Contents

Abstract

This is the course note for Math 206A at University of California, Riverside during the quarter Fall 2021. The course is about Probability theories.

Graduate Probability Course note -Fall 2021
Instructor: Dr.Heyrim Cho

Prepared by: Khoi Vo *

July 22, 2022

1 Day 1: Probability measure

1.1 About this course note

This course note was transcribed from his own hand-written note to Latex file by Khoi during the Summer 2022, while he was trying to review the notes and study for the Qualifying exam. There is no doubt that there will be errors and typos lying around in this note (almost surely all the mistakes made here are due to Khoi alone). If you spot any, and kind enough to share, please let Khoi know and update this document via email at kvo020@ucr.edu. Thank you in advance and good luck in whatever you are doing.

1.2 Introduction

Course Goals:

1. Develop language of probability based on measure theory
2. Major results in probability:
   (a) Law of large numbers
   (b) Central Limit Theorem
   (c) Large Deviation Theorem
3. Advanced topics

Recall undergraduate probability:

1. Discrete probability: Example Dice Rolling $\{1, 2, \dots, 6\}$: $\frac{1}{6}$
2. Continuous probability: Example: Gaussian distribution.

Definition 1. A probability space/triple $(\Omega, \mathcal{F}, \ )$ will have:

1. Ω sample space: all outcomes of the experiments.
   (a) Example 1: Dice rolling once $\Omega = \{1, 2, 3, 4, 5, 6\}$

*email: kvo020@ucr.edu

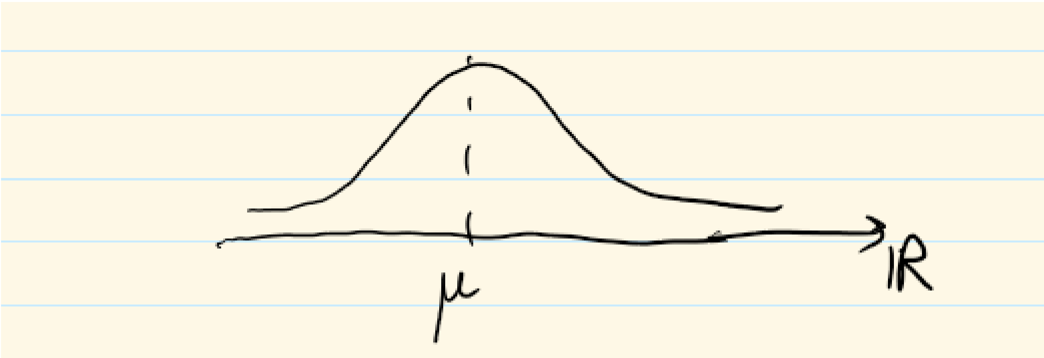


Figure 1: Gaussian probability

(b) Example 2: Dice rolling twice $\Omega = \{(1,1), \dots, (6,6)\}$

(c) Example 3: Randomly pick a number from 0 to 1. So $\Omega = [0,1)$

2. $\mathcal{F}$ set of events, which is set of subsets of Ω.

   (a) Examples of events are:

      i. $A = \{1\}$, B= odd numbers $= \{1, 3, 5\}$

      ii. A= same number twice $= \{(1,1), \dots, (6,6)\}$

      iii. A= a number larger than $0.9 = (0.9, 1)$

   (b) Examples of event sets:

      i. $\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \dots, \{6\}, \{1,2\}, \dots, \{1, \dots, 6\}\} = 2^{\Omega}$ complete σ-algebra.

3. μ: probability function (or probability measure): $\mathcal{F} \to [0,1]$

   (a) Example 1: $\mu(\{1\}) = \mu(A) = \frac{\text{number of elements in A}}{6}$

   (b) Example 3: $\mu((a,b)) = \frac{|b-a|}{1}$, for $a, b \in [0,1)$

   Properties that we want for μ:

   (a) If $A_i \in \mathcal{F}$ are disjoint, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

Note that this property is satisfied in Example 1 and Example 2, but **not** in real line for nasty sets (events), such as the **Vitalli** set:

Example 2. Here we construct the Vitalli set as followed:
Equivalent class $x \sim y$ if $x - y \in \mathbb{Q}$. For instanct, the equivalence class of z_1 is $x = z_1 + q$ for all $q \in \mathbb{Q}$. Let $A := \{$pick one element from each equivalent class$\}$. Consider $A_q = (A + q) \mod 1$ a translation of A. Then note that $\mu(A) = \mu(A_q)$ and that $\bigcup_{q \in \mathbb{Q}} A_q = [0,1)$. Then by additivity,

$$\mu\left(\bigcup_{q \in \mathbb{Q}} (A_q)\right) = \sum_{q \in \mathbb{Q}} \mu(A_q)$$

But we get trouble:

(a) If $(A_q) = 0$ then the RHS is 0, while the LHS is $([0, 1)) = 1$

(b) If $(A_q) > 0$, then the RHS diverges to infinity, while LHS is 1.

To resolve that problem, we have to introduce more terminologies.

Definition 3. [-algebra of $\Omega = F$] A sigma algebra of Ω must satisfy:

1. $\Omega \in F$
2. If $A \in \Omega$ then $A^C \in \Omega$
3. If $A_i \in F$ then $\bigcup_{i=1}^{\infty} A_i \in F$

Note that property 3 above has an equivalence of $\bigcap_{i=1}^{\infty} A_i \in F$.

Example 4. Here is some example of F sigma algebra:

1. $F = 2$, the complete sigma algebra
2. $F = \{;, \Omega\}$
3. $F = \{;, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\}$

Definition 5. [sigma algebra generated by $A = (A)$]

$$(A) = \bigcap_{A \in F; F \text{ is a sigma algebra}} F$$

Definition 6. [Borel sigma algebra on $\mathbb{R}$] is the sigma algebra on $\mathbb{R}$ generated by all open intervals, denoted $B(\mathbb{R})$.

Properties of $B(\mathbb{R})$:

1. $(a, \infty) \in B(\mathbb{R})$ because $(a, \infty) = \bigcup_{i=1}^{\infty}(a, a + i)$
2. $\{a\} \in B(\mathbb{R})$ because $\{a\} = \bigcap_{i=1}^{\infty}(a \quad \frac{1}{n}, a + \frac{1}{n})$
3. $[a, b] \in B(\mathbb{R})$ because $\{a\} \cup (a, b) \cup \{b\} = [a, b]$

Remark 7. $B(\mathbb{R})$ can also be generated by $[a, b], (a, b], (a, \infty), \ldots$

Definition 8. $: F \to [0, 1]$ is the **probability measure/function** for any F a sigma algebra of Ω if it satisfy:

1. $(\Omega) = 1$
2. For any $A \in F$, $(A) \quad 0$
3. If $A_i \in F$ **pairwise disjoint**, then:

$$(\bigcup_{i}^{\infty} A_i) = \sum_{i=1}^{\infty} (A_i)$$

Example 9. Here are some examples of probability measure :

1. Dirac measure at $\omega_0 \in \Omega$:

$$\mu(A) = \begin{cases} 0 & \text{if } \omega_0 \in A \\ 1 & \text{if } \omega_0 \notin A \end{cases}$$

   We check that condition (3) in the measure definition is satisfied:
   If $\omega_0 \notin A_i$ for all i, then $\omega_0 \notin \cup A_i$, so $\mu(A_i) = 0$ for all i, so $\mu(\cup A_i) = 0 = \sum_i \mu(A_i)$.
   If $\omega_0 \in A_1$ and not any other A_i, then $\omega_0 \in \cup_{i=1}^{\infty} A_i$, so $1 = \mu(\cup A_i) = 1 + 0 + \ldots 0 = 1$.
   If more than one A_i contains ω_0, then they are not pairwise disjoint!

2. Discrete probability measure: There is a **countable set** S such that $\mu(S^C) = 0$.

   (a) Bernoulli with success probability p, $\Omega = \{0, 1\}$ and $\mathcal{F} = 2^{\Omega}$:
   Setting $p_0 = 1 - p$ and $p_1 = p$, then $\mu(\{k\}) = p_k$. Note that $\sum_k p_k = 1$. ($k = 0, 1$).

   (b) Binomial probability n trials with success probability p:
   $\Omega = \{1, 2, 3, \ldots, n\}$ and $\mathcal{F} = 2^{\Omega}$. Then $p_k = \binom{n}{k}(1-p)^{n-k}p^k$ and $\mu(\{k\}) = p_k$ ($k = 1, 2, \ldots, n$).

   (c) Geometric probability with p:
   $\Omega = \{1, 2, 3, \ldots\}$ and $\mathcal{F} = 2^{\Omega}$. Then $p_k = (1-p)^{k-1}p^k$

   (d) Poisson probability with parameter λ:
   $p_k = e^{-\lambda}\frac{\lambda^k}{k!}$ (limit of binomial case with $\lambda = np$ and $n \to \infty$).

3. Continuous probability on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$:
   If $f(x)$ continuous such that $\mu(A) = \int_A f(x)dx$ and that $\int_{\mathbb{R}} f(x)dx = 1$:

   (a) Gaussian probability with mean μ, standard deviation σ and we have the function (later we call this the distributive function):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

   (b) Exponential probability with parameter λ

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & ; x \geq 0 \\ 0 & ; x < 0 \end{cases}$$

Theorem 10 (properties of probability measures)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a triple of probability space. Let $A, B \in \mathcal{F}$. Then:*

1. *Monotonicity: If $A \subseteq B$ then $\mu(A) \leq \mu(B)$*

2. *Subadditivity: $A_i \in \mathcal{F}$ with **no disjoint condition**, then:*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i)$$

3. *Continuity from below: If $A_1 \subseteq A_2 \subseteq \ldots$ and if $A := \bigcup_{i=1}^{\infty} A_i$, then*

$$\mu(A) = \lim_{n \to \infty} \mu(A_n)$$

4. *Continuity from above: If $A_1 \supseteq A_2 \supseteq \ldots$ and if $A := \bigcap_{i=1}^{\infty} A_i$, then*

$$\mu(A) = \lim_{n \to \infty} \mu(A_n)$$

2 Day 2: pdf and cdf

We begin by the proof of theorem from last time, Theorem 10, which is theorem 1.1 in Durett book:

Proof. Let $A, B \in \mathcal{F}$.

1. Monotonicity: $\mu(B) = \mu(A \cup (B - A)) = \mu(A) + \mu(B - A) \geq \mu(A)$ (since $\mu(B - A) \geq 0$).

2. Sub-additivity (This is also called Boole's inequality). Let $A_i \in \mathcal{F}$. We define the following:
   $B_1 = A_1$
   $B_2 = A_2 - A_1$
   $B_3 = A_3 - (A_1 \bigcup A_2)$
   $\ldots$
   Then
   $$\begin{aligned} \mu(\bigcup_{i=1}^{\infty} A_i) &= \mu(\bigcup_{i=1}^{\infty} B_i) \\ &= \sum_{i=1}^{\infty} \mu(B_i) \text{ (since the } B_i \text{ are pairwise disjoint)} \\ &\leq \sum_{i=1}^{\infty} \mu(A_i) \text{ (since the } B_i \subset A_i) \end{aligned}$$

3. (Continuity from below). With $A_1 \subset A_2 \subset \ldots$ Same technique:
   We define the following:
   $B_1 = A_1$
   $B_2 = A_2 - A_1$
   $B_3 = A_3 - (A_1 \bigcup A_2)$
   $\ldots$
   Note that $A = \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$
   $$\begin{aligned} \mu(A) &= \mu(\bigcup_i B_i) \\ &= \sum_{i=1}^{\infty} \mu(B_i) \\ &= \mu(A_1) + \sum_{i=2}^{\infty} \mu(A_i - A_{i-1}) \\ &= \mu(A_1) + \sum_{i=2}^{\infty} [\mu(A_i) - \mu(A_{i-1})] \\ &= \lim_{n \to \infty} \mu(A_n) \end{aligned}$$

4. (Continue from below).Same technique

   Remark 11. For general measure, we need at least one of the A_i has finite measure. For instance, it could fail if $A_i = [i, \infty)$

□

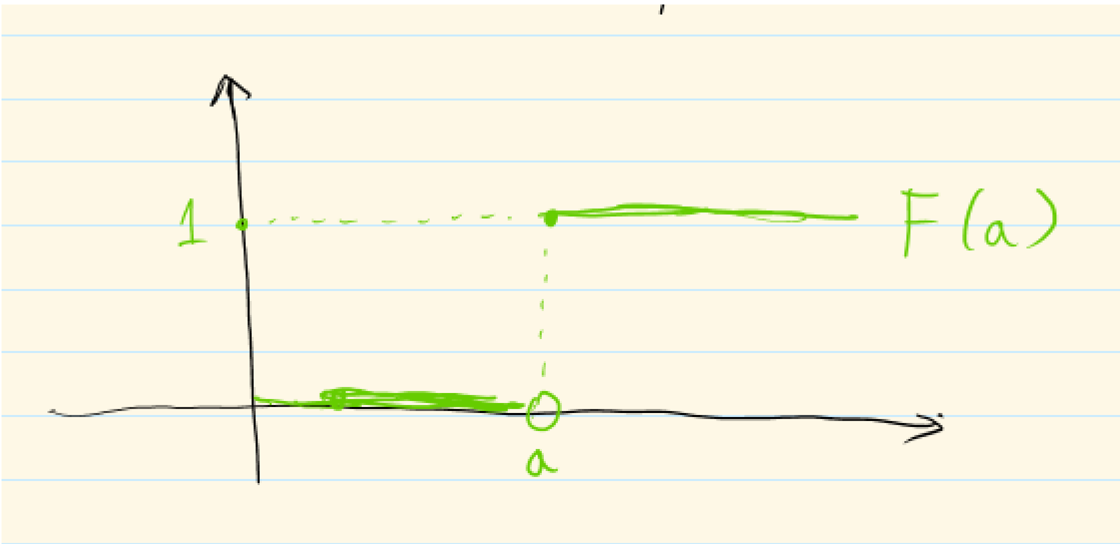


Figure 2: Dirac mass CDF

2.1 Cumulative Distribution function- cdf

Definition 12 (Cumulative distribution function c.d.f/distribution)**.** Given a probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$, the cumulative distribution function (CDF) is

$$F(a) := \mu((-\infty, a])$$

Example 13. Here are some examples of CDF:

1. Dirac mass at $0 \in \mathbb{R}$:

$$\delta_0(A) = \begin{cases} 1 & ; 0 \in A \\ 0 & ; 0 \notin A \end{cases}$$

   If $a < 0$ then $F(a) = \delta_0((-\infty, a]) = 0$ since $0 \notin (-\infty, a]$.
   If $a \geq 0$, then then $F(a) = \delta_0((-\infty, a]) = 1$ since $0 \in (-\infty, a]$

2. Uniform probability on $(0, 1]$:

$$\lambda((a, b]) = \begin{cases} b - a & ; (a, b] \subseteq (0, 1] \\ 0 & ; \text{otherwise} \end{cases}$$

   - If $a < 0$, then $F(a) = \lambda((-\infty, a]) = 0$, since $(-\infty, a] \cap (0, 1] = \emptyset$
   - If $0 \leq a < 1$ then $F(a) = \lambda((-\infty, a]) = \lambda((-\infty, 0]) + \lambda((0, a]) = 0 + a = a$.
   - If $1 \leq a$, then $F(a) = \lambda((-\infty, a]) = 0 + \lambda((0, 1]) + 0 = 1$

Theorem 14. *(Theorem 1.2.1- Durett) Any distribution function F has the following properties:*

1. *non-decreasing*

2. $$\lim_{a \to -\infty} F(a) = 0$$

   and

$$\lim_{a \to \infty} F(a) = 1$$

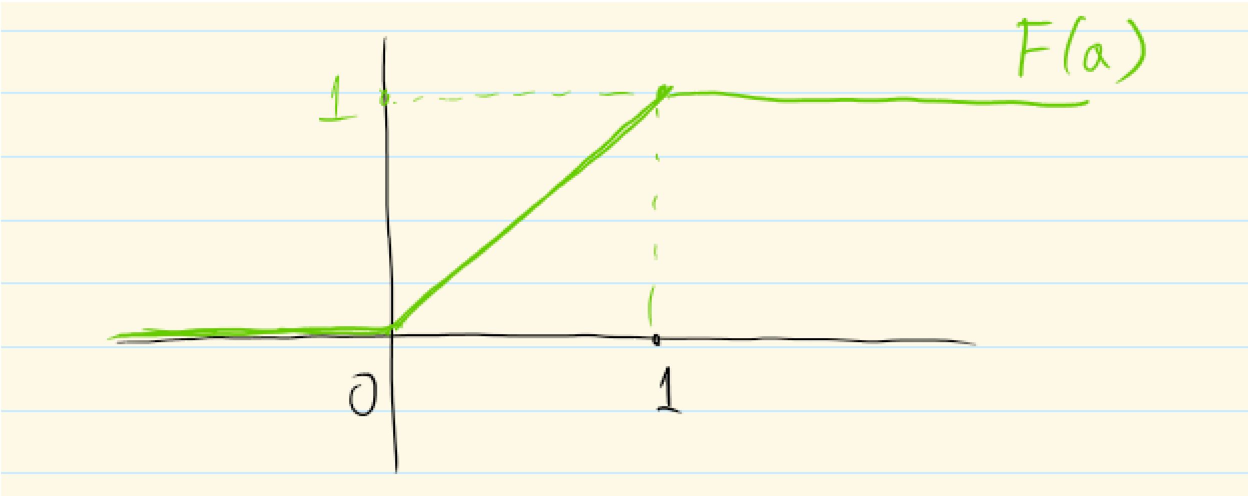


Figure 3: Uniform prob on $(0,1]$ CDF

3. *right-continuous:*
$$\lim_{y \downarrow a} F(y) = F(a)$$
where $F(a^+) = \lim_{y \downarrow a} F(y) = \lim_{y \to a, y > a} F(y)$.

4.
$$\lim_{y \uparrow a} F(y) = P((-\infty, a])$$

5.
$$P(\{a\}) = \lim_{y \downarrow a} F(y) - \lim_{y \uparrow a} F(y)$$

Remark 15. $F(x)$ is **not** necessarily continuous (it is just right continuous) but it can have **at most** countable jumps- discontinuity points (because of monotonicity), and the jump at $x = a$ is the probability at a.

Proof. We now give the proof of previous theorem (14):

1. Let $a < b$. Then $F(a) = P((-\infty, a]) \le P((-\infty, b]) = F(b)$. Since $(-\infty, a] \subseteq (-\infty, b]$, so by monotonicity of measure, we get the result.

2. Consider $A_n = (-\infty, -n)$, then $A_1 \supseteq A_2 \supseteq \dots$ and note that $\bigcap_{n=1}^{\infty} A_n = \emptyset$. then:
$$\begin{aligned}\lim_{a \to -\infty} F(a) &= \lim_{n \to \infty} P(A_n); \text{ using continuity from above} \\ &= P(A) = P(\emptyset) = 0\end{aligned}$$
A similar argument can be used to show $\lim_{a \to \infty} F(a) = 1$ when using continuity from above with the sets $B_n = (n, \infty)$ and that $B_n \supseteq B_{n+1}$.

3. Consider $A_n = (-\infty, a + \frac{1}{n})$. We also have $A_1 \supseteq A_2 \supseteq \dots$. And
$$\bigcap_{n=1}^{\infty} A_n = (-\infty, a]$$
. Thus using continuity from below argument:
$$F(a) = P((-\infty, a]) = \lim_{n \to \infty} P(A_n) = \lim_{y \downarrow a} F(y)$$

4. Consider $A_n = (-\infty; a - \frac{1}{n})$. We also have $A_1 \subseteq A_2 \subseteq \dots$.

□

Theorem 16 (Backward direction from CDF to measure)**.** *If $F(x)$ is a function on $\mathbb{R}$ that is **nondecreasing, right continuous** and $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$, then:*
There exists a "unique" probability measure:

$$\mu((a; b]) := F(b) - F(a); \forall a; b \in \mathbb{R}$$

Remark 17.

1. This μ is called Lebesgue-Stieljes measure in general.
2. If $F(x) = x$, then this μ is called Lebesgue measure.
3. $\mu((a; b)) = F(b-) - F(a)$
4. $\mu(\{a\}) = F(a) - F(a-)$

2.2 Probability density function - pdf

Definition 18 (Probability density function pdf)**.** If there exists a $f(x)$ which is **non-negative** and satisfy

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Then we called $f(x)$ the **p.d.f**.

Question 19. When will we have pdf? (from a cdf?)

Theorem 20 (existence of pdf)**.** *If $F(x)$ is a **absolutely continuous** cdf on $\mathbb{R}$, then there exists a pdf $f(x)$.*
In addition, if f is continuous, we have $f(x) = \frac{d}{dx}F(x)$.

Remark 21. The condition **absolutely continuous** is stronger than uniform condition. It guarantees the integral exists except for a set of measure zero.

Example 22. Given a cdf F, can we find measure μ and pdf $f(x)$?. For example, consider the uniform c.d.f:

$$F(x) = \begin{cases} 0 & ; \text{ if } x < 0 \\ x & ; \text{ if } 0 \le x \le 1 \\ 1 & ; \text{ if } x > 1 \end{cases}$$

The pdf can be found by taking derivative of F, so:

$$f(x) = \begin{cases} \frac{d}{dx}F(x) = \frac{d}{dx}x = 1 & ; \text{ on } 0 < x < 1 \\ 0 & ; \text{ otherwise} \end{cases}$$

And the measure:

$$\mu((a; b]) = \int_a^b f(x)dx$$

Special case:

$$\mu(\{a\}) = 0 = \lim_{t \to 0} \int_{a-t}^{a+t} f(x)dx$$

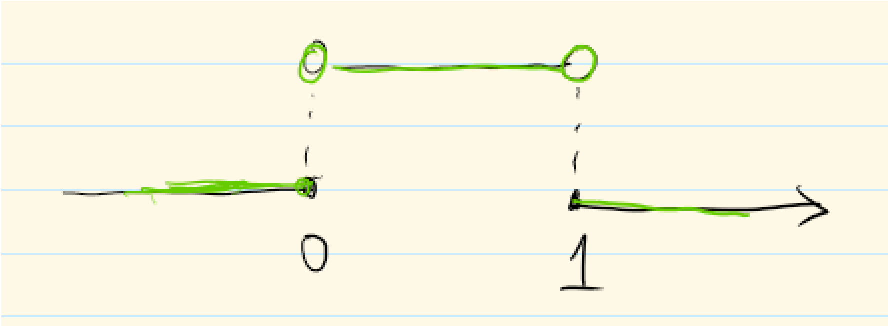


Figure 4: pdf of the uniform cdf

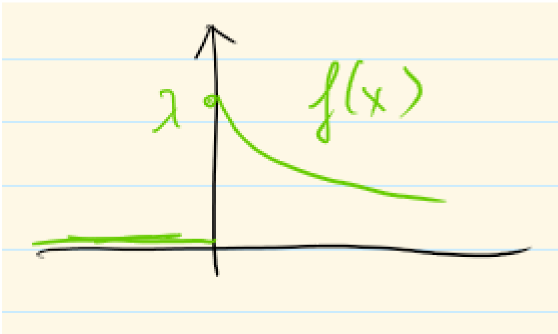


Figure 5: exponential distribution prob with paramenter

Remark 23. The pdf of continuous density can get the relative fractional at points.

Example 24. From a pdf f can we find the cdf F? Consider the exponential distribution probability with parameter :

$$f(x) = \begin{cases} e^{\quad x} & ; x \quad 0 \\ 0 & ; x < 0 \end{cases}$$

Now to find the cdf:

$$F(x) = \int_{1}^{x} f(y)dy = \int_{0}^{x} e^{\quad y} dy = \quad e^{\quad y}\Big|_{0}^{x} = 1 \quad e^{\quad x}; x > 0$$

Definition 25 (probability mass function pmf for discrete probability)**.** This is defined as

$$p(x) := \quad (fxg)$$

Note that $p(x) \quad 0$ and that $\sum_{x} p(x) = 1$ and that the x are countable.

Remark 26. A distribution function can yield a **singular probability measure**, that is, a measure such that it is **non-zero** at **zero measure set**!!.

Example 27. The example of such singular probability from a distribution function is the **Cantor set** K. The Cantor set is defined as followed:

$$K := fx\ 2\ [0;1] : x = \sum_{k=1}^{1} \frac{a_k}{3^k}; a_k = 0; 2g$$

Figure 6: cdf of exponential pdf with paramenter

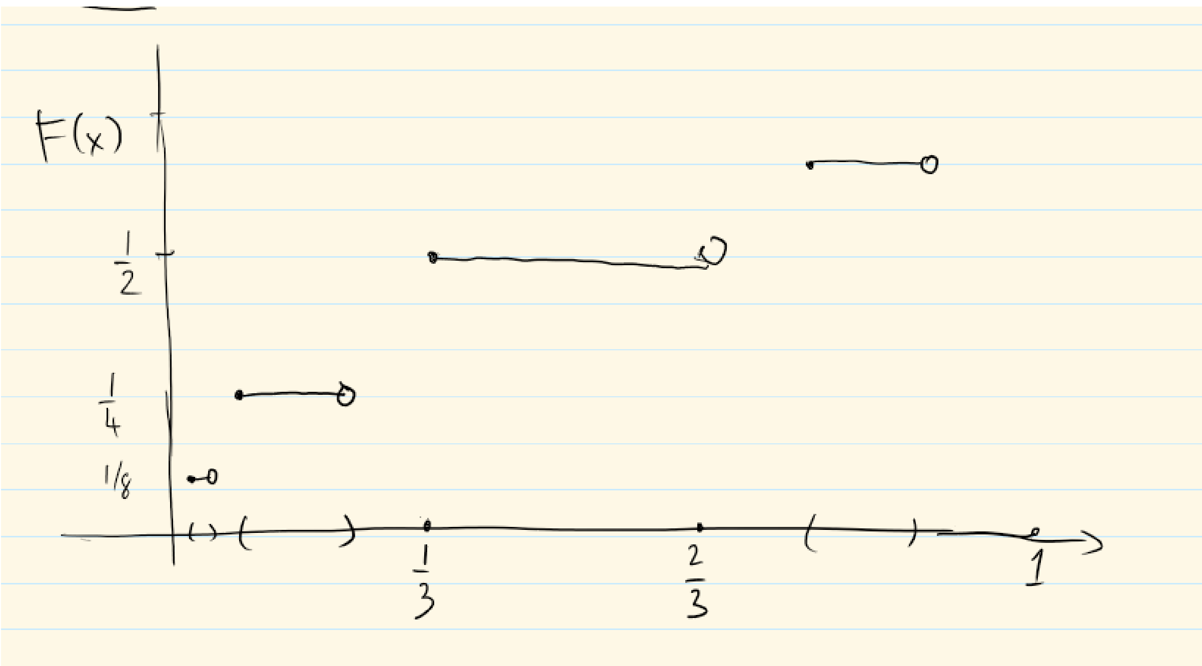


Figure 7: Devil staircase

which is the remaining after taken out all the middle 1/3.
The Lebsgue measure $(K) = 1 \quad \frac{1}{3} \quad \frac{2}{3} \quad \frac{1}{3} \quad (\frac{2}{3})^2 \quad \frac{1}{3} \quad ::: = 0$
However, if we consider $y = \sum_{k=1}^{1} \frac{b_k}{2^k}$ with $b_k = 0; 1$, these are all the numbers in $[0; 1]$. This gives a one to one relation from each $x\ 2\ K$ to each $y\ 2\ [0; 1]$. This means K is uncountable with the same cardinality of $[0; 1]$. But K has Lebesuge measure zero!

Example 28. An example that cdf F does **not** have a density pdf f. Consider the **Cantor distribution - The devil staircase**: The $F(x)$ is defined as follow: For any $x\ 2\ [0; 1]$ we write x in base 3. Then the value $F(x)$ in base 3 at that location will be determined by:

$$F(x) = \begin{cases} 0 & ; \text{ if value of x is 0} \\ 1 & ; \text{ if value of x is 2} \\ 1 & ; \text{ if value of x is 1, then truncate after 1} \end{cases}$$

Note that the value of $F(x)$ is in base 2.

For example, if $x = 0.73 = (0.20122\ldots)_3$, then $F(x) = (0.101)_2$.
Another example is that if $\frac{1}{3} < x < \frac{2}{3}$, we can write $x = (0.1111\ldots)_3$, so $F(x) = (0.1)_2 = \frac{1}{2}$.

In formula, if $x = \sum_{k=1}^{\infty} \frac{a_k}{3^k} \in K$, then

$$F(x) = \sum_{k=1}^{\infty} \frac{b_k}{2} \text{ with } b_k = \frac{a_k}{2}$$

Outside of K, we just extend $F(x)$ continuously.
Then $F'(x)$ of the Cantor distribution is zero except on the Cantor set. The density does not exists. This is because $F(x)$ is **not absolutely continuous**.

3 Day 3: Random Variables

3.1 Measurable functions and Random Variables

Definition 29 (measurable function)**.** Let $f : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ is called **measurable function** if for any $B \in \mathcal{F}_2$, we have $f^{-1}(B) \in \mathcal{F}_1$, where

$$f^{-1}(B) := \{x \in \Omega_1 : f(x) \in B\}$$

Definition 30 (random variable)**.** $X : (\Omega, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ is a (real-valued) **random variable** on probability space $(\Omega, \mathcal{F}, \mu)$ if:

$$\forall a \in \mathbb{R} : X^{-1}((-\infty, a]) \in \mathcal{F}$$

Remark 31. We also note the following notation:

$$\mathbf{P}(x \in B) = \mathbf{P}(X(\omega) \in B) = \mu(\{\omega \in \Omega : \omega \in X^{-1}(B)\}) = \mu(X^{-1}(B))$$

Theorem 32 (Theorem 1.3.1 Durett)**.** *If $f^{-1}(B) \in \mathcal{F}_1$ for all $B \in \mathcal{A}$ and that if $\sigma(\mathcal{A}) = \mathcal{F}_2$, then f is a measurable function. Note that $f : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ as a function.*

Remark 33. This theorem allows us to check measurable function only on generating sets of σ-algebra $\mathcal{F}$ instead of on every elements in $\mathcal{F}$.

Proof. Consider $\mathcal{C} = \{B \subset \Omega_2 : f^{-1}(B) \in \mathcal{F}_1\}$. We will show that $\mathcal{C}$ is a sigma algebra:

1. Show if $B \in \mathcal{C}$ then $\Omega_2 \setminus B \in \mathcal{C}$:

$$\begin{aligned} f^{-1}(\Omega_2 \setminus B) &= \{\omega \in \Omega_1 : f(\omega) \in \Omega_2 \setminus B\} \\ &= \Omega_1 \setminus \{\omega \in \Omega_1 : f(\omega) \in B\} \\ &= \Omega_1 \setminus f^{-1}(B) \in \mathcal{C} \end{aligned}$$

2. Show if $B_i \in \mathcal{C}$ then their union must also be in $\mathcal{C}$:

$$\begin{aligned} f^{-1}\Big(\bigcup_{i=1}^{\infty} B_i\Big) &= \{\omega \in \Omega_1 : f(\omega) \in \bigcup_{i=1}^{\infty} B_i\} \\ &= \bigcup_{i=1}^{\infty} \{\omega \in \Omega_1 : f(\omega) \in B_i\} \\ &= \bigcup_{i=1}^{\infty} f^{-1}(B_i) \in \mathcal{C} \end{aligned}$$

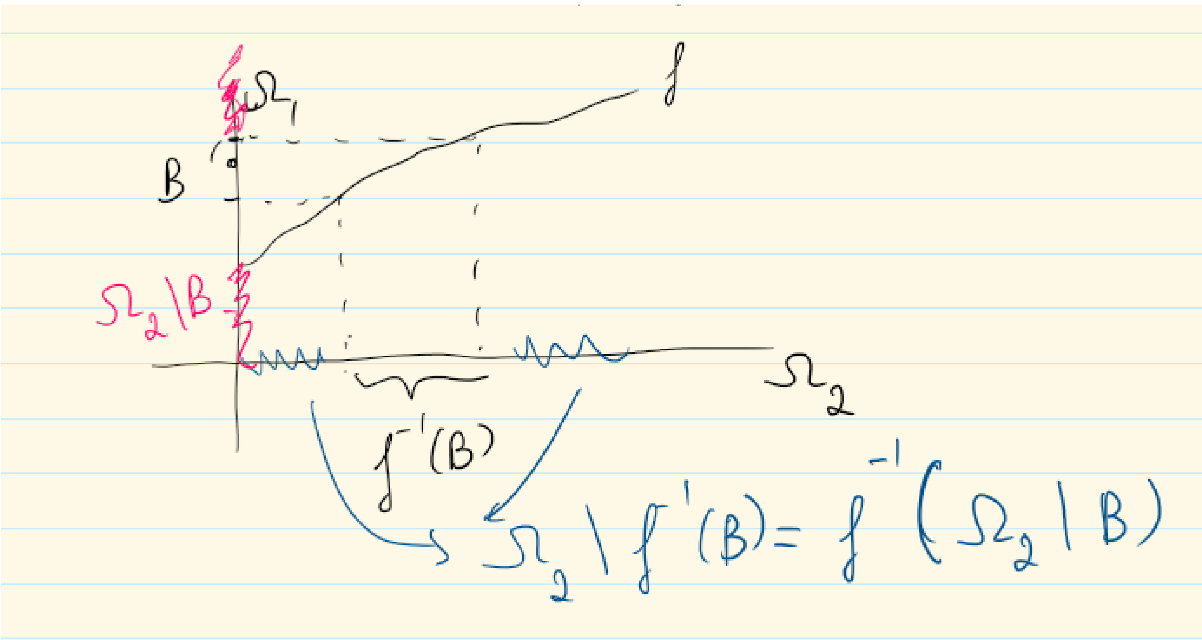


Figure 8: Illustrative Theorem 1.1.3 proof

So C is a sigma algebra, and C also contains A. Now since

$$F_2 = \sigma(A) = \bigcup_{A \subseteq F; F \text{ is a sigma algebra}} F$$

We must have that $F_2 \subseteq C$. Hence it is true on all F_2. □

Definition 34 (Random vector)**.** Let $X : (\Omega; F; \mathbf{P}) \to (\mathbb{R}^d; B(\mathbb{R}^d); \mathbf{P})$ be a measurable function from the probability space $(\Omega; F; \mathbf{P})$ and $d > 1$. Then X is called a **random vector**.

Example 35. Here are some examples of random variables:

1. Tossing coins 5 times. Then $\Omega = \{(\omega_1; \ldots; \omega_5) : \omega_i = H \text{ or } T\}$. Take $F = 2^\Omega$.

   Take $\mathbf{P}$ is such that $\mathbf{P}(\{H\}) = p$ and $\mathbf{P}(\{T\}) = 1 - p$. An example of $\mathbf{P}$ is such as $\mathbf{P}(\{HHHHH\}) = p^5$. Another example is that $\mathbf{P}(\{HHTTT\}) = p^2(1-p)^3$.

   We then can define a random variable that counts the number of H's:

$$X(\omega) = \sum_{i=1}^{5} \mathbf{1}_{\{\omega_i = H\}}$$

   Then

$$\mathbf{P}(X(\omega) = k) = \binom{5}{k} p^k (1-p)^{5-k}$$

   Why do we have $\binom{5}{k}$? For example, if we are taking $k = 1$, we could have the choices $\{HTTTT\}$,$\{THTTT\}; \{TTHTT\} \ldots$. We are choosing $k = 1$ of H out of all the 5 choices of tossing.

