to μ, denotes $\nu \ll \mu$ if

$$[\mu(A) = 0] \implies [\nu(A) = 0].$$

Example: The binomial distribution ν (with n trials an success probability p) is not absolutely continuous with respect to the Gaussian distribution, since $\mu(1) = 0$, while $\nu(1) = np(1-p)^{n-1} > 0$.

Example: On the other hand, if λ is the Lebesgue measure on $[0,1]$, then $\lambda \ll \mu$, but $\mu \not\ll \lambda$, because, for example, $\lambda((-1,0)) = 0$ but $\mu((-1,0)) > 0$.

Remark: Any two Gaussian measures are mutually absolutely continuous in finite dimensions, but it is possible for them to be mutually singular in infinite dimension.

Definition: We say that ν concentrates on $A \in \mathcal{F}$ if $\nu(A^C) = 0$. We say that ν and μ are mutually singular, denotes $\mu \perp \nu$, if there exist events $A, B \in \mathcal{F}$ such that $A \cap B = \emptyset$, and ν concentrates on A, while μ concentrates on B.

Theorem 10.1.1: (Lebesgue Decomposition Theorem) Suppose that μ and ν are positive, bounded measures on $(\Omega, \mathcal{F})$.

(a) There exists a unique pair of positive bounded measures λ_a and λ_s on $\mathcal{F}$ such that

$$\lambda = \lambda_a + \lambda_s$$

and $\lambda_a \ll \mu$, while $\lambda_s \perp \mu$. Furthermore $\lambda_a \perp \lambda_s$.

(b) There exists a nonnegative $\mathcal{F}$-measurable function X with

$$\int X d\mu < \infty$$

such that

$$\lambda_a(E) = \int_E X d\mu$$

for all $E \in \mathcal{F}$, and X is unique up to sets of μ-measure 0.

Radon-Nikodym Theorem Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Suppose ν is a positive bounded measure, and $\nu \ll \mathbb{P}$. Then, there exists an integrable random variable $X \in \mathcal{F}$ such that

$$\nu(E) = \int_E X d\mathbb{P}$$

for all $E \in \mathcal{F}$. X is $\mathbb{P}$-almost-surely unique, and is written

$$X = \frac{d\nu}{d\mathbb{P}},$$

or

$$d\nu = X d\mathbb{P}.$$

Sketch of Proof: Consider the following proposition.

Proposition: Let $\mathbb{H}$ be a Hilbert space with inner product $(\cdot,\cdot)$. If $L : \mathbb{H} \to \mathbb{R}$ is a lienar functional on $\mathbb{H}$, then there exists a unique $y \in \mathbb{H}$ such that $L(x) = (x, y)$.

Proof: If $L(x) = 0$, then $y = 0$, and the proof is complete. Otherwise, without loss of generality, suppose $\mathbb{H}$ is real and define

$$M := x \in \mathbb{H} : L(x) = 0.$$

Since L is linear, M is a subspace. Since L is continuous, M is closed. Since $L \not\equiv 0$, $M \neq \mathbb{H}$.

Therefore, for all $z' \notin M$, by the **Projection Theorem** we have that there exists $z_1 \in M$ and $z_2 \in M^\perp$ such that

$$z' = z_1 + z_2.$$

Hence $M^\perp$ is nontrivial.

For some $z \notin M$, define

$$y := \frac{L(z)}{(z,z)} z.$$

It follows that

$$L(y) = \frac{L(z)}{(z,z)} L(z).$$

Meanwhile,

$$(y,y) = \frac{L(z)^2}{(z,z)^2}(z,z) = \frac{L(z)^2}{(z,z)} = L(y).$$

Define x_1 and x_2 such that

$$x = \underbrace{\left(x - \frac{L(x)}{(y,y)} y\right)}_{=:x_1} + \underbrace{\left(\frac{L(x)}{(y,y)} y\right)}_{=:x_2}.$$

Now,

$$\begin{aligned}
L(x_1) &= L(x) - \frac{L(x)}{(y,y)}(y,y) = 0,\\
L(x_2) &= L(x)\frac{L(y)}{(y,y)} = L(x),\\
(x,y) &= (x_1,y) + (x_2,y) = 0 + (x_2,y) = \frac{L(x)(y,y)}{(y,y)} = L(x).
\end{aligned}$$

Uniqueness can alao be shown. □

Lemma: (Integral Comparison) Consider $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathcal{G} \subseteq \mathcal{F}$ a σ-subfield. Suppose $X, Y \in \mathcal{G}$ are integrable. Then, $X = Y$ almost surely if and only if

$$\int_A X d\mathbb{P} = \int_A Y d\mathbb{P}$$

for all $A \in \mathcal{G}$.

Proof: Supose $\nu \ll \mathbb{P}$ and define

$$\mathbb{Q}(A) := \frac{\nu(A)}{\nu(\mathbb{P})}.$$

Then, $\mathbb{Q}$ is a probability measure and $\mathbb{Q} \ll \mathbb{P}$. Set

$$\mathbb{P}^* = \frac{1}{2}(\mathbb{P} + \mathbb{Q}).$$

Define $\mathbb{H} := L_2(\mathbb{P}^*)$, i.e., $\mathbb{H}$ contains all $X \in \mathcal{F}$ such that

$$\int_\Omega X^2 d\mathbb{P}^* < \infty.$$

Consider the inner product

$$(Y_1, Y_1) := \int_\Omega Y_1 Y_2 d\mathbb{P}^*.$$

Note by **Cauchy-Schwarz** that

$$\int_\Omega Y_1 Y_2 d\mathbb{P}^* \leq \left(\int_\Omega Y_1^2 d\mathbb{P}^*\right)^{1/2} \left(\int_\Omega Y_2^2 d\mathbb{P}^*\right)^{1/2} < \infty.$$

Define for $Y \in L_2(P^*)$,

$$L(Y^*) := \int_\Omega Y dQ.$$

By the previous **Proposition**, there exists $Z \in L_2(\mathbb{P}^*)$ such that

$$L(Y) = (Y, Z) = \int_\Omega YZ d\mathbb{P}^* = \frac{1}{2}\int_\Omega YZ d\mathbb{P} + \frac{1}{2}\int_\Omega YZ dQ.$$

By the definition of Z,

$$\frac{1}{2}\int_\Omega YZ d\mathbb{P} = \int_\Omega Y dQ.$$

So,

$$\int_\Omega Y\left(1 - \frac{Z}{2}\right) dQ = \int_\Omega \frac{YZ}{2} d\mathbb{P}.$$