2. Random variable that count the first H: $\Omega = \{(\omega_1, \ldots, \omega_n, \ldots) : \omega_i = H \text{ or } T\}$ and $\mathcal{F} = 2^{\Omega}$ and μ as in previous example. Note that $\omega \in \Omega$ can have **infinite** components. We define the random variable:

$$T = \inf\{n : \omega_n = H\}$$

And we also have:

$$\mathbf{P}(T = k) = p(1-p)^{k-1}$$

Definition 36 (Distribution function induced by random variable)**.** Let $X : (\Omega, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ be a random variable. Note that $\mathbf{P}(X \in B) = \mu(X^{-1}(B))$. We can define the **distribution function cdf** of X by:

$$F_X(x) := \mathbf{P}(X \in (-\infty, x]) = \mu(X^{-1}((-\infty, x]))$$

We remark that this cdf F_X also have the property of being increasing, right continuous, and $F_X(-\infty) = 0$ and $F_X(\infty) = 1$.

Remark 37. We will see later a concept that if 2 random variables X and Y has identical distributions, that is, $F_X(x) = F_Y(x)$, then we said X and Y are **equal in distribution**.

Question 38. How about the other direction, **random variable induced by a distribution function**?

Consider F a function on $\mathbb{R}$ that is increasing, right-continuous and that $F(\infty) = 1$ and $F(-\infty) = 0$. Then you can define a random variable on $([0,1]\mathcal{B}([0,1]), \lambda)$ (note that $\lambda((a,b)) = b - a$ as follow:

$$X^+(\omega) = \inf\{x : F(x) > \omega\}$$

and

$$X^-(\omega) = \inf\{x : F(x) \geq \omega\}$$

and

$$F_{X^-}(x) = F(x)$$

Theorem 39 (Theorem 1.3.4 Durett's book: compose function of random variable)**.** *If $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ and $f : (S, \mathcal{S}) \to (T, \mathcal{T})$ are measurable functions, then*

$$f(X) : (\Omega, \mathcal{F}) \to (T, \mathcal{T}) \text{ is also a measurable function}$$

Proof. We need to show that for all $B \in \mathcal{T}$, we also have $[f(X)]^{-1}(B) \in \mathcal{F}$.

$$\begin{aligned}
[f(X)]^{-1}(B) &= \{\omega \in \Omega : f(X(\omega)) \in B\} \\
&= \{\omega \in \Omega : X(\omega) \in f^{-1}(B)\}; \text{ since } f \text{ is measurable, } f^{-1}(B) \in \mathcal{S} \\
&= \{\omega \in \Omega : \omega \in X^{-1}(Y)\} \text{ such that } Y \in \mathcal{S} \\
&\in \mathcal{F}; \text{ since } X \text{ is measurable}
\end{aligned}$$

□

Application of the previous theorem:

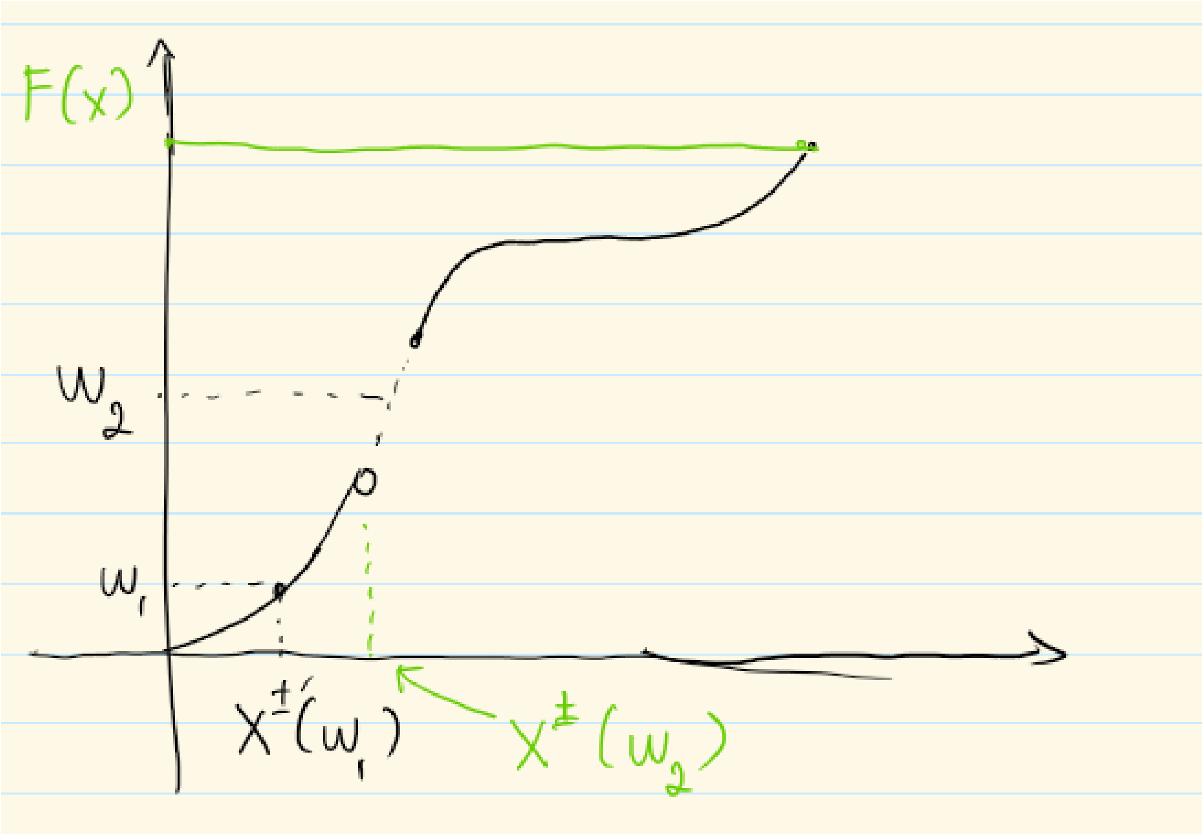


Figure 9: random variable induced by cdf

Example 40. If $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a random variable, then cX is also a random variable, for any $c > 0$ real number.

Proof. By theorem (39), it suffices to show that $f(x) = cx$ is measurable from $\mathbb{R} \to \mathbb{R}$. To check this, we need to check for any $a \in \mathbb{R}$, does $f^{-1}((-\infty, a]) \in \mathcal{B}(\mathbb{R})$?

$$\begin{aligned} f^{-1}((-\infty, a]) &= \{x \in \mathbb{R} : cx \in (-\infty, a]\} \\ &= \{x \in \mathbb{R} : cx \leq a\} \\ &= \{x \in \mathbb{R} : x \leq \frac{a}{c}\} \in \mathcal{B}(\mathbb{R}) \end{aligned}$$

□

Example 41. Any "nice" function of a random variable X is also a random variable. For instance, $\sin(X), X^2, X^3, e^X \ldots$ are all random variables.

Example 42 (Theorem 1.35 in Durett book)**.** If $X_1, \ldots, X_n$ are random variables and $f : (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a measurable function, then:

$$f(X_1, \ldots, X_n) \text{ is also a random variable.}$$

Example 43 (Theorem 1.36 Durett book)**.** $X_1 + \ldots X_n$ is a random variable.

Proof. It suffices to show that $f(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$ is a measurable function.

$$\begin{aligned} f^{-1}((-\infty, a]) &= \{x \in \mathbb{R}^n : f(x) \leq a\} \\ &= \{x \in \mathbb{R}^n : x_1 + x_2 + \ldots + x_n \leq a\}, \text{ which is a closed set, therefore } \in \mathcal{B}(\mathbb{R}). \end{aligned}$$

□

Remark 44. The $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}((a_1, b_1] \times \ldots \times (a_n, b_n])$

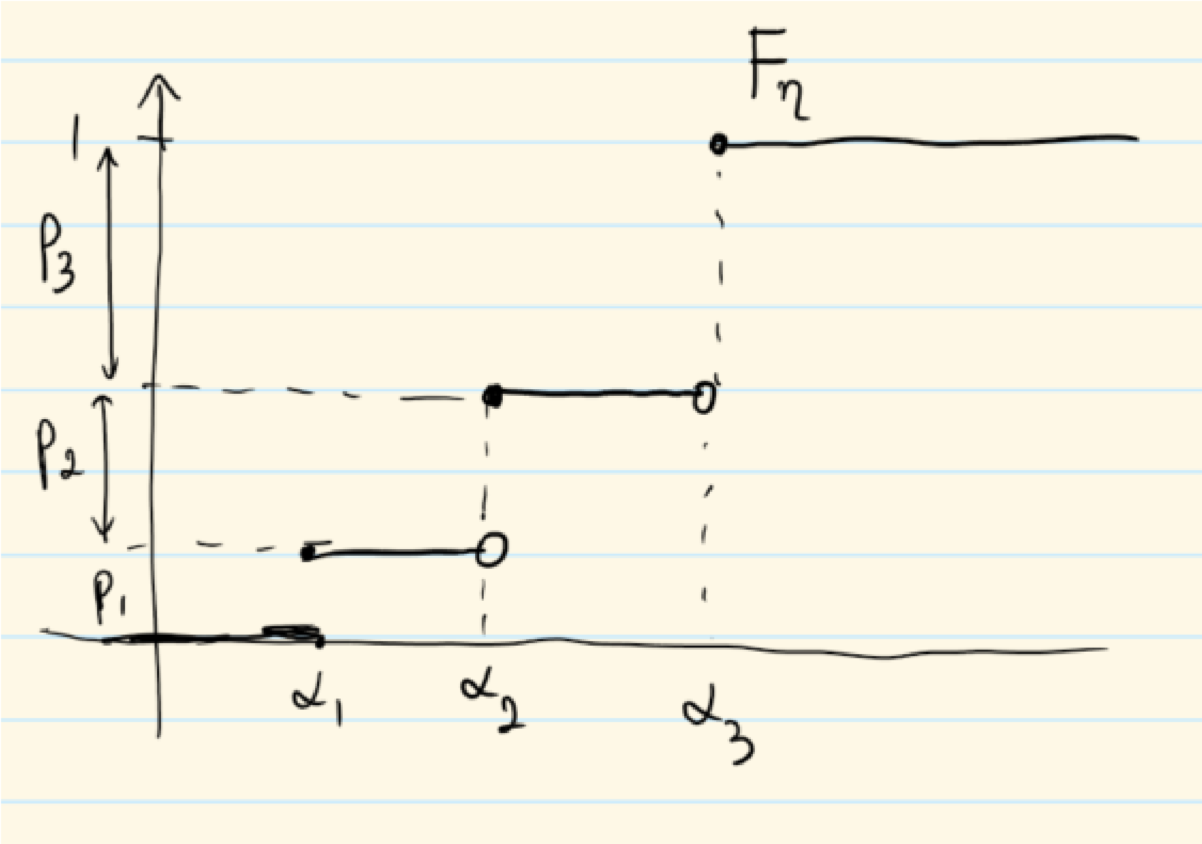


Figure 10: random sample generating

3.2 side note: how to generate random samples of random variables?

Note that random samples have graph look like uniform distribution.
Consider $\ (!)$ on $(\mathrm{R}; B(\mathrm{R}); \ )$. Then consider the cdf $F\ (x) = \ (\ \ ^{1}((\ \ 1\,; x]))$.
We can generate samples of $\ _{!_i} = \ _i$ by:

$$F\ ^{1}(\ _i)\ \text{,where the}\ \ _i\ \text{is a sample of uniform random variable on}\ [0; 1].$$

Example 45.

$$_{(0;1)}(\ ) = \begin{cases} \ _1 & ; \ (\ = \ _1) = p_1 \\ \ _2 & ; \ (\ = \ _2) = p_2 \\ \ _3 & ; \ (\ = \ _3) = p_3 \end{cases}$$

where $\ _1 < \ _2 < \ _3$ and that $p_1 + p_2 + p_3 = 1$.

1. If $\ _i\ 2\ [0; p_1)$ then $\ _i = \ _1$
2. If $\ _i\ 2\ [p_1; p_1 + p_2)$ then $\ _i = \ _2$
3. If $\ _i\ 2\ [p_2; p_1 + p_2 + p_3)$ then $\ _i = \ _3$

In general we need to check the following:
Define $Y = F^{\ 1}(\ )$ where is $U[0;1]$. Then the distribution of Y is $F_Y(y)$ defined as:

$$\begin{aligned} F_Y(y) &= \ (\{Y \quad y\}) \\ &= \ (\{F^{\ 1}(\ ) \quad y\}) \\ &= \ (\quad F(y)) \\ &= F(F(y)) \\ &= F(y) \end{aligned}$$

The last equation above is because $F(y) = y$ for a cdf of a uniform random variable.

Example 46 (generating random samples of exponential random variable X)**.** The pdf density function is

$$f_X(x) = \begin{cases} e^{\ x} & ; x \quad 0 \\ 0 & ; x < 0 \end{cases}$$

while the cdf is

$$F_X(x) = \begin{cases} 1 \quad e^{\ x} & ; x \quad 0 \\ 0 & ; x < 0 \end{cases}$$

To generate unique $[0;1]$ sample $_i$:
Set $_i = F_X^{\ 1}(\ _i)$. That is, $F_X(\ _i) = \ _i$. Using formula of F_X above, we get: $1 \quad e^{\ \ i} = \ _i$. That is $1 \quad _i = e^{\ \ i}$. Now take log gives:

$$\log(1 \quad _i) = \quad _i$$

So we deduce

$$_i = \frac{1}{\ } \log(1 \quad _i) = \frac{1}{\ } \log(\ _i)$$

The last equality is because if $_i \quad U(0;1]$ then $1 \quad _i \quad U(0;1]$.

4 Day 4: Expectation of random variables, some useful inequalities

4.1 Expectation

Remark 47. Expectation is defined as integral, and is denoted as $\mathbf{E}[X]$. It is an indication of average/mean value.

Definition 48 (expectation of random variable)**.** The **expectation of a random variable** X on $(\Omega; F; \ )$ is defined as:

$$\mathbf{E}[X] = \int X(!)d \ (!)$$

And it is also on $(\mathsf{R}; B(\mathsf{R}); \mathbf{P})$:

$$\mathbf{E}[X] = \int_{\mathsf{R}} X d\mathbf{P}(x)$$

And if the pdf density $f(x)$ exists:

$$\mathbf{E}[X] = \int_{\mathsf{R}} x f(x) dx$$

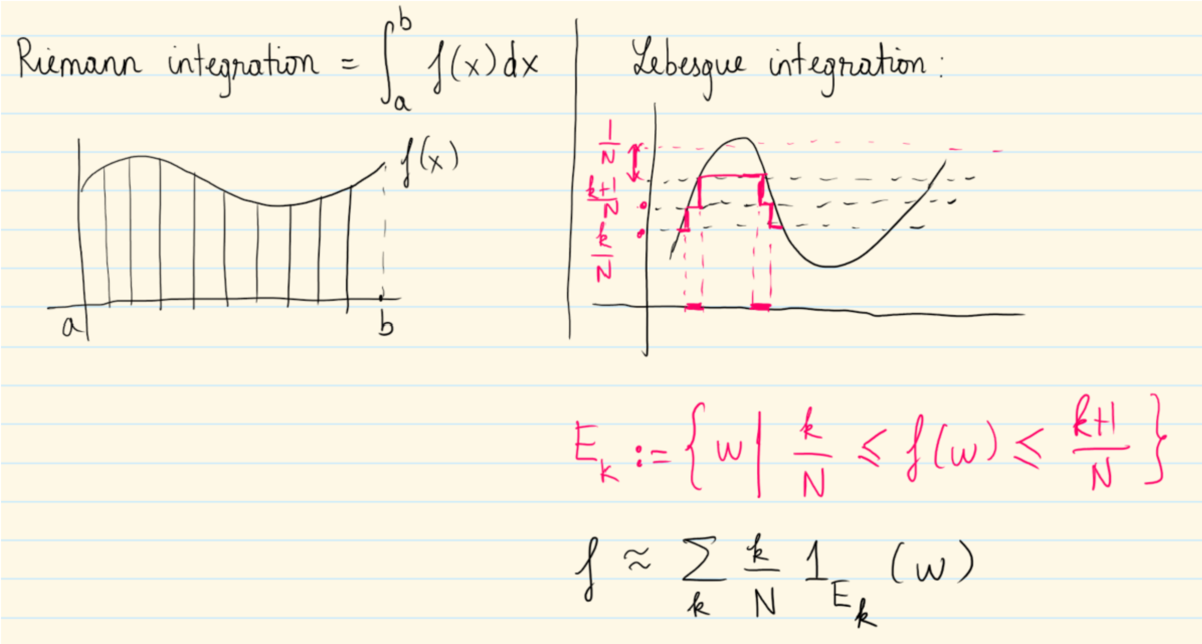


Figure 11: Riemann vs Lebsgue integral

We now need to define **Lebesgue Integration** for measurable function (so that we can use it for expectation of random variables). A good illustrative compare between traditional Riemann and Lebsgue integration can be found in Figure (11).

Steps to define Expectation of random variables (or integral of measurable functions):

1. Step 1: Simple function random variable: $\varphi(\omega) = \sum_{j=1}^{n} a_j \mathbf{1}_{A_j}(\omega)$, where the A_j are pairwise disjoint.
2. Step 2: bounded function random variables.
3. Step 3: non-negative function random variables.
4. Step 4: measurable function random variables.

We will now begin each step:

Step 1: (Integral of step function) Expectation of step function random variable $\varphi(\omega)$:

$$\begin{aligned}\int \varphi(\omega) d\mu(\omega) &= \int \sum_{j=1}^{n} a_j \mathbf{1}_{A_j}(\omega) d\mu(\omega) \\ &= \sum_{j=1}^{n} a_j \int \mathbf{1}_{A_j}(\omega) d\mu(\omega) \\ &= \sum_{j=1}^{n} a_j \mu(A_j)\end{aligned}$$

Properties of Expectation of step functions random variables:

1. If $\varphi(\omega) \geq 0$ except measure zero, then $\mathbf{E}[\varphi] \geq 0$.

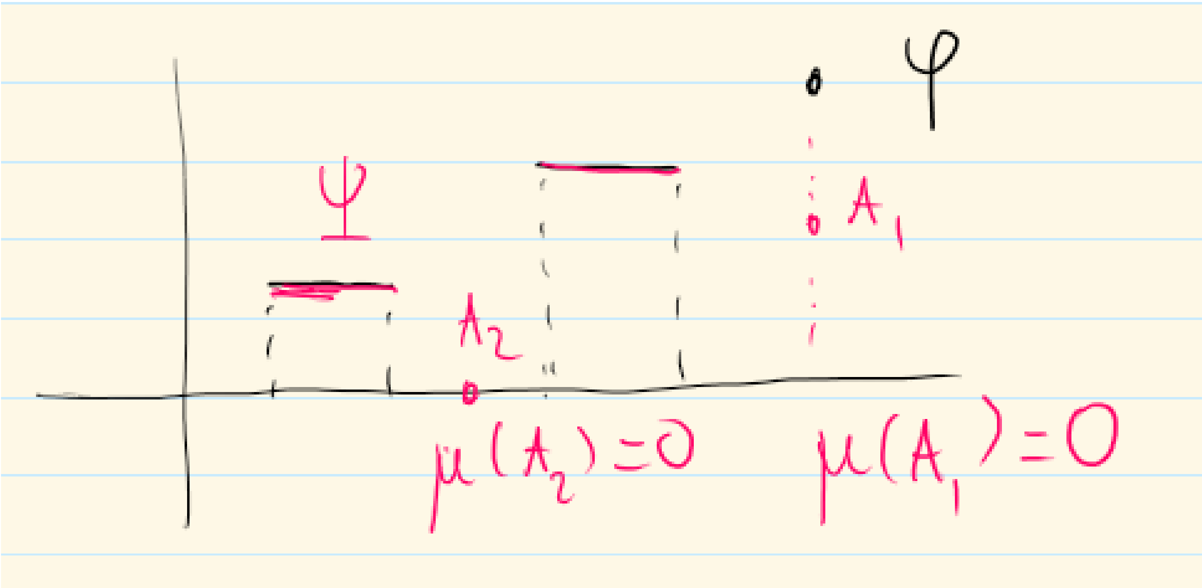


Figure 12: equal of expectation of function equal except measure zero

2. For $a \in \mathbb{R}$, then $a\mathbf{E}[\varphi] = \mathbf{E}[a\varphi]$.

3. $\mathbf{E}[\varphi + \psi] = \mathbf{E}[\varphi] + \mathbf{E}[\psi]$

   Proof. Define $\varphi + \psi$ as simple functions on $A_i \cup B_j$, where $\varphi(\omega) = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(\omega)$ and $\psi(\omega) = \sum_{j=1}^{n} a_j \mathbf{1}_{B_j}(\omega)$. □

4. (monotonicity) If $\varphi \leq \psi$ except measure zero, then $\mathbf{E}[\varphi] \leq \mathbf{E}[\psi]$.

   Proof. Since $\psi - \varphi \geq 0$, so $\mathbf{E}[\psi - \varphi] \geq 0$. Writing $\psi = \varphi + (\psi - \varphi)$ and take expectation (integration) gives:

$$\mathbf{E}[\psi] = \mathbf{E}[\varphi] + \mathbf{E}[\psi - \varphi] \geq \mathbf{E}[\varphi]$$

   The last inequality is true because $\mathbf{E}[\varphi - \psi] \leq 0$ □

5. If $\varphi = \psi$ except on a measure set, then $\mathbf{E}[\varphi] = \mathbf{E}[\psi]$. See figure(12) for illustration.

6. $\mathbf{E}[|\varphi|] \geq |\mathbf{E}[\varphi]|$

Step 2: bounded function random variable X, that is, there exists simple function φ and ψ such that $\varphi(\omega) \leq X(\omega) \leq \psi(\omega)$. Then $\mathbf{E}[\varphi] \leq \mathbf{E}[x] \leq \mathbf{E}[\psi]$. Taking sup and inf:

$$\sup_{\varphi \leq X} \mathbf{E}[\varphi] \leq \mathbf{E}[X] \leq \inf_{X \leq \psi} \mathbf{E}[\psi]$$

And that is how we define

$$\mathbf{E}[X] := \sup_{\varphi \leq X} \mathbf{E}[\varphi] = \inf_{X \leq \psi} \mathbf{E}[\psi]$$

when the $\sup_{\varphi \leq X} \mathbf{E}[\varphi] = \inf_{X \leq \psi} \mathbf{E}[\psi]$.

Step 3: non-negative functions X:. We take $X_n(\omega) = \max\{X(\omega), n\}$. Then each X_n is bounded function random variable and $X_1 \leq X_2 \leq \dots$ and $X_n \uparrow X$ as $n \to \infty$. Then we define

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

Step 4: measurable function X: We write $X = X^+ - X^-$, where:

$$X^+ = \max\{0, X\} = \begin{cases} 0 & \text{if } X(\omega) \leq 0 \\ X(\omega) & \text{if } X(\omega) > 0 \end{cases}$$

and

$$X^- = \max\{-X, 0\} = \begin{cases} 0 & \text{if } X(\omega) \geq 0 \\ -X(\omega) & \text{if } X(\omega) < 0 \end{cases}$$

Then both X^+ and X^- are non-negative, and we define:

$$\mathbf{E}[X] = \mathbf{E}[X^+] - \mathbf{E}[X^-]$$

Definition 49 (integrable)**.** A measurable function X is **integrable** if $\int |X| d\mu < \infty$. Note that $|X| = X^+ + X^-$. So $\int |X| d\mu < \infty$ is equivalent to $\int X^+ d\mu + \int X^- d\mu < \infty$, that is, at least one of $\int X^+ d\mu$ or $\int X^- d\mu$ is finite.

Remark 50. Interchange the expectation and the limit is **not** always possible! That is, if $\lim_{n \to \infty} X_n = X$, it is not always the case that

$$\lim_{n \to \infty} E[X_n] \stackrel{?}{=} \mathbf{E}[\lim_{n \to \infty} X_n] = \mathbf{E}[X]$$

Example 51. Consider $X_n(\omega)$ on $([0, 1], B([0, 1]), \lambda)$ as followed:

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

Then $\lim_{n \to \infty} X_n(\omega) = 0 := X$ except on a measure zero set (the singleton $\{\omega = 0\}$). But

$$\mathbf{E}[X_n] = n \cdot \lambda([0, \frac{1}{n}]) = n \cdot \frac{1}{n} = 1 \neq \mathbf{E}[X] = 0.$$

Remark 52. When can we switch expectation and limit: in **Monotone Convergence Theorem- MCT** and in **Dominated Convergence Theorem -DCT**.

Theorem 53 (Monotone Convergence Theorem)**.** *(Theorem 1.6.6 and 1.5.7) Consider X_n are random variables on (Ω, F, P) such that $0 \leq X_1 \leq X_2 \leq \dots$ and that $X_n \uparrow X$. Then:*

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

Remark 54. It is important that these (measurable functions) random variables are all non-negative.

Proof. We will prove by showing 2 inequalities:

1. For $\leq$: Since $X_n \leq X$ for all n, so $\mathbf{E}[X_n] \leq \mathbf{E}[X]$ for all n. Therefore taking the limit gives:

$$\lim_{n \to \infty} \mathbf{E}[X_n] \leq \mathbf{E}[X]$$

2. For $\geq$: There exists simple function φ such that $X \geq \varphi \geq 0$. In addition, there exists a $t \in (0, 1)$ such that:

$$E_n := \{\omega : X_n(\omega) \geq t\varphi(\omega)\}$$

   and that

$$E_1 \subseteq E_2 \subseteq \dots \subseteq E_n \to \Omega \text{ as } n \to \infty.$$

   Note that such a t exists because of $X_n \to X$ (if no such t exists, the X_n will then $< \varphi$ for all n, and so their limit which is X will also $< \varphi$). Then:

$$\begin{aligned}
\mathbf{E}[X_n] &\geq \mathbf{E}[X_n \mathbf{1}_{E_n}], \text{ since } X_n \text{ is greater than truncated } X \wedge n \text{ on } E_n \\
&\geq \mathbf{E}[t\varphi \mathbf{1}_{E_n}], \text{ since on } E_n,\ X_n \geq t\varphi \\
&= t\mathbf{E}[\varphi \mathbf{1}_{E_n}]
\end{aligned}$$

   Thus

$$\lim_{n \to \infty} \mathbf{E}[X_n] \geq t\mathbf{E}[\varphi]$$

   Letting $t \uparrow 1$ and take the $\sup_{X \geq \varphi}$ we get:

$$\lim_{n \to \infty} \mathbf{E}[X_n] \geq \sup_{X \geq \varphi} \mathbf{E}[\varphi] = \mathbf{E}[X]$$

□

Definition 55 (Limit infimume definition)**.**

$$\liminf_{n \to \infty} = \lim_{n \to \infty} \inf_{k \geq n} a_k = \sup_{n} \inf_{k \geq n} a_k$$

Note that the $s_n = \inf_{k \geq n} a_k$ is a sequence of non-decreasing for each n. Pictorially, we can think of lim inf as the limit going on the vertical direction. As a reminder, $\inf A$ is the greatest lower bound in A. See below figure (13) for illustration.

Theorem 56 (Fatou's Lemma)**.** *Let $X_n \geq 0$. Then:*

$$\mathbf{E}[\liminf_{n \to \infty} X_n] \leq \liminf_{n \to \infty} \mathbf{E}[X_n]$$

Proof. Recall $\liminf_{n \to \infty} X_n = \sup_n \inf_{k \geq n} X_k$. Denote $Y_n = \inf_{k \geq n} X_k$. Then the Y_n is non-decreasing and non-negative, so we can apply MCT to the Y_n:

$$\begin{aligned}
\liminf_{n \to \infty} \mathbf{E}[X_n] &= \lim_{n \to \infty} \mathbf{E}[X_n]; \text{ just definition of limit, } \lim = \liminf \\
&\geq \lim_{n \to \infty} \mathbf{E}[\inf_{k \geq n} X_k]; \text{ since } X_n \geq Y_n = \inf_{k \geq n} X_k \text{ for each } n \\
&= \lim_{n \to \infty} \mathbf{E}[Y_n] \\
&= \mathbf{E}[\lim_{n \to \infty} Y_n]; \text{ use MCT for } Y_n : \lim_{n \to \infty} \mathbf{E}[Y_n] = \mathbf{E}[\lim_{n \to \infty} Y_n] \\
&= \mathbf{E}[\liminf_{n \to \infty} X_n]
\end{aligned}$$

□

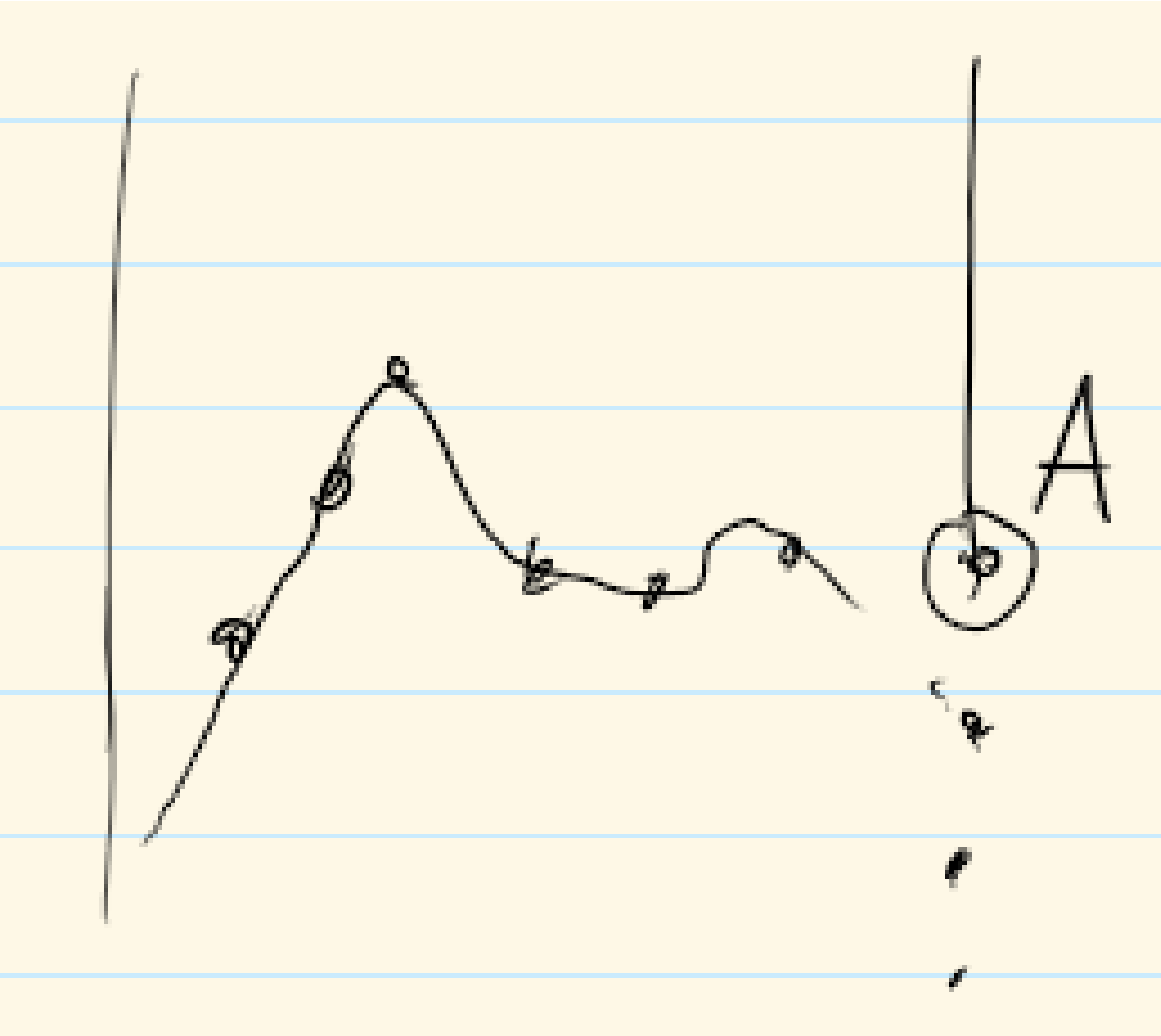


Figure 13: $\liminf$ illustration

Remark 57. We use Fatou's Lemma to prove Dominated Convergence Theorem DCT.

Theorem 58 (Dominated Convergence Theorem)**.** *(This is Theorem 1.6.7 or 1.5.8) Consider X_n random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\lim_{n \to \infty} X_n = X$. If there is a Y random variable such that $|X_n| \leq Y$ for all n, and that $\mathbf{E}[Y] < \infty$, then:*

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

Proof. Since $|X_n| \leq Y$, so $-Y \leq X_n \leq Y$ for all n. That is, $Y + X_n \geq 0$ and $Y - X_n \geq 0$ for all n. Then we can apply Fatou's Lemma for both $Y + X_n$ and $Y - X_n$:

1. Apply Fatou's lemma to $Y + X_n$ gives:

   $$\mathbf{E}[\liminf(Y + X_n)] \leq \liminf \mathbf{E}[Y + X_n]$$

   For the LHS:

   $$\liminf(Y + X_n) = \liminf(Y) + \liminf(X_n) = Y + \lim(X_n) = Y + X$$

   where the last equality is true because $\liminf(X_n) = \lim(X_n) = X$. Thus LHS $= \mathbf{E}[Y] + \mathbf{E}[X]$.

   For the RHS:

   $$\liminf_{n \to \infty} \mathbf{E}[Y + X_n] = \liminf_{n \to \infty} (\mathbf{E}[Y] + \mathbf{E}[X_n]) = \mathbf{E}[Y] + \liminf_{n \to \infty} \mathbf{E}[X_n]$$

   In conclution we get:

   $$\mathbf{E}[Y] + \mathbf{E}[X] \leq \mathbf{E}[Y] + \liminf_{n \to \infty} \mathbf{E}[X_n]$$

   And because $\mathbf{E}[Y] < \infty$ we can subtract it from both sides of the inequality and get:

   $$E[X] \leq \liminf_{n \to \infty} \mathbf{E}[X_n]$$

2. Apply Fatou's lemma to $Y - X_n$ we get:

   $$\mathbf{E}[\liminf(Y - X_n)] \leq \liminf \mathbf{E}[Y - X_n]$$

   In a similar argument as part (1), we get:

   $$\mathbf{E}[-X] \leq \liminf_{n \to \infty} \mathbf{E}[-X_n]$$

   So

   $$\mathbf{E}[X] \geq -\liminf_{n \to \infty} \mathbf{E}[-X_n] = \limsup_{n \to \infty} \mathbf{E}[X_n]$$

   where the last equality is a property of lim sup and lim inf, that $\limsup(x_n) = -\liminf(-x_n)$.