Let $A \in \mathcal{F}$, and define $Y = I_A$. Then,

$$\int_A \left(1 - \frac{Z}{2}\right) dQ = \int_A \frac{Z}{2} d\mathbb{P}.$$

Thus,

$$\int_A Y dQ = Q(A) = \int_A Z d\mathbb{P}^*.$$

Next (see book for details),

$$Y = \left(\frac{Z}{2}\right)^n 1_A$$

and so

$$Q(A) = \int_A \frac{Z}{2 - Z} d\mathbb{P}.$$

Corollary: Suppose $\mathbb{Q}$ and $\mathbb{P}$ are probability measures on $(\Omega, \mathcal{F})$, and suppose that $\mathbb{Q} \ll \mathbb{P}$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-subfield, and let $\mathbb{Q}\big|_\mathcal{G}$ and $\mathbb{P}\big|_\mathcal{G}$ be restrictions of $\mathbb{Q}$ and $\mathbb{P}$ to $\mathcal{G}$. Then, in the space $(\Omega, \mathcal{G})$,

$$\mathbb{Q}\big|_\mathcal{G} \ll \mathbb{P}\big|_\mathcal{G}$$

and

$$\frac{d\mathbb{Q}\big|_\mathcal{G}}{d\mathbb{P}\big|_\mathcal{G}} \text{ is } \mathcal{G}\text{-measurable.}$$

Remark: This relates to conditional expectation in the following way. Let $X \in L_1$. Define

$$\nu(A) := \int_A X d\mathbb{P}$$

for all $A \in \mathcal{F}$. Then, ν is finite and $\nu \ll \mathbb{P}$. We want it to be true that

$$\int_G X d\mathbb{P} = \int_G \mathbb{E}\left[X \mid \mathcal{G}\right] d\mathbb{P}$$

for all $G \in \mathcal{G}$, so

$$\mathbb{E}\left[X \mid \mathcal{G}\right] = \frac{d\nu\big|_{\mathcal{G}}}{d\mathbb{P}\big|_{\mathcal{G}}},$$

which exists by the **Radon-Nikodym Theorem**.

Sections 10.2 / 10.3 - Definition / Properties of Conditional Expectation

Definition: Conditional Probability is defined as follows, on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and σ-subfield $\mathcal{G}$.

$$\mathbb{P}\left\{A \mid \mathcal{G}\right\} = \mathbb{E}\left[1_A \mid \mathcal{G}\right].$$

(a) $\mathbb{P}\left\{A \mid \mathcal{G}\right\}$ is $\mathcal{G}$-measurable and integrable.

(b) $\displaystyle\int_G \mathbb{P}\left\{A \mid \mathcal{G}\right\} d\mathbb{P} = \mathbb{P}\left\{A \cap \mathcal{G}\right\}$.

Properties: (for more details, read Section 10.3 of the textbook)

(i) Linearity

(ii) If $X \in \mathcal{G}$ and $X \in L_1$, then $\mathbb{E}\left[X \mid \mathcal{G}\right] = X$ almost surely.

(iii) $\mathbb{E}\left[X \mid \{\emptyset, \Omega\}\right] = \mathbb{E}\left[X\right]$.

(iv) Monotonicity

(v) Modulus Inequality:

$$\mathbb{E}\left[X \mid \mathcal{G}\right] \leq \mathbb{E}\left[|X| \mid \mathcal{G}\right].$$

(vi) Monotone Convergence Theorem

(vii) Fatou's Lemma

(viii) Dominated Convergence Theorem

(ix) Product Rule: Suppose X and Y are random variables such that $X, XY \in L_1$. If $Y \in \mathcal{G}$, then

$$\mathbb{E}\left[XY \mid G\right] = Y\mathbb{E}\left[X \mid G\right] \text{ almost surely.}$$

Let $A \in \mathcal{G}$. Suppose we know that for any $A \in \mathcal{G}$, we have

$$\int_A Y\mathbb{E}\left[X \mid \mathcal{G}\right] d\mathbb{P} = \int_A XY d\mathbb{P}. \tag{$\bigstar$}$$

Then,

$$\int_A Y\mathbb{E}\left[X \mid \mathcal{G}\right] d\mathbb{P} = \int_A XY d\mathbb{P} = \int_A \mathbb{E}\left[XY \mid \mathcal{G}\right] d\mathbb{P}.$$

Now, assume $Y = 1_\Delta$, where $\Delta \in \mathcal{G}$. Then, $A \cap \Delta \in \mathcal{G}$, and

$$\begin{aligned}
\int_A Y\mathbb{E}[X \mid \mathcal{G}]\, d\mathbb{P} &= \int_{A\cap\Delta} \mathbb{E}[X \mid \mathcal{G}]\, d\mathbb{P} && \text{(by the definition of } Y\text{)} \\
&= \int_{A\cap\Delta} X d\mathbb{P} && \text{(by conditional expectation)} \\
&= \int_A XY d\mathbb{P}.
\end{aligned}$$

This shows that $(*)$ holds for simple random variables.

Next, suppose that

$$Y = \sum_{i=1}^{k} c_i 1_{\Delta_i}$$

holds by linearity. The, suppose that X and Y are nonnegative. Then, there exists a sequence of simple random variables Y_n such that $Y_n \nearrow Y$. Use Monotone Convergence. (The proof in the general case follows the typical line of reasoning.)

(x) Smoothing: Suppose $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$. Then, if $X \in L_1$¡ we have

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \mathbb{E}[X \mid \mathcal{G}_1] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_1] \mid \mathcal{G}_2].$$

Colloquially, the smaller σ-field always wins. In particular,

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_1]] = \mathbb{E}[X].$$

Proof: Take $A \in \mathcal{G}_1$. Then, $\mathbb{E}[X \mid \mathcal{G}_1]$ is $\mathcal{G}_1$-measurable, and

$$\begin{aligned}
\int_A \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1]\, d\mathbb{P} &= \int_A \mathbb{E}[X \mid G_2]\, d\mathbb{P} && \text{(by definition)} \\
&= \int_A X d\mathbb{P} && \text{(since } A \subseteq \mathcal{G}_1 \subseteq \mathcal{G}_2\text{)} \\
&= \int_A \mathbb{E}[X \mid \mathcal{G}_1]\, d\mathbb{P}. \ \square && \text{(by definition)}
\end{aligned}$$

Theorem: Suppose $\mathbb{E}\left[X^2\right] < \infty$. Then, $\mathbb{E}[X \mid \mathcal{G}]$ is the random variable $Y \in \mathcal{G}$ that minimizes the mean-squared error $\mathbb{E}\left[(X-Y)^2\right]$.