Combining both (1) and (2):

$$\limsup \mathbf{E}[X_n] \leq \mathbf{E}[X] \leq \liminf \mathbf{E}[X_n]$$

But we always have

$$\liminf \mathbf{E}[X_n] \leq \limsup \mathbf{E}[X_n]$$

So this gives:

$$\limsup \mathbf{E}[X_n] = \liminf \mathbf{E}[X_n] = \lim \mathbf{E}[X_n] = \mathbf{E}[X]$$

□

4.2 some useful inequalities regarding expectation

Theorem 59 (Jensen's Inequality)**.** *(This is theorem 1.6.2 or 1.5.1) Consider random variable X on $(\Omega, F, \mathbb{P})$ such that $\mathbf{E}[X] < \infty$, and a convex function φ. Note that convex function mean for any $\lambda \in (0, 1)$: $\lambda\varphi(x) + (1-\lambda)\varphi(y) \geq \varphi(\lambda x + (1-\lambda)y)$. Then we get:*

$$\mathbf{E}[\varphi(X)] \geq \varphi(\mathbf{E}[X])$$

Proof. Denote $c_0 = \mathbf{E}[X]$. Since φ is convex, we can find B such that $\varphi(x) - \varphi(c_0) \geq B(x - c_0)$. Now this work for all $x \in \mathbb{R}$, so we replace x by $X(\omega) \in \mathbb{R}$ and get:

$$\varphi(X) - \varphi(c_0) \geq B(X - c_0)$$

Taking expectation gives:

$$\mathbf{E}[\varphi(X)] - \mathbf{E}[\varphi(c_0)] \geq B\mathbf{E}[X - c_0] = B\left(\mathbf{E}[X] - \mathbf{E}(c_0)\right)$$

But since c_0 is a constant, then $\mathbf{E}[c_0] = c_0$. And similarly, since $\varphi(c_0)$ is a constant, so $\mathbf{E}[\varphi(c_0)] = \varphi(c_0)$. So the above inequality is:

$$\mathbf{E}[\varphi(X)] - \varphi(c_0) \geq B(\mathbf{E}[X] - c_0) = B(\mathbf{E}[X] - \mathbf{E}[X]) = 0$$

since $c_0 = \mathbf{E}[X]$. Hence we get:

$$\mathbf{E}[\varphi(X)] \geq \varphi(c_0) = \varphi(\mathbf{E}[X])$$

□

Remark 60. When we apply Jensen inequality to some useful function we get:

1. $\varphi(X) = |X|$:

$$\mathbf{E}\left[|X|\right] \geq \left|\mathbf{E}[X]\right|$$

2. $\varphi(X) = X^2$:

$$\mathbf{E}[X^2] \geq \left(\mathbf{E}[X]\right)^2$$

   This is relation between moments.

Theorem 61 (Holder Inequality)**.** *Let $p, q \in (1, \infty) : \frac{1}{p} + \frac{1}{q} = 1$. X, Y random variables, then:*

$$\mathbf{E}[|XY|] \leq \|X\|_p\|Y\|_q$$

where

$$\|X\|_p := (\mathbf{E}[X^p])^{\frac{1}{p}}$$

Proof. If $\|X\|_p$ or $\|Y\|_q$ is zero and infinity, the inequality is trivial. So consider $\|X\|_p, \|Y\|_q \in (0, \infty)$. Define:

$$a := \frac{|X|^p}{\|X\|_p^p}$$

and

$$b := \frac{|Y|^q}{\|Y\|_q^q}$$

Using the convexity of the function $-\log(x)$, for any $\lambda \in (0,1)$, we have:

$$-\log(\lambda a + (1-\lambda)b) \leq \lambda(-\log(a)) + (1-\lambda)(-\log(b)) = -\log(a^{\lambda} b^{1-\lambda})$$

So

$$a^{\lambda} b^{1-\lambda} \leq \lambda a + (1-\lambda)b$$

With the above inequality, consider $\lambda = \frac{1}{p}$ and a, b as we defined (note that $1-\lambda = \frac{1}{q}$):

$$\left(\frac{|X|^p}{\|X\|_p^p}\right)^{\frac{1}{p}} \left(\frac{|Y|^q}{\|Y\|_q^q}\right)^{\frac{1}{q}} \leq \frac{1}{p}\left(\frac{|X|^p}{\|X\|_p^p}\right) + \frac{1}{q}\left(\frac{|Y|^q}{\|Y\|_q^q}\right)$$

Take expectation:

$$\frac{\mathbf{E}[|XY|]}{\|X\|_p\|Y\|_q} \leq \frac{1}{p}.1 + \frac{1}{q}.1 = 1$$

□

Remark 62 (Cauchy-Schwarz Inequality)**.** A special case of Holder inequality is when $p = q = 2$ we get the **Cauchy-Schwarz Inequality**:

$$\left(\mathbf{E}[XY]\right)^2 \leq \mathbf{E}[X^2]\mathbf{E}[Y^2]$$

A relation between covariance and variance.

5 Day 5: Modes of convergence

5.1 almost surely convergence/ almost everywhere

Recall last time MCT and DCT allow us to interchange limits and Expectation (integral). In doing this, we have an assumption that $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ except on a measure zero set. What does this mean?

Definition 63 (except on a measure zero)**.** $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ except on a measure zero set means the set

$$N := \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\}$$

has measure zero, i.e, $\mu(N) = 0$.

Definition 64 (almost surely convergence)**.** We says X_n converges to X **almost surely** if

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\}) = 0$$

We denoted this as $X_n \xrightarrow{a.s.} X$.

Remark 65. The condition

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\}) = 0$$

in a probability measure $\mathbb{P}$ also mean that

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$

So we sometimes say this as X_n converges to X **with probability 1**. Keep in mind this is not the same with convergence in probability!

Remark 66. Sure convergence is pointwise convergence, that is $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for all $\omega \in \Omega$. We will later show that pointwise $\Rightarrow$ almost surely $\Rightarrow$ in probability.

5.2 Converge in probability- converge in measure

Definition 67 (converges in probability)**.** X_n converges to X **in probability**, denoted $X_n \xrightarrow{\mathbf{P}} X$ (or converges in measure, $X_n \xrightarrow{\mu} X$) if:
For any $\varepsilon > 0$:

$$\lim_{n \to \infty} \mathbf{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0$$

(Replace $\mathbf{P}$ by μ to get converge in measure)

Theorem 68 (Convergence a.s. implies convergence in probability)**.** *If $X_n \xrightarrow{as} X$, then $X_n \xrightarrow{P} X$*

Proof. Denote $N = \{\lim_{n \to \infty} X_n(\omega) \neq X(\omega)\}$. Then since $X_n \xrightarrow{as} X$ we have $\mathbf{P}(N) = 0$. Let $\varepsilon > 0$ be given and define the following sets:

$$A_n := \{|X_j(\omega) - X(\omega)| \geq \varepsilon; \text{ for some } j \geq n\}$$

Note that

$$A_n = \bigcup_{j=n}^{\infty} \{|X_j(\omega) - X(\omega)| \geq \varepsilon\}$$

Then we also have $A_1 \supseteq A_2 \supseteq \dots$. Denote $A := \bigcap_{n=1}^{\infty} A_n$.Then $\lim_{n \to \infty} \mathbf{P}(A_n) = \mathbf{P}(A)$. We claim that $A \subseteq N$ since if $\omega \in A$, then $\omega \in N$. Thus by monotonicity, $\mathbf{P}(A) \leq \mathbf{P}(N) = 0$. So we must have $\mathbf{P}(A) = 0$. That is:

$$\lim_{n \to \infty} \mathbf{P}(A_n) = 0$$

Now since

$$A_n = \bigcup_{j=n}^{\infty} \{|X_j(\omega) - X(\omega)| \geq \varepsilon\} \supseteq \{|X_n(\omega) - X(\omega)| \geq \varepsilon\}$$

Taking probability measure:

$$\mathbf{P}(A_n) \geq \mathbf{P}(\{|X_n(\omega) - X(\omega)| \geq \varepsilon\})$$

And letting limit $n \to \infty$ we get:

$$0 = \lim_{n \to \infty} \mathbf{P}(A_n) \geq \lim_{n \to \infty} \mathbf{P}(\{|X_n(\omega) - X(\omega)| \geq \varepsilon\})$$

Therefore:

$$\mathbf{P}(\{|X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0$$

And we concluded that $X_n \xrightarrow{\mathbf{P}} X$. □

Remark 69. In general, $X_n \xrightarrow{as} X$ does **not** imply $X_n \xrightarrow{\mu} X$, because we need the condition that $\mu(\Omega) < \infty$.

Question 70. Does $X_n \xrightarrow{\mathbf{P}} X$ implies $X_n \xrightarrow{as} X$? **NO.**

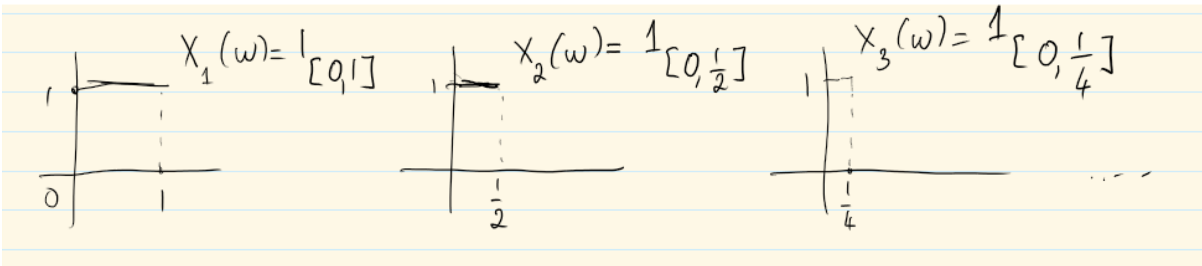


Figure 14: converge in probability does not imply converge as

Example 71 (Counter example to show that $X_n \xrightarrow{P}$ does NOT imply $X_n \xrightarrow{as} X$)**.** Consider:

$$X_n(\omega) = \mathbf{1}_{[0, \frac{1}{2^n}]}$$

For example, $X_2(\omega) = \mathbf{1}_{[0, \frac{1}{4}]}(\omega)$, see figure (14 for illustration. With these X_n, we got $X_n \xrightarrow{P} 0$, since, let $\varepsilon > 0$ be given. For large enough $N(\delta)$ depends on δ, we can find $N(\delta)$ such that with all $n \geq N(\delta)$ and for all $\delta > 0$:

$$\mathbf{P}(\{|X_n - 0| > \varepsilon\}) < \delta$$

This is because

$$\mathbf{P}(\{|X_n - 0| > \varepsilon\}) = \mathbf{P}(\{|X_n| > \varepsilon\}) = \mathbf{P}(\mathbf{1}_{[1, 1/2^n]} > \varepsilon)$$

and when we find large enough N, for $n > N$, then interval $[0, \frac{1}{2^n}]$ will have small length, which gives small measure and therefore is less then whichever δ that you choose.
Now, note that $X_n(\omega)$ does not converges a.s. to 0. This is because for any $\omega \in (0, 1)$, the set $A_n = \{\omega : \mathbf{1}_{[0, \frac{1}{2^n}]}(\omega) > 0\}$ will contains $[0, \frac{1}{N}]$ such that $\frac{1}{N} \geq \omega$ and $\mathbf{P}([1, 1/N]) > 0$, therefore $\mathbf{P}(A_n) \neq 0$ for all n. So $\lim_{n \to \infty} \mathbf{P}(A_n) > 0$, and that means X_n cannot converge to 0 a.s.

However, we have the following theorem (not full sequence, but a subsequence):

Theorem 72. *If $X_n \xrightarrow{P} X$, then there exists a subsequence n_k such that $X_{n_k} \xrightarrow{as} X$.*

Proof. By definition of convergence in probability, we have: $\mathbf{P}(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) \to 0$. That is, for any k, (letting $\varepsilon = \frac{1}{k}$), we have:

$$\mathbf{P}(\{\omega : |X_n(\omega) - X(\omega)| > \frac{1}{k}\}) \to 0$$

Since the limit goes to zero, we can take n_k such that:

$$\mathbf{P}(\{\omega : |X_n(\omega) - X(\omega)| > \frac{1}{k}\}) < \frac{1}{2^k}, \forall n > n_k$$

Claim: $X_{n_k} \xrightarrow{as} X$.
If at some ω, we have $X_{n_k}(\omega) \nrightarrow X(\omega)$ with $j > N > k$, then:

$$\begin{aligned} B_k &= \{|X_{n_j}(\omega) - X(\omega)| > \frac{1}{k} \text{ for some } j \geq N\} \\ &= \bigcup_{j=N}^{\infty} \{|X_{n_j}(\omega) - X(\omega)| > \frac{1}{k}\} \\ &\subseteq \bigcup_{j=N}^{\infty} \{|X_{n_j}(\omega) - X(\omega)| > \frac{1}{j}\}, \text{ since } j > k \end{aligned}$$

Now taking probability measure:

$$\mathbf{P}\Big(\bigcup_{j=N}^{\infty}\{|X_{n_j}(\omega) - X(\omega)| > \frac{1}{k}\Big) \leq \sum_{j=N}^{\infty}\mathbf{P}(\{|X_{n_j}(\omega) - X(\omega)| > \frac{1}{j}\})$$
$$< \sum_{j=N}^{\infty}\frac{1}{2^j} = \frac{1}{2^{N-1}} \to 0 \text{ as } N \to \infty \text{ (since } k \to \infty)$$

So $\mathbf{P}(\lim_{k\to\infty} B_k) = 0$, that is

$$\mathbf{P}(\{\lim_{k\to\infty} X_{n_k}(\omega) \neq X(\omega)\}) = 0$$

And this means $X_{n_k} \xrightarrow{as} X$. □

5.3 Converge in distribution

Definition 73 (converges in distribution/converges in law/weakly converges)**.** We say that X_n **converges in distribution** to X, denoted $X_n \xrightarrow{d} X$ if the cdf F_{X_n} and F_X satisfy:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all x such that F is continuous.

Remark 74. Later, we will also say this is **weakly converges**, which have the following definition: X_n **weakly converge** to X if and only if:

$$\text{For all bounded continuous function } f(x) : \mathbf{E}[f(X_n)] \to \mathbf{E}[f(X)].$$

We have the following property: convergence in probability imply convergence in distribution:

Theorem 75 (convergence in probability implies convergence in distribution)**.** *If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$.*

Proof. To prove this, we will need the following lemma:

Lemma 76. For any random variable X and Y, take any $a \in \mathbb{R}$ and $\varepsilon > 0$), then:

$$\mathbf{P}(Y \leq a) \leq \mathbf{P}(X \leq a + \varepsilon) + \mathbf{P}(|X - Y| > \varepsilon)$$

By lemma we get these:

$$\mathbf{P}(X_n \leq a) \leq \mathbf{P}(X \leq a + \varepsilon) + \mathbf{P}(|X_n - X| > \varepsilon) \tag{1}$$

and

$$\mathbf{P}(X_n \leq a - \varepsilon) \leq \mathbf{P}(X \leq a) + \mathbf{P}(|X_n - X| > \varepsilon) \tag{2}$$

From second equation we get:

$$\mathbf{P}(X_n \leq a - \varepsilon) - \mathbf{P}(|X_n - X| > \varepsilon) \leq \mathbf{P}(X \leq a) \tag{3}$$

Combining (1) and (3) gives:

$$\mathbf{P}(X_n \leq a - \varepsilon) - \mathbf{P}(|X_n - X| > \varepsilon) \leq \mathbf{P}(X \leq a) \leq \mathbf{P}(X \leq a + \varepsilon) + \mathbf{P}(|X_n - X| > \varepsilon) \tag{4}$$

Since $X_n \xrightarrow{P} X$, we can choose large N so that for all $n \geq N$:

$$\mathbf{P}(|X_n - X| > \epsilon) \to \delta$$

where we let $\delta \to 0$.
Then taking cdf definition in equation (4) gives:

$$F_{X_n}(a - \epsilon) - \delta \leq F_X(a) \leq F_{X_n}(a + \epsilon) + \delta$$

Finally, letting $\epsilon \to 0$ (and note that $\delta \to 0$ as well) gives:

$$\lim_{n \to \infty} F_{X_n}(a) = F_X(a)$$

And that concluded $X_n \xrightarrow{d} X$. □

Question 77. Does convergence in distribution implies convergence in probability? NO.

Example 78 (a counter example that shows convergence in distribution does NOT imply convergence in probability)**.** Consider the space $([0,1], \mathcal{B}([0,1]), \lambda)$. And consider the random variables:

$$X_{2n}(\omega) = \omega$$

and

$$X_{2n+1}(\omega) = 1 - \omega$$

Claim 1: $F_{X_{2n}} = F_{X_{2n+1}} = F_{\mathbf{1}_{[0,1]}}$. This is true because, for any $a \in \mathbb{R}$:

$$\begin{aligned} F_{X_{2n}}(a) &= \mathbf{P}(\{\omega \in [0,1] : X_{2n}(\omega) \in (-\infty, a]\}) \\ &= \mathbf{P}(\{\omega \in [0,1] : \omega \in (-\infty, a]\}) \\ &= \mathbf{P}(0 \leq \omega \leq a) = a \end{aligned}$$

On the other hand:

$$\begin{aligned} F_{X_{2n+1}}(a) &= \mathbf{P}(\{\omega \in [0,1] : X_{2n++1}(\omega) \in (-\infty, a]\}) \\ &= \mathbf{P}(\{\omega \in [0,1] : 1 - \omega \in (-\infty, a]\}) \\ &= \mathbf{P}(0 \leq 1 - \omega \leq a) \\ &= \mathbf{P}(1 - a \leq \omega \leq 1) \\ &= a \end{aligned}$$

However, we do **not** have convergence in probability:

$$\{|X_{2n} - X_{2n+1}| > \epsilon\} = \{|\omega - (1 - \omega)| > \epsilon\} = \{|2\omega - 1| > \epsilon\}$$

Take $\epsilon = \frac{1}{10}$ for example, then $|2\omega - 1|$, for $\omega \in [0,1]$, this means $2\omega < 0.9$ or $2\omega > 1.1$, which will not have probability measure zero !

5.4 Converges in mean

Definition 79 (convergence in Mean). X_n converges in mean to X if:

$$\lim_{n \to \infty} \mathbf{E}\Big[|X_n - X|\Big] = 0$$

We sometimes says this is convergence in L^1 norm and denote it as $X_n \xrightarrow{L^1} X$.

Remark 80. The general version of this is **Convergence in L^p norm**: f_n converges to f in L^p norm if:

$$\lim_{n \to \infty} \int |f_n - f|^p d\mu = 0$$

We also have convergence in L^1 implies convergence in probability as the theorem below:

Theorem 81 (convergence in mean implies convergence in probability). *If $X_n \xrightarrow{L^1} X$, then $X_n \xrightarrow{P} X$.*

Proof. Let $\delta > 0$ be given. By definition, for a given $\varepsilon > 0$:

$$\begin{aligned}
\mathbf{E}\Big[|X_n - X|\Big] &= \int |X_n - X| d\mathbf{P} \\
&= \int_{|X_n - X| > \varepsilon} |X_n - X| d\mathbf{P} + \int_{|X_n - X| \leq \varepsilon} |X_n - X| d\mathbf{P} \\
&\geq \int_{|X_n - X| \geq \varepsilon} \varepsilon d\mathbf{P} + \int_{|X_n - X| \leq \varepsilon} |X_n - X| d\mathbf{P}; \text{ since on domain } |X_n - X| \geq \varepsilon \\
&= \int_{|X_n - X| \geq \varepsilon} \varepsilon d\mathbf{P} + 0; \text{ since } \mathbf{E}\Big[|X_n - X|\Big] \to 0 \\
&= \varepsilon \mathbf{P}(\{|X_n - X| > \varepsilon\})
\end{aligned}$$

In short, we just showed:

$$\mathbf{P}(\{|X_n - X| > \varepsilon\}) \leq \frac{\mathbf{E}\Big[|X_n - X|\Big]}{\varepsilon}$$

If we take large enough N, the quantity $\mathbf{E}\Big[|X_n - X|\Big] < \varepsilon\delta$, then:

$$\mathbf{P}(\{|X_n - X| > \varepsilon\}) < \delta$$

And since δ is arbitrary, we conclude that $X_n \xrightarrow{P} X$. □

5.5 A revisit to Expectation- how to compute them

Recall the expectation of X on $(\Omega, \mathcal{F}, \mu)$ or on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ is defined as:

$$\mathbf{E}[X] = \int X(\omega) d\mu(\omega)$$

or

$$\mathbf{E}[X] = \int_{\mathbb{R}} X(\omega) d\mathbf{P}(\omega)$$

Definition 82 (Variance)**.** The **variance** of a random variable X is defined as:

$$\mathrm{Var}(X) = \int (X \quad \mathbf{E}[X])^2 \, d\mathbf{P} = \mathbf{E}[X^2] \quad \Big(\mathbf{E}[X]\Big)^2$$

Note that this quantity is non-negative by Jensen inequality ($\mathbf{E}[X^2] \quad \Big(\mathbf{E}[X]\Big)^2 \quad 0$.

Definition 83 (standard deviation)**.** The **standard deviation** of a random variable X is defined as:

$$(X) = \sqrt{\mathrm{Var}(X)}$$

Definition 84 (p-th moment (L^p norm))**.** The p**-th moment** if a random variable X is defined as:

$$\mathbf{E}[X^p] = \int X^p d\mathbf{P}$$

To compute expectations:

1. Continuous random variable with pdf $f(x)$:

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

Example 85. Exponential random variable with $f(x) = \quad e^{\quad x}$ for $x \in [0, \infty)$, then:

$$\mathbf{E}[X] = \frac{1}{\ }$$

and

$$\mathrm{Var}(X) = (\frac{1}{\ })^2$$

To see this we need to do an integral, here $g(x) = x$, so

$$\begin{aligned}
\mathbf{E}[X] &= \int_{\mathbb{R}} x \quad e^{\quad x} dx = \int_0^{\infty} x \quad e^{\quad x} dx \\
&= \int_0^{\infty} x e^{\quad x} dx \\
&= \left[\frac{x e^{\quad x}}{\ } \Big|_0^{\infty} \quad \int_0^{\infty} \frac{e^{\quad x}}{\ } dx \right] \\
&= \quad x e^{\quad x} \Big|_0^{\infty} + \int_0^{\infty} \frac{e^{\quad x}}{\ } dx \\
&= \quad x e^{\quad x} \Big|_0^{\infty} + \frac{e^{\quad x}}{\ } \Big|_0^{\infty} \\
&= \quad \infty e^{\quad \infty} \quad 0 + \frac{e^{\quad \infty}}{\ } \quad \frac{e^0}{\ } \\
&= \frac{1}{\ }
\end{aligned}$$

For the variance, note that $\mathbf{E}[X^2] = \quad \int_0^{\infty} (x^2 e^{\quad x}) =$ usual integration by parts $= \frac{1}{2} + (E[X])^2$.

2. Discrete random variable with pmf:

$$\mathbf{E}[g(X)] = \sum_x g(x)\mathbf{P}(X = x)$$

Example 86. Bernoulli random variable with $\mathbf{P}(X = 1) = p$ and $\mathbf{P}(X = 0) = 1 - p$:

$$\mathbf{E}[X] = p$$

and

$$\mathrm{Var}(X) = p(1 - p)$$

5.6 More useful Inequality

In addition to Jensen, Holder, and Cauchy Schwartz Inequality, we have an important inequality called Chebychev Inequality:

Theorem 87 (Chebychev Inequality)**.** *For random variable X with finite mean $\mathbf{E}[X] = m < \infty$ and finite non-zero $Var(X)$ (that is, the standard deviation $0 \neq \sigma(X) = \sqrt{Var(X)} < \infty$). Then, for any $k > 0$:*

$$\mathbf{P}(|X - m| \geq k\sigma) \leq \frac{1}{k^2}$$

Proof. Denote $Y = (X - m)^2$. We will show that

$$\mathbf{P}(Y \geq k^2\sigma^2) \leq \frac{1}{k^2}$$

We consider the following:

$$\begin{aligned}
\mathbf{E}[Y] &= \int Y\,d\mathbf{P} \\
&= \int_{\{Y \geq k^2\sigma^2\}} Y\,d\mathbf{P} + \int_{\{Y < k^2\sigma^2\}} Y\,d\mathbf{P} \\
&\geq \int_{\{Y \geq k^2\sigma^2\}} Y\,d\mathbf{P} \geq \int_{\{Y \geq k^2\sigma^2\}} k^2\sigma^2 d\mathbf{P} = k^2\sigma^2\mathbf{P}(\{Y \geq k^2\sigma^2\})
\end{aligned}$$

Also note that

$$\mathbf{E}[Y] = \mathbf{E}[(X - m)^2] = \sigma^2 \; (= \mathrm{Var}(X))$$

In short, we showed:

$$\sigma^2 \geq k^2\sigma^2\mathbf{P}(\{Y \geq k^2\sigma^2\})$$

Then divide by $k^2\sigma^2$ gives:

$$\mathbf{P}(\{Y \geq k^2\sigma^2\}) \leq \frac{1}{k^2}$$

□

Example 88. When $k = 2$ we get (see figure (15):

$$\mathbf{P}(|X - m| \geq 2\sigma) \leq 2 \quad < \frac{1}{4} = 25\%$$

Theorem 89. *For all $p > 0$, and for all $a > 0$, and any measurable function f, we get:*

$$\mu(\{\omega : |f(\omega)| \geq a\}) \leq \frac{\int |f|^p d\mu}{a^p}$$

Proof. HW □

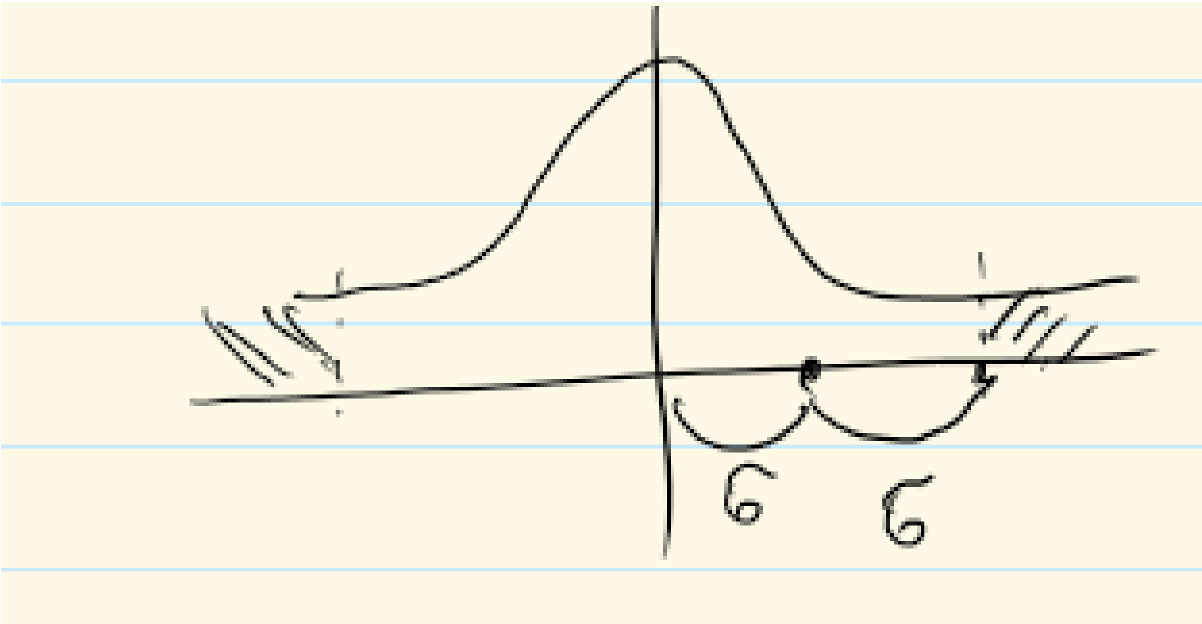


Figure 15: $k = 2$ in Chebychev inequality (the shaded is 25%)

6 Day 6: product measures, and Independence

6.1 Product measures

Consider two spaces (Ω_1, F_1, μ_1) and (Ω_2, F_2, μ_2). How do we construct a probability space:

1. $\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$.
2. $S = \{A_1 \times A_2 : A_1 \in F_1, A_2 \in F_2\}$
3. $F = \sigma(S)$

How can we define the measure $\mu = \mu_1 \times \mu_2$? We define it on rectangle set $A_1 \times A_2$ by (see figure 16):

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

Such existence and uniqueness of the product measure is proved in Theorem 1.1.9 and the Appendix in Durett's book.

Theorem 90 (product measure definition -Theorem 1.7.1)**.** *On the product measure space (Ω, F), there exists a unique measure μ such that*

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

Similarly, we can extend this to multi-dimensional space:

1. $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$
2. $F = \sigma(A_1 \times A_2 \times \dots \times A_n)$
3. $\mu = \mu_1(A_1) \dots \mu_n(A_n)$

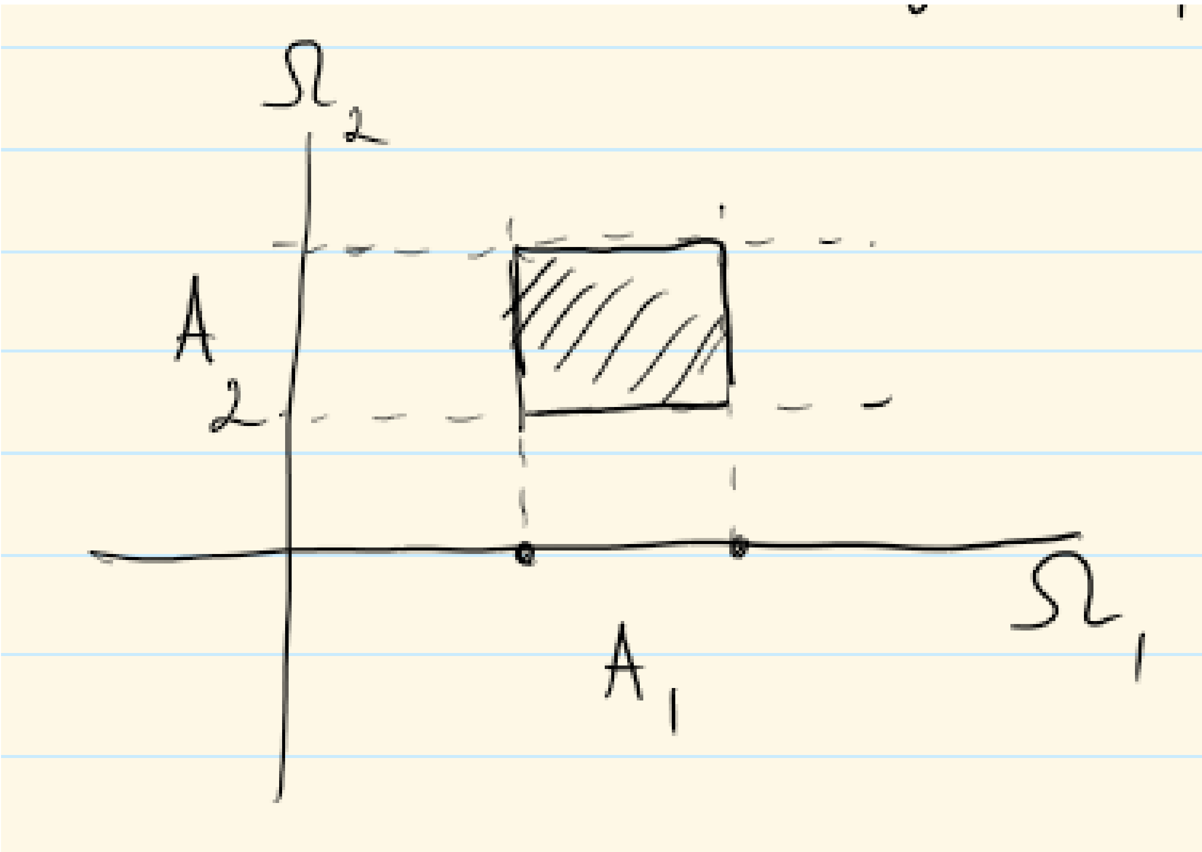


Figure 16: product measure $\mu = \mu_1 \times \mu_2$

Next, the integration on product space is the tool in computing expectation on product space. The theorems that can be used in product space integration are **Tonelli Theorem** and **Fubini Theorem**.

Theorem 91 (Tonelli Theorem)**.** *Let $X(\omega_1, \omega_2) : \Omega_1 \times \Omega_2 \to [0, \infty)$ be* ***non-negative.*** *Then the following are true:*

1. *The function $\omega_1 \mapsto \int X(\omega_1, \cdot) d\mu_2$ is $\mathcal{F}_1$-measurable function, i,e., a random variable on Ω_1.*

2. *The function $\omega_2 \mapsto \int X(\cdot, \omega_2) d\mu_1$ is $\mathcal{F}_2$-measurable function, i,e., a random variable on Ω_2.*

3. ***Most importantly, you can interchange the order of integration:***

$$\int X(\omega_1, \omega_2) d\mu = \int_{\Omega_1} \int_{\Omega_2} X d\mu_1 d\mu_2 = \int_{\Omega_2} \int_{\Omega_1} X d\mu_2 d\mu_1$$

Theorem 92 (Fubini Theorem)**.** *Let $\mathbf{E}[X] = \int |X| d\mu < \infty$. Then the Tonelli Theorem holds, that is:*

1. *The function $\omega_1 \mapsto \int X(\omega_1, \cdot) d\mu_2$ is $\mathcal{F}_1$-measurable function, i,e., a random variable on Ω_1.*

2. *The function $\omega_2 \mapsto \int X(\cdot, \omega_2) d\mu_1$ is $\mathcal{F}_2$-measurable function, i,e., a random variable on Ω_2.*

3. ***Most importantly, you can interchange the order of integration:***

$$\int X(\omega_1;\omega_2)d \;= \int_{1}\int_{2} X d\ {}_1 d\ {}_2 = \int_{2}\int_{1} X d\ {}_2 d\ {}_1$$

6.2 Independence

We begin with the definitions of independence of 2 events, 2 random variables, and 2 sigma algebras:

Definition 93 (2 events independence)**.** 2 events A and B are independence if

$$\mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(A \cup B)$$

Definition 94 (2 random variables independence)**.** 2 random variables X and Y is independence if: for any $E;F \in B(\mathrm{R})$:

$$\mathbf{P}(X \in E)\mathbf{P}(Y \in F) = \mathbf{P}(X \in E \text{ and } Y \in F)$$

Definition 95 (2 sigma algebras independence)**.** We say the 2 sigma algebras F_1 and F_2 (which are both $\;F$) are independence if: for all $A \in F_1$ and $B \in F_2$, the A and B are independent (events) with regards to definition (93).

Example 96. twice dice rolling. Consider A the event of getting a 1 in the first round. Then $A = \{1;\ \}$ where $\ = 1;2;3;4;5;6$ possible choices. There are 6 possible elements in A out of total Σ which has 36 possible elements. Thus $\mathbf{P}(A) = \frac{1}{6}$.
Consider B the event of getting a 1 in the second round. Similarly, $B = \{\ ;1\}$ where $\ = 1;2;3;4;5;6$ possible choices. There are 6 possible elements in B out of total Σ which has 36 possible elements. Thus $\mathbf{P}(B) = \frac{1}{6}$.
In addition, the probability of getting 1 in the first round and 1 in the second round is: $\mathbf{P}(A \cap B) = \mathbf{P}(\{1;1\}) = \frac{1}{36}$. So $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. Hence A and B are independent.

Example 97 (an example of **not** independence)**.** Toss a dice twice. Let A be the event which the number of is 8. For instance, $A = \{(2;6);(3;5);(4;4);(5;3);(6;2)\}$. So A has 5 elements. So $\mathbf{P}(A) = \frac{5}{36}$.
Let B be the event of getting a 2 in the first round. As in earlier example, $\mathbf{P}(B) = \frac{1}{6}$. Now $\mathbf{P}(A \cap B) = \mathbf{P}(\{(2;6)\}) = \frac{1}{6}$. It is clear then that $\mathbf{P}(A)\mathbf{P}(B) \neq \mathbf{P}(A \cap B)$, so A and B is not independence.