Proof: Let $Z \in L^2(\mathcal{G})$, where

$$L^2(\mathcal{G}) = \{Y \in \mathcal{G} \mid \mathbb{E}\left[Y^2\right] < \infty\}.$$

Then, by the Product Rule,

$$Z\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[ZX \mid \mathcal{G}].$$

Note that $\mathbb{E}[|ZX|] < \infty$ by **Cauchy-Schwarz**. Taking expectations,

$$\mathbb{E}[Z\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[\mathbb{E}[ZX \mid \mathcal{G}]] = \mathbb{E}[ZX].$$

Thus,

$$\mathbb{E}[Z(X - \mathbb{E}[X \mid \mathcal{G}])] = 0$$

for all $Z \in L^2(\mathcal{G})$. Now suppose that $Y \in L^2(\mathcal{G})$ and define

$$Z = Y - \mathbb{E}[X \mid \mathcal{G}].$$

Thus,

$$\begin{aligned}\mathbb{E}\left[(X-Y)^2\right] &= \mathbb{E}\left[(X-(Z+\mathbb{E}[X\mid\mathcal{G}]))^2\right]\\ &= \mathbb{E}\left[((X-\mathbb{E}[X\mid\mathcal{G}])-Z)^2\right]\\ &= \mathbb{E}\left[(X-\mathbb{E}[X\mid\mathcal{G}])^2\right]-2\underbrace{\mathbb{E}\left[Z\left(X-\mathbb{E}[X\mid\mathcal{G}]\right)\right]}_{=0}+\mathbb{E}\left[Z^2\right].\end{aligned}$$

This quantity is minimized when $\mathbb{E}\left[Z^2\right]=0$. □

Sections 10.4 / 10.5 - Martingales and Examples

Common Betting Scheme: Double-down until you win. We now express this formally.

Let $\{X_i\}_{i=1}^{\infty}$ be the outcomes of a fair game, taking values in $\{-1,1\}$ with probability $\frac{1}{2}$.

Let $\{B_i\}_{i=1}^{\infty}$ be the bets placed on each game.

Let $\{S_i\}_{i=1}^{\infty}$ be the running total money in hand, so that

$$S_n = S_0 + \sum_{i=1}^{n} B_i X_i.$$

Define $S_0 = 0$ (thus assuming we can go into debt).

Define $\tau := \min\{i : X_i = 1\}$. We set $B_1 := 1$ and we have the recurrence

$$B_i = \begin{cases} 2^{i-1}, & \tau > i-1 \\ 0, & \text{otherwise} \end{cases}.$$

By definition,

$$S_\tau = 1.$$

In this construction, we've made the following assumptions:

- The X_i are iid.
- The B_i are finite (possibly unbounded) and are based entirely on past events, i.e., $B_i \in \sigma(X_1,\ldots,X_{i-1})$.
- The S_i are finite (possibly unbounded) and are based entirely on past events.

Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space. $\{X_i\}_{i=1}^{\infty}$ are integrable $\mathcal{F}$-measurable random variables. The sequence of σ-fields

$$\mathcal{F}_i := \sigma(X_1, X_2, \ldots, X_i)$$

satisfies the condition

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

Any such sequence of σ-subfields is called a filtration. A sequence of random variables $\{B_i\}_{i=1}^{\infty}$ that satisfies $B_i \in \mathcal{F}_i$ is called adapted to the filtration $\{\mathcal{F}_i\}$ (in our construction, we need to shift an index by 1).

Definition: Let $\{M_i\}_{i=1}^{\infty}$ be a sequence of random variables satisfying:

(1) $\mathbb{E}[|M_i|] < \infty$, for all i.

(2) $\{M_i\}_{i=1}^{\infty}$ is adapted to some filtration $\{\mathcal{F}_i\}_{i=1}^{\infty}$.

(3) $\mathbb{E}[M_{i+1} \mid \mathcal{F}_i] = M_i$.

We call $\{M_i\}_{i=1}^{\infty}$ a martingale with respect to the filtration $\{\mathcal{F}_i\}_{i=1}^{\infty}$.

Exercise: Let $X_i \in \{-1, 1\}$ with $\mathbb{P}\{X_i = 1\} = \dfrac{1}{2}$ and X_i iid. Let $B_i \in \mathcal{F}_{i-1}$. Let

$$S_n = S_0 + \sum_{i=1}^{n} B_i X_i.$$

Then, $\{S_n\}_{n=1}^{\infty}$ is a martingale with respect to $\mathcal{F}_n$.

Observation: Let N be an integer. Suppose $M_1 = m$ is given. Then,

$$\mathbb{E}[M_N] = m.$$

(If the game is fair, your balance shouldn't change.) This can be shown formally using the **Smoothing Property**:

$$\mathbb{E}[M_N] = \mathbb{E}[\mathbb{E}[M_N \mid \mathcal{F}_{N-1}]] = \mathbb{E}[M_{N-1}] = \cdots = \mathbb{E}[M_1] = m.$$

Exercise: If $k < n$, then $\mathbb{E}[M_n \mid \mathcal{F}_k] = M_k$.

Remark: We've found an apparent contradiction. By the previous logic,

$$\mathbb{E}[S_\tau] = \mathbb{E}[S_0] = 0,$$

but we computed that

$$S_\tau = 1.$$

Optional Sampling Theorem: If τ is a certain kind of random time (called a stopping time) and we have a martingale $(M_n, \mathcal{F}_n)$ which is "nice" (for now, this means bounded), then

$$\mathbb{E}[M_i] = \mathbb{E}[M_1].$$

Definition: From here on out, we use the definitions:

$$\begin{aligned} \text{Martingale: } & \mathbb{E}[X_n \mid \mathcal{F}_k] = X_k, \ \ \forall k \le n. \\ \text{Supermartingale: } & \mathbb{E}[X_n \mid \mathcal{F}_k] \le X_k, \ \ \forall k \le n. \\ \text{Submartingale: } & \mathbb{E}[X_n \mid \mathcal{F}_k] \ge X_k, \ \ \forall k \le n. \end{aligned}$$

Sections 10.6 / 10.7 / 10.8 - Theorems on Martingales

Theorem: If $(X_n, \mathcal{F}_n)$ is a martingale and ϕ is a convex function with $\mathbb{E}[|\phi(X_n)|] < \infty$ for all $n \in \mathbb{N}$, then $(\phi(X_n), \mathcal{F}_n)$ is a submartingale.