Question 98. How will we extend this to more than 2 events, random variables, sigma algebras? Will pairwise independence be good enough? **NO**.

Example 99 (Example showing pairwise independence is not good enough)**.** Let $X_1;X_2;X_3$ be random variables with $\mathbf{P}(X_i = 0) = \frac{1}{2} = \mathbf{P}(X_i = 1)$. Consider the following events:

$$A_1 = \{X_2 = X_3\}$$

$$A_2 = \{X_1 = X_3\}$$

$$A_3 = \{X_1 = X_2\}$$

For example, $A_1 = \{(\ ;0;1);(\ ;0;0);(\ ;1;0);(\ ;1;1)\}$, where $\ = 0;1$.
They are pairwise independent because:

$$\mathbf{P}(A_i)\mathbf{P}(A_j) = \frac{1}{2}\frac{1}{2} = \mathbf{P}(A_i \cap A_j) = \mathbf{P}(X_1 = X_2 = X_3) = \frac{1}{4}$$

However, they are not three event independent because:

$$\mathbf{P}(A_1)\mathbf{P}(A_2)\mathbf{P}(A_3) = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$$

while

$$\mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(X_1 = X_2 = X_3) = \frac{1}{4}$$

Definition 100 (Independence of Multiple events)**.** A collection of events $A_1, \ldots, A_n$ are independent if:
For any $I \subseteq \{1, 2, \ldots, n\}$:

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

Definition 101 (Independence of multiple random variables)**.** A collection of random variables $X_1, \ldots, X_n$ are independent if:
For all $B_1, B_2, \ldots, B_n \in \mathcal{B}(\mathbb{R})$ and for any $I \subseteq \{1, 2, \ldots, n\}$:

$$\mathbf{P}\left(\bigcap_{i \in I} \{X_i \in B_i\}\right) = \prod_{i \in I} \mathbf{P}(\{X_i \in B_i\})$$

Definition 102 (independence of multiple sigma algebras)**.** The collection of sigma algebras $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$ are independent if: For all $A_i \in \mathcal{F}_i$, and for any $I \subseteq \{1, 2, \ldots, n\}$, the $\{A_i\}_{i \in I}$ are independent events.

Question 103. To check if random variables are independence, do we need to check for all the sets in $\mathcal{B}(\mathbb{R}^n)$? Can we only check for the generating set of the sigma algebra? The answer is NO. The generating set is not enough.

Example 104. An example showing that generating set is not good enough. Consider $\Omega = \{1, 2, 3, 4\}$ and $\mathbf{P}(\{i\}) = \frac{1}{4}$ for $i = 1, 2, 3, 4$. Let

$$A = \{1, 2\}$$

$$B = \{1, 3\}$$

$$C = \{1, 4\}$$

Then we can see that $\mathcal{F} = 2^{\Omega} = \sigma(A, B, C)$. That is, A, B, C generated the sigma algebra 2^{Ω}. A, B, C are also independent because:

$$\mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(A \cap B) = \mathbf{P}(\{1\}) = \frac{1}{4}$$

and

$$\mathbf{P}(A)\mathbf{P}(C) = \mathbf{P}(A \cap C) = \mathbf{P}(\{1\}) = \frac{1}{4}$$

and

$$\mathbf{P}(B)\mathbf{P}(C) = \mathbf{P}(B \cap C) = \mathbf{P}(\{1\}) = \frac{1}{4}$$

However, A and $D = B \cap C = \{1\}$ is not independent, because:

$$\mathbf{P}(A) = \frac{1}{2}$$

$$\mathbf{P}(D) = \frac{1}{4}$$

So

$$\mathbf{P}(A)\mathbf{P}(D) = \frac{1}{8}$$

while

$$\mathbf{P}(A \cap D) = \mathbf{P}(A \cap B \cap C) = \mathbf{P}(\{1\}) = \frac{1}{4}$$

Question 105. What conditions can be added to the **generating sets** so that we only need to check on the generating sets for independence? The π-system condition.

Definition 106 (π-system)**.** A set S is a π-system if: For any $A, B \in S$, then $A \cap B \in S$.

Example 107 (An example of a π-system)**.** On the real line $\mathbb{R}$, the set $S = \{(-\infty, a)\}$ is a π-system, since, given $a_1 < a_2 \in \mathbb{R}$, if $(-\infty, a_n) \in S$, then

$$(-\infty, a_1] \cap (-\infty, a_2] = (-\infty, a_2] \in S$$

The following theorem is about the condition on the generating sets:

Theorem 108 (sigma algebra generated from π system sets)**.** *Suppose $A_1, \ldots, A_n$ are independent sets of events such as each A_i is a π-system. Then: $\sigma(A_1), \sigma(A_2), \ldots, \sigma(A_n)$ are independent.*

Proof. (partially) The proof of this theorem relies on another theorem called the π-λ system theorem, see theorem 112 below (and of course, see also the definition 110 of a lambda system) for reference. To show this, Without loss of generality, we would need to show that $(\sigma(A_1), A_2, \ldots, A_n)$ are independent.
To show that, we need, for any $C \in \sigma(A_1)$, we wish to show that:

$$\begin{aligned}
\mathbf{P}(C \cap A_2 \cap \ldots \cap A_n) &= \mathbf{P}(C)\mathbf{P}(A_2)\ldots\mathbf{P}(A_n) \\
&= \mathbf{P}(C)\mathbf{P}(A_2 \cap \ldots \cap A_n)
\end{aligned}$$

Consider L to be the set of B such that if $B \in L$, then $\mathbf{P}(B \cap L) = \mathbf{P}(B)\mathbf{P}(F)$ where $F = A_2 \cap A_2 \cap \ldots \cap A_n$. We know that $A_1 \subset L$, we wish to show that $\sigma(A_1) \subset L$ as well. To achieve this goal, we will use the π-λ theorem and need to show that L is a λ-system. Check the 3 conditions of the λ-system for L:

1. Is $\Omega \in L$? Yes, because

$$\mathbf{P}(\Omega \cap F) = \mathbf{P}(F) = 1\mathbf{P}(F) = \mathbf{P}(\Omega)\mathbf{P}(F)$$

2. Let $A, B \in L$ and that $A \subset B$, is $B - A \in L$? Yes:

$$\begin{aligned}
\mathbf{P}((B - A) \cap F) &= \mathbf{P}((B \cap F) - (A \cap F)) \\
&= \mathbf{P}(b \cap F) - \mathbf{P}(A \cap F) \\
&= \mathbf{P}(B)\mathbf{P}(F) - \mathbf{P}(A_{\mathbf{P}}(F) \\
&= [\mathbf{P}(B) - \mathbf{P}(A)]\,\mathbf{P}(F) \\
&= \mathbf{P}(B - A)\mathbf{P}(F)
\end{aligned}$$

3. Let $A_n \in \mathcal{L}$ such that $A_n \subseteq A_{n+1}$. Is $\lim_{n\to\infty} A_n \in \mathcal{L}$? Yes, by the monotone convergence theorem for probability measure with $B_n = A_n \cap F$:

$$\begin{aligned}\mathbf{P}(\lim_{n\to\infty} B_n) &= \mathbf{P}(\lim_{n\to\infty}(A_n \cap F)) \\ &= \mathbf{P}\left([\lim_{n\to\infty} A_n] \cap F\right)\end{aligned}$$

while by the monotone convergence theorem,

$$\mathbf{P}(\lim_{n\to\infty} B_n) = \lim_{n\to\infty} \mathbf{P}(B_n) = \lim_{n\to\infty} \mathbf{P}(A_n \cap F) \in \mathcal{L}$$

□

The Example (107) and the theorem (108) gives an interesting (and useful) condition to check two random variables are independent:

Corollary 109 (useful way to check independence of random variables)**.** *X and Y are random variables. Then, X and Y are independent if and only if:*

$$\boldsymbol{P}(X \le s, Y \le t) = \boldsymbol{P}(X \le s)\boldsymbol{P}(Y \le t) \text{ , for all } s, t \in \mathbb{R}$$

In the language of cdf function F, the above is equivalent to

$$F_{XY}(s, t) = F_X(s)F_Y(t), \ \forall s, t \in \mathbb{R}$$

And furthermore, if the density pdf function f exists, we have the equivalent:

$$f_{XY}(s, t) = f_X(s)f_Y(t)$$

Keep in mind that we need to do the partial derivative $\frac{\partial}{\partial s}\frac{\partial}{\partial t}$ to get F_{XY}.

Definition 110 (λ-system)**.** A set $\mathcal{L}$ is a λ system if:

1. $\Omega \in \mathcal{L}$
2. If $A, B \in \mathcal{L}$, then $B - A \in \mathcal{L}$
3. Let $A_n \in \mathcal{L}$ such that $A_n \subseteq A_{n+1}$, then $\lim_{n\to\infty} A_n \in \mathcal{L}$

Remark 111. If $\mathcal{F}$ is a σ-algebra, then $\mathcal{F}$ is both a π system and a λ-system.

Theorem 112 (π-λ system theorem)**.** *If $\mathcal{S}$ is a π-system and $\mathcal{S} \subseteq \mathcal{L}$ where $\mathcal{L}$ is a λ-system, then $\sigma(\mathcal{S}) \subseteq \mathcal{L}$.*

7 Day 7: More on dependence and begin laws of large number

7.1 More properties of Independence

Theorem 113 (Theorem 2.1.3 in Durett book - Independence of complement of 2 events)**.** *If A and B are independent events, then so are A and B^C, A^C and B, and A^C and B^C.*

Proof.

$$\begin{aligned}\mathbf{P}(A) &= \mathbf{P}\left[(A\cap B)\cup(A\cap B^C)\right]\\ &= \mathbf{P}(A\cap B)+\mathbf{P}(A\cap B^C)\\ &= \mathbf{P}(A)\mathbf{P}(B)+\mathbf{P}(A\cap B^C)\end{aligned}$$

So

$$\mathbf{P}(A\cap B^C)=\mathbf{P}(A)-\mathbf{P}(A)\mathbf{P}(B)=\mathbf{P}(A)\left[1-\mathbf{P}(B)\right]=\mathbf{P}(A)\mathbf{P}(B^C)$$

Hence A and B^C are independent. □

Theorem 114 (Theorem 2.1.3 Durett book - independence of complement of multiple events)**.** *If $A_1, A_2, \dots, A_n$ are independent, then $A_1^C, \dots, A_n$ are also independent. Here note that A_i are events (not set of events).*

Proof. Consider $I \subseteq \{1, 2, 3, \dots, n\}$.

1. If $\{1\} \notin I$: then

$$\mathbf{P}\Big(\bigcap_{i\in I} A_i\Big)=\prod_{i\in I}\mathbf{P}(A_i);\ \text{since the } A_i \text{ are independent}$$

2. If $\{1\} \in I$:
   Define $B := \bigcap_{i\in I-\{1\}} A_i$. Note that $B=\{B\cap A_1^c\}\cup\{B\cap A_1\}$. Therefore:

$$\begin{aligned}\mathbf{P}(B) &= \mathbf{P}(B\cap A_1^C)+\mathbf{P}(B\cap A_1)\\ &= \mathbf{P}(B\cap A_1^C)+\mathbf{P}(B)\mathbf{P}(A_1^C)\end{aligned}$$

   Therefore

$$\mathbf{P}(B\cap A_1^C)=\mathbf{P}(B)\left[1-\mathbf{P}(A_1)\right]=\mathbf{P}(B)\mathbf{P}(A_1^C)$$

   So from the definition of B, we get:

$$\mathbf{P}\Big(\big(\bigcap_{i\in I-\{1\}} A_i\big)\cap A_1^C\Big)=\mathbf{P}\Big(\bigcap_{i\in I-\{1\}} A_i\Big)\left[1-\mathbf{P}(A_1)\right]=\prod_{i\in I-\{1\}}\mathbf{P}(A_i)\mathbf{P}(A_1^C)$$

By the arbitrariness of I, we conclude that $A_1^C, A_2, \dots, A_n$ are independent. □

Theorem 115 (Theorem 2.1.11 Durett book - probablity measure on independence random vector)**.** *Suppose $X_1, \dots, X_n$ are independent random variables and have probability measures $\mu_i(B_i) := \mathbf{P}(X_i \in B_i)$. Then the $(X_1, \dots, X_n)$ is a random vector and have probability measure $\mu_1 \times \mu_2 \times \cdots \times \mu_n$.*

Proof. Consider the tublet $S = \{A_1 \times A_2 \times \cdots A_n\}$ where each $A_i \in \mathcal{B}(\mathbb{R})$. Denote $\mathcal{F} = \sigma(S)$ the sigma algebra of $X_1 \times \cdots \times X_n$.Then

$$\begin{aligned}\mathbf{P}((X_1, X_2, \dots, X_n) \in A_1 \times \cdots \times A_n) &= \mathbf{P}(X_1 \in A_1, \dots, X_n \in A_n)\\ &= \mathbf{P}(X_1 \in A_1)\cdots\mathbf{P}(X_n \in A_n), \text{ since the } X_i \text{ are independent}\\ &= \mu_1(A_1)\cdots\mu_n(A_n) =\end{aligned}$$

The last equality is due to theorem 1.7.1 where there exists a unique measure in $\Omega_1 \times \Omega_n$ such that that happens.
Now since $A_1 \times \cdots \times A_n$ is a π-system, this will hold for all $\mathcal{F}$ (By the $\pi-\lambda$-system theorem). □

Theorem 116 (theorem 2.1.12 - product measures of independent random variables)**.** *Suppose X and Y are independent random variables, with probability measures μ and ν. If $h : \mathbb{R}^2 \to \mathbb{R}$ is a measurable function such that either $h \geq 0$ or $\mathbf{E}[|f(X;Y)|] < \infty$, then:*

$$\mathbf{E}[h(X;Y)] = \int \int h(x;y)\ \mu(dx)\ \nu(dy)$$

Proof. Tonelli or Fubini Theorem application. □

Theorem 117 (Theorem 1.6.9 - a useful theorem of product measure)**.** *Suppose X is a random variable on $(\Omega; F)$ with measure μ_X. If $f : \Omega \to \mathbb{R}$ is measurable, with $f \geq 0$ or $\mathbf{E}[|f(x)|] < \infty$, then $E[f(X)] = \int f(x)\ \mu_X(dx)$.*
In addition, let $h(x;y) = f(x)g(y)$ with either $f;g \geq 0$ OR $\mathbf{E}|f(x)| < \infty; \mathbf{E}|g(y)| < \infty$, then

$$\mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)]\mathbf{E}[g(Y)]$$

Proof. Tonelli or Fubini application. □

Definition 118 (covariance and correlation of 2 random variables)**.** Let $X;Y$ be 2 random variables. The **covariance** of 2 random variables X and Y are defined as

$$\mathrm{Covar}(X;Y) := \mathbf{E}\Big[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])\Big]$$

But it is also equivalent to:

$$\mathrm{Covar}(X;Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$$

And the correlation between X and Y is defined as:

$$\mathrm{Cor}(X;Y) := \frac{\mathrm{Covar}(X;Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

Some properties: $\mathrm{Covar}(aX;Y) = a\mathrm{Covar}(X;Y)$ and $\mathrm{Covar}(Y;X) = \mathrm{Covar}(X;Y)$

Definition 119. We called X and Y **uncorrelated** if $\mathrm{Covar}(X;Y) = 0$, that is, if $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.

Proposition 120. *If $X;Y$ are independent, then $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.*

Proof.

$$\mathbf{P}(X \in E; Y \in F) = \mathbf{P}(X \in E)\mathbf{P}(Y \in F)$$

Then carry it over to cdf and then it get to expectation definition:

$$\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$$

□

Remark 121. The reverse is NOT true. X and Y are uncorrelated, but they are not independent.

Example 122. This is an example of X and Y are uncorrelated, but they are not independent. Consider:

$$\mathbf{P}((X,Y) = (-1,1)) = \frac{1}{4}$$
$$\mathbf{P}((X,Y) = (0,0)) = \frac{1}{2}$$
$$\mathbf{P}((X,Y) = (1,1)) = \frac{1}{4}$$

Then:

$$\mathbf{E}[XY] = -1(-1)\frac{1}{4} + 0 \cdot 0 \cdot \frac{1}{2} + 1 \cdot 1\frac{1}{4} = 0$$

And

$$\mathbf{E}[X] = -1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 0$$
$$\mathbf{E}[Y] = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}$$

So

$$\mathbf{E}[XY] = 0 = \mathbf{E}[X]\mathbf{E}[Y]$$

However:

$$\mathbf{P}(X = -1, Y = 1) = \frac{1}{4}$$

while

$$\mathbf{P}(X = -1)\mathbf{P}(X = 1) = \frac{1}{4}\frac{1}{2} = \frac{1}{8}$$

So $\mathbf{P}(X = -1, Y = 1) \neq \mathbf{P}(X = -1)\mathbf{P}(Y = 1)$, and therefore X and Y are not independent.

7.1.1 How to compute variance of a sum

Definition 123 (variance of sum)**.** For $S_n = X_1 + X_2 + \dots + X_n$, we compute:

$$\begin{aligned}
\mathrm{Var}(S_n) &= \mathbf{E}\left[(X_1 + \dots X_n)^2\right] - \left(\mathbf{E}[X_1 + \dots X_n]\right)^2 \\
&= \mathbf{E}\left[\sum_{i=1}^{n} X_i^2 + \sum_{i<j} 2X_iX_j\right] - \left(\sum_{i=1}^{n} \mathbf{E}[X_i]^2 + \sum_{i<j} 2\mathbf{E}[X_i]\mathbf{E}[X_j]\right) \\
&= \sum_{i=1}^{n}\left(\mathbf{E}[X_i^2] - \mathbf{E}[X_i]^2\right) + 2\sum_{i<j}\left(\mathbf{E}[X_iX_j] - \mathbf{E}[X_i]\mathbf{E}[X_j]\right) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Covar}(X_i, X_j)
\end{aligned}$$

Since X_i are independent, then $\mathrm{Covar}(X_i, X_j) = 0$ whenever $i \neq j$. So

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

7.2 Laws of large numbers (LLN) - intro

7.2.1 Introduction

Repeat a (same) experiment X_i with $\mathbf{E}[X_i] = \mu$ and $\mathbf{E}[|X_i|] < \infty$. Denote:

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

Then

$$\frac{S_n}{n} \to \mu = \mathbf{E}[X_i]$$

But what type of convergence the above is true and what condition is required for it?

1. **The Weak LLN**

$$\frac{S_n}{n} \xrightarrow{p} \mu$$

   meaning: For all $\varepsilon > 0$, then

$$\mathbf{P}(|\frac{S_n}{n} - \mu| > \varepsilon) \to 0$$

2. **The Strong LLN**

$$\frac{S_n}{n} \xrightarrow{as} \mu$$

   meaning:

$$\mathbf{P}(\frac{S_n}{n} = \mu) = 1$$

Example 124 (An example of LLN)**.** Repeating flipping a coin, the i-th coin is denoted as X_i, that is $X_i = \{1_{\omega=H}\}$. That means we have:

$$\mathbf{P}(X_i = 1) = \frac{1}{2} = \mathbf{P}(X_i = 0), \forall i.$$

Also

$$\mathbf{E}[X_i] = 1\frac{1}{2} + 0\frac{1}{2} = \frac{1}{2}$$

With $S_n = \sum_{i=1}^{n} X_i$, we have $\frac{S_n}{n} \to \frac{1}{2}$.

Definition 125 (**identically distributed**)**.** The X_n are identically distributed if their distribution are the same, that is, for any $a > 0$, $\mathbf{P}(|X_n| \geq a)$ is the same for all n. (Or equivalently, we can also have $\mathbf{P}(|X_n| \leq a)$ is the same for all n.)

7.2.2 Weak Law of Large Numbers -L2 version:

Theorem 126 (Weak LLN with L^2 norm)**.** *Suppose X_i are independent (Note that **uncorrelated** condition is good enough here), and **identically distributed i.i.d**, and that $\mathbf{E}[|X|^2] \leq \infty$. (Note that this condition also gives $\mu = \mathbf{E}[X_i] < \infty$, for all i, and that $\sigma^2 = Var(X_i) < \infty$).[we could also replace this condition by the condition that $Var(X_i) \leq C < \infty$, bounded by some C.] Then:*

$$\frac{S_n}{n} \xrightarrow{p} \mu$$

Where $S_n = X_1 + ::: + X_n$. Note that in some case, we denote

$$\overline{X_n} = \frac{S_n}{n}$$

and we called $\overline{X_n}$ the sample mean.

Proof. [**The proof idea is to use Chebychev inequality for $X = \frac{S_n}{n}$ and its corresponding mean and standard deviation.**] We know the following:

$$\mathbf{E}\left[\frac{S_n}{n}\right] = \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}n \quad =$$

This is because all the X_i has the same expectation value $E[X_i] =$.
In addition, we also know:

$$\mathrm{Var}(\frac{S_n}{n}) = \left(\frac{1}{n}\right)^2 \mathrm{Var}(S_n) = \frac{1}{n^2}\left(n\mathrm{Var}(X_i)\right) = \frac{1}{n} \; ^2 = \frac{^2}{n}$$

The first equality is from the definition of $\mathrm{Var}(aZ) = a^2\mathrm{Var}(Z)$, while the second equality is because the X_i are independent so $\mathrm{Var}(S_n) = \mathrm{Var}(X_1 + ::: + X_n) = n\mathrm{Var}(X_i)$.
Now let $" > 0$ be given. In addition, let $\; > 0$ be given. We wish to find an N such that for all $n \quad N$, the

$$\mathbf{P}(j\frac{S_n}{n} \qquad j > ") <$$

Now

$$\mathbf{P}(j\frac{S_n}{n} \qquad j > ") \quad \mathbf{P}(j\frac{S_n}{n} \qquad j \quad ") \text{ ,replace } < \text{ by}$$

Recall Chebychev Inequality 87:

$$\mathbf{P}(jX \quad mj \quad k \;) < \frac{1}{k^2}$$

This inequality works for any random variable X with finite mean m. We apply this inequality to the random variable $X = \frac{S_n}{n}$ and the mean $m = \mathbf{E}[X] = \mathbf{E}[\frac{S_n}{n}] = \quad < 1$. Moreover, note that the in Chebychev inequality is the standard deviation of the variable X. So in above, we calculated $\mathrm{Var}(\frac{S_n}{n})$, we should get the standard deviation of $\frac{S_n}{n}$ as $SD(\frac{S_n}{n}) = \sqrt{\mathrm{Var}(\frac{S_n}{n})} < 1$. In addition, it also works for any k, so we set our k as below:

$$k = \frac{"}{SD(\frac{S_n}{n})}$$

Thus

$$\frac{1}{k^2} = \frac{SD^2(\frac{S_n}{n})}{"^2} = \frac{\mathrm{Var}(\frac{S_n}{n})}{"^2} = \frac{\frac{^2}{n}}{"^2} = \frac{^2}{n^2 "^2}$$

Note that with this k, we also have $k{:}SD(\frac{S_n}{n}) = "$. Putting everything back into the Chebychev inequality gives:

$$\mathbf{P}(j\frac{S_n}{n} \qquad j \quad ") < \frac{^2}{n^2 "^2}$$

We can choose large N such that $N \quad \frac{^2}{"^2}$, so then for all $n \quad N$, the quantity $\frac{^2}{n^2 "^2} < \quad ! \; 0$ as $\; ! \; 0$. And this completed the proof. □

Example 127 (An application of Weak LLN - **Weierstrass Approximation Theorem-** Polynomials approximated continuous function)**.** Let f be a continuous function on $[0,1]$. Then the polynomials:

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

will satisfy

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

Proof. Let $S_n = \sum_{i=1}^{n} X_i$ where $\mathbf{P}(X_i = 1) = p$, and $\mathbf{P}(X_i = 0) = 1 - p$ for some $p \in [0,1]$. Then we calculate:

$$\mathbf{E}[X_i] = p$$

$$\mathrm{Var}(X_i) = p(1-p) \leq \frac{1}{4} < \infty$$

$$\mathbf{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Then in the probability language,

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

Then we need to show that

$$f_n(p) \to f(p)$$

Recall that since $f(p)$ is constant, so $\mathbf{E}[f(p)] = f(p)$. The above convergence is now:

$$\mathbf{E}\left[f(\frac{S_n}{n})\right] \to \mathbf{E}[f(p)]?$$

Indeed, we have:

$$\left|\mathbf{E}\left[f(\frac{S_n}{n})\right] - \mathbf{E}[f(p)]\right| = \left|\mathbf{E}\left[f(\frac{S_n}{n}) - f(p)\right]\right|$$
$$\leq E\left[\left|f(\frac{S_n}{n}) - f(p)\right|\right] \text{ ,by Jensen Inequality, bring abs value inside}$$

Since f is continuous in $[0,1]$, so f is uniform continuous on $[0,1]$. That is, for any $\varepsilon > 0$, there is a unique $\delta > 0$ such that if $|x - y| < \delta$; then $|f(x) - f(y)| < \frac{\varepsilon}{2}$.
In addition, f is also bounded in $[0,1]$, so $\sup_{x \in [0,1]} |f(x)| \leq K$.
Now then returning to the above Expectation, we break the domain into two sets $\{|\frac{S_n}{n} - p| \leq \delta\}$ and $\{|\frac{S_n}{n} - p| > \delta\}$:

$$E\left[\left|f(\frac{S_n}{n}) - f(p)\right|\right] = E\left[\left|f(\frac{S_n}{n}) - f(p)\right| \mathbf{1}_{\{|\frac{S_n}{n} - p| \leq \delta\}}\right] + E\left[\left|f(\frac{S_n}{n}) - f(p)\right| \mathbf{1}_{\{|\frac{S_n}{n} - p| > \delta\}}\right]$$
$$\leq \frac{\varepsilon}{2} + 2K\mathbf{P}(\{|\frac{S_n}{n} - p| > \delta\})$$

where we used the uniform condition in the first:

$$E\left[\left|f(\frac{S_n}{n}) - f(p)\right| \mathbf{1}_{\{|\frac{S_n}{n} - p| \leq \delta\}}\right] \leq \mathbf{E}[\frac{\varepsilon}{2}] = \frac{\varepsilon}{2}$$

and the bounded K in the second:

$$E\left[\left|f\left(\frac{S_n}{n}\right) - f(p)\right| \mathbf{1}_{\{|\frac{S_n}{n} - p| > \delta\}}\right] \leq \mathbf{E}[2K\mathbf{1}_{\{|\frac{S_n}{n} - p| > \delta\}}] = 2K\mathbf{P}(\{|\frac{S_n}{n} - p| > \delta\}$$

Now by the proof technique in Weak LLN, we can find N_0 such that $\forall n \geq N_0$,

$$\mathbf{P}(\{|\frac{S_n}{n} - p| > \delta\} < \frac{\varepsilon}{4K}$$

and therefore

$$2K\mathbf{P}(\{|\frac{S_n}{n} - p| > \delta\}) < \frac{\varepsilon}{2}$$

Hence we get the final inequality:

$$E\left[\left|f\left(\frac{S_n}{n}\right) - f(p)\right|\right] \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

And therefore we proved the required inequality. □

8 Day 8: Laws of Large Numbers

8.1 The strong law of large numbers -proof strategies

In order to prove the Strong LLN, we need many definitions. We begin with:

Definition 128 (lim inf and lim sup definition of sequence of sets (events))**.** Let $\{A_n\}_{n=1}^{\infty} \in \mathcal{F}$ be a collection of events.

1. The lim inf definition:

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\Big[\underbrace{\bigcap_{k\geq n} A_k}_{\text{the intersection is in all } k \geq n}\Big] = \{\omega \in \Omega : \liminf_{n\to\infty} \mathbf{1}_{A_n}(\omega) = 1\}$$

   If $\omega \in \liminf_{n\to\infty} A_n$, then there exists an N such that for all $n \geq N$, $\omega \in A_n$, that is:

$$\underbrace{A_1, \ldots, A_{N-1}}_{\text{not included}}, \underbrace{A_N, A_{N+1}, \ldots, A_n, \ldots,}_{\text{included}}$$

   Or we say ω_n is **in all except finitely many A_n's.**

2. The lim sup definition:

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\Big[\underbrace{\bigcup_{k\geq n} A_k}_{\text{for all } n\text{, at least one } k \geq n \text{ that is included}}\Big] = \{\omega \in \Omega : \limsup_{n\to\infty} \mathbf{1}_{A_n}(\omega) = 1\}$$

   If $\omega \in \limsup_{n\to\infty} A_n$, then there exists an N such that for all $n \geq N$, $\omega \in A_n$ **for infinitely many n's**, that is:

$$\underbrace{A_1, \ldots, A_{N-1}, A_N, \ldots, A_n, \ldots,}_{\text{included in infinitely many, just don't know where}}$$

   Or we say ω_n is **in infinitely many A_n's.**

Example 129 (An example of limsup and liminf)**.** Consider

$$A_n = \begin{cases} \{0\} & \text{if } n \text{ is odd} \\ \{1\} & \text{if } n \text{ is even} \end{cases}$$

Then

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

   The second equality is true because the intersection of the A_k are empty.

2.
$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k = \bigcap_{n=1}^{\infty} \{0,1\} = \{0,1\}$$

   The second equality is true because the union of the A_k are the set $\{0,1\}$.

Example 130 (Another example of limsup and liminf)**.** Let $A_n = (-\frac{1}{n}, 1-\frac{1}{n}]$. Then:

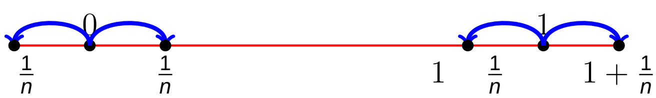
1.
$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k = \bigcup_{n=1}^{\infty} [0, 1-\frac{1}{n}] = [0,1)$$

2.
$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k = \bigcap_{n=1}^{\infty} (-\frac{1}{n}, 1) = [0,1)$$

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

   And this is also the $\lim_{n\to\infty} A_n = [0,1)$

Example 131 (More Example of limsup and liminf)**.** Consider $A_n = (\frac{(-1)^n}{n}, 1-\frac{(-1)^n}{n}]$.

1. Pictorially: 

2.
$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k = \bigcup_{n=1}^{\infty} (\frac{1}{n}, 1-\frac{1}{n}) = (0,1)$$

3.
$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k = \bigcap_{n=1}^{\infty} (-\frac{1}{n}, 1+\frac{1}{n}) = [0,1]$$

4. We see that
$$\liminf_{n\to\infty} A_n \neq \limsup_{n\to\infty} A_n$$

Remark 132. It is always true that

$$\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$$

Definition 133 (infinitely often=i.o). We say ω occurs **infinitely often, i.o.** if:

$$\omega \in \{A_n, i.o.\} = \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k$$

The following Theorem is the backbone in proving LLN, the **Borel-Cantelli Lemma**:

Theorem 134 (Borel-Cantelli Lemma and its inverse, sometimes called the first Borel-Cantelli Lemma and the second Borel-Cantelli Lemma. Theorem 2.2.1 and 2.3.7 Durett's book)**.**

1. ***Borel-Cantelli Lemma:*** *If*

$$\sum_{n=1}^{\infty} \boldsymbol{P}(A_n) < \infty$$

then

$$\boldsymbol{P}(A_n, i.o) = 0$$

2. ***Inverse of Borel-Cantelli Lemma*** *If*

$$\sum_{n=1}^{\infty} \boldsymbol{P}(A_n) = \infty$$

AND the $\{A_n\}$ are ***independent****, then*

$$\boldsymbol{P}(A_n, i.o) = 1.$$

Proof.

1. For the first part, this is what we wish to show:

$$\mathbf{P}\Big(\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k\Big) = 0 \tag{5}$$

We consider

$$B_n = \bigcup_{k\geq n} A_k$$

Then

$$\bigcup_{k=1} A_k \supseteq \bigcup_{k=2} A_k \supseteq \bigcup_{k=3} A_k \supseteq \dots$$

Meaning

$$B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots$$

We now use **Continuity from above** and get:

(a)

$$\lim_{n\to\infty} \mathbf{P}(B_n) = \mathbf{P}(\lim_{n\to\infty} B_n)$$

(b) Let's clarify the terms in the equation above:

$$\lim_{n\to\infty} B_n = \lim_{n\to\infty} \bigcup_{k \quad n}^{\infty} A_k = \bigcap_{n=1}^{\infty} \bigcup_{k \quad n}^{\infty} A_k$$

So the RHS is

$$\mathbf{P}(\lim_{n\to\infty} B_n) = \mathbf{P}(\bigcap_{n=1}^{\infty} \bigcup_{k \quad n}^{\infty} A_k)$$

We wish to show the RHS is zero, now we can look at the LHS term. But first, looking at the term $\lim_{n\to\infty} \mathbf{P}(B_n)$ in the RHS without the limit:

(c)

$$\mathbf{P}(B_n) = \mathbf{P}(\bigcup_{k=n}^{\infty} A_k)) \quad \sum_{k=n}^{\infty} \mathbf{P}(A_k)$$

where the last inequality is true by the monotonicity of probability measure. Now we take the limit, which will give us:

$$\text{LHS} = \lim_{n\to\infty} \mathbf{P}(B_n) \quad \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbf{P}(A_k)$$

Now since $\sum_{k=1}^{\infty} \mathbf{P}(A_k) < \infty$, it must be the case that the end terms in the series all go to zero, that is $\lim_{k\to\infty} \mathbf{P}(A_k) = 0$. But this also means that the summation of these end terms will be zero as well, that is, there exists a big n such that:

$$\sum_{k=n}^{\infty} \mathbf{P}(A_k) = 0$$

(d) So

$$\text{LHS} = \lim_{n\to\infty} \mathbf{P}(B_n) \quad \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbf{P}(A_k) = 0$$

(e) Therefore the RHS must also be zero and we get:

$$\text{RHS} = \mathbf{P}(\lim_{n\to\infty} B_n) = \mathbf{P}(\bigcap_{n=1}^{\infty} \bigcup_{k \quad n}^{\infty} A_k) = 0$$

This is what we wanted for this part.