Proof: By **Jensen's Inequality**,

$$\mathbb{E}[\phi(X_{n+1}) \mid \mathcal{F}_n] \ge (\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]) = \phi(X_n). \ \square$$

Corollary: If $p \ge 1$ and $(X_n, \mathcal{F}_n)$ is a martingale with $\mathbb{E}[|X_n|^p] < \infty$, then $(X_n^p, \mathcal{F}_n)$ is a submartingale.

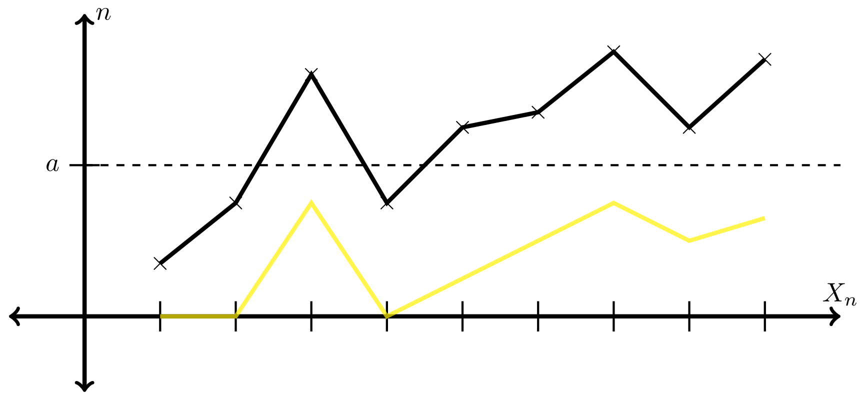
Theorem: If $(X_n, \mathcal{F}_n)$ is a submartingale and ϕ is an increasing convex function with $\mathbb{E}[\phi(X_n)] < \infty$ for all $n \in \mathbb{N}$, then $(\phi(X_n), \mathcal{F}_n)$ is a submartingale.

Exercise: Find a submartingale X_n such that X_n^2 is a submartingale.

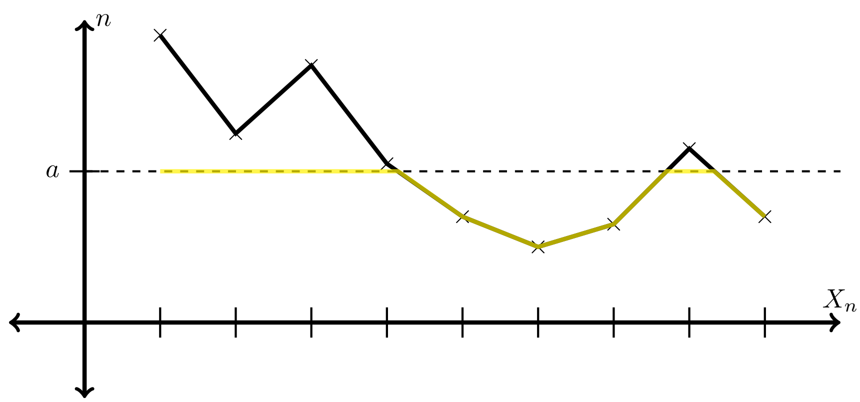
Solution: Define $X_n := -\frac{1}{n}$. This is an increasing sequence. However, $X_n^2 = \frac{1}{n^2}$ is decreasing.

Corollary:

(a) If $(X_n, \mathcal{F}_n)$ is a submartingale, then $(X_n - a)^+$ is a submartingale.



(b) If $(X_n, \mathcal{F}_n)$ is a supermartingale, then $X_n \wedge a$ is a supermartingale.



Definition: Define

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

If $(X_n, \mathcal{F}_n)$ is a supermartingale and H_n is $\mathcal{F}_n$-predictable (i.e., $H_n \in \mathcal{F}_{n-1}$) and H_n is nonnegative and bounded, then $((H \cdot X)_n, \mathcal{F}_n)$ is a supermartingale.

Remark: Observe that

$$\begin{aligned}\mathbb{E}\left[(H \cdot X)_{n+1} \mid \mathcal{F}_n\right] &= \mathbb{E}\left[(H \cdot X)_n + H_{n+1}(X_{n+1} - X_n) \mid \mathcal{F}_n\right] \\ &= \mathbb{E}\left[(H \cdot X)_n \mid \mathcal{F}_n\right] + \mathbb{E}\left[H_{n+1}(X_{n+1} - X_n) \mid \mathcal{F}_n\right] \\ &= (H \cdot X)_n + H_n\left(\mathbb{E}\left[X_{n+1} - X_n\right]\right).\end{aligned}$$

Now, $H_{n+1} \geq 0$ by hypothesis, and so

$$\mathbb{E}\left[X_{n+1} - X_n \mid \mathcal{F}_n\right] \leq 0.$$

Thus

$$\mathbb{E}\left[(H \cdot X)_{n+1} \mid \mathcal{F}_n\right] \leq (H \cdot X)_n. \ \square$$

Corollary: If $(X_n, \mathcal{F}_n)$ is a martingale, and H_n is nonnegative, bounded, and $\mathcal{F}_n$ predictable, then $((H \cdot X)_n, \mathcal{F}_n)$ is a martingale.

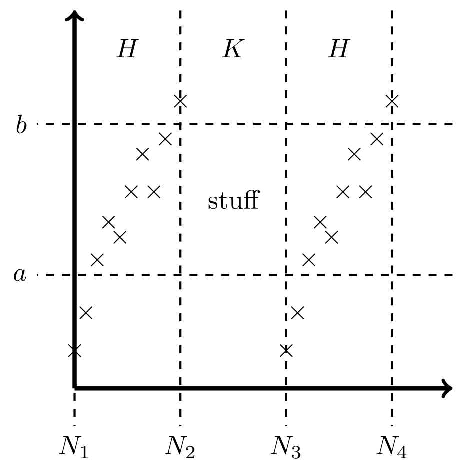
Definition: We say that a random variable τ is a stopping time of $\{\mathcal{F}_n\}_{n\geq 1}$ if $\{\tau = n\} \in \mathcal{F}_n$.

Example: Define $F := \min\{n : X_n \geq 10\}$ and $L := \max\{n : X_n \geq 10\}$, with $\mathcal{F}_n = \sigma(X_1, X_2, \ldots)$. Then, F is a stopping time of $\mathcal{F}_n$ and L is not.

The Upcrossing Lemma: Let $(X_n, \mathcal{F}_n)$ be a submartingale. Define a sequence of upcrossings $\{U_n\}$ on an interval (a, b) to be

$$U_n := \sup\{k : N_{2k} \leq n\}$$

where $\{N_k\}$ is defined by this picture:



Then,

$$(b-a)\mathbb{E}\left[U_n\right] \leq \mathbb{E}\left[(X_n - a)^+\right]\mathbb{E}\left[(X_0 - a)^+\right]$$

where

$$X_n^+ := \max(X_n, 0).$$

Proof: Define $Y_n := a + (X_n - a)^+$. Y_n is a submartingale. Now define

$$(H \cdot Y)_n := \sum_{k=1}^{n} H_k(Y_k - Y_{k-1})$$

where $H_n = \mathbf{1}_{\{n \in (N_{\text{odd}}, N_{\text{even}})\}}$. Note

$$(b-a)U_n \leq (H \cdot Y)_n.$$

Define $K_n := 1 - H_n$. Then,

$$Y_n - Y_0 = (H \cdot Y)_n + (K \cdot y)_n.$$

Now, because $(K \cdot Y)^n$ is also a submartingale,

$$\mathbb{E}\left[(K \cdot Y)_n\right] \geq \mathbb{E}\left[(K \cdot Y)_0\right] = 0.$$

Hence,

$$\mathbb{E}\left[Y_n - Y_0\right] \geq \mathbb{E}\left[(H \cdot Y)_n\right] \geq (b-a)\mathbb{E}\left[U_n\right].$$

Martingale Convergence Theorem: If $(X_n, \mathcal{F}_n)$ is a submartingale with $\sup_n \mathbb{E}[X_n^+] < \infty$, then as $n \to \infty$, X_n converges almost surely to a random variable X with $\mathbb{E}[|X|] < \infty$.

Proof: Since $(X-a)^+ \leq X^+ + |a|$, the **Upcrossing Inequality** implies that

$$\mathbb{E}[U_n] \leq \frac{\mathbb{E}[X_n^+] + |a|}{b-a}. \qquad (*)$$

As $n \to \infty$, we have that $U_n \nearrow U$, which is the total number of upcrossings over $[a,b]$ by the whole sequence $\{X_n\}_{n \geq 1}$.

A priori, U may be ∞. However, by hypothesis,

$$\sup_n \frac{\mathbb{E}[X_n^+] + |a|}{b-a} = \frac{\sup_n(\mathbb{E}[X_n^+] + |a|}{b-a} < \infty.$$

Therefore,

$$\mathbb{E}[U] < \infty$$

and

$$\mathbb{P}\{U < \infty\} = 1.$$

This holds for all pairs $a, b \in \mathbb{Q}$ with $a < b$.

Now consider the event

$$\Omega_{\text{NL}} := \bigcup_{a,b \in \mathbb{Q}} \left\{ \omega : \liminf_{n\to\infty} X_n(\omega) < a < b < \limsup_{n\to\infty} X_n(\omega) \right\}.$$

(NL stands for "no limit".) Note that $\mathbb{P}\{\Omega_{\text{NL}}\} = 0$, since such a path visits values above some b and below some a infinitely often, meaning $U(\omega) = \infty$, i.e., for all $\omega \in \Omega_{\text{NL}}$, we have $U(\omega) = \infty$.

Therefore,

$$\liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n$$

almost surely, i.e.,

$$X := \lim_{n\to\infty} X_n$$

exists. It remains to show that $\mathbb{E}\left[\left|\lim_{n\to\infty} X_n\right|\right] < \infty$.

By **Fatou's Lemma**, $\mathbb{E}[X^+] \leq \liminf_{n\to\infty} \mathbb{E}\left[X_n^+\right] < \infty$. Hence, $X < \infty$ almost surely.

On the other side,

$$\begin{aligned} \mathbb{E}\left[X_n^-\right] &= \mathbb{E}\left[X_n^+\right] - \mathbb{E}[X_n] \\ &\leq \mathbb{E}\left[X_n^+\right] - \mathbb{E}[X_0]. \qquad \text{(since } X_n \text{ is submartingale)} \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{E}\left[X^-\right] &\leq \liminf_{n\to\infty} \mathbb{E}\left[X_n^+\right] - \mathbb{E}[X_0] \\ &\leq \sup_n \mathbb{E}\left[X_n^+\right] - \mathbb{E}[X_0] \\ &< \infty. \quad \square \end{aligned}$$

Corollary: If $(X_n, \mathcal{F}_n)$ is a supermartingale with $X_n \geq 0$, then as $n \to \infty$, we have $X_n \to X$ almost surely and $\mathbb{E}[X] \leq \mathbb{E}[X_0]$.

Doob's Decomposition: Any submartingale $(X_n, \mathcal{F}_n)$ can be written uniquely as $X_n = M_n + A_n$ where $(M_n, \mathcal{F}_n)$ is a martingale and A_n is $\mathcal{F}_n$-predictable with $A_n = 0$.

Proof: Define $A_n - A_{n-1} := \mathbb{E}[X_n \mid \mathcal{F}_{n-1}] - X_{n-1}$. Then, $M_n := X_n - A_n$. To show that A_n is $\mathcal{F}_n$-predictable, observe that

$$A_n = \sum_{k=1}^{n} \mathbb{E}[X_n \mid F_{k-1}] - X_{k-1}$$

and for each k the summand is $\mathcal{F}_{k-1} \subseteq \mathcal{F}_{n-1}$. Thus, $A_n \in \mathcal{F}_{n-1}$.

Meanwhile,

$$\begin{aligned}
\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[X_{n+1} - A_{n+1} \mid \mathcal{F}_n] \\
&= \mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - \mathbb{E}[A_{n+1} \mid \mathcal{F}_n] \\
&= (A_{n+1} - A_n + X_n) - A_{n+1} \\
&= X_n - A_n \\
&= M_n. \ \square
\end{aligned}$$

Sections 10.9 - 10.16 - More on Martingales

Definition: Let ξ_i^n for $i, n \geq 0$ be iid nonnegative integer-valued random variables (a doubly-indexed family). Define $Z_0 := 1$ and

$$Z_{n+1} := \begin{cases} \xi_1^{n+1} + \cdots + \xi_{Z_n}^{n+1}, & Z_n > 0 \\ 0, & \text{otherwise} \end{cases}.$$

Definition: Let ξ_i^n for $i, n \geq 0$ be iid nonnegative integer-valued random variables (a doubly-indexed family). Define $Z_0 := 1$ and

$$Z_{n+1} := \begin{cases} \xi_1^{n+1} + \cdots + \xi_{Z_n}^{n+1}, & Z_n > 0 \\ 0, & \text{otherwise} \end{cases}.$$

We say that the family $\{Z_n\}$ is a branching process.