2. For this part, we wish to show:

$$\mathbf{P}(\bigcap_{n=1}^{\infty} \bigcup_{k \quad n} A_k) = 1 \tag{6}$$

This is the same as showing

$$1 \quad \mathbf{P}(\bigcap_{n=1}^{\infty} \bigcup_{k \quad n} A_k) = 0 \tag{7}$$

A similar strategy to the first part:

$$1 - \mathbf{P}(\bigcap_{n=1}^{\infty}\bigcup_{k \geq n} A_k) = \mathbf{P}(\Big(\bigcap_{n=1}^{\infty}\bigcup_{k \geq n} A_k\Big)^C)$$
$$= \mathbf{P}(\Big(\bigcup_{n=1}^{\infty}\bigcap_{k \geq n} A_k^C\Big))$$

(a) We now let

$$B_n = \bigcap_{k \geq n} A_k^C$$

and we get:

$$\bigcap_{k=1} A_k \subseteq \bigcap_{k=2} A_k \subseteq \bigcap_{k=3} A_k \subseteq \dots$$

Meaning

$$B_1 \subseteq B_2 \subseteq B_3 \subseteq \dots$$

So we can use **Continuity from below** and get:

$$\lim_{n \to \infty} \mathbf{P}(B_n) = \mathbf{P}(\lim_{n \to \infty} B_n)$$

(b) We again will need to clarify the terms, the RHS is the one that we need it to equal to zero, because:

$$\lim_{n \to \infty} B_n = \lim_{n \to \infty} \bigcap_{k \geq n} A_k^C = \bigcup_{n=1}^{\infty}\bigcap_{k \geq n}^{\infty} A_k^C$$

So

$$\text{RHS} = \mathbf{P}(\Big(\bigcup_{n=1}^{\infty}\bigcap_{k \geq n} A_k^C\Big))$$

(c) Now let's take a look at the LHS, $\lim_{n \to \infty} \mathbf{P}(B_n)$. We will proceed by showing this term is indeed zero. Let's take a look at the $\mathbf{P}(B_n)$:

$$\lim_{n \to \infty} \mathbf{P}(B_n) = \lim_{n \to \infty} \mathbf{P}(\bigcap_{k \geq n} A_k^C)$$
$$= \prod_{k=n}^{\infty} \mathbf{P}(A_k^C); \text{ since the } A_k \text{ are independent, and so are the } A_k^C$$
$$= \prod_{k=n}^{\infty} \Big(1 - \mathbf{P}(A_k)\Big)$$

Now the trick here is to use this inequality to estimate the product:

$$1 - x \leq e^{-x}$$

So we get:

$$\prod_{k=n}^{\infty} \Big(1 - \mathbf{P}(A_k)\Big) \leq \prod_{k=n}^{\infty} e^{-\mathbf{P}(A_k)}$$
$$= e^{-\sum_{k=n}^{\infty} \mathbf{P}(A_k)} = e^{-\infty} = 0$$

The last equality is true because $\sum_{k=1}^{\infty} \mathbf{P}(A_k) = \infty$, so we can find a big (fixed) n such that

$$\sum_{k=n}^{\infty} \mathbf{P}(A_k) = \infty$$

So we concluded that

$$\text{LHS} = \lim_{n \to \infty} \mathbf{P}(B_n) = 0$$

And passing it to the RHS:

$$\text{RHS} = \mathbf{P}(\left(\bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k^C \right)) = \text{LHS} = 0$$

(d) So

$$1 - \mathbf{P}(A_n, i.o.) = 0$$

And therefore

$$\mathbf{P}(A_n, i.o.) = 1$$

This completed the proof. □

Example 135 (An example of Borel-Cantelli Lemma)**.** Consider A_n = winning at the n-th bet. So the A_n are independent. And also,

$$\mathbf{P}(A_n) = \frac{1}{n}$$

Now then

$$\sum_{i=1}^{\infty} \mathbf{P}(A_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

, so we use the **second Borel-Cantelli Lemma** and get

$$\mathbf{P}(A_n, i.o.) = 1 = \text{ probability winning infinitely often}$$

But if we consider the set $A_{n^2} = A_1, A_4, A_9, \ldots$, then:

$$\sum_{n=1}^{\infty} \mathbf{P}(A_{n^2}) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

Therefore by the first Borel-Cantelli Lemma,

$$\mathbf{P}(\text{winning i.o.}) = 0$$

BINGO!! SHOULD WE BET?

Example 136 (Another **seem to counter?** Borel-Cantelli Lemma Example)**.** Consider the space $([0,1], B([0,1]), \mathbf{P})$ with $\mathbf{P}((a,b)) = b - a$. Consider then the events $A_n = [0, \frac{1}{n}]$, so $\mathbf{P}(A_n) = \frac{1}{n}$.Then:

1.

$$\limsup_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} [0, \frac{1}{k}] = \bigcap_{n=1}^{\infty} [0, \frac{1}{n}] = \{0\}$$

2. So

$$\mathbf{P}(A_n, i.o.) = \mathbf{P}(\{0\}) = 0$$

3. But we have

$$\sum_{i=1}^{\infty} \mathbf{P}(A_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

So using second Borel-Cantelli Lemma, we should have:

$$\mathbf{P}(A_n, i.o.) = 1$$

4. This is because the A_n are **NOT independent**, because:

$$\mathbf{P}(A_{10} \cap A_2) = \mathbf{P}([0, \frac{1}{10}] \cap [0, \frac{1}{2}]) = \mathbf{P}([0, \frac{1}{10}]) = \frac{1}{10} \neq \mathbf{P}(A_{10})\mathbf{P}(A_2) = \frac{1}{10}\frac{1}{2} = \frac{1}{20}$$

Remark 137 (**Important Relationship between i.o and a.s**)**.** Now this is why we introduce **infinitely often, i.o.**, because it is related to **almost surely** convergence. Recall the definition of convergence almost surely (a.s):

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

if:

$$\mathbf{P}(\{\omega \in \Omega : \lim_{n \to \infty} X_n = X\}) = 1$$

Looking at the limit definition, that translated to, for any $\epsilon > 0$, there will exists a k such that for all $n \geq k$, $|X_n(\omega) - X(\omega)| < \epsilon$ is the set that we are considering now. Meaning:

$$\forall \epsilon > 0, \mathbf{P}(\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \{|X_n - X| < \epsilon\}) = 1$$

If we apply the complement probability to the above, and note that complement of union is intersection (and via versa), we get:

$$\forall \epsilon > 0, \mathbf{P}(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} \{|X_n - X| < \epsilon\}) = 0$$

And this by definition of infinitely often, we have:

$$\forall \epsilon > 0, \mathbf{P}(\{|X_n - X| \geq \epsilon\}, i.o.) = 0$$

And here is where we use the First Borel-Cantelli Lemma, if we can show

$$\sum_{n=1}^{\infty} \mathbf{P}(\{|X_n - X| \geq \epsilon\}) < \infty$$

then the above is true:

$$\mathbf{P}(\{|X_n - X| \geq \epsilon\}, i.o.) = 0.$$

and therefore, equivalently, we could show:

$$X_n \xrightarrow{a.s.} X$$

Remark 138. It is true that

$$X_n \xrightarrow{a.s.} X$$

is equivalent to:

$$\mathbf{P}(\{\omega \in \Omega : \lim_{n \to \infty} X_n \neq X\}) = 0$$

And in turns, equivalent to: $\exists \varepsilon > 0$ such that for all N, the quantity $|X_n(\omega) - X(\omega)| > \varepsilon$ is true for all $n \geq N$.
However, this is NOT useful. And therefore is not used. Rather we use the earlier remark and apply Borel-Cantelli Lemma.

9 Day 9: More Laws of Large Numbers

9.1 Strong law of large numbers in L2 and L4, and an anti LLN

Recall an important inequality, Chebychev Inequality, here we have a variation of it:

Theorem 139 (Chebychev Inequality variation)**.**

$$P(|X| \geq c) \leq \frac{\mathbf{E}[|X|^p]}{c^p}$$

is true for some positive real numbers c, p.

Proof. Strategy: split the domain $|X| > c$ as before. □

Theorem 140 (Strong Law of Large Numbers in L^4)**.** *Let X_i be i.i.d* ***(independent identical distributed)*** *where $\mathbf{E}[X_i^4] < \infty$, and $S_n = \sum_{i=1}^n X_i$. Then:*

$$\frac{S_n}{n} \xrightarrow{a.s.} \mu = \mathbf{E}[X_1]$$

Proof. WLOG, we can assume $\mu = 0$ by considering $\overline{X_n} = X_n - \mu$. We will use the infinitely often technique to prove convergence almost surely (see Remark 137 for more details). So it suffices to show that for all $\varepsilon > 0$:

$$\mathbf{P}(|\frac{S_n}{n}| > \varepsilon \text{ i.o.}) = 0$$

We can use Borel-Cantelli Lemma and try to show:

$$\sum_{i=1}^{\infty} \mathbf{P}(|\frac{S_n}{n}| > \varepsilon) < \infty$$

We use chebychev Inequality to obtain the following estimates:

$$\mathbf{P}(|\frac{S_n}{n}| > \varepsilon) \leq \frac{\mathbf{E}\left[|\frac{S_n}{n}|^4\right]}{\varepsilon^4} = \frac{\mathbf{E}\left[|S_n|^4\right]}{n^4 \varepsilon^4}$$

Now we would try to estimate $\mathbf{E}[S_n^4]$:

$$\mathbf{E}[S_n^4] = \mathbf{E}[(\sum_{i=1}^n X_i)^4] = \mathbf{E}[\sum_{i,j,k,l} X_i X_j X_k X_l]$$

Among those terms, because of independence of the X_i, we note the following:

1. $\mathbf{E}[X_i X_j^3] = \mathbf{E}[X_i]\mathbf{E}[X_j^3] = 0$, for $i \neq j$.
2. $\mathbf{E}[X_i X_j X_k X_l] = \mathbf{E}[X_i]\mathbf{E}[X_j]\mathbf{E}[X_k]\mathbf{E}[X_l] = 0$, for $i \neq j \neq l \neq k$ distinct.
3. $\mathbf{E}[X_j X_j^2 X_k] = 0$, for $i \neq j \neq k$ distinct.

Thus the $\mathbf{E}[S_n^4]$ term is now:

$$\begin{aligned}\mathbf{E}[S_n^4] &= \mathbf{E}\Bigg[\sum_{i=1}^{n} X_i^4 + \text{constant}\sum_{i,j} X_i^2 X_j^2\Bigg] \\ &= n\mathbf{E}[X_i^4] + Cn^2\Big(\mathbf{E}[X_i^2]\Big)^2 \leq Kn^2, \text{ since } \mathbf{E}[X_i^4], \mathbf{E}[X_i^2] < \infty\end{aligned}$$

Therefore we combine and get:

$$\mathbf{P}(|\frac{S_n}{n}| \geq \epsilon) \leq \frac{Kn^2}{n^4 \epsilon^4} = \frac{K}{n^2 \epsilon^4}$$

Taking the summation $n = 1$ to infinity of the above gives:

$$\sum_{n=1}^{\infty}\mathbf{P}(|\frac{S_n}{n}| \geq \epsilon) \leq \sum_{i=1}^{\infty}\frac{K}{n^2\epsilon^4} = \frac{K}{\epsilon^4}\sum_{i=1}^{\infty}\frac{1}{n^2} < \infty$$

And Borel-Cantelli Lemma gives the result. □

Theorem 141 (**Anti LLN**)**.** *Suppose X_i are i.i.d, and that $\mathbf{E}[|X_i|] = \infty$. Denote $S_n = \sum_{i=1}^n X_i$. Then:*

$$\boldsymbol{P}(\frac{S_n}{n} \textit{ converges to a finite limit}) = 0$$

Proof. We have:

$$\begin{aligned}\infty = \mathbf{E}[|X_1|] &\leq \sum_{n=0}^{\infty}\mathbf{P}(|X_1| \geq n) \text{ , definition of expectation } \int |X_1| \leq \int_{\{\bigcup_{n=0}^{\infty} |X_1| \geq n\}} \mathbf{1} \\ &= \sum_{n=0}^{\infty}\mathbf{P}(|X_n| \geq n); \text{iid of the } X_i\end{aligned}$$

So we must have $\sum_{n=0}^{\infty}\mathbf{P}(|X_n| \geq n) = \infty$. Since the X_n are independent, by the Second Borel-Cantelli lemma, this gives:

$$\mathbf{P}(|X_n| \geq n, i.o.) = 1$$

Equivalently:

$$\mathbf{P}(\frac{|X_n|}{n} \geq 1, i.o.) = 1$$

Let us denote the set

$$C = \{\omega : \frac{S_n(\omega)}{n} \text{ converges to a finite limit}\}$$

We wish to show that $\mathbf{P}(C) = 0$. Now for any ω, we have:

$$\left|\frac{S_n(\omega)}{n} - \frac{S_{n-1}(\omega)}{n-1}\right| = \left|\frac{S_{n-1}+X_n}{n} - \frac{S_{n-1}}{n-1}\right| = \left|-\frac{S_{n-1}}{n(n-1)} + \frac{X_n}{n}\right|$$

The first equality is from $S_n = S_{n-1} + X_n$. The second equality is just by grouping the S_{n-1} together. Now we consider this for $\omega \in C$. On this set C, since $\lim_{n\to\infty} \frac{S_n}{n} = c < \infty$, we must have:

$$\left|-\frac{S_{n-1}}{n(n-1)} + \frac{X_n}{n}\right| \xrightarrow{n\to\infty} 0$$

In addition:

$$\frac{S_{n-1}}{n(n-1)} \xrightarrow{n\to\infty} \frac{c}{n-1} \xrightarrow{n\to\infty} 0$$

Therefore we must have:

$$\frac{X_n}{n} \xrightarrow{n\to\infty} 0$$

Thus the intersection of C and $\{\frac{|X_n|}{n} \geq 1; i.o.\}$ must be empty:

$$C\bigcap\{\frac{|X_n|}{n} \geq 1; i.o.\} = \emptyset$$

In other words, C and $\{\frac{|X_n|}{n} \geq 1; i.o.\}$ are independent. That is

$$\mathbf{P}(C\bigcap\{\frac{|X_n|}{n} \geq 1; i.o.\}) = \mathbf{P}(C)\mathbf{P}(\{\frac{|X_n|}{n} \geq 1; i.o.\})$$

But

$$\mathbf{P}(C\bigcap\{\frac{|X_n|}{n} \geq 1; i.o.\}) = \mathbf{P}(\emptyset) = 0$$

Therefore:

$$\mathbf{P}(C)\mathbf{P}(\{\frac{|X_n|}{n} \geq 1; i.o.\}) = 0$$

As we showed earlier

$$\mathbf{P}(\{\frac{|X_n|}{n} \geq 1; i.o.\}) = 1 \neq 0$$

So it must be the case that

$$\mathbf{P}(C) = 0$$

And this concluded the proof. □

Theorem 142 (Strong Law of Large Numbers with L^2 condition)**.** *Let X_i be i.i.d. with $E[|X_i|^2] < \infty$. Then:*

$$\frac{S_n}{n} \xrightarrow{a.s.} \mu = 0$$

Here we assume without loss of generality that $\mu = 0$*. Also note that* $\sigma^2 = Var(X_i) < \infty$*.*

Proof. We tend to use the same strategy with the Borel Cantelli Lemma and Chebychev inequality and start with:

$$\begin{aligned}
\mathbf{P}\left(\frac{|S_n|}{n} > \varepsilon\right) &\leq \frac{\mathbf{E}[S_n^2]}{n^2\varepsilon^2} \\
&= \frac{\mathrm{Var}(S_n^2)}{n^2\varepsilon^2} \text{ ,because } \mathrm{Var}(S_n^2) = \mathbf{E}[S_n]^2 - n^2(\mu)^2 = \mathbf{E}[S_n]^2 \text{ ,since } \mu = 0. \\
&= \frac{n\mathrm{Var}(X_i)}{n^2\varepsilon^2} = \frac{\mathrm{Var}(X_i)}{n\varepsilon^2}
\end{aligned}$$

But if we want to use Borel-Cantelli Lemma here, by taking summation, we get inconclusive inequality:

$$\sum_{n=0}^{\infty} \mathbf{P}\left(\frac{|S_n|}{n} > \varepsilon\right) \leq \sum_{n} \frac{c}{n} = \infty$$

Here is an important technique from Real Analysis, **passing to a subsequence:** Consider the subsequence X_{n^2}. For this subsequence we have:

$$\mathbf{P}\left(\frac{|S_{n^2}|}{n^2} \geq \varepsilon\right) \leq \frac{\mathrm{Var}(X_i)}{n^2 \varepsilon^2}$$

Therefore taking the sum:

$$\sum_{n=1}^{\infty} \mathbf{P}\left(\frac{|S_{n^2}|}{n^2} \geq \varepsilon\right) \leq \sum_{n=1}^{\infty} \frac{\mathrm{Var}(X_i)}{n^2 \varepsilon^2} = \sum_{n=1}^{\infty} \frac{c}{n^2} < \infty$$

That is, the first Borel-Cantelli Lemma can be applied for this subsequence (because $\sum_{n=1}^{\infty} \mathbf{P}(\frac{|S_{n^2}|}{n^2} \geq \varepsilon) < \infty$) and we get:

$$\mathbf{P}\left(\frac{S_{n^2}}{n^2}; i.o.\right) = 0$$

So we just showed:

$$\frac{S_{n^2}}{n^2} \xrightarrow{a.s.} 0$$

What about for the original sequence? Here we have **an important observation**, for any k with $n^2 < k < (n+1)^2$, we have:

$$\begin{aligned} 0 \leq \left|\frac{S_k}{k}\right| \leq \left|\frac{S_k}{n^2}\right| &= \left|\frac{S_{n^2} + \sum_{i=n^2+1}^{k} X_i}{n^2}\right| \\ &= \left|\frac{S_{n^2}}{n^2} + \frac{\sum_{i=n^2+1}^{k} X_i}{n^2}\right| \\ &\leq \left|\frac{S_{n^2}}{n^2}\right| + \left|\frac{\sum_{i=n^2+1}^{k} X_i}{n^2}\right| \end{aligned}$$

In short, we have:

$$0 \leq \left|\frac{S_k}{k}\right| \leq \left|\frac{S_{n^2}}{n^2}\right| + \left|\frac{\sum_{i=n^2+1}^{k} X_i}{n^2}\right|$$

We already know that $\frac{S_{n^2}}{n^2} \xrightarrow{a.s.} 0$, so if we can show $\frac{\sum_{i=n^2+1}^{k} X_i}{n^2} \xrightarrow{a.s.} 0$, then we are done. We define

$$D_n := \max_{n^2 < k < (n+1)^2} \sum_{i=n^2+1}^{k} X_i$$

Same strategies before, we obtained:

$$\mathbf{P}\left(\frac{D_n}{n^2} > \varepsilon\right) \leq \frac{\mathbf{E}[D_n^2]}{n^4 \varepsilon^2}$$

And we will now try to estimate $\mathbf{E}[D_n^2]$. Indeed we have:

$$\mathbf{E}[D_n^2] = \mathbf{E}[\Big(\max_{n^2<k<(n+1)^2} \sum_{i=n^2+1}^{k} X_i\Big)^2]$$

$$\leq \mathbf{E}[\Big(\sum_{i=n^2+1}^{(n+1)^2} |X_i|^2\Big)] = (2n+1)^2\sigma^2$$

Then we take summation:

$$\sum_{n=1}^{\infty} \mathbf{P}(\frac{D_n}{n^2} > \varepsilon) \leq \sum_{n=1}^{\infty} \frac{(2n+1)^2\sigma^2}{n^4\varepsilon^2}$$

$$\leq \sum_{n=1}^{\infty} \frac{K(2n+1)^2}{n^4} \leq \sum_{n=1}^{\infty} \frac{K(2n+2)^2}{n^2} = \sum_{n=1}^{\infty} \frac{4K(n+1)^2}{n^4} < \infty$$

The last inequality is because the series $\sum_{n=1}^{\infty} \frac{(n+1)^2}{n^4}$ converges by integral test. In short, because

$$\sum_{n=1}^{\infty} \mathbf{P}(\frac{D_n}{n^2} > \varepsilon) \leq \infty$$

, so Borel-Cantelli lemma gives:

$$\mathbf{P}(\frac{D_n}{n^2}; i.o.) = 0$$

That is, $\frac{D_n}{n^2} \xrightarrow{a.s.} 0.$ And from the way we defined $D_n = \max_{n^2<k<(n+1)^2} \sum_{i=n^2+1}^{k} X_i$, we must have

$$\frac{\sum_{i=n^2+1}^{k} X_i}{n^2} \xrightarrow{a.s.} 0$$

And this completed the proof. □

9.2 Weak Law of Large Number in L1

This is one of the main point. The proof of this theorem employs many great techniques and must do in many steps. (Half of these are done in Day 10 in real lecture time, but I combined it in one place here. So it might seems lengthy.)

Theorem 143 (Weak Law of Large Numbers, final version in L^1). *The X_i are i.i.d. with $E[X_i] < \infty$. Then:*

$$\frac{S_n}{n} \xrightarrow{P} \mu = 0$$

Proof. The proof will be showed in 4 steps. First we define

$$Y_n := X_n \mathbf{1}_{|X_n| \leq n}$$

The four steps are:

1. Show

$$Y_n \overset{a.s}{=} X_n$$

2. Show
$$\mathbf{E}[Y_n] \xrightarrow[as]{n\to\infty} \quad = 0$$

3. Show
$$\frac{\sum_{i=1}^n Y_i}{n} \xrightarrow[p]{n\to\infty} 0$$

4. Show
$$\frac{S_n}{n} \quad \frac{\sum_{i=1}^n Y_i}{n} \xrightarrow{n} 0$$

We will now begin showing the first step.

1. **Step 1:** Show $Y_n \overset{a.s}{=} X_n$. This again due to almost surely, we will use the infinitely often strategy (the First Borel-Catelli Lemma) and we will need to show that:
$$\sum_{i=1}^{\infty} \mathbf{P}(\{X_n \neq Y_n\}) < \infty$$
We start with the set:
$$\{X_n \neq Y_n\} = \{\mathbf{1}_{|X_n| \quad n}\}$$
The above is true because of the definition of Y_n. Now then we take the summation (and probability):
$$\sum_{i=1}^{\infty} \mathbf{P}(\{X_n \neq Y_n\}) = \sum_{i=1}^{\infty} \mathbf{P}(|X_n| > n) \quad 1 + \mathbf{E}[|X_1|] < \infty$$
The last inequality is from Homework.

2. **Step 2:** Show $\mathbf{E}[Y_n] \xrightarrow[as]{n\to\infty} \quad = 0$. Observe the following bound:
$$|Y_n| = |X_n \mathbf{1}_{|X_n| \quad n}| \quad |X_n|$$
and also $\mathbf{E}[|X_n|] = \mathbf{E}[|X_1|] < \infty$. We then can use the Dominated Convergence Theorem with X_1 bounding all the Y_n:
$$\lim_{n\to\infty} \mathbf{E}[Y_n] = \mathbf{E}[\lim_{n\to\infty} Y_n] = \mathbf{E}[\lim_{n\to\infty} X_n \mathbf{1}_{|X_n| \quad n}] = \mathbf{E}[X_n] = \quad = 0$$

3. **Step 3:** Show $\frac{\sum_{i=1}^n Y_i}{n} \xrightarrow[p]{n\to\infty} 0$. From definition of convergence in probability, we wish to show that, for all $\varepsilon > 0$, the:
$$\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i\right| > \varepsilon\right) \to 0$$
We could at first use the Chebychev inequality to estimate:
$$\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i\right| > \varepsilon\right) \quad \frac{\mathbf{E}\left[\left(\sum_{i=1}^n Y_i\right)\right]}{n^2\varepsilon^2}$$

However, this idea will not work because the numerator is:

$$
\begin{aligned}
\mathbf{E}[\left(\sum_{i=1}^{n} Y_i\right)] &= \sum_{i=1}^{n} \mathbf{E}[Y_i^2] \\
&= \sum_{i=1}^{n} \mathbf{E}[X_i^2 \mathbf{1}_{|X_i| \le n}] \\
&\le \sum_{i=1}^{n} i\,\mathbf{E}[X_i \mathbf{1}_{|X_i| \le n}], \text{ taking one of the } X_i \text{ inside of the expectation to be } \le i \\
&\le \frac{n(n+1)}{2}
\end{aligned}
$$

Putting this into the denominator:

$$
\mathbf{P}(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| > \varepsilon) \le \frac{n(n+1)}{2n^2\varepsilon^2} = \frac{n+1}{2n\varepsilon^2}
$$

And the fraction $\frac{n+1}{2n\varepsilon^2}$ will not go to zero when n goes to infinity.
We then explore another strategy, which is also common in Real Analysis, the diagonal slicing technique for summing over 2 indices:

$$
\begin{aligned}
\mathbf{E}[\left(\sum_{i=1}^{n} Y_i\right)] &= \sum_{i=1}^{n} \mathbf{E}[Y_i^2] \\
&= \sum_{i=1}^{n} \mathbf{E}[X_i^2 \mathbf{1}_{|X_i| \le n}] \\
&= \sum_{i=1}^{n}\sum_{j=0}^{i-1} \mathbf{E}[X_i^2 \mathbf{1}_{j \le |X_i| \le j+1}]; \text{ , here is the summation break into i and j} \\
&= \sum_{j=0}^{i-1} (n-j)\mathbf{E}[X_i^2 \mathbf{1}_{j \le |X_i| \le j+1}]; \text{ see figure 17 below}
\end{aligned}
$$

Now in summary we get:

$$
\begin{aligned}
\mathbf{P}(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| > \varepsilon) &\le \frac{1}{n^2\varepsilon^2}\sum_{j=0}^{i-1}(n-j)\mathbf{E}[X_i^2 \mathbf{1}_{j \le |X_i| \le j+1}] \\
&\le \frac{1}{\varepsilon^2}\sum_{j=0}^{i-1}\frac{(n-j)(j+1)}{n^2}\mathbf{E}[|X_i| \mathbf{1}_{j \le |X_i| \le j+1}]
\end{aligned}
$$

, we pull one of $|X_i| \le j+1$ out of the expectation

Now we will take care of that product in the following way. Let $\delta > 0$ be given (and we want this to be small). Then:

(a) For "small" j, that is, there exists a large N_2 such that for all $n \ge N_2$, the quantity

$$
\sum_{j=1}^{N_2}\frac{(n-j)(j+1)}{n^2} < \delta
$$

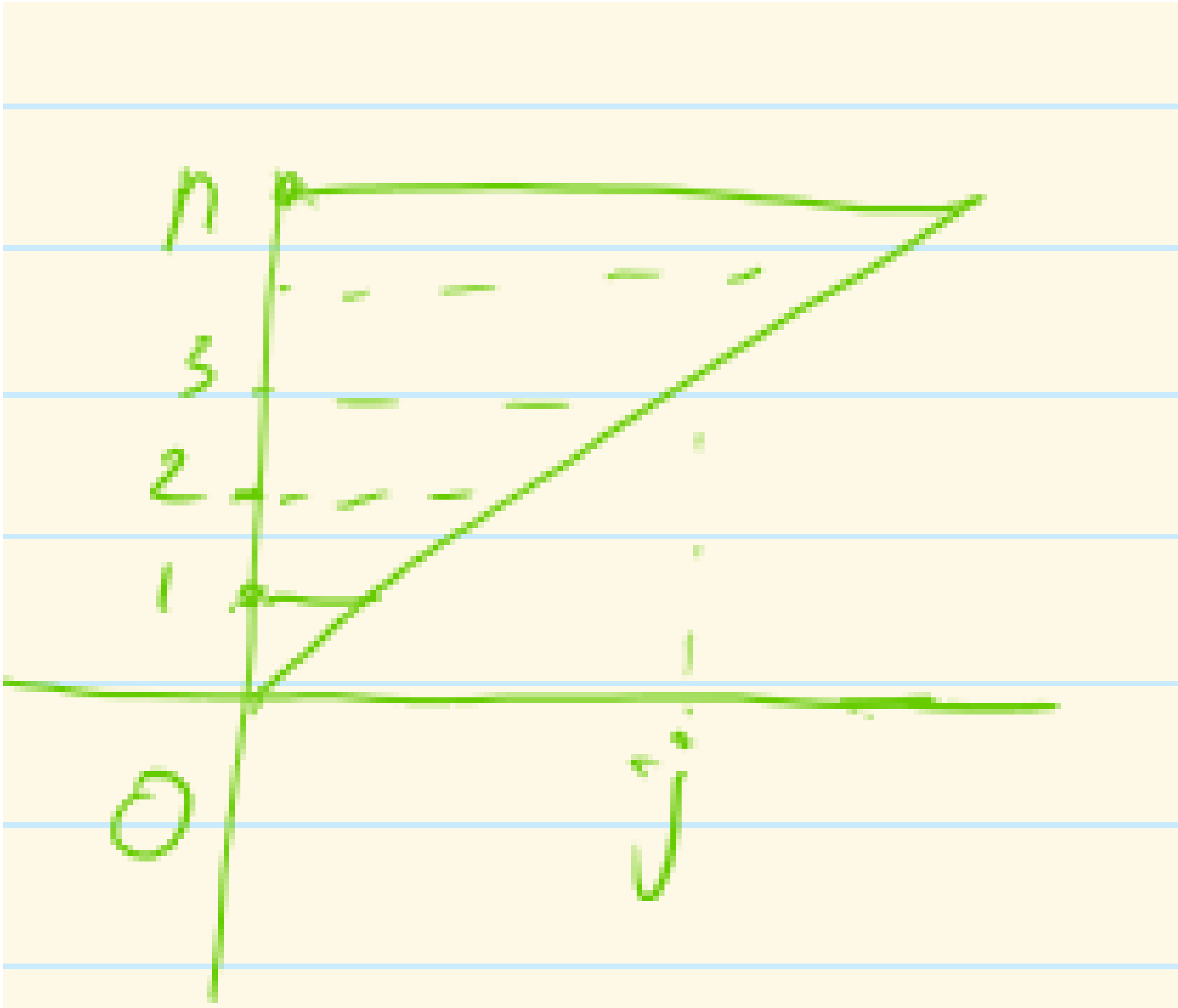


Figure 17: diagonal slicing index

(b) For large j, we can also find a large N_1 such that:

$$\sum_{j=N_1}^{\infty} \mathbf{E}[|X_i|\mathbf{1}_{j \leq |X_i| \leq j+1}] < \varepsilon$$

The existence of such N_1 is guaranteed because

$$\mathbf{E}[|X_1|] = \sum_{j=1}^{\infty} \mathbf{E}[|X_i|\mathbf{1}_{j \leq |X_i| \leq j+1}] < \infty$$

(and so the sum of the tail of the series must be small)

(c) Now we choose $N = \max\{N_1, N_2\}$, so that whenever $n \geq N$, both part (a) and part (b) must be true. Then:

$$\begin{aligned}
\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| > \varepsilon\right) &\leq \frac{1}{\varepsilon^2}\sum_{j=0}^{i-1} \frac{(n-j)(j+1)}{n^2}\mathbf{E}[|X_i|\mathbf{1}_{j \leq |X_i| \leq j+1}] \\
&= \frac{1}{\varepsilon^2}\sum_{j=0}^{N} \frac{(n-j)(j+1)}{n^2}\mathbf{E}[|X_i|\mathbf{1}_{j \leq |X_i| \leq j+1}] \\
&+ \frac{1}{\varepsilon^2}\sum_{j=N+1}^{n} \frac{(n-j)(j+1)}{n^2}\mathbf{E}[|X_i|\mathbf{1}_{j \leq |X_i| \leq j+1}]
\end{aligned}$$

Let us name

$$A_n = \frac{1}{\varepsilon^2}\sum_{j=0}^{N} \frac{(n-j)(j+1)}{n^2}\mathbf{E}[|X_i|\mathbf{1}_{j \leq |X_i| \leq j+1}]$$

and

$$B_n = \frac{1}{\varepsilon^2}\sum_{j=N+1}^{n} \frac{(n-j)(j+1)}{n^2}\mathbf{E}[|X_i|\mathbf{1}_{j \leq |X_i| \leq j+1}]$$

i. For A_n, we will use part (a): We take an upper bound for the

$$\mathbf{E}[|X_i|\mathbf{1}_{j \leq |X_i| \leq j+1}] \leq M$$

and get:

$$\begin{aligned}
A_n &= \frac{1}{\varepsilon^2}\sum_{j=0}^{N} \frac{(n-j)(j+1)}{n^2}\mathbf{E}[|X_i|\mathbf{1}_{j \leq |X_i| \leq j+1}] \\
&\leq \frac{1}{\varepsilon^2}\sum_{j=0}^{N} \frac{(n-j)(j+1)}{n^2}M \\
&= \frac{M}{\varepsilon^2}\sum_{j=0}^{N} \frac{(n-j)(j+1)}{n^2} \\
&\leq \frac{M}{\varepsilon^2} = C\varepsilon \text{ , use part (a), } \sum_{j=1}^{N} \frac{(n-j)(j+1)}{n^2} < \varepsilon
\end{aligned}$$

So when $n \to \infty$, the $A_n < C\varepsilon$.

ii. For B_n, we will use part (b): And we note that

$$\frac{(n-j)(j+1)}{n^2} \le 1$$

Then:

$$\begin{aligned} B_n &= \frac{1}{\varepsilon^2}\sum_{j=N+1}^{n} \frac{(n-j)(j+1)}{n^2}\mathbf{E}[|X_i|\mathbf{1}_{j \le |X_i| \le j+1}] \\ &\le \frac{1}{\varepsilon^2}\sum_{j=N+1}^{n} \mathbf{E}[|X_i|\mathbf{1}_{j \le |X_i| \le j+1}]; \text{ , by the note} \end{aligned}$$

Therefore when taking limit $n \to \infty$:

$$\begin{aligned} \lim_{n\to\infty} B_n &\le \lim_{n\to\infty} \frac{1}{\varepsilon^2}\sum_{j=N+1}^{n} \mathbf{E}[|X_i|\mathbf{1}_{j \le |X_i| \le j+1}] \\ &= \frac{1}{\varepsilon^2}\sum_{j=N+1}^{\infty} \mathbf{E}[|X_i|\mathbf{1}_{j \le |X_i| \le j+1}] \\ &\le \frac{1}{\varepsilon^2}\delta; \text{ using part (b), } \sum_{j=N}^{\infty} \mathbf{E}[|X_i|\mathbf{1}_{j \le |X_i| \le j+1}] < \delta \quad = D\delta \end{aligned}$$

In short, when taking $n \to \infty$, we have:

$$\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| > \varepsilon\right) \le \lim_{n\to\infty} A_n + \lim_{n\to\infty} B_n \le C\delta + D\delta = K\delta$$

Letting $\delta \to 0$ we obtained:

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[p]{n\to\infty} 0.$$

4. **Step 4:** Show $\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) \xrightarrow{p} 0$. To show this, let $\varepsilon > 0$ be given and we wish to show

$$\mathbf{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}(X_i - Y_i)\right|\right) \to 0$$

Indeed we have:

$$\begin{aligned} \mathbf{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}(X_i - Y_i)\right|\right) &\le \frac{1}{n}\sum_{i=1}^{n}\mathbf{P}(X_i \neq Y_i) \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{P}(|X_i| > i); \text{ by definition of } Y_i \\ &\le \frac{1}{n}\Big(1 + \mathbf{E}[|X_1|]\Big) \xrightarrow{n\to\infty} 0 \end{aligned}$$

where the last inequality $\mathbf{P}(|X_i| > i) \le \Big(1 + \mathbf{E}[|X_1|]\Big)$ is proven in Homework.

Then we can use a small lemma, that is if we have $f_n \xrightarrow{p} f$, and $g_n \xrightarrow{p} g$, then $f_n + g_n \xrightarrow{p} f + g$. Using this lemma and combine step 3 and step 4 together to get the result:

1. Step 3 showed that:
$$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{p} 0$$

2. Step 4 showed that:
$$\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) \xrightarrow{p} 0$$

3. Adding both step 3 and step 4 showed:
$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} 0$$

And that concluded the proof. □

10 Day 10: Strong law of large number in L1 and applications

10.1 Strong LLN in L1

Theorem 144. *Let X_n be i.i.d. and that $E[|X_1|] < \infty$. Then:*

$$\frac{S_n}{n} \xrightarrow[as]{n \to \infty} \mathbf{E}[X_1] = \mu = 0$$

Proof. As before, we can always assume $\mu = 0$ without loss of generality. In addition, by breaking $X_1 = X_1^+ - X_1^-$ like in the definition of expectation (the last step in the definition), we may also assume X_1 is non-negative. The strategy is similar to the proof in the L^1 case of the Weak LLN. First we define:

$$Y_n = X_n \mathbf{1}_{|X_n| < n}$$

And we follow the earlier steps 1 and 2 in the proof of Weak LLN case and obtained:

1. The first part is that (by Borel-Cantelli Lemma):
$$Y_n \overset{a.s}{=} X_n$$
However, we would like to write this result in a bit different (but equivalent) way:
$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - Y_i\right) \xrightarrow[n \to \infty]{as} 0 \tag{8}$$

2. The second part is the same (by DCT):
$$\mathbf{E}[Y_n] \xrightarrow[as]{n \to \infty} \mu = 0$$

Dividing that by n gives:

$$\frac{1}{n}\mathbf{E}[Y_n] \xrightarrow[as]{n\to\infty} = 0 \tag{9}$$

Then if we can show:

$$\frac{1}{n}\sum_{i=1}^{n}\Big(Y_i - \mathbf{E}[Y_i]\Big) \xrightarrow[n\to\infty]{as} 0 \tag{10}$$

Adding all the equations 8 and 9 and 10 will get the desired result:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[as]{n\to\infty} 0$$

It remains to show 10. As usual, this is not easy.

3. Let us define:

$$T_n := \frac{1}{n}\sum_{i=1}^{n}\Big(Y_i - \mathbf{E}[Y_i]\Big)$$

So using the same infinitely often technique again, tt's suffice to show, for all $\varepsilon > 0$ be given:

$$\sum_{n=1}^{\infty}\mathbf{P}(|T_n| > \varepsilon) < \infty$$

$$\begin{aligned}
\mathbf{P}(|T_n| > \varepsilon) &\le \frac{\mathbf{E}[|T_n|^2]}{\varepsilon^2} = \frac{\mathbf{E}[\sum_{i=1}^{n}\Big(Y_i - \mathbf{E}[Y_i]\Big)^2]}{\varepsilon^2} \\
&= \frac{\sum_{i=1}^{n}\mathbf{E}[\Big(Y_i - \mathbf{E}[Y_i]\Big)^2]}{n^2\varepsilon^2}; \text{ since all the } Y_i \text{ are independent} \\
&= \sum_{i=1}^{n}\frac{\mathbf{E}[\Big(Y_i - \mathbf{E}[Y_i]\Big)^2]}{n^2\varepsilon^2} \\
&\le \frac{\sum_{i=1}^{n}\mathbf{E}[Y_i^2]}{n^2\varepsilon^2}
\end{aligned}$$

We will now try to estimate $\mathbf{E}[Y_i^2]$. Just like in the proof of weak law of large number in L^1, a naive Chebychev Inequality strategy will not work any more. As such we come to the subsequence strategy.

(a) We take the subsequence:

$$a_n := \lfloor \alpha^n \rfloor; \text{ for any fixed } \alpha > 1$$

Where $\lfloor \alpha^n \rfloor$ is the largest integer below α^n. For example, when $\alpha = 1.5$, then $\lfloor \alpha \rfloor = 1$, $\lfloor \alpha^2 \rfloor = 2$ (since $\alpha^2 = 2.25$). With a fixed α, we can have a sequence

a_n and proceed:

$$\begin{aligned} \mathbf{P}(|T_n| > \varepsilon) &\le \frac{\sum_{i=1}^{a_n} \mathbf{E}[Y_i^2]}{a_n^2 \varepsilon^2} \\ &= \sum_{i=1}^{a_n} \frac{\mathbf{E}[X_i^2 \mathbf{1}_{\{|X_i| \le i\}}]}{a_n^2 \varepsilon^2} \quad \text{,using definition of } Y_i \\ &\le \sum_{i=1}^{a_n} \frac{\mathbf{E}[X_i^2 \mathbf{1}_{\{|X_i| \le a_n\}}]}{a_n^2 \varepsilon^2} \\ &\le \frac{\mathbf{E}[X_i^2 \mathbf{1}_{\{|X_i| \le a_n\}}]}{a_n \varepsilon^2} \end{aligned}$$

We drop $1/a_n$ and the summation because, on the numerator, each $\mathbf{E}[X_i^2 \mathbf{1}_{|X_i| \le a_n}] \le a_n$, so summing them up have a_n factor in the front, and this factor got canceled with one of the a_n on the denominator. So in short:

$$\begin{aligned} \mathbf{P}(|T_n| > \varepsilon) &\le \frac{\mathbf{E}[X_i^2 \mathbf{1}_{\{|X_i| \le a_n\}}]}{a_n \varepsilon^2} \\ &= \frac{1}{\varepsilon^2} \mathbf{E}\left[X_1^2 \frac{1}{a_n} \mathbf{1}_{\{|X_1| \le a_n\}} \right] \end{aligned}$$

We can choose N such that the smallest integer $a_N = \alpha^N > X_1$, and then taking infinite sum on the above:

$$\begin{aligned} \sum_{i=1}^{\infty} \mathbf{P}(|T_n| > \varepsilon) &\le \sum_{i=1}^{\infty} \frac{1}{\varepsilon^2} \mathbf{E}\left[X_1^2 \frac{1}{a_n} \mathbf{1}_{\{|X_1| \le a_n\}} \right] \\ &\le \frac{1}{\varepsilon^2} \mathbf{E}\left[\sum_{i=1}^{\infty} X_1^2 \frac{1}{a_n} \mathbf{1}_{\{|X_1| \le a_n\}} \right] \\ &= \frac{1}{\varepsilon^2} \mathbf{E}\left[\sum_{i=1}^{N-1} X_1^2 \frac{1}{a_n} \mathbf{1}_{\{|X_1| \le a_n\}} \right] + \frac{1}{\varepsilon^2} \mathbf{E}\left[\sum_{i=N}^{\infty} X_1^2 \frac{1}{a_n} \mathbf{1}_{\{|X_1| \le a_n\}} \right] \\ &= A + B \end{aligned}$$

where

$$A = \frac{1}{\varepsilon^2} \mathbf{E}\left[\sum_{i=1}^{N-1} X_1^2 \frac{1}{a_n} \mathbf{1}_{\{|X_1| \le a_n\}} \right]$$

and

$$B = \frac{1}{\varepsilon^2} \mathbf{E}\left[\sum_{i=N}^{\infty} X_1^2 \frac{1}{a_n} \mathbf{1}_{\{|X_1| \le a_n\}} \right]$$

For the A term, since it is a finite sum (less than $N-1$), the sum is definitely finite.
Now for the B term, we have the following observation, with the choice of N ($X_1 < \alpha^N$) we picked earlier:

$$\sum_{n=N}^{\infty} \frac{1}{a_n} \le \sum_{n=N}^{\infty} \frac{1}{\alpha^n} = \frac{1}{\alpha^N} \frac{1}{1 - \frac{1}{\alpha}} \le \frac{1}{X_1} \frac{1}{1 - \frac{1}{\alpha}} = K \frac{1}{X_1}$$

Putting this observation into B term:

$$B = \frac{1}{\varepsilon^2}\mathbf{E}\left[\sum_{i=N}^{\infty} X_1^2 \frac{1}{a_n} \mathbf{1}_{\{|X_1| \leq a_n\}}\right] \leq \frac{1}{\varepsilon^2}\mathbf{E}\left[X_1^2 K \frac{1}{X_1}\right] = \frac{1}{\varepsilon^2}\mathbf{E}[K X_1] < \infty$$

So together, the sum $A + B$ is finite, hence we showed:

$$\sum_{i=1}^{\infty} \mathbf{P}(|T_n| > \varepsilon) < \infty$$

And as a result (technique of infinitely often), for the subsequence a_n we have $\sum_{i=1}^{\infty} \mathbf{P}(|T_{a_n}| > \varepsilon) < \infty$, and therefore:

$$\frac{S_{a_n}}{a_n} \xrightarrow{as} 0$$

(b) Now for the intermediate k's with $a_n \leq k \leq a_{n+1}$:

$$\frac{a_n}{a_{n+1}} \frac{S_{a_n}}{a_n} \leq \frac{1}{a_{n+1}} S_k \leq \frac{S_k}{k} \leq \frac{S_{a_{n+1}}}{k} \leq \frac{S_{a_{n+1}}}{a_{n+1}} \frac{a_{n+1}}{a_n}$$

Then note that $\lim_{n \to \infty} a_n = 1$, and the fraction $\frac{a_n}{a_{n+1}} \to 1$ (and also the fraction $\frac{a_{n+1}}{a_n} \to 1$), so the above is now:

$$\frac{S_{a_n}}{a_n} \leq \frac{S_k}{k} \leq \frac{S_{a_{n+1}}}{a_{n+1}}$$

And taking $n \to \infty$, the two quantities $\frac{S_{a_n}}{a_n} \xrightarrow{as} 0$ and $\frac{S_{a_{n+1}}}{a_{n+1}} \xrightarrow{as} 0$, therefore, squeezing the middle must get: $\frac{S_k}{k} \xrightarrow{as} 0$.

Hence it is true for all the n. And we completed the proof. □

10.2 Applications

10.2.1 Monte-Carlo Integration

We can use uniform random variables U_i on $[0, 1]$ so that:

$$\int_0^1 f(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} U_i$$

Why can we do this? Observe the following:

$$\mathbf{E}[f(U_i)] = \int_0^1 f(x) f_U(x) dx \text{ ,where } f_U \text{ is the pdf, and equal to } \mathbf{1}_{[0,1]} = \int_0^1 f(x)dx$$

By Law of Large Numbers:

$$\frac{1}{n}\sum_{i=1}^{n} f(U_i) \xrightarrow[p]{as} \int_0^1 f(x)dx$$

Note that we still need to check $\mathbf{E}[f(U_i)] < \infty$.

10.2.2 Higher dimensional cube is almost a boundary of a ball

Take uniform random variables X_i on $[-1, 1]$, then their corresponding pdf is

$$p_{X_i} = \frac{1}{2}\mathbf{1}_{[-1,1]}$$

We consider the $Y_i = X_i^2$, then:

$$\mathbf{E}[Y_i] = \mathbf{E}[X_i^2] = \int_{-1}^{1} x^2 p_{X_i}(x) dx = \int_{-1}^{1} x^2 \frac{1}{2} dx = \frac{1}{3}$$

By Law of Large Numbers:

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[p]{as} \frac{1}{3}$$

Equivalently:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow[p]{as} \frac{1}{3}$$

So: for any $\varepsilon > 0$. we get:

$$\mathbf{P}\left(|\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{3}| < \varepsilon\right) \to 1$$

That is:

$$\mathbf{P}\left(\frac{1}{3} - \varepsilon < \frac{1}{n}\sum_{i=1}^{n} X_i^2 < \frac{1}{3} + \varepsilon\right) \to 1$$

And so

$$\mathbf{P}\left(\sqrt{\frac{n}{3} - n\varepsilon} < \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2} < \sqrt{\frac{n}{3} + n\varepsilon}\right) \to 1$$

The square-root $\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$ term is the "length" in n dimensional-cube.

11 Day 11: Sequence of events and Tail events

11.1 Tail sigma algebra

Recall the sigma algebra generated by random variables definition:

1. For one random variable $X : (\mathbb{R}, \mathcal{B}(\mathbb{R}), P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

$$\begin{aligned} \sigma(X) &:= \sigma(\{X^{-1}(B)\}), \forall B \in \mathcal{B}(\mathbb{R}) \\ &= \sigma(\{X(\omega) \leq a\}), \forall a \in \mathbb{R} \end{aligned}$$

2. Now for multiple random variables: Let $X_i : (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_i) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for $i = 1, \ldots, n$. Then the sigma algebra is defined:

$$\begin{aligned} \sigma(X_1, \ldots, X_n) &= \sigma(\underbrace{A_1 \times \cdots \times A_n}_{\text{cylinder box for sets}}) \text{ where the } A_i \in \mathcal{B}(\mathbb{R}) \\ &= \sigma(X_1 \leq a_1, \ldots, X_n \leq a_n), \forall (a_1, \ldots, a_n) \in \mathbb{R}^n \end{aligned}$$

3. And the above sigma algebra is used to define for random vector $X = (X_1, \ldots, X_n)$ defined on (Ω, F, μ) where $\Omega = \mathbb{R}^n$, $F = \sigma(X_1, \ldots, X_n)$ and $\mu = \mu_1 \times \ldots \times \mu_n$.

We see that these are finite case n. We wish to extend the above definition to the infinite case n.

Theorem 145 (Kolmogorov's extenstion Theorem). *Suppose that for each $n \geq 1$ and each $k \geq 1$, we always have:*

$$\mu_{n+k}(A_1 \times \ldots \times A_n \times \mathbb{R}^k) = \mu_n(A_1 \times \ldots \times A_n)$$

Then there exists a unique probability measure μ on $(\mathbb{R}^\infty, F)$ with

$$F = \bigcup_{n=1}^{\infty} \{\text{cylider sets in } \mathbb{R}^n\}$$

satisfying, for all $n \geq 1$, and $B \in \mathcal{B}(\mathbb{R}^n)$:

$$\mu(B \times \mathbb{R} \times \mathbb{R} \times \ldots) = \mu_n(B)$$

In words, we "extend" from n dimension to n+ to infinite dimensions.

Definition 146 (Tail sigma algebra). Let X_n be random variables with $n = 1, 2, \ldots, \infty, \ldots$. We define first:

$$\tilde{F}_n := \sigma(X_n, X_{n+1}, \ldots,)$$

be the events that do not depend on the first $(n-1)$ random variables. Next, the tail sigma algebra is defined as

$$T = \bigcap_{n=1}^{\infty} \tilde{F}_n$$

the events that does not depend on a finite number of initial random variables, for arbitrary large number n.

Example 147. Here are some examples of tail sigma algebra. Denote $S_n = X_1 + X_2 + \ldots + X_n$. Then:

1. The set $\{\lim_{n\to\infty} S_n \text{ exists as a finite number}\} \in T$ because, if $\lim_{n\to\infty}(X_1 + X_2 + \ldots + X_n) < \infty$, then:

$$\tilde{X}_1 + \tilde{X}_2 + \ldots + X_n + X_{n+1} + \ldots + X_\infty = K$$

and this will still converges, therefore is in T. When we change a finite number of random variables does not affect the event.

2. The set $\{\lim_{n\to\infty} X_n \text{exists and} \lim_{n\to\infty} X_n < \infty\} \in T$.

3. The set $\{\lim_{n\to\infty} X_n \text{ exists and } \lim_{n\to\infty} S_n > 0\} \notin T$ with considering the $\{X_n\}_{n=1}^{\infty} = \{100, -1^2, -\frac{1}{2^2}, -\frac{1}{3^2}, \ldots\}$

4. The set $\{\lim_{n\to\infty} \frac{S_n}{n} = 0\} \in T$, same reason of the finite case.

5. The set $\{A_n \text{ i.o.}\} \in T$ for events A_n, because:

$$\{A_n \text{ i.o.}\} = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k$$

where the $\tilde{F}_n = \bigcup_{k \geq n} A_k$.

11.2 Kolmogorov 0-1 laws

Theorem 148 (Kolmogorov 0-1 laws)**.** *Let X_n be a sequence of independent random variables. Then for any event in the tail sigma algebra $A \in \mathcal{T}$, we have*

$$\boldsymbol{P}(A) = 0 \text{ OR } \boldsymbol{P}(A) = 1$$

Proof. The main idea is to show that A is independent to itself, that is

$$\mathbf{P}(A \cap A) = \mathbf{P}(A)\mathbf{P}(A)$$

So

$$\mathbf{P}(A) = \mathbf{P}(A)^2$$

And therefore $\mathbf{P}(A) = 0$ or $\mathbf{P}(A) = 1$. But how do we show A is independent to itself?

1. Step 1: We claim that for any $A \in \sigma(X_1, X_2, \ldots, X_n)$ and $B \in \sigma(X_{n+1}, \ldots)$, then A and B are independent. In fact, we will show that the 2 sigma algebras $\sigma(X_1, X_2, \ldots, X_n)$ and $\sigma(X_{n+1}, \ldots)$ are independent from definition. Take any $j \in \mathbb{N}$ with $j \geq 1$, then:

   $$\begin{aligned} &\mathbf{P}(X_1 \in B_1, X_2 \in B_2, \ldots, X_k \in B_k, X_{k+1} \in B_{k+1}, \ldots, X_{k+j} \in B_{k+j}) \\ &= \mathbf{P}(\{X_1 \in B, \ldots, X_k \in B\})\mathbf{P}(\{X_{k+1} \in B_{k+1}, \ldots, X_{k+j} \in B_{k+j}\}) \end{aligned}$$

   Therefore $\sigma(X_1, \ldots, X_k)$ and $\sigma(X_{k+1}, \ldots, X_{k+j})$ are independent. Then taking the union over j gives: $\sigma(X_1, \ldots, X_k)$ and $\bigcup_{j=1}^{\infty} \sigma(X_{k+1}, \ldots, X_{k+j}) = \sigma(X_{k+1}, X_{k+2}, \ldots)$ are independent.

2. Step 2: If $A \in \sigma(X_1, \ldots,)$ and $B \in \mathcal{T}$, then A and B are independent. Recall the definition of $\mathcal{T}$:

   $$\mathcal{T} = \bigcap_{k=1}^{\infty} \sigma(X_{k+1}, X_{k+2}, \ldots,)$$

   And also

   $$\sigma(X_1, \ldots,) = \bigcup_{k=1}^{\infty} \sigma(X_1, \ldots, X_k)$$

   For $A \in \sigma(X_1, \ldots,) = \bigcup_{k=1}^{\infty} \sigma(X_1, \ldots, X_k)$, there exists an $N < \infty$ such that

   $$A \in \sigma(X_1, \ldots, X_{N+1})$$

   In addition, $\mathcal{T} \subset \sigma(X_{N+1}, X_{N+2}, \ldots)$, so by step 1, A and $\mathcal{T}$ must be independent.

3. Step 3: Since $\mathcal{T} \subset \sigma(X_1, \ldots)$, step 2 gives any $A \in \mathcal{T}$, then $A \in \sigma(X_1, \ldots)$ and $A \in \mathcal{T}$, so by step 2, A will be independent to itself A.

□

Remark 149 (Comparision between Borel-Cantelli Lemma and Kolmogorov 0-1 law)**.** The Borel-Cantelli Lemma has an additional condition $\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$ or $= \infty$, then we can tell if the probability is 0 or 1.
Kolmogorov law does not require such additional condition, however, we cannot tell if the probability is 0 or 1.

Question 150. Can we tell if a tail event occur with probability 1 or 0? Answer: Use Kolmogorov's Maximal Inequality.

Theorem 151 (Kolmogorov's Maximal Inequality)**.** *Let X_n be independent random variables, and $S_n = X_1 + \dots + X_n$, and X_n are zero mean,i.e., $\mathbf{E}[X_i] = 0$ for all i, and X_n has finite variance, i.e., $Var(X_i) < \infty$ for all i. Then:*

$$\boldsymbol{P}(\max_{1 \le k \le n} |S_k| \ge a) \le \frac{Var(S_n)}{a^2}$$

Remark 152. Note that this is different than Chebychev Inequality. The Chebychev Inequality is

$$\mathbf{P}(|S_n| \ge a)$$

while in this Kolmogorov Inequality, we considered all partial sum:

$$\mathbf{P}\underbrace{(\max_{1 \le k \le n} |S_k| \ge a)}_{\text{all partial sums}}$$

Proof. Take $A_j = \{|S_j| \ge a$ but $|S_i| < a$ for all $1 \le i < j\}$. This means we break things down according to the time that $|S_j|$ first exceeds j. Note that A_j are disjoint and that $(S_n - S_j)^2 \ge 0$, we break the set into:

$$\begin{aligned}
\mathbf{E}[S_n^2] &\ge \sum_{j=1}^{n} \int_{A_j} S_n^2 d\mathbf{P} \\
&= \sum_{j=1}^{n} \int_{A_j} S_j^2 + 2S_j(S_n - S_j) + (S_n - S_j)^2 d\mathbf{P} \\
&\ge \sum_{j=1}^{n} \int_{A_j} S_j^2 d\mathbf{P} + \sum_{j=1}^{n} \int 2S_j \mathbf{1}_{A_j}(S_n - S_j) d\mathbf{P}
\end{aligned}$$

Since $S_j \mathbf{1}_{A_j} \in \sigma(X_1, \dots, X_k)$ and $S_n - S_k \in \sigma(X_{k+1}, \dots, X_n)$ are independent, and note that $\mathbf{E}[S_n - S_k] = 0$ we get:

$$\int 2S_j \mathbf{1}_{A_j}(S_n - S_j) d\mathbf{P} = \mathbf{E}[2S_j \mathbf{1}_{A_j}]\mathbf{E}[S_n - S_k] = 0$$

In summary we now have:

$$\begin{aligned}
\mathbf{E}[S_n^2] &\ge \sum_{j=1}^{n} \int_{A_j} S_j^2 d\mathbf{P} \\
&\ge \sum_{j=1}^{n} a^2 \mathbf{P}(A_j); \text{ on } A_j : S_j^2 \ge a^2 \\
&= a^2 \left(\max_{1 \le j \le n} |S_j| \ge a \right)
\end{aligned}$$

And dividing by a^2 gives the desired result. □

12 Day 12: Central Limit Theorem: beginning

12.1 Central Limit Theorem - the statement

Theorem 153 (Central Limit Theorem). *Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d, and $\mu = \mathbf{E}[X_1]$, $\sigma^2 = Var(X_1) < \infty$, and $S_n = X_1 + \dots + X_n$. Then:*

$$\boldsymbol{P}\left(a \le \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \le b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \Phi(b) - \Phi(a)$$

where the $\Phi(x)$ is the cdf (cumulative distribution) of standard normal random variable. Equivalently, in the LHS, the probability can be written as:

$$\boldsymbol{P}\left(a\frac{\sigma}{\sqrt{n}} \le \frac{S_n}{n} - \mu \le b\frac{\sigma}{\sqrt{n}}\right)$$

And we may write:

$$\boldsymbol{P}\left(a\frac{\sigma}{\sqrt{n}} \le \frac{S_n}{n} - \mu \le b\frac{\sigma}{\sqrt{n}}\right) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \Phi(b) - \Phi(a)$$

Recall some definitions for **normal distribution random variable** with parameters μ and σ:

1. the pdf
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
2. the cdf
$$F_X(x) = \int_{-\infty}^{x} f_x(y)dy = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

And with standard normal/Gaussian random variable, $\mu = 0$, $\sigma = 1$:

1. the pdf is
$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$
2. the cdf is:
$$\Phi_Z(z) = \int_{-\infty}^{z} f_Z(y)dy = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

Example 154 (A pictorial example). Consider X_n with $\mathbf{P}(X_n = 1) = \mathbf{P}(X_n = 0) = \frac{1}{2}$: coin flipping. Then $\mathbf{E}[X_n] = \frac{1}{2}$ and $\mathrm{Var}(X_n) = \frac{1}{4}$. Then, by Law of Large Numbers:

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \dots + X_n}{n} \overset{a.s.}{\underset{p}{\to}} \frac{1}{2} = \mathbf{E}[X_n]$$

See figure 18 for a pictorial increasing numbers of n.

Remark 155. Recall in the **Monte-Carlo Simulation**, the converge with $O(\frac{1}{\sqrt{n}})$ come from CLT above.

Remark 156. In the CLT, the mode of convergence is convergence **in distribution**. It might be a good idea to remind the definition of convergence in distribution.

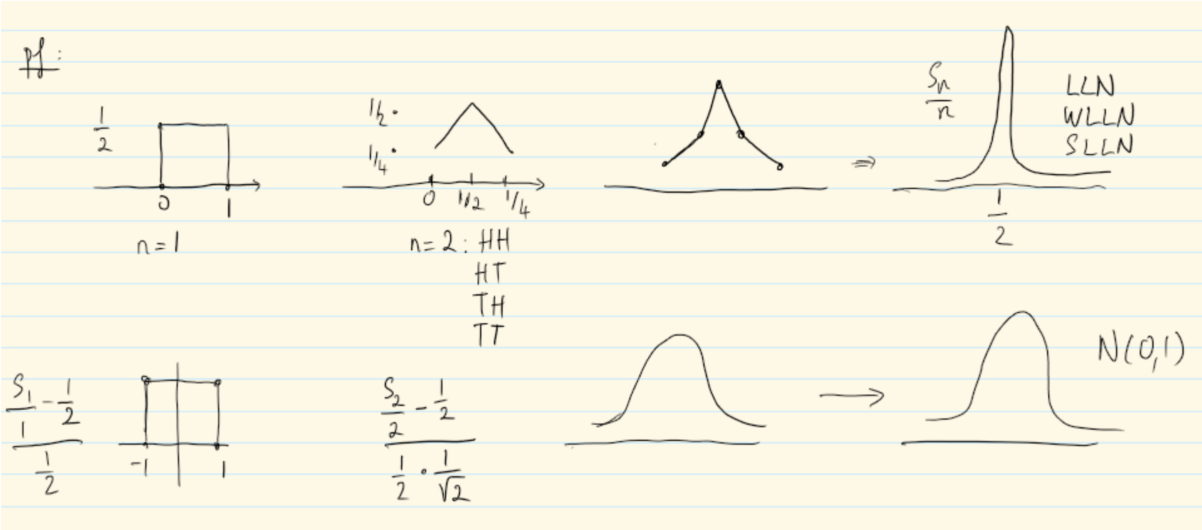


Figure 18: Pictorial example CLT

12.2 two definitions of convergence in distribution is equivalent

Here are the 2 definitions of convergence in distribution:

1. The first definition is a reminder from Definition 73:

   Definition 157 (First definition of convergence in distribution)**.** We say that X_n **converges in distribution** to X, denoted $X_n \xrightarrow{d} X$ if the cdf F_{X_n} and F_X satisfy:

   $$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

   at all x such that F is continuous.

2. The second definition is similar to the Remark right after the definition. It is also the definition of the **weakly convergence:**

   Definition 158. For any bounded and continuous function $h(x) : \mathbb{R} \to \mathbb{R}$, the following is true:

   $$\mathbf{E}[f(X_n)] \to \mathbf{E}[f(X)].$$

   where the convergence in the above is almost surely convergence.

It is crucial to see that the two definitions are indeed equivalent:

Theorem 159 (equivalence of 2 definitions of convergence in distribution - Theorem 3.2.9)**.** *Let X_n be a sequence of random variables. $F_{X_n}(t) \to F_X(t) \iff \mathbf{E}[h(X_n)] \to \mathbf{E}[h(X)]$, for any bounded continuous function $h : \mathbb{R} \to \mathbb{R}$.*

Proof. For this direction (1) $\Rightarrow$ (2): We will need a lemma:

Lemma 160. X_n are random variables such that $X_n \xrightarrow{d} X$, then there exists Y_n and Y satisfying:

1. Y_n has same distribution as X_n.

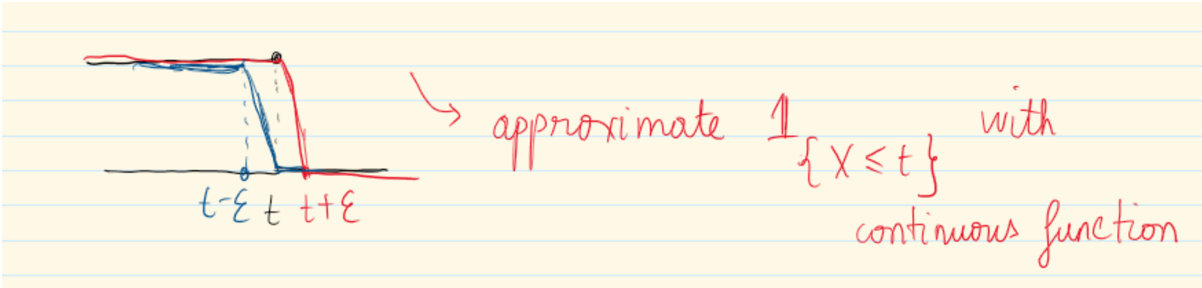


Figure 19: approximation of $\mathbf{1}_{\{X \le t\}}$ by $g_{t,\varepsilon}(x)$

2. Y has same distribution as X
3. $Y_n \xrightarrow{a.s.} X$

This Lemma can be shown by setting $Y(z) := \sup\{y : \mathbf{P}(X \le y) < z\}$. Now assume we can use this lemma. Then we have Y_n and Y as in the conclusion of the lemma, that is $Y_n \overset{d}{=} X_n$, and $Y \overset{d}{=} X$ and that $Y_n \xrightarrow{a.s.} Y$. Let h be a bounded real valued continuous function. We apply DCT to $h(Y_n) \xrightarrow{a.s.} h(Y)$ (since continuity preserve convergence a.s.) and $|h| \le M < \infty$:

$$\begin{aligned}
\lim_{n \to \infty} \mathbf{E}[h(Y_n)] &\overset{DCT}{=} \mathbf{E}[\lim_{n \to \infty} h(Y_n)] \\
&= \mathbf{E}[h(\lim_{n \to \infty} Y_n)] \text{ by continuity of } h \\
&= \mathbf{E}[h(Y)]; \text{ by Lemma } Y_n \xrightarrow{a.s.} Y \\
&= \mathbf{E}[h(X)], \text{ since } Y \text{ and } X \text{ agree in distribution}
\end{aligned}$$

And note that since X_n and Y_n agree in distribution, we also have $\mathbf{E}[h(Y_n)] = \mathbf{E}[h(X_n)]$ (so we get $\lim_{n \to \infty} \mathbf{E}[h(Y_n)] = \lim_{n \to \infty} \mathbf{E}[h(X_n)]$). Therefore we concluded:

$$\lim_{n \to \infty} \mathbf{E}[h(X_n)] = \mathbf{E}[h(X)]$$

For the other direction $\Leftarrow$, we suppose $\mathbf{E}[h(X_n)] \to \mathbf{E}[h(X)]$, for any bounded continuous function $h : \mathbb{R} \to \mathbb{R}$. Then by definition:

$$F_X(t) = \mathbf{P}(X \le t) = \mathbf{E}[X\mathbf{1}_{\{X \le t\}}]$$

The general strategy now will be approximating $\mathbf{1}_{\{X \le t\}}$ by continuous function $g_{t,\varepsilon}(x)$ defined below (see figure 19):

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

Then for the $X_n \xrightarrow{a.s.} X$, we get:

$$\mathbf{E}[X_n g_{t,\varepsilon}] \xrightarrow{a.s.} \mathbf{E}[X g_{t,\varepsilon}]$$

And therefore:

$$\begin{aligned}
F_{X_n}(t) &= \mathbf{E}[X_n \mathbf{1}_{\{X_n \le t\}}] \\
&\le \mathbf{E}[X_n g_{t,\varepsilon}] \le \mathbf{E}[X_n \mathbf{1}_{\{X_n \le t+\varepsilon\}}] = F_{X_n}(t + \varepsilon)
\end{aligned}$$

In short we have:

$$F_{X_n}(t) \leq F_{X_n}(t+\varepsilon) \tag{11}$$

A similar argument gives:

$$F_{X_n}(t-\varepsilon) \leq \mathbf{E}[X_n g_{t,\varepsilon}] \leq F_{X_n}(t) \tag{12}$$

In 11 we take lim sup (and note that since $\mathbf{E}[X_n g_{t,\varepsilon}] \xrightarrow{as} \mathbf{E}[X g_{t,\varepsilon}]$, we deduce $\limsup_{n\to\infty} \mathbf{E}[X_n g_{t,\varepsilon}] = \lim_{n\to\infty} \mathbf{E}[X_n g_{t,\varepsilon}] = \mathbf{E}[X g_{t,\varepsilon}]$) and get:

$$\limsup_{n\to\infty} F_{X_n}(t) \leq \lim_{n\to\infty} \mathbf{E}[X_n g_{t,\varepsilon}] = \mathbf{E}[X g_{t,\varepsilon}] \leq \limsup_{n\to\infty} F_{X_n}(t+\varepsilon) \tag{13}$$

Similarly, in 12 we take lim inf and get:

$$\lim_{n\to\infty} F_{X_n}(t-\varepsilon) \leq \lim_{n\to\infty} \mathbf{E}[X_n g_{t,\varepsilon}] = \mathbf{E}[X g_{t,\varepsilon}] \leq \liminf_{n\to\infty} F_{X_n}(t) \tag{14}$$

Combining 13 and 14 we get:

$$\mathbf{E}[X g_{t,\varepsilon}] \leq \liminf_{n\to\infty} F_{X_n}(t) \leq \limsup_{n\to\infty} F_{X_n}(t) \leq \mathbf{E}[X g_{t,\varepsilon}]$$

Letting $\varepsilon \downarrow 0$, we get equality must happen and therefore the limit converges to $F_X(t)$:

$$\liminf_{n\to\infty} F_{X_n}(t) = \limsup_{n\to\infty} F_{X_n}(t) = \lim_{n\to\infty} F_{X_n}(t) = \mathbf{E}[X \mathbf{1}_{\{X \leq t\}}] = F_X(t)$$

And this concluded the proof. □

12.3 examples of convergence in distribution

Now let's take a look at reviewing some example of convergence in distribution:

Example 161. Let's take a look at several examples:

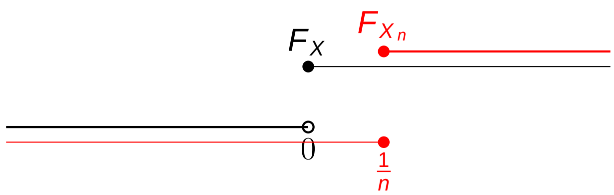
1. Here is a simple example of convergence in distribution. Let X be a random variable such that $\mathbf{P}(X = c) = 1$. Then:

$$F_X(t) = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}$$

   Take random variables X_n such that $\mathbf{P}(X_n = \frac{1}{n}) = 1$, then:

$$F_{X_n}(t) = \begin{cases} 1 & \text{if } t \geq \frac{1}{n} \\ 0 & \text{if } t < \frac{1}{n} \end{cases}$$

   Then for $t \neq 0$, we see that $F_{X_n}(t) \xrightarrow{as} F_X(t)$ and therefore $X_n \xrightarrow{d} X$.



   Note that at $t = 0$, $F_X(0) = 1$, but $F_{X_n}(0) = 0$ for all n. So $\lim_{n\to\infty} F_{X_n}(0) \neq F_X(0)$.

2. Here is an **important example**: binomial random variables $\xrightarrow{d}$ Poisson random variable. Recall the

   (a) binomial with n trials is:

   $$\mathbf{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}; k = 1; 2; \dots; n$$

   (When $n \to \infty$, setting $np = \lambda$, that is $n = \frac{\lambda}{p}$, we will see:)

   (b) Poisson random variable with parameter λ:

   $$\mathbf{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

   Now how can we show $X_n \xrightarrow{d} X$? The corresponding distribution functions:

   $$F_{X_n}(t) \to F_X(t)$$

   which are:

   $$\sum_{i \geq 0}^{\lfloor t \rfloor} \mathbf{P}(X_n = i) \to \sum_{i \geq 0}^{\lfloor t \rfloor} \mathbf{P}(X = i)$$

   In other words, we can try to show:

   $$\mathbf{P}(X_n = k) \to \mathbf{P}(X = k)$$

   We will show the above by many computations, let's start with the definition:

   $$\begin{aligned} \mathbf{P}(X_n = k) &= \frac{n(n-1)\dots(n-k+1)}{k!} p^k (1-p)^n (1-p)^{-k} \\ &= \frac{n(n-1)\dots(n-k+1)}{k!} \frac{\lambda^k}{n^k} (1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-k}, \text{ letting } np = \lambda; \text{ i.e. } p = \frac{\lambda}{n} \end{aligned}$$

   In the above, we will keep the terms $\frac{\lambda^k}{k!}$, and let's deal with the other terms as followed:

   (a) the term $(1 - \frac{\lambda}{n})^{-k}$. When $n \to \infty$, this term goes to 1.