Theorem: $\left\{\dfrac{Z_n}{\mu^n}\right\}$ is a martingale. We use

$$\mathcal{F}_n = \bigcup_{\ell=0}^{n} \{\xi_i^\ell\}_{i=0}^{\infty}.$$

Proof: We need to show that

$$\mathbb{E}\left[\frac{Z_{n+1}}{\mu^{n+1}} \mid \mathcal{F}_n\right] = \frac{Z_n}{\mu^n}$$

which is equivalent to showing that

$$\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] = \mu Z_n.$$

Observe that

$$Z_{n+1} = \sum_{k=0}^{\infty} Z_{n+1} \mathbf{1}_{\{Z_n = k\}} = \sum_{k=1}^{\infty} Z_{n+1} \mathbf{1}_{\{Z_n = k\}}.$$

Hence,

$$\begin{aligned}
\mathbb{E}\left[Z_{n+1} \mid \mathcal{F}_n\right] &= \mathbb{E}\left[\sum_{k=1}^{\infty} Z_{n+1}\mathbf{1}_{\{Z_n=k\}} \mid \mathcal{F}_n\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[Z_{n+1}\mathbf{1}_{\{Z_n=k\}} \mid \mathcal{F}_n\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\mathbf{1}_{\{Z_n=k\}} Z_{n+1} \mid \mathcal{F}_n\right] \\
&= \sum_{k=1}^{\infty} \mathbf{1}_{\{Z_n=k\}}\mathbb{E}\left[\left(\xi_1^{n+1} + \cdots + \xi_k^{n+1}\right) \mid \mathcal{F}_n\right] \\
&= \sum_{k=1}^{\infty} k\mu\mathbf{1}_{\{Z_n=k\}} \\
&= \mu Z_n.
\end{aligned}$$

Thus, $\dfrac{Z_n}{\mu^n}$ is a martingale. □

Theorem: A nonnegative martingale converges almost surely.

Theorem 3.7 / 3.8: If $\mu \leq 1$, then $\lim_{n\to\infty} Z_n = 0$ almost surely.

Proof: If $\mu < 1$, then $\dfrac{Z_n}{\mu^n}$ blows up when $Z_n \neq 0$ (recall Z_n is nonnegative-integer-valued). So, if $Z_n \neq 0$ on a set of positive measure, we must have that $\dfrac{Z_n}{\mu^n}$ diverges to infinity, which is a contradiction. Hence $Z_n = 0$ almost surely, and we're done.

Now consider the case $\mu = 1$. By the previous theorem, we have $Z_n \to Z_\infty$. Therefore $Z_n = Z_\infty$ for all n larger than some sufficiently large N_ω. If $Z_n = k$ then

$$Z_{n+1} = \xi_1^{n+1} + \cdots + \xi_k^{n+1} = k. \ \square$$

Theorem 3.9: If $\mu > 1$, then

$$\mathbb{P}\left\{Z_n > 0, \text{ for all n}\right\} > 0.$$

Define $\phi(s) = \sum_{k\geq 0} p_k s^k$.

(a) Let $\theta_m := \mathbb{P}\left\{Z_m = 0\right\}$ (the probability of extinction as of generation m). Then,

$$\theta_m = \sum_{k=0}^{\infty} p_k \left(\theta_{m-1}\right)^k.$$

Proof: To see this, we can think about the recursive one-step analysis: We have $Z_0 = 1$ and

$$\mathbb{P}\left\{Z_m = 0\right\} = \sum_{k=0}^{\infty} \mathbb{P}\left\{Z_1 = k\right\} \theta_{m-1}^k. \ \square$$

(b) Note $\theta_m = \phi(\theta_{m-1})$. It's always true that $\phi'(1) = \mu$. If $\mu > 1$, then there is a unique $\rho < 1$ such that $\phi(\rho) = \rho$.

Proof: $\phi(0) \geq 0$ and $\phi(1) = 1$. By hypothesis, $\phi'(1) > 1$. Therefore, for sufficiently small $\epsilon > 0$,

$$\phi(1-\epsilon) < 1-\epsilon.$$

This implies the existence of at least one fixed point by the Intermediate Value Theorem.

We want to calculate $\phi''(\theta)$. We see that

$$\phi''(\theta) = \sum_{k=2}^{\infty} k(k-1)p_k s^{k-2}.$$

There exists $k > 2$ such that $p_k > 0$. ϕ is convex. Hence there is only 1 fixed point. □

(c) $\theta_m \to \rho$

Proof: We know $\theta_0 = 0$ and $\phi(\rho) = \rho$. ϕ is increasing (when $\phi(0) < \rho$) and so

$$\theta_{n+1} = \phi(\theta_n) \geq \theta_n$$

and so $\{\theta_n\}_{n\in\mathbb{N}}$ is an increasing sequence. Use Newton's method (also known as cobwebbing analysis). □

Remark: We are working toward **Doob's Inequality** and the **Maximal Inequality**. We need some preliminaries.

Proposition: If $(X_n, \mathcal{F}_n)$ is a submartingale and N is a stopping time with

$$\mathbb{P}\{N \leq k\} = 1,$$

then

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_N] \leq \mathbb{E}[X_k].$$

Example: Let $S_k = 1 + \sum_{k=1}^{n} \xi_k$, with $\xi_k \in \{\pm 1\}$, each with probability $\frac{1}{2}$. Define

$$N := \int \{n : S_n = 0\}.$$

Then, $\mathbb{E}[S_0] = 1$, but $\mathbb{E}[S_N] = 0$, because N is an unbounded stopping time, and so the previous proposition does not hold.

Proof: Recall that if $(X_n, \mathcal{F}_n)$ is a submartingale, then so is $(X_{n\wedge N}, \mathcal{F}_n)$. So,

(i) $\mathbb{E}[X_0] = \mathbb{E}[X_{N\wedge 0}] \leq \mathbb{E}[X_{N\wedge k}] = \mathbb{E}[X_N]$.

For (ii), define

$$H_n = \mathbf{1}_{\{N<n\}} = 1_{\{N\leq n-1\}}$$

which is predictable. Now, $(H \cdot X)_n$ is a submartingale. □

Doob's Inequality: Let $(X_n, \mathcal{F}_n)$ be a submartingale. Define $\widehat{X}_n := \max_{0 \le k \le n} X_k^+$. Let $\lambda > 0$ and $A := \{\widehat{X}_n \ge \lambda\}$. Then,

$$\lambda \mathbb{P}\{A\} \le \mathbb{E}[X_n \mathbf{1}_A] \le \mathbb{E}[X_n+].$$

The term on the left is a "path property" and the term on the right is an "endpoint evaluation".

Proof: Define

$$N := \inf\{k : X_n \ge \lambda \text{ or } k = n\}.$$

Since $X_N \ge \lambda$ on A, we have

$$\lambda \mathbb{P}\{A\} \le \mathbb{E}[X_N \mathbf{1}_A].$$

On A^C, $X_N = X_n$. The second inequality is obvious. □

Example: Let $S_n = \xi_1 + \cdots + \xi_n$, where the ξ_i are iid with $\mathbb{E}[\xi_i] = 0$ and $\mathbb{E}\left[\xi_i^2\right] < \infty$. Then, defining $X_n := S_n^2$, we have that $(X_n, \sigma(\xi_1, \cdots, \xi_n))$ is a submartingale. So, **Doob's Inequality** tells us that

$$\mathbb{P}\left\{\max_{1 \le k \le n} |S_n| \ge \lambda\right\} \le \frac{\mathrm{Var}(S_n)}{\lambda^2}.$$

***L^p* Maximum Inequality:** Let $(X_n, \mathcal{F}_n)$ be a submartingale. Let $p > 1$. Then,

$$\mathbb{E}\left[\widehat{X}_n^p\right] \le \frac{p}{p-1} \mathbb{E}\left[\left(X_n^+\right)^p\right].$$

Proof: Consider $\widehat{X}_n \wedge M$ for some $M > 0$. Note that when $M > \lambda$¡ we have

$$\{\widehat{X}_n \wedge M \ge \lambda\} = \{\widehat{X}_n \ge \lambda\}.$$

So,

$$\begin{aligned}
\mathbb{E}\left[(\widehat{X}_n \wedge M)^p\right] &= \int_0^\infty p\lambda^{p-1} \mathbb{P}\left\{\widehat{X}_n \wedge M \ge \lambda\right\} d\lambda \\
&\le \int_0^\infty p\lambda^{p-1} \left(\lambda^{-1} \int_\Omega X_n^+ \mathbf{1}_{\{\widehat{X}_n \wedge M \ge \lambda\}} d\mathbb{P}\right) d\lambda \\
&= \int_\Omega X_n^+ \int_0^{\widehat{X}_n \wedge M} p\lambda^{p-2} d\lambda d\mathbb{P} \qquad \text{(by \textbf{Fubini's Theorem})} \\
&= \frac{p}{p-1} \int_\Omega X_n^+ \left(\widehat{X}_n \wedge M\right)^{p-1} d\mathbb{P} \\
&\le \frac{p}{p-1} \mathbb{E}\left[\left(X_n^+\right)^p\right]^{1/p} \mathbb{E}\left[\left(\widehat{X}_n \wedge M\right)^{q(p-1)}\right]^{1/q} \\
&\qquad \text{(by \textbf{Hölder's Inequality}, with } q := \frac{p}{p-1}) \\
&= \frac{p}{p-1} \mathbb{E}\left[\left(X_n^+\right)^p\right]^{1/p} \mathbb{E}\left[\left(\widehat{X}_n \wedge M\right)^p\right]^{1-1/p} \\
&\le \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[\left(X_n^+\right)^p\right].
\end{aligned}$$

Now let $M \to \infty$ and use the **Monotone Convergence Theorem**. □

Corollary: Define $X_n^* := \max_{0 \le k \le n} |X_k|$. Then,

$$\mathbb{E}\left[(X_n^*)^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[|X_n|^p\right].$$

Definition: A family $\{X_t : t \in T\}$ of L^1 random variables indexed by T is uniformly integrable, or ui, if

$$\sup_{t\in T} \mathbb{E}\left[|X_t|\mathbf{1}_{\{|X_t|>a\}}\right] \to 0$$

as $a \to \infty$. In other words,

$$\int_{\{|X_t|>a\}} |X_t| d\mathbb{P} \to 0$$

as $a \to \infty$, uniformly in $t \in T$.

Simple Sufficient Criteria for Uniform Integrability:

(1) If $T = \{1\}$, then

$$\int_{\{|X_1|>a\}} |X_1| d\mathbb{P} \to 0$$

since $X_1 \in L^1$.

(2) (Dominated Families) If there exists a dominating random variable $Y \in L^1$ such that $|X_t| < Y$ for all t, then $\{X_t\}$ is uniformly integrable.

(3) (Finite Families) If $X_i \in L^1$ for all $i \in [n]$, then $\{X_i\}_{i=1}^n$ is uniformly integrable.

(4) (More Domination) Suppose that for each $t \in T$, we have $X_t \in L_1$ and $Y_t \in L_1$, and $|X_t| \leq |Y_t|$. Then, if $\{Y_t\}$ is uniformly integrable, so is $\{X_t\}$.

(5) (Crystal Ball Condition) Let $p > 0$ and suppose

$$\sup_n \mathbb{E}\left[|X_n|^{p+\delta}\right] < \infty$$

for some $\delta > 0$. Then, the family $\{|X_n|^p\}$ is uniformly integrable.

Proof:

$$\begin{aligned}
\sup_n \int_{\{|X_n|^p>a\}} |X_n|^p d\mathbb{P} &= \sup_n \int_{\left\{\left|\frac{X_n}{a^{1/p}}\right|\geq 1\right\}} |X_n|^p \, d\mathbb{P} \\
&= \sup_n \int_{\left\{\frac{|X_n|^\delta}{a^{\delta/p}}\right\}} |X_n|^p \cdot 1 \, d\mathbb{P} \\
&\leq \sup_n \int_\Omega |X_n|^p \cdot \frac{|X_n|^\delta}{a^{\delta/p}} d\mathbb{P}.
\end{aligned}$$

Now apply the given condition. $\square$

Example: Let X_n be a sequence with $\mathbb{E}[X_n] = 0$ and $\text{Var}(X_n) = 1$ for all n. Then, $\{X_n\}$ is uniformly integrable.

Theorem: If $X_n \to X$ in probability (with $X_n, X \in L^1$), then the following are equivalent.

(i) $\{X_n\}$ is uniformly integrable.

(ii) $X_n \to X$ in L_1.

(iii) $\mathbb{E}[|X_n|] \to \mathbb{E}[|X|]$.