   (b) the term $\frac{n(n-1)\dots(n-k+1)}{n^k}$, this term also goes to 1 when $n \to \infty$.

   (c) the term $(1 - \frac{\lambda}{n})^n$. We claim that when $n \to \infty$:

   $$\lim_{n \to \infty} (1 - \frac{\lambda}{n})^n = e^{-\lambda}$$

   The claim follows from this Lemma:

   Lemma 162. If $c_j \to 0$ and $a_j \to \infty$ with $c_j a_j = \lambda$, then:

   $$(1 + c_j)^{a_j} \to e^{\lambda}$$

Proof of lemma: Recall from Calculus L'Hospital rule:

$$\lim_{x \to 0} \frac{\log(1+x)}{x} = \lim_{x \to 0} \frac{\frac{1}{1+x}}{1} = 1$$

Here we take $x = \frac{\lambda}{a_j}$, then:

$$\lim_{c_j \to 0} \frac{\log(1+c_j)}{\lambda / a_j} = 1$$

And so:

$$\log(1+c_j)^{a_j} \to \lambda$$

And raising exponent:

$$(1+c_j)^{a_j} \to e^{\lambda}$$

This completed the lemma proof.

We will now back to our original probability, using all the information above:

$$\lim_{n \to \infty} \mathbf{P}(X_n = k) = \lim_{n \to \infty} (1)(1)\frac{\lambda^k}{k!}(1 - \frac{\lambda}{n})^n = \frac{\lambda^k}{k!}e^{-\lambda} = \mathbf{P}(X = k)$$

3. Another example: We also take the X_i to be i.i.d with $\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = -1) = \frac{1}{2}$. Also $\mathbf{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = 1$. Letting $S_n = X_1 + \dots + X_n$. When we choose $X_i = 1$ for $n+k$ choices and $X_i = -1$ for another $(n-k)$ choices, we get:

$$\mathbf{P}(S_{2n} = 2k) = \frac{(2n)!}{(n+k)!(n-k)!}\left(\frac{1}{2}\right)^{2n}$$

Applying **Stirlings formula** $n! \sim n^n e^{-n}\sqrt{2\pi n}$ when $n \to \infty$ to $(2n)!$ and $(n+k)!$ and $(n-k)!$ in the above, we get:

$$\begin{aligned}
\mathbf{P}(S_{2n} = 2k) &= \frac{(2n)^{2n}e^{-2n}\sqrt{2\pi 2n}}{(n+k)^{n+k}e^{-n-k}\sqrt{2\pi(n+k)}(n-k)^{n-k}e^{-n+k}\sqrt{2\pi(n-k)}}\frac{1}{2^{2n}} \\
&= \left(\frac{n}{n+k}\right)^{n+k}\left(\frac{n}{n-k}\right)^{n-k}\frac{\sqrt{2\pi 2n}}{\sqrt{2\pi(n+k)}\sqrt{2\pi(n-k)}}\frac{\sqrt{n}}{\sqrt{n}} \\
&= \left(1+\frac{k}{n}\right)^{-n-k}\left(1-\frac{k}{n}\right)^{-n+k}\frac{\sqrt{2}}{\sqrt{2\pi}\sqrt{n}}\left(1+\frac{k}{n}\right)^{-\frac{1}{2}}\left(1-\frac{k}{n}\right)^{-\frac{1}{2}} \\
&= \left(1-\frac{k^2}{n^2}\right)^{-n}\frac{1}{\sqrt{\pi n}}\left(1+\frac{k}{n}\right)^{-k}\left(1-\frac{k}{n}\right)^{k}\left(1-\frac{k^2}{n^2}\right)^{-\frac{1}{2}}
\end{aligned}$$

In the above, leaving only the $\frac{1}{\sqrt{\pi n}}$ alone, and also note that the term $\left(1-\frac{k^2}{n^2}\right)^{-\frac{1}{2}}$ goes to 1 when n goes to infinity. We also let $2k = x\sqrt{2n}$, i.e, $K = \frac{x\sqrt{n}}{\sqrt{2}}$:

(a) The term $\left(1-\frac{k^2}{n^2}\right)^{-n}$:

$$\left(1-\frac{k^2}{n^2}\right)^{-n} = \left(1-\frac{x^2}{2n}\right)^{-n} \to e^{\frac{x^2}{2}}$$

(b) The term $\left(1+\frac{k}{n}\right)^{-k}$:

$$\left(1+\frac{k}{n}\right)^{-k} \to e^{-\frac{k^2}{n}} = e^{-\frac{x^2}{2}}$$

(c) The term $\left(1-\frac{k}{n}\right)^{k}$:

$$\left(1-\frac{k}{n}\right)^{k} \to e^{-\frac{k^2}{n}} = e^{-\frac{x^2}{2}}$$

Multiplying all terms together we get:

$$\mathbf{P}(S_{2n}=2k) = e^{-\frac{x^2}{2}}\frac{1}{\sqrt{\pi n}}$$

Now we changed the probability a little bit:

$$\mathbf{P}(a\sqrt{2n} \le S_n \le b\sqrt{2n}) = \sum_{m\in[a\sqrt{2n},b\sqrt{2n}]\cap 2\mathbb{Z}} \mathbf{P}(S_{2n}=m)$$

letting $x=\frac{m}{\sqrt{2n}}$, so $m=2k=x\sqrt{2n}$ we get the above summation of probability is

$$\begin{aligned}\mathbf{P}(a\sqrt{2n} \le S_n \le b\sqrt{2n}) &= \sum_{m\in[a\sqrt{2n},b\sqrt{2n}]\cap 2\mathbb{Z}} \mathbf{P}(S_{2n}=m)\\ &= \sum_{x\in[a,b]\cap\frac{2\mathbb{Z}}{\sqrt{2n}}} \mathbf{P}(S_{2n}=x\sqrt{2n})\end{aligned}$$

Taking limit $n\to\infty$, the right sum will get to be the integral from a to b, and use the

$$\mathbf{P}(S_{2n}=2k) = e^{-\frac{x^2}{2}}\frac{1}{\sqrt{\pi n}}$$

we get:

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

13 Day 13: More on Central Limit Theorem- proof strategies

To be able to prove CLT, we need 2 important concepts: **Tightness** and **Characteristic Function**.

13.1 Tightness

Question 163. The main question here is that can we extract a converging subsequence from a sequence of distribution function $F_n(x)$?

Theorem 164 (Helly's Selection Theorem -Theorem 3.2.12)**.** *For any sequence of distribution F_n, there exists a subsequence n_k such that $\lim_{n\to\infty} F_{n_k}(x)=F(x)$ for all continuity points x of F, and F is non-decreasing and right continuous.*

Proof. The strategy is to construct the F at $\mathbb{Q}$ and then extend it to be right continuous. First, reorder all the rational numbers:

$$q_1 < q_2 < \dots$$

Then we follow this procedure (diagonalize):

1. For $F_n(q_1) \in [0,1]$ and for all n, this is a bounded sequence, so we can extract a converging subsequence n_k such that

$$\lim_{k\to\infty} F_{1,n_k}(q_1) = G(q_1)$$

2. For $F_{n_k}(q_2) \in [0,1]$ and for all n_k, this is a bounded sequence, so we can extract a converging subsequence n_k^2 such that

$$\lim_{k\to\infty} F_{2,n_k}(q_1) = G(q_2)$$

3. continue in this manner

4. We can define $G(q_i)$ for all $q_i \in \mathbb{Q}$.

5. Now we define:

$$F(x) := \inf\{G(q_i) : x \le q_i\}$$

Now we will show that F is non-decreasing: For $x < y$, the set:

$$\{G(q) : q \le x\} \subseteq \{G(q) : q \le y\}$$

And therefore

$$F(x) = \inf\{G(q) : q \le x\} \le \inf\{G(q) : q \le y\} = F(y)$$

Next we show that it is right-continuous:

$$\begin{aligned}\lim_{x_n \downarrow x} F(x_n) &= \inf\{G(q) : q > x_n\} \\ &= \inf\{G(q) : q > x\}, \text{ for large enough } n : \{q \le x_n\} = \{q \le x\} \\ &= F(x)\end{aligned}$$

This completed the proof. □

Remark 165. The $F(x)$ found in the proof of Helly Selection theorem (164) may **not** be a distribution. There are 2 additional conditions required for F to be a distribution, they are $\lim_{x\to 0} F(x) \stackrel{?}{=} 0$ and the condition $\lim_{x\to\infty} F(x) \stackrel{?}{=} 1$.

Example 166. Here is an example that the F_X is not a distribution. The **point-mass** at $X_n = n$: $\mathbf{P}(X_n = n) = 1$ have the distributions:

$$F_{X_n}(x) = \begin{cases} 1 & \text{if } x \geq n \\ 0 & \text{if } x < n \end{cases}$$

But $\lim_{n \to \infty} F_{X_n}(x) = 0$, and clearly not a distribution.

In order for F_X to be a distribution, we require the condition **tightness** on the sequence F_n.

Definition 167 (tightness)**.** A sequence X_n is **tight** if for any $\epsilon > 0$, there exists an M_ϵ (independent of n) such that:

$$\mathbf{P}(|X_n| \leq M_\epsilon) > 1 - \epsilon$$

Or equivalently:

$$\mathbf{P}(|X_n| > M_\epsilon) < \epsilon$$

In terms of distribution F_{X_n}:

$$F_{X_n}(M_\epsilon) - F_{X_n}(-M_\epsilon) > 1 - \epsilon$$

Example 168. The point-mass at n sequence $\mathbf{P}(X_n = n) = 1$ earlier is **not tight**. This is because if we fix an $\epsilon > 0$, and for large enough N, we always have $\mathbf{P}(|X_n| \leq N) = 0$. So we cannot find any M_ϵ that is independent of n.

Example 169. Here is an example of tightness. Consider the pdf of X_n:

$$f_{X_n}(x) = \begin{cases} ne^{-nx} & \text{if } x \geq 0 \\ 0 & x < 0 \end{cases}$$

See figure(20). For any fixed $\epsilon > 0$ (small), we can:

$$\begin{aligned} \mathbf{P}(|X_n| > M_\epsilon) &= \int_{M_\epsilon}^{\infty} ne^{-nx} dx \\ &= -e^{nx}\Big|_{M_\epsilon}^{\epsilon} = e^{-nM_\epsilon} < \infty \end{aligned}$$

In particular, it is always the case (biggest when $n = 1$):

$$e^{-nM_\epsilon} \leq e^{-M_\epsilon}$$

So if we set

$$e^{-M_\epsilon} < \epsilon$$

equivalently:

$$-M_\epsilon < \log(\epsilon)$$

Then we found our M_ϵ that is independent of n:

$$M_\epsilon > -\log(\epsilon)$$

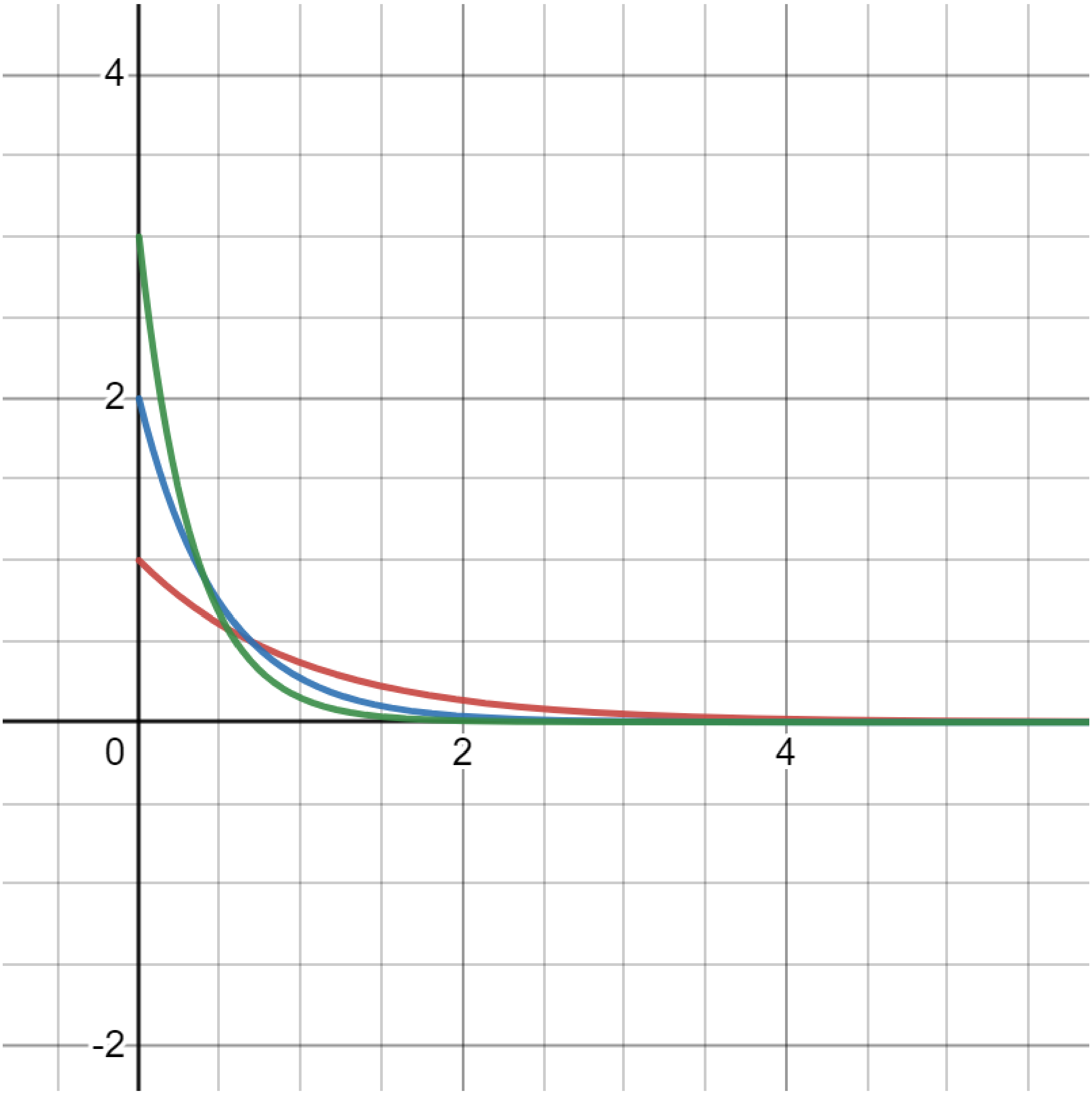


Figure 20: tightness example

Theorem 170 (Poronar Theorem). *If $F_n(x)$ a tight sequence of distribution, then $F(x)$ found in Helly's Selection Theorem (164) is a distribution.*

Proof. For any $\varepsilon > 0$, we can find M_ε such that:

$$F_{X_n}(M_\varepsilon) - F_{X_n}(-M_\varepsilon) > 1 - \varepsilon$$

Then we get:

1. taking limit to infinity:

$$\lim_{x \to \infty} F(x) \geq F(M_\varepsilon) \geq F(M_\varepsilon) - F(-M_\varepsilon) > 1 - \varepsilon$$

   Letting $\varepsilon \downarrow 0$ we get

$$\lim_{x \to \infty} F(x) = 1$$

2. taking limit to negative infinity:

$$\lim_{x \to -\infty} F(x) \leq F(-M_\varepsilon) < \varepsilon$$

   Letting $\varepsilon \downarrow 0$ we get

$$\lim_{x \to -\infty} F(x) = 0$$

   This completed the proof.

□

13.2 Characteristic Function

Here is an important function that will help us in the proof of CLT: the **Characteristic function**.

Definition 171 (Characteristic function). The **Characteristic function** of a random variable X (and density μ_X) is:

$$\begin{aligned}\varphi_X(t) := \mathbf{E}[e^{itX}] &= \int_{\mathbb{R}} e^{itx}\,\mu_X(dx) \\ &= \mathbf{E}[\cos(tX) + i\sin(tX)]\end{aligned}$$

Example 172 (coin flip characteristic function). $\mathbf{P}(X = 1) = \mathbf{P}(X = -1) = \frac{1}{2}$, then:

$$\begin{aligned}\varphi(t) &= e^{it}\frac{1}{2} + e^{-it}\frac{1}{2} \\ &= (\cos t + i\sin t)\frac{1}{2} + (\cos(-t) + i\sin(-t))\frac{1}{2} \\ &= \cos t; \text{ since } \cos(t) = \cos(-t)\end{aligned}$$

Example 173 (uniform random variable characteristic function)**.** Let X be a uniform random variable on $[0, h]$, that is $\rho_X(x) = \frac{1}{h}\mathbf{1}_{[0,h]}(x)$. Then the characteristic function is:

$$
\begin{aligned}
\varphi(t) = \mathbf{E}[e^{itX}] &= \int_0^h e^{itx}\frac{1}{h}dx \\
&= \int_0^h (\cos(tx) + i\sin(tx))\frac{1}{h}dx \\
&= \frac{1}{h}\left[\frac{1}{t}\sin(tx) - \frac{1}{h}\cos(tx)\right]\Big|_0^h \\
&= \frac{1}{h}\left[\frac{1}{t}\sin(th) - \frac{1}{h}\cos(th)\right] - \frac{1}{h}\left[\frac{1}{t}\sin(0) - \frac{1}{h}\cos(0)\right] \\
&= \frac{1}{iht}[i\sin(th) + \cos(th)] - \frac{1}{iht} \\
&= \frac{1}{iht}[e^{iht} - 1]
\end{aligned}
$$

Example 174 (Standard Normal Distribution Characteristic Function)**.** $N(0, 1)$ with $\rho_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Then the characteristic function is:

$$
\begin{aligned}
\varphi(t) &= \int_{\mathbb{R}} e^{itx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}(\cos(tx) + i\sin(tx))e^{-\frac{x^2}{2}}dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\cos(tx)e^{-\frac{x^2}{2}}dx; \text{ by symmetry, } \int_{\mathbb{R}}\sin(tx)e^{-\frac{x^2}{2}}dx = 0 \\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} -x\sin(tx)e^{-\frac{x^2}{2}}dx; \text{ using integration by part 1 time where } u = e^{-\frac{x^2}{2}}; dv = \cos(tx)dx \\
&= -\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} t\cos(tx)e^{-\frac{x^2}{2}}dx; \text{ integration by part again, with } v = -x\sin(tx); du = e^{-\frac{x^2}{2}}dx
\end{aligned}
$$

So $\varphi'(t) = -t\varphi(t)$ and $\varphi(0) = 1$, and this gives $\varphi(t) = e^{-\frac{t^2}{2}}$.

13.3 Properties of Characteristic Function

Theorem 175 (Properties of Characteristic function - Theorem 3.3.1)**.**

1. $\varphi(0) = 1$
2. $\varphi(-t) = \overline{\varphi(t)}$
3. $|\varphi(t)| \le 1$ *for all* t.
4. $\varphi(t)$ *is uniformly continuous*
5. $\mathbf{E}[e^{it(aX+b)}] = e^{ibt}\varphi(at)$

Proof. We sketch some of the proof:

1. This is true because:

$$\varphi(0) = \mathbf{E}[e^{i0X}] = \mathbf{E}[\cos(0) + i\sin(0)] = \mathbf{E}[1] = 1$$

2. This is true because:

$$\varphi(-t) = \mathbf{E}[\cos(-tx) + i\sin(-tx)] = \ldots = \overline{\varphi(t)}$$

3. $|\varphi(t)| \leq 1$ for all t. This is true because:

$$|\varphi(t)| = |\mathbf{E}[e^{itX}]| \leq \mathbf{E}\left[\left|e^{itX}\right|\right] \leq \mathbf{E}[1] = 1$$

4. $\varphi(t)$ is uniformly continuous because:

$$\begin{aligned} |\varphi(t+h) - \varphi(t)| &= |\mathbf{E}[e^{i(t+h)X}] - \mathbf{E}[e^{itX}]| \\ &= \left|\mathbf{E}[e^{i(t+h)X} - e^{itX}]\right| \\ &= \left|\mathbf{E}[e^{itX}(e^{ihX} - 1)]\right| \\ &\leq \mathbf{E}\left[|e^{itX}|.|e^{ihX} - 1|\right] \\ &= \mathbf{E}[|e^{ihX} - 1|], \text{ since } |e^{itX}| = 1 \end{aligned}$$

And because $\mathbf{E}[|e^{ihX} - 1|]$ does not depend on t, so we get $\varphi(t)$ is uniformly continuous.

□

For the last property, we have an interesting application:

Example 176. The char function of $X = N(\mu, \sigma^2)$. Note that $X = \mu + \sigma Z$ where Z is $N(0, 1)$. Then:

$$\varphi_X(t) = \varphi(e^{it(\mu+\sigma Z)}) = e^{i\mu t}\varphi(\sigma t) = e^{i\mu t}e^{-\frac{\sigma^2 t^2}{2}} = e^{i\mu t - \frac{\sigma^2 t^2}{2}}$$

Next Ultimate Goal: If the char function equal, then the distribution equal. This is the content of the Uniqueness Theorem.
To achieve this goal, we have to proceed with many steps.

14 Day 14: Yet more on on Central Limit Theorem

14.1 The ultimate goal preparation

We begin with lots of preparation for the Ultimate goal, starting with some useful identities and Lemmas.

Theorem 177 (Parseval Identity)**.** *If μ_X, μ_Y are probability measures, and $\varphi_X(t)$ and $\varphi_Y(t)$ are char functions. Then:*

$$\int_{\mathbb{R}} e^{-ity}\varphi_X(t)\,\mu_Y(dt) = \int_{\mathbb{R}} \varphi_Y(x-y)\,\mu_X(dx)$$

This identity can be thought of as "change of variable" identity for integrals.

Proof. We use Fubini's theorem:

$$\begin{aligned}\int_{\mathbb{R}} e^{-ity}\varphi_X(t)\,\mu_Y(dt) &= \int_{\mathbb{R}} e^{-ity}\int_{\mathbb{R}} e^{itx}\,\mu_X(dx)\,\mu_Y(dy)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{it(x-y)}\,\mu_Y(dt)\,\mu_X(dt)\\ &= \int_{\mathbb{R}} \varphi_Y(x-y)\,\mu_X(dx)\end{aligned}$$

where the interchange of integrals were done using Fubini Theorem. □

Example 178. An application of Parseval Identity (177) when $Y \sim N(0, \frac{1}{a^2})$. Then:

$$d\mu_Y(t) = \frac{a}{\sqrt{2\pi}}e^{-\frac{a^2t^2}{2}}dt$$

We also need to find $\varphi_Y(t)$. Recall when $X \sim N(0,1)$, we get $\varphi_X(t) = e^{-\frac{t^2}{2}}$. Using property (5) of char function, we have when $X \sim N(\mu, \sigma^2)$:

$$\varphi_X(t) = e^{i\mu t - \frac{\sigma^2t^2}{2}}$$

Here with Y that have $\mu = 0$ and $\sigma = \frac{1}{a}$:

$$\varphi_Y(t) = e^{-\frac{a^2t^2}{2a^2}}$$

So by Parselval identity:

$$\int e^{-ity}\frac{a}{\sqrt{2\pi}}e^{-\frac{a^2t^2}{2}}\varphi_X(t)dt = \int e^{-\frac{(x-y)^2}{2a^2}}d\mu_X(x)$$

Mutiplying by $\frac{1}{\sqrt{2\pi}a}$ we get (note the cancellation of a and no more square-root in the constant term $\frac{1}{2\pi}$ on the LHS):

$$\frac{1}{2\pi}\int e^{-ity-\frac{a^2t^2}{2}}\varphi_X(t)dt = \int \frac{1}{\sqrt{2\pi}a}e^{-\frac{(x-y)^2}{2a^2}}d\mu_X(x)$$

Lemma 179 (a useful lemma for density of linear combination of random variables and normal distribution)**.** If X is a random variable with μ_X be its associated probability measure, and Z is $N(0,1)$, and X and Z is independent. Then the density of $X + aZ$, for any $a \in \mathbb{R}$ is:

$$\mathbf{P}(X + aZ \in B) = \int_B\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}a}e^{-\frac{(y-x)^2}{2a^2}}dxdy$$

Proof. The proof relies on Fubini Theorem again:

$$\begin{aligned}\mathbf{P}(X + aZ \in B) &= \int\int \mathbf{1}_{\{X+aZ\in B\}}\,\mu_X\otimes\mu_Z(dx\ dz)\\ &= \int\int \mathbf{1}_{\{aZ\in B-X\}}\,\mu_X\otimes\mu_Z(dx\ dz)\end{aligned}$$

Recall the pdf of aZ is $\frac{1}{\sqrt{2\pi}a}e^{-\frac{z^2}{2a^2}}$, so the above now become:

$$\mathbf{P}(X + aZ \in B) = \int_B\int \frac{1}{\sqrt{2\pi}a}e^{-\frac{(x-y)^2}{2a^2}}d\mu_X(x)dy$$

which completed the proof. □

Remark 180. In the above lemma, if we let $a \downarrow 0$, we get:

$$\mathbf{P}(X \in B) = \lim_{a\downarrow 0}\mathbf{P}(X + aZ \in B) = \lim_{a\downarrow 0}\int_B\int \frac{1}{\sqrt{2\pi}a}e^{-\frac{(x-y)^2}{2a^2}}d\mu_X(x)dy$$

14.2 The Ultimate Goal: Uniqueness Theorem

Recall our ultimate goal is that if $\varphi_X(t) = \varphi_Y(t)$ then $X \stackrel{d}{=} Y$. This is called Uniqueness Theorem. There are 2 ways to prove this theorem, using 2 Inversion Theorems. Let's begin with the first inversion theorem:

Theorem 181 (First Inversion Theorem- Distribution in terms of char functions)**.** *For a random variable X, we have this relationship:*

$$\begin{aligned} F_X(u) = \mathbf{P}(X \le a) &= \lim_{a\downarrow 0} \mathbf{P}(X + aZ \le u) \\ &= \lim_{a\downarrow 0} \int_{-\infty}^{u} \int \frac{1}{\sqrt{2\pi}\, a} e^{-\frac{(x-y)^2}{2a^2}} d\mu_X(x) dy \\ &= \lim_{a\downarrow 0} \int_{-\infty}^{u} \frac{1}{2\pi} \int e^{-ity - \frac{a^2t^2}{2}} \varphi_X(t) dt dy \end{aligned}$$

Proof. This is a combination of the previous section, example (178) and lemma (179). ☐

Remark 182. We do not know that if the limit exists.

Theorem 183 (Uniqueness Theorem)**.** *If 2 char functions equal, say $\varphi_X(t) = \varphi_Y(t)$ for all t, then the corresponding random variable have the same distribution, i.e., $X \stackrel{d}{=} Y$.*

Before proving this theorem, we noted that the converse is always true:

Remark 184. If $X \stackrel{d}{=} Y$, then $\varphi_X(t) = \varphi_Y(t)$. Why? Suppose $X \stackrel{d}{=} Y$. Then by definition $\mathbf{E}[g(X)] = \mathbf{E}[g(Y)]$ for any bounded continuous function g. Note that $g(x) = e^{itx}$ is bounded and continuous, therefore:

$$\varphi_X(t) = \mathbf{E}[e^{itX}] = \mathbf{E}[e^{itY}] = \varphi_Y(t).$$

Back to proof of First Uniqueness theorem:

Proof. The proof is a direct application of First Inversion Theorem (181): If $\varphi_X(t) = \varphi_Y(t)$, then:

$$F_X(u) = \ldots = \varphi_X(t) = \ldots = \varphi_Y(t) = F_Y(u)$$

And that is the proof. ☐

Next we lay the work for the Second Inversion Theorem (a second way to prove Uniqueness Theorem). We begin by the following theorem:

Theorem 185 (density from a char function)**.** *If char function $\varphi_X(t)$ satisfying $\int_{\mathbb{R}} |\varphi_X(t)| dt < \infty$ (i.e., $\varphi_X \in L^1$), then:*
There exists a density function pdf of X as defined:

$$f_X(x) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_X(t) dt$$

Proof. We use the First Inversion Theorem. Since $\varphi_X(t)$ is in L^1 so $e^{-ity-\frac{\varepsilon^2 t^2}{2}}\varphi_X(t)$ is also in L^1 and bounded by $\varphi_X(t)$ (take absolute value and note $|e^{-ity-\frac{\varepsilon^2 t^2}{2}}| = 1$), this mean we can use DCT:

$$\begin{aligned}
\mu_X(a,b) = F_X(b) - F_X(a) &= \lim_{\varepsilon \downarrow 0} \int_a^b \frac{1}{2\pi} \int e^{-ity-\frac{\varepsilon^2 t^2}{2}} \varphi_X(t)\,dt\,dy \\
&= \int_a^b \frac{1}{2\pi} \lim_{\varepsilon \downarrow 0} \int e^{-ity-\frac{\varepsilon^2 t^2}{2}} \varphi_X(t)\,dt\,dy; \text{ use DCT to bring the limit inside} \\
&= \int_a^b \frac{1}{2\pi} \int e^{-ity} \varphi_X(t)\,dy
\end{aligned}$$

Note that we defined the $\int e^{-ity}\varphi_X(t)dy$ as the density of random variable X from char function φ_X. □

Remark 186. Do we have the converse? That is, suppose density pdf exists, will that imply char function be integrable? NO. A counter example is the random variable exponential.

Theorem 187 (second inversion theorem)**.** *X random variable with μ_X as probability measure, and $\varphi_X(t)$ is a char function. Then:*

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

Recall the Uniqueness Theorem (183), here we will provide a second way to prove that theorem using Second Inverse Theorem.

Proof of Uniqueness Theorem. Since $\varphi_X(t) = \varphi_Y(t)$, by second inverse theorem, we get:

$$\mu_X((a,b)) + \frac{1}{2}\mu_X(\{a\}) + \frac{1}{2}\mu_X(\{b\}) = \mu_Y((a,b)) + \frac{1}{2}\mu_Y(\{a\}) + \frac{1}{2}\mu_Y(\{b\})$$

Now we choose $a_n \downarrow a$ such that $\mu_X(\{a_n\}) = \mu_Y(\{a_n\}) = 0$ for all n. We can do this because there are only countably many $\mu_X(\{c\}) > 0$.
Next we take $b = a_n$, then:

$$\mu_X((a,a_n)) + \frac{1}{2}\mu_X(\{a\}) = \mu_Y((a,a_n)) + \frac{1}{2}\mu_Y(\{a\})$$

Now note that by continuity from above:

$$\lim_{n \to \infty} \mu_X((a,a_n)) = \mu_X(\lim_{n \to \infty}(a,a_n)) = \mu_X(\emptyset) = 0$$

and similarly for μ_Y we get

$$\lim_{n \to \infty} \mu_Y((a,a_n)) = 0$$

Therefor taking limit $n \to \infty$ we get:

$$\mu_X(\{a\}) = \mu_Y(\{a\}), \forall a \in \mathbb{R}$$

And we deduced

$$\mu_X((a,b)) = \mu_Y((a,b))$$

for all open interval (a,b). Thus:

$$\mu_X(B) = \mu_Y(B), \forall B \in \mathcal{B}(\mathbb{R})$$

And therefore X and Y has same distribution. □

Corollary 188 (sum of Gaussian distribution)**.** *Let* $X \quad N(\ _1;\ ^2_1)$ *and* $Y \quad N(\ _2;\ ^2_2)$ *be Gaussian. Then:*

$$X + Y \quad N(\ _1 + \ _2;\ ^2_1 + \ ^2_2)$$

Proof. The Char function of $X + Y$ is

$$\begin{aligned} '(t) = \mathbf{E}[e^{it(X+Y)}] &= \mathbf{E}[e^{itX}e^{itY}] \\ &= \mathbf{E}[e^{itX}]\mathbf{E}[e^{itY}]; \text{ independence} \\ &= e^{i\ _1t\ \ \frac{^2_1t^2}{2}}e^{i\ _2t\ \ \frac{^2_2t^2}{2}} \\ &= e^{\ \ it(\ _1+\ _2)\ \ \frac{(\ ^2_1+\ ^2_2)t^2}{2}} \end{aligned}$$

And that is the char function of $N(\ _1 + \ _2;\ ^2_1 + \ ^2_2)$. By Uniqueness Theorem, we must have

$$X + Y \quad N(\ _1 + \ _2;\ ^2_1 + \ ^2_2)$$

' And that concluded the proof. □

14.3 Levy's Continuity Theorem: convergence in char function implies convergence in distribution

Theorem 189 (Levy Continuity Theorem)**.** *Let X_n be a sequence of random variables with char functions $'_n(t) = \mathbf{E}[e^{itX_n}]$.*
Suppose that $'_n(t)$ is ***continuous at*** *$t = 0$ and that $'_n(t)\ !\ '(t)$ for all t. Then there exists a random variable X such that its char function is $'(t)$ and that $X_n \overset{q}{!} X$.*

Proof. The proof is divided into 3 steps:

1. Step 1: show that there exists a subsequence X_{n_k} that converges in distribution to X by using Tightness argument.

2. Step 2: the $'_X(t)$ that we get from the limit of the subsequence $'_{n_k}(t)$ is the same with the $'(t)$ in the assumption of the theorem.