Proof:

(i) $\Rightarrow$ (ii) Define

$$\phi_M(x) := \begin{cases} M, & x \geq M \\ x, & x \in (-M, M) \\ -M, & x \leq -M \end{cases} .$$

Then,

$$|X_n - X| \leq |X_n - \phi_M(X_n)| + |\phi_M(X_n) - \phi_M(X)| + |\phi_M(X) - X|,$$

Well,

$$|X_n - \phi_M(X_n)| = (|X_n| - M)^+ \leq |X_n| \mathbf{1}_{\{|X_n| > M\}}$$

and

$$|X - \phi_M(X)| \leq |X| \mathbf{1}_{\{|X| > M\}}.$$

For sufficiently large M,

$$\mathbb{E}[|X_n - \phi_M(X_n)|] \leq \epsilon$$

by uniform integrability, and

$$\mathbb{E}[|X - \phi_M(X)|] \leq \epsilon$$

Since $X \in L^1$. Since $X_n \xrightarrow{\mathbb{P}} X$ and since ϕ_M is bounded, we have

$$\mathbb{E}[|\phi_M(X_n) - \phi_M(X)|] \to 0$$

which shows that (i) $\Rightarrow$ (ii). $\square$

(ii) $\Rightarrow$ (iii) Observe that $|X|$ is convex. So,

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

(iii) $\Rightarrow$ (i) Now define

$$\psi_M(x) := \begin{cases} 0, & x \geq M \\ x, & x \in [0, M-1] \\ \text{linear interpolation}, & x \in (M-1, M) \end{cases} .$$

By the **Dominated Convergence Theorem**, for M sufficiently large,

$$\mathbb{E}[|X|] - \mathbb{E}[\psi_M(|X|)] \leq \frac{\epsilon}{2}.$$

By the **Bounded Convergence Theorem**,

$$\mathbb{E}[\psi_M(|X_n|)] \to \mathbb{E}[\psi_M(|X|)] /$$

Using (iii), we have that for all $n > N$ sufficiently large,

$$\begin{aligned} \mathbb{E}\left[|X_n| \mathbf{1}_{\{|X_n| \geq M\}}\right] &= \mathbb{E}[|X_n|] - \mathbb{E}[\psi_M(|X_n|)] \\ &\leq \mathbb{E}[|X|] - \mathbb{E}[\psi_M(|X|)] + \frac{\epsilon}{2} \\ &< \epsilon. \end{aligned}$$

By taking M sufficiently large, we can show that

$$\mathbb{E}\left[|X_n| \mathbf{1}_{\{|X_n| > M\}}\right] < \epsilon$$

for all $n \in [N]$. $\square$

Theorem: For a submartingale $(X_n, \mathcal{F}_n)$, the following are equivalent.

(i) $\{X_n\}$ is uniformly integrable.

(ii) X_n converges almost surely to some $X \in L^1$.

(iii) $X_n \to X$ in L^1.

Lemma: If $\{X_n\} \subseteq L^1$ and $X_n \xrightarrow{L^1} X$, then

$$\mathbb{E}\left[X_n \mathbf{1}_A\right] \to \mathbb{E}\left[X 1_A\right]$$

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

Proof:

$$\begin{aligned}\left|\mathbb{E}\left[X_n 1_A\right] - \mathbb{E}\left[X 1_A\right]\right| &\leq \mathbb{E}\left[\left|X_n - X\right| \mathbf{1}_A\right] \\ &\leq \mathbb{E}\left[\left|X_n - X\right|\right] \\ &\to 0. \ \square\end{aligned}$$

Lemma: If $(X_n, \mathcal{F}_n)$ is a martingale and $X_n \xrightarrow{L^1} X$, then,

$$X_n = \mathbb{E}\left[X \mid \mathcal{F}_n\right].$$

Proof: The martingale property implies that if $n \geq k$, then

$$\mathbb{E}\left[X_n \mid \mathcal{F}_k\right] = X_k.$$

Now suppose $\mathbf{1}_A \in \mathcal{F}_k$. Then,

$$\mathbb{E}\left[X_n \mathbf{1}_A \mid \mathcal{F}_k\right] = X_k \mathbf{1}_A.$$

Taking expectations,

$$\mathbb{E}\left[\mathbb{E}\left[X_n \mathbf{1}_A \mid \mathcal{F}_k\right]\right] = \mathbb{E}\left[X_k \mathbf{1}_A\right]$$

and so

$$\mathbb{E}\left[X_n \mathbf{1}_A\right] = \mathbb{E}\left[X_k \mathbf{1}_A\right].$$

By the previous lemma,

$$\left[X_n \xrightarrow{L^1} X\right] \implies \left[X_n \mathbf{1}_A \xrightarrow{L^1} X 1_A\right].$$

Thus, for all $A \in \mathcal{F}_k$,

$$\mathbb{E}\left[X_k \mathbf{1}_A\right] = \mathbb{E}\left[X \mathbf{1}_A\right],$$

i.e.,

$$\int_A X_k d\mathbb{P} = \int_A X d\mathbb{P}.$$

Recall the definition of conditional expectation: $\mathbb{E}\left[X \mid \mathcal{F}_k\right]$ is (unique) random variable satisfying

$$\int_A X d\mathbb{P} = \int_A \mathbb{E}\left[X \mid \mathcal{F}_k\right] d\mathbb{P}.$$

Hence,

$$\mathbb{E}\left[X \mid \mathcal{F}_k\right] = X_k. \ \square$$

Remark: The two results below follow from the previous lemma.

Theorem: Suppose $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, where $\mathcal{F}_\infty := \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n\right)$. Then, as $n \to \infty$, we have

$$\mathbb{E}\left[X \mid \mathcal{F}_n\right] \to \mathbb{E}\left[X \mid \mathcal{F}_\infty\right].$$

Theorem: (Lévy's 0-1 Law) If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and $A \in \mathcal{F}_\infty$, then

$$\mathbb{E}\left[\mathbf{1}_A \mid \mathcal{F}_n\right] \to \mathbf{1}_A.$$

Optional Stopping Times

Theorem: If L, M are stopping times with $L \leq M$ almost surely and $(Y_{M\wedge n}, \mathcal{F}_n)$ is a uniformly integrable submartingale. Then,

$$\mathbb{E}\left[Y_L\right] \leq \mathbb{E}\left[Y_M\right]$$

and

$$Y_L \leq \mathbb{E}\left[Y_M \mid \mathcal{F}_L\right].$$

Index