3. Step 3: Show $X_n \overset{q}{!} X$.

We will now proceed in showing the steps.
Step 1: show that there exists a subsequence X_{n_k} that converges in distribution to X by using Tightness argument.
It suffices to show that X_n is tight. For a fixed $u\ 2\ \mathsf{R}$:

$$\begin{aligned} \int_u^u (1 \quad e^{itx})dt &= \int_u^u \Big(1 \quad \cos(tx) \quad i\sin(tx)\Big)dt \\ &= \int_u^u \Big(1 \quad \cos(tx)\Big)dt; \text{ since } \int_u^u (\sin(x)) = 0 \\ &= t \quad \frac{\sin(tx)}{x}\Big|_u^u \\ &= 2u \quad \frac{2\sin(ux)}{x} \end{aligned}$$

Now we multiply by $\frac{1}{u}$ and get:

$$\frac{1}{u}\int_{-u}^{u}(1 - e^{itx})dt = 2 - \frac{2\sin(ux)}{ux}$$

Then we take x as X_n and take the expected values of the above:

$$\int \frac{1}{u}\int_{-u}^{u}(1 - e^{itx})dt d\mu_{X_n}(x) = \mathbf{E}[2 - \frac{2\sin(uX_n)}{uX_n}]$$

Now the LHS term above, using Fubini:

$$\begin{aligned}\int \frac{1}{u}\int_{-u}^{u}(1 - e^{itx})dt d\mu_{X_n}(x) &= \frac{1}{u}\int_{-u}^{u}\left(\int (1 - e^{itx})d\mu_{X_n}(x)\right)dt \\ &= \frac{1}{u}\int_{-u}^{u}\left(1 - \varphi_n(t)\right)dt\end{aligned}$$

Note the following argument for the bracket above:

$$\begin{aligned}\int (1 - e^{itx})d\mu_{X_n}(x) &= \int 1 d\mu_{X_n}(x) - \int e^{itx}d\mu_{X_n}(x) \\ &= \mu_{X_n}(\mathbb{R}) - \varphi_n(t) \\ &= 1 - \varphi_n(t)\end{aligned}$$

Therefore we now have:

$$\frac{1}{u}\int_{-u}^{u}\left(1 - \varphi_n(t)\right)dt = \mathbf{E}[2 - \frac{2\sin(uX_n)}{uX_n}] \tag{15}$$

For the RHS of (15), we will now find a lower bound for it as followed:

$$\begin{aligned}\mathbf{E}[2 - \frac{2\sin(uX_n)}{uX_n}] &= 2\int\left(1 - \frac{\sin(uX_n)}{uX_n}\right)d\mu_{X_n}(x) \\ &\geq 2\int\left(1 - \frac{1}{uX_n}\right)d\mu_{X_n}(x); \text{ since } 1 - \frac{\sin(uX_n)}{uX_n} \geq 1 - \frac{1}{uX_n}; (\sin(uX_n) \leq 1) \\ &\geq 2\int_{\{|X_n| \geq \frac{2}{u}\}}\left(1 - \frac{1}{uX_n}\right)d\mu_{X_n}(x) \text{ ,we consider a smaller domain } \{|X_n| \geq \frac{2}{u}\} \\ &\geq 2\int_{\{|X_n| \geq \frac{2}{u}\}}\left(1 - \frac{1}{u\frac{2}{u}}\right)d\mu_{X_n}(x) \\ &\text{since } |X_n| \geq \frac{2}{u} \Rightarrow uX_n \geq u\frac{2}{u} \Rightarrow \frac{1}{uX_n} \leq \frac{1}{u\frac{2}{u}} \Rightarrow -\frac{1}{uX_n} \geq -\frac{1}{u\frac{2}{u}} \\ &= 2\frac{1}{2}\mathbf{P}(|X_n| \geq \frac{2}{u}) = \mathbf{P}(|X_n| \geq \frac{2}{u})\end{aligned}$$

In short, we get:

$$\mathbf{E}[2 - \frac{2\sin(uX_n)}{uX_n}] \geq \mathbf{P}(|X_n| \geq \frac{2}{u})$$

Now back to the LHS of (15), since $\varphi_n(0) = 1$ and is continuous at $t = 0$, we deduce that, for any $\varepsilon > 0$, there exists an U_ε st for all $u > U_\varepsilon$:

$$\varepsilon > \frac{1}{u}\int_{-u}^{u}\left(1 - \varphi_n(t)\right)dt$$

Combining this with the above bound we get:

$$\epsilon \geq \mathbf{P}(|X_n| \geq \frac{2}{u})$$

And setting $M_\epsilon = \frac{2}{U_\epsilon}$ for the tightness argument, we showed that X_n is tight. Hence we showed step 1. In particular, we known that there exists a subsequence X_{n_k} such that $X_{n_k} \xrightarrow{d} X$ (and existence of such X is guaranteed) and that $\varphi_{X_{n_k}}(t) \to \varphi_X(t)$.

Step 2: We need to show that the $\varphi_X(t)$ that we get from the limit of the subsequence $\varphi_{n_k}(t)$ is the same with the $\varphi(t)$ in the assumption of the theorem. But this is true because when the sequence converges, the limit of all the subsequence must be the same with the limit of the sequence, that is:

$$\lim_{k \to \infty} \varphi_{n_k}(t) = \lim_{n \to \infty} \varphi_n(t)$$

And because

$$\lim_{k \to \infty} \varphi_{n_k}(t) = \varphi_X(t)$$

while

$$\lim_{n \to \infty} \varphi_n(t) = \varphi(t)$$

we must have:

$$\varphi_X(t) = \varphi(t)$$

This completed step 2.

Step 3: We must show $X_n \xrightarrow{d} X$. Suppose not, say

$$X_{n_j} \xrightarrow{d} Y$$

and that $Y \neq X$. Then:

$$\varphi_{n_j}(t) \to \varphi_Y(t) = \varphi(t)$$

So

$$\varphi_Y(t) = \varphi_X(t) (= \varphi(t))$$

By Uniqueness Theorem, we must have $Y \stackrel{d}{=} X$. This completed the proof. □

Theorem 190 (derivative of char function and moments)**.** *If $\varphi_X(t)$ is k-th time differentiable, then for $j = 0, \ldots, k$:*

$$\varphi_X^{(j)}(0) = (i)^j \mathbf{E}[X^j]$$

Proof. For $k = 0$:

$$\varphi_X(t) = \mathbf{E}[e^{itX}]$$

Then

$$\begin{aligned} \varphi'_X(t) &= \frac{\partial}{\partial t} \int e^{itx} d\mu_X(x) \\ &= \int \frac{\partial}{\partial t} e^{itx} d\mu_X(x) \\ &= \int iX e^{itX} d\mu_X(x) \end{aligned}$$

So $\varphi'_X(0) = \int iX d\mu_X(x) = i\mathbf{E}[X]$. Continue in this manner.... □

Remark 191. The reverse direction of the above theorem said that if k-th moments exists, then char function is k times differentiable at $t = 0$.

15 Day 15: Central Limit Theorem: final

15.1 Central Limit Theorem version 1, with L3 norm finite

Theorem 192 (Central Limit Theorem version 1, with L3 norm finite). *Let X_i be random variables, i.i.d, and $\mu = \mathbf{E}[X_i] < \infty$, and $\sigma^2 = Var(X_i) < \infty$, and $\mathbf{E}[|X|^3] < \infty$. Denote $S_n = X_1 + \dots + X_n$. Then:*

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

OR equivalently:

$$\lim_{n \to \infty} \boldsymbol{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Proof. WLOG, assume $\mathbf{E}[X_i] = \mu = 0$ and $\mathrm{Var}(X_i) = \sigma^2 = 1$. We let $Z \sim N(0,1)$ be the standard Gaussian random variable. Then the char function of Z is $\varphi_Z(t) = e^{-\frac{t^2}{2}}$. Now let us denote:

$$\begin{aligned} Z_n &:= \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n}{1\sqrt{n}}, \text{ because } \mu = 0 \text{ and } \sigma = 1 \\ &= \frac{S_n}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \end{aligned}$$

These Z_n will have the char functions $\varphi_n(t)$. The idea is to use Levy Continuity Theorem (189), and we will show that $\varphi_n \to \varphi_Z(t)$ to get the result. So here is the computation:

$$\begin{aligned} \varphi_n(t) &= \mathbf{E}[e^{i\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k\right)t}] \\ &= \mathbf{E}[\prod_{k=1}^{n} e^{iX_k\frac{t}{\sqrt{n}}}] \\ &= \prod_{k=1}^{n} \mathbf{E}[e^{iX_k\frac{t}{\sqrt{n}}}]; \text{ by independence} \\ &= \left(\mathbf{E}[e^{iX\frac{t}{\sqrt{n}}}]\right)^n \text{ by identical distribution that we let } X_k = X \\ &= \varphi_X\left(\frac{t}{\sqrt{n}}\right)^n \end{aligned}$$

Now we use Taylor expansion to $\varphi_X(t)$ at $t = 0$:

$$\varphi_X(t) = \varphi_X(0) + \varphi'_X(0)t + \frac{\varphi''_X(0)}{2!}t^2 + \frac{\varphi'''_X(s)}{3!}t^3$$

for some $s \in [-t, t]$.
Next we use the hypothesis that $E[|X|^3] < \infty$, therefore $\varphi'''_X(s)$ will be bounded by some constant M. So we get its approximated value is $\varphi'''_X(0) = (i)^3\mathbf{E}[X^3] = \frac{C}{\sqrt{n}}$:

$$\frac{\varphi'''_X(s)}{3!}t^3 = \frac{M}{6}\left[\frac{1}{\sqrt{n}}\right]^3 t^3 = \frac{M}{6}\frac{t^3}{n^{3/2}}$$

And for the other terms:

$$\varphi_X(0) = 1$$

And

$$\varphi'_X(0) = i\mathbf{E}[X] = 0$$

and

$$\varphi''_X(0) = (i)^2\mathbf{E}[X^2] = (-1)(1) = -1$$

Therefore:

$$\varphi_X\left(\frac{t}{\sqrt{n}}\right) = 1 + 0\frac{t}{\sqrt{n}} - \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + \frac{M}{6}\frac{t^3}{n^{3/2}}$$

Simplifying and we get:

$$\varphi_X\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{-\frac{t^2}{2} + \frac{Mt^3}{6\sqrt{n}}}{n}$$

We write $C_n = -\frac{t^2}{2} + \frac{Mt^3}{6\sqrt{n}}$, so we can simplify:

$$\varphi_X\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{C_n}{n}$$

Now we need some lemma to deal with raising the above term to power n. Note that $\lim_{n\to\infty} C_n = -\frac{t^2}{2}$.
We claimed that: for complex numbers c_n such that $c_n \to c$, we get:

$$\left(1 + \frac{c_n}{n}\right)^n \to e^c$$

Using that claim:

$$\varphi_Z(t) = \left(\varphi_X\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \frac{C_n}{n}\right)^n \to e^{-\frac{t^2}{2}}$$

because our C_n converges to $-\frac{t^2}{2}$. Now the $e^{-\frac{t^2}{2}}$ is the char function of Z. And we are done by Levy Theorem, we must have:

$$Z_n \xrightarrow{d} Z$$

It remains to prove the claim. See Appendix section for the proof (as the proof is very technical). □

15.2 Central Limit Theorem version 2, without L3 norm finite

Theorem 193 (Central Limit Theorem version 2, without L3 norm finite)**.** *Let X_i be random variables, i.i.d, and $\mu = \mathbf{E}[X_i] < \infty$, and $\sigma^2 = Var(X_i) < \infty$. Denote $S_n = X_1 + \dots + X_n$. Then:*

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

OR equivalently:

$$\lim_{n\to\infty} \boldsymbol{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Proof. Following the proof of the finite L^3 norm version, we need to show that:

$$\varphi_X(t) \approx 1 + it\mathbf{E}[X] + \frac{(it)^2}{2}\mathbf{E}[X^2]$$

regardless of the boundedness of $\mathbf{E}[X^2]$. Let's try to estimate this:

$$\begin{aligned}
\left|\varphi_X(t) - (1 + it\mathbf{E}[X] + \frac{(it)^2}{2}\mathbf{E}[X^2])\right| &= \left|\mathbf{E}[e^{itX}] - (\mathbf{E}[1] + it\mathbf{E}[X] + \frac{(it)^2}{2}\mathbf{E}[X^2])\right| \\
&= \left|\mathbf{E}\left[e^{itX} - (1 + itX + \frac{(it)^2}{2}X^2)\right]\right| \\
&\leq \mathbf{E}\left[\left|e^{itX} - (1 + itX + \frac{(it)^2}{2}X^2\right|\right]
\end{aligned}$$

Now it would be a great idea if we can estimate (upper bound) the intergrand of the expectation:

$$\left|e^{itX} - (1 + itX + \frac{(it)^2}{2}X^2\right|$$

Indeed, we have the following lemma to bound it:

Lemma 194. For any real number y, we always have:

$$|e^{iy} - (1 + iy + \frac{i^2y^2}{2})| \leq \min\{\frac{|y|^3}{6}; |y|^2\}$$

The proof of this lemma can be found in the Appendix section.
Now applying this Lemma with $y = tX$ we get:

$$\left|e^{itX} - (1 + itX + \frac{(it)^2}{2}X^2\right| \leq \min\{\frac{|tX|^3}{6}; |tX|^2\}$$

Combining with earlier expectation estimate we get:

$$\begin{aligned}
\left|\varphi_X(t) - (1 + it\mathbf{E}[X] + \frac{(it)^2}{2}\mathbf{E}[X^2])\right| &\leq \mathbf{E}\left[\left|e^{itX} - (1 + itX + \frac{(it)^2}{2}X^2\right|\right] \\
&\leq \mathbf{E}\left[\min\{\frac{|tX|^3}{6}; |tX|^2\}\right] \\
&= t^2\mathbf{E}\left[\min\{\frac{t|X|^3}{6}; |X|^2\}\right] \\
&= t^2\,(t)
\end{aligned}$$

Where we denoted $(t) = \mathbf{E}\left[\min\{\frac{t|X|^3}{6}; |X|^2\}\right]$. In short we showed:

$$\left|\varphi_X(t) - (1 + it\mathbf{E}[X] + \frac{(it)^2}{2}\mathbf{E}[X^2])\right| \leq t^2\,(t)$$

In addition, WLOG, we can assume $\mathbf{E}[X] = \mu = 0$ and $\sigma^2 = \mathrm{Var}(X) = 1$ (and therefore $\mathbf{E}[X^2] = \mathrm{Var}(X) - \mu^2 = \mathrm{Var}(X) = 1$). Hence the above is now simplified into:

$$\left|\varphi_X(t) - (1 + \frac{(it)^2}{2})\right| \leq t^2\,(t)$$

And $i^2 = -1$ gives:

$$\left|\varphi_X(t) - \left(1 - \frac{t^2}{2}\right)\right| \le t^2 \delta(t)$$

That means, approximately, we have:

$$\varphi_X(t) \approx 1 - \frac{t^2}{2} + t^2 \delta(t)$$

We now also claim that $\delta(t) \to 0$ when $t \to 0$. To see this claim, we recall the definition of $\delta(t)$:

$$\delta(t) = \mathbf{E}\left[\min\{\frac{t|X|^3}{6}, |X|^2\}\right]$$

Then note that $\min\{\frac{t|X|^3}{6}, |X|^2\} \le |X|^2$ and because $\mathbf{E}[|X|^2] < \infty$, we can apply DCT with the bound $|X|^2$ and get:

$$\begin{aligned}\lim_{t\to 0} \mathbf{E}\left[\min\{\frac{t|X|^3}{6}, |X|^2\}\right] &= \mathbf{E}\left[\lim_{t\to 0} \min\{\frac{t|X|^3}{6}, |X|^2\}\right] \\ &= \mathbf{E}[0] = 0, \\ &\text{since the limit} \lim_{t\to 0} \min\{\frac{t|X|^3}{6}, |X|^2\} = 0\end{aligned}$$

So we showed the claim is true, that is, $\delta(t) \xrightarrow{t\to 0} 0$. Now back to the approximation:

$$\varphi_X(t) \approx 1 - \frac{t^2}{2} + t^2 \delta(t)$$

Apply this approximate to $\frac{t}{\sqrt{n}}$ in place of t we get (why do we do this? wait in a moment):

$$\varphi_X\left(\frac{t}{\sqrt{n}}\right) \approx 1 - \frac{1}{2}\frac{t^2}{n} + \frac{t^2}{n}\delta\left(\frac{t}{\sqrt{n}}\right)$$

Recall that (same from the proof of CLT version 1):

$$\begin{aligned}\varphi_n(t) &= \left(\varphi_X\left(\frac{t}{\sqrt{n}}\right)\right)^n \\ &\approx \left(1 - \frac{1}{2}\frac{t^2}{n} + \frac{t^2}{n}\delta\left(\frac{t}{\sqrt{n}}\right)\right)^n, \text{ use the approximation earlier of } \varphi_X\left(\frac{t}{\sqrt{n}}\right) \\ &= \left(1 + \frac{-\frac{t^2}{2} + t^2\delta\left(\frac{t^2}{\sqrt{n}}\right)}{n}\right)^n \\ &= \left(1 + \frac{c_n}{n}\right)^n\end{aligned}$$

where we denoted $c_n = \frac{-\frac{t^2}{2} + t^2\delta\left(\frac{t^2}{\sqrt{n}}\right)}{n}$. Now:

$$\lim_{n\to\infty} c_n = \lim_{n\to\infty}\left(\frac{-\frac{t^2}{2} + t^2\delta\left(\frac{t^2}{\sqrt{n}}\right)}{n}\right) = -\frac{t^2}{2} = c$$

So using the same Lemma which stated:

Lemma: Suppose $c_n \to c$. Then we always have:

$$\left(1 + \frac{c_n}{n}\right)^n \to e^c$$

Therefore $\varphi_n(t) \to e^{-\frac{t^2}{2}} = \varphi_Z$ so the conclusion follows as in CLT version 1. □

15.3 CLT without identical distribution- Lindeberg CLT

Theorem 195 (CLT without identical distribution- Lindeberg CLT)**.** *Let X_i be independent and $\mu_i = \mathbf{E}[X_i] < \infty$, and $\sigma_i^2 = Var(X_i) < \infty$ for all i. Let us denote*

$$B_n^2 = \sum_{i=1}^{n} \sigma_j^2$$

In addition, we suppose that, for any $\varepsilon > 0$, we have this hypothesis:

$$\lim_{n \to \infty} \frac{1}{B_n^2} \sum_{j=1}^{n} \mathbf{E}\Big[|X_j - \mu_j|^2 \mathbf{1}_{\{|X_j - \mu_j| \geq \varepsilon B_n\}}\Big] = 0$$

Then:

$$\frac{\sum_{j=1}^{n}(X_j - \mu_j)}{B_n} \xrightarrow{d} N(0,1)$$

Remark 196. The extra hypothesis for any $\varepsilon > 0$, we have this hypothesis:

$$\lim_{n \to \infty} \frac{1}{B_n^2} \sum_{j=1}^{n} \mathbf{E}\Big[|X_j - \mu_j|^2 \mathbf{1}_{\{|X_j - \mu_j| \geq \varepsilon B_n\}}\Big] = 0$$

can be satisfied if:

1. the X_i are identical
2. $|X_i| \leq K$ for all i (uniformly bounded) and $\lim_{n \to \infty} B_n = 0$

A quite technical question:

Question 197. How fast does convergence of CLT?

$$F_{Z_n}(t) \overset{?}{\to} \Phi(t)$$

This question is answered by

Theorem 198 (Berry-Esseen -Theorem 3.4.17)**.** *Let X_i be iid with $E[X_i] = 0$ and $\mathbf{E}[X_i^2] = \sigma^2$, and $\mathbf{E}[|X_i|^3] = \rho < \infty$. If F_{Z_n} is the distribution of $\frac{S_n}{\sigma\sqrt{n}}$ and $\Phi(t)$ be the char function of standard normal distribution, then:*

$$|F_n(t) - \Phi(t)| \leq \frac{3\rho}{\sigma^3\sqrt{n}}$$

15.4 Large Deviation

Note that in CLT, we know the convergence of

$$\mathbf{P}\left(\frac{S_n}{\sigma\sqrt{n}} \geq a\right)$$

Here we will be interested in how fast does

$$\mathbf{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq a\right) \to 0$$

Example 199. Take X_i iid with $X_i \sim N(\mu, \sigma^2)$. Then

$$\frac{S_n}{n} \sim N(\mu, \frac{\sigma^2}{n})$$

So

$$\begin{aligned}
\mathbf{P}(|\frac{S_n}{n} - \mu| \geq a) &= \mathbf{P}\left(\left|\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}}\right| \geq \frac{a\sqrt{n}}{\sigma}\right) \\
&= 2\mathbf{P}\left(\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{a\sqrt{n}}{\sigma}\right) \\
&\xrightarrow[CLT]{n \to \infty} 2\mathbf{P}(Z \geq \frac{a\sqrt{n}}{\sigma}) \text{ ,where } Z \sim N(0,1) \\
&= 2\Phi(\frac{-a\sqrt{n}}{\sigma}) \text{ , where } \Phi \text{ is cdf of } N(0,1)
\end{aligned}$$

Remark 200. In general, when $x \to \infty, X > 0$:

$$\frac{1}{x + \frac{1}{x}} e^{-x^2} \leq \Phi(x) \leq \frac{1}{x} e^{-\frac{x^2}{2}}$$

Using this remark, we get

$$2\Phi(\frac{-a\sqrt{n}}{\sigma}) \approx 2\frac{\sigma}{a\sqrt{n}} e^{-\frac{a^2 n}{2\sigma^2}}$$

In short:

$$\mathbf{P}(|\frac{S_n}{n} - \mu| \geq a) \approx 2\frac{\sigma}{a\sqrt{n}} e^{-\frac{a^2 n}{2\sigma^2}}$$

taking log gives:

$$\frac{\log\left(\mathbf{P}(|\frac{S_n}{n} - \mu| \geq a)\right)}{n} \approx -\frac{a^2}{2\sigma^2} \quad \text{exponential decay}$$

So the tail event probability decay exponentially with respect to n with rate $\frac{a^2}{2\sigma^2} = I(a)$ (we call $I(a)$ the rate function).

Example 201. Take X_i iid with $\mathbf{P}(X_i = 0) = \mathbf{P}(X_i = 1) = \frac{1}{2}$. Then;

$$\mathbf{P}(S_n > a) \approx e^{-nI(a)}$$

for any $0.5 < a < 1$. And we can compute and get $I(a) = a\log(a) + (1 - a)\log(1 - a) + \log 2$.

16 Day 16: Brownian Motion

Another name for Brownian Motion is **Wiener Process**.

16.1 Stochastic Process and First definition of Brownian Motion

Definition 202 (Stochastic Process)**.** A stochastic process is $X(t, \omega)$ for which $\omega \in (\Omega, \mathcal{F}, \mathbb{P})$ and $t \in T$ satisfying:

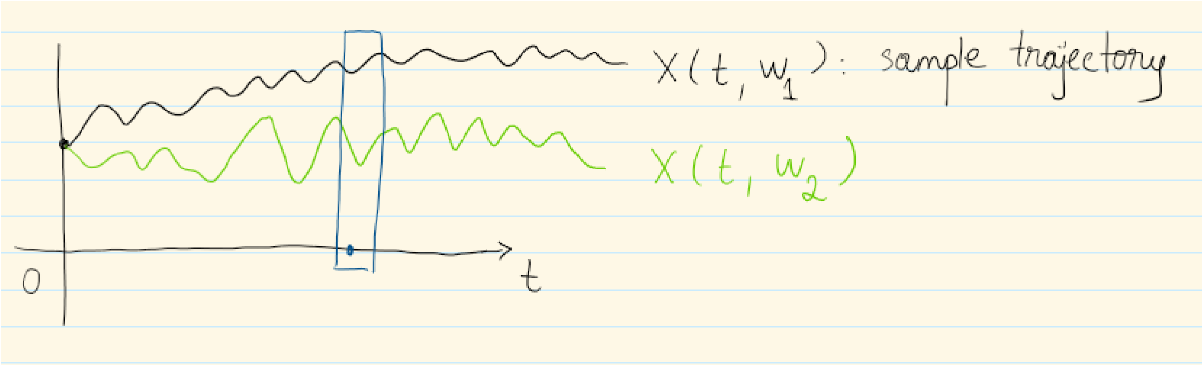


Figure 21: Stochastic Process example

1. For fixed t, the $X(\cdot;\omega)$ is a random variable.
2. For fixed ω, the $X(t;\cdot)$ is a function on T.

We called

1. **Discrete time stochastic process** if $T = \{1, 2, 3, 4, \ldots\}$
2. **Continuous time stochastic process** if $T = [0, \infty)$.

See figure (21) for reference.

With discrete time, it is a good idea to recall a theorem in the past, the Kolmogorov extension Theorem, which stated:

Theorem 203 (Kolmogorov Extension Theorem). *When $T \subseteq \{t_1, \ldots, t_k\}$.*

1. *Permutation π of $\{1, 2, \ldots, k\}$:*

$$\mu_{1,2,\ldots,k}(A_1 \times \cdots \times A_k) = \mu_{\pi_1,\ldots,\pi_k}(A_{\pi_1} \times \cdots \times A_{\pi_k})$$

2. *for any $m \in \mathbb{N}$:*

$$\mu_{1,2,\ldots,k}(A_1 \times \cdots \times A_k) = \mu_{1,\ldots,k,k+1,\ldots,k+m}(A_1 \times \cdots \times A_k \times \mathbb{R}^m)$$

there exists a unique probability measure on $(\mathbb{R}^n)^T$ that agrees with the finite dimensional restricted definition.

Definition 204 (Brownian Motion First definition). A Brownian Motion (BM) $B(t;\omega)$ is a real-valued (continuous) stochastic process satisfying:

1. $B(0;\omega) = 0$
2. $B(t;\omega)$ is continous in t for a fixed $\omega \in \Omega$.
3. For any $0 \leq s \leq t$:
$$B(t;\omega) - B(s;\omega) \sim N(0, t - s)$$
4. Independent increment: For $t_0 = 0 < t_1 < \cdots < t_k$, we have:
$$B(t_i;\omega) - B(t_{i-1};\omega)$$
are independent.

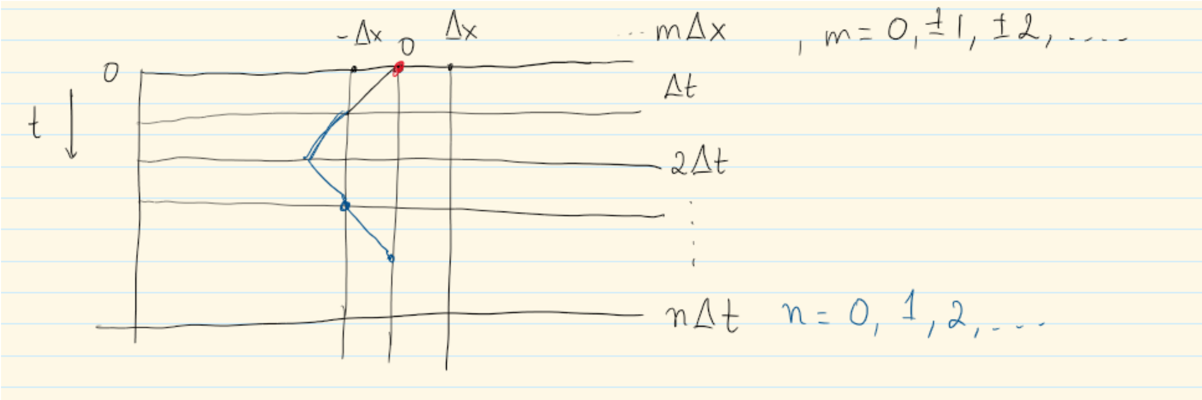


Figure 22: Random Walk

Remark 205 (The intuition behind 1-3 in the definition of BM (especially 3))**.** It is the Random Walk. See figure (22) for reference. We start with these notations:

1. $p(m;n)$: probability that particle is at $m\Delta x$ at time $n\Delta t$
2. Suppose the probability of moving left or right is always $\frac{1}{2}$
3. Then $p(m;n+1) = \frac{1}{2}p(m-1;n) + \frac{1}{2}p(m+1;n)$

And then by algebra we get the **Master Equation**:

$$p(m;n+1) - p(m;n) = \frac{1}{2}\Big[p(m-1;n) - 2p(m;n) + p(m+1;n)\Big]$$

It is customary to describe the deviations (in space) by the quadratic deviation Δx^2 (distance square). In addition, time deviation is only in first order. Hence we choose

$$D = \frac{\Delta x^2}{\Delta t}$$

or we may have equivalently:

$$\frac{1}{\Delta t} = D\frac{1}{\Delta x^2}$$

Now in the master equation, we divided by Δt and get (use $D\frac{1}{\Delta x^2}$ for the RHS):

$$\begin{aligned}\frac{p(m;n+1) - p(m;n)}{\Delta t} &= \frac{D}{2}\frac{p(m-1;n) - 2p(m;n) + p(m+1;n)}{\Delta x^2} \\ &= \frac{D}{2}\Big[\Big(\frac{p(m-1;n) - p(m;n)}{\Delta x^2}\Big) - \Big(\frac{p(m;n) - p(m-1;n)}{\Delta x^2}\Big)\Big]]\end{aligned}$$

When letting $\Delta x \to 0$ and $\Delta t \to 0$ we obtained the Heat equation:

$$\frac{\partial p}{\partial t}(x;t) = \frac{D}{2}\frac{\partial^2 p}{\partial x^2}$$

With the initial condition:

$$p(x;0) = \delta_{\{x=0\}} = \begin{cases} 1 & \text{if } x = 0 \\ 0 & x \neq 0 \end{cases}$$

This Heat equation has the exact solution:

$$p(t;x) = \frac{1}{\sqrt{2\pi Dt}}e^{-\frac{x^2}{2Dt}} \sim N(0;Dt)$$

Remark 206. Another way to see part (3) in the definition is as followed: We write $X(t)$ random variable of position at time $t = n\Delta t$. We define X_i such that $\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = 0) = \frac{1}{2}$. So

$$S_n = \sum_{i=1}^{n} X_i \text{ is the counting of right moves}$$

We get these calculation:

1. When you make S_n right moves and $n - S_n$ left moves:
$$X(t) = S_n\Delta x + (n - S_n)(-\Delta x) = (2S_n - n)\Delta x$$
2. For each i:
$$\mathbf{E}[X_i] = \frac{1}{2}; \mathrm{Var}(X_i) = \frac{1}{4}$$
3. Then:
$$\mathbf{E}[S_n] = n\frac{1}{2}$$
and
$$\mathrm{Var}(S_n) = n\frac{1}{4}$$
4. Therefore:
$$\mathbf{E}[X(t)] = 0$$
And
$$\mathrm{Var}(X(t)) = 4\Delta x^2\frac{n}{4} = n\Delta x^2 = nD\Delta t = Dt$$

For the probability measure:

$$\begin{aligned}
\mathbf{P}(a \leq X(t) \leq b) &= \mathbf{P}(a \leq (2S_n - n)\Delta x \leq b) \\
&= \mathbf{P}\left(\frac{a}{\frac{\sqrt{n}}{2}} \leq \frac{2(S_n - \frac{n}{2})\Delta x}{\frac{\sqrt{n}}{2}} \leq \frac{b}{\frac{\sqrt{n}}{2}}\right) \\
&= \mathbf{P}\left(\frac{2a}{\sqrt{n}} \leq \frac{2(S_n - \frac{n}{2})\Delta x}{\frac{\sqrt{n}}{2}} \leq \frac{2b}{\sqrt{n}}\right) \\
&= \mathbf{P}\left(\frac{a}{\sqrt{n}} \leq \frac{(S_n - \frac{n}{2})\Delta x}{\frac{\sqrt{n}}{2}} \leq \frac{b}{\sqrt{n}}\right) \\
&= \mathbf{P}\left(\frac{a}{\sqrt{n}} \leq \frac{(S_n - \frac{n}{2})\sqrt{D\Delta t}}{\frac{\sqrt{n}}{2}} \leq \frac{b}{\sqrt{n}}\right), \text{ since} \Delta x = \sqrt{D\Delta t} \\
&= \mathbf{P}\left(\frac{a}{\sqrt{nD\Delta t}} \leq \frac{(S_n - \frac{n}{2})}{\frac{\sqrt{n}}{2}} \leq \frac{b}{\sqrt{nD\Delta t}}\right)
\end{aligned}$$

Note that $\frac{(S_n - \frac{n}{2})}{\frac{\sqrt{n}}{2}}$ is $\sim N(0;1)$ by CLT (as $n \to \infty$). So the above probability give:

$$\int_{\frac{a}{\sqrt{nD\Delta t}}}^{\frac{b}{\sqrt{nD\Delta t}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_a^b \frac{1}{\sqrt{2\pi nD\Delta t}} e^{-\frac{y^2}{2Dn\Delta t}} dy \text{ ,change of variable}$$
$$N(0; Dn\Delta t) = N(0; Dt)$$

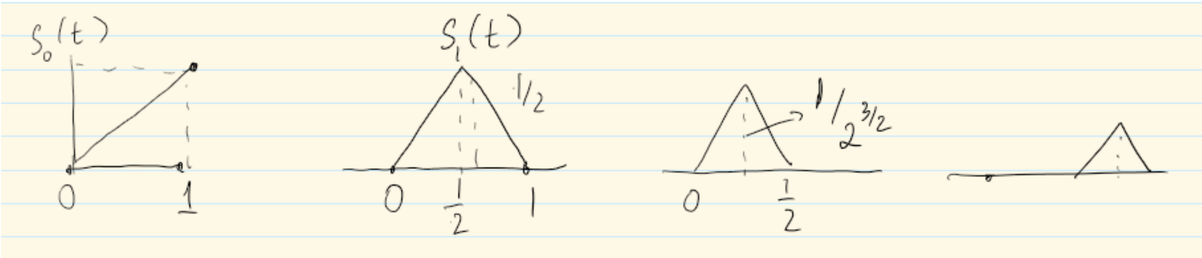


Figure 23: Heat functions in construction of BM

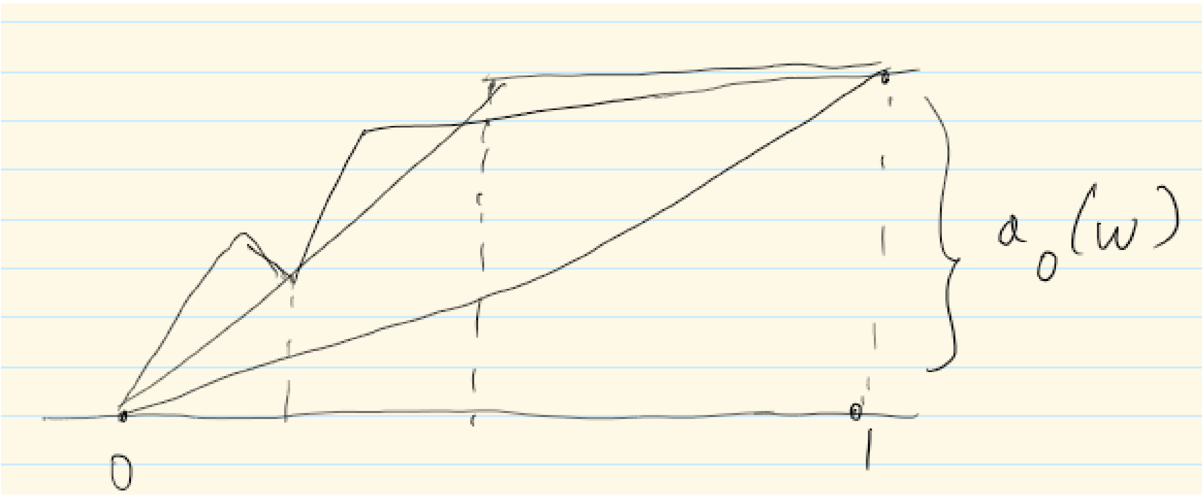


Figure 24: The BM constructed using heat functions

16.2 Construction of Brownian Motion

There are 2 ways:

1. Wiener use:

$$B(t;\omega) := \frac{}{2\sqrt{2}}\sum_{k=1}^{\infty}\frac{a_k(\omega)}{k}\sin(\frac{kt}{2})$$

   where the $a_k(\omega) \sim N(0;1)$

2. Using Heat functions $S_k(t)$:

$$B(t;\omega) := \sum_{k=0}^{\infty} a_k(\omega)S_k(t)$$

   where the $a_k(\omega) \sim N(0;1)$. See figure (23) for pictorial of heat functions and figure (24) for the BM constructed from such heat functions.

16.3 Second Definition of Brownian Motion

Theorem 207 (covariance of Brownian Motions)**.** *The Covariance of Brownian Motions is:*

$$Covar(B(t;\omega);B(s;\omega)) = \mathbf{E}[B(t;\omega)B(s;\omega)] = \min(t;s) = t \wedge s$$

Proof. Suppose $s < t$ (then $\min(s;t) = s$). For simplified notation, we write $B(t;\omega) = B(t)$ and $B(s;\omega) = B(s)$, then:

$$\begin{aligned}
\mathbf{E}[B(t)B(s)] &= \mathbf{E}[B(t)B(s) - B(s)^2 + B(s)^2] \\
&= \mathbf{E}[B(s)^2] + \mathbf{E}[B(s)(B(t) - B(s))] \\
&= \mathbf{E}[B(s)^2] + \mathbf{E}[B(s)]\mathbf{E}[(B(t) - B(s))]; \text{ , by independence of } B(t); B(t) - B(s) \\
&= \mathbf{E}[B(s)^2] \text{ ,since } \mathbf{E}[B(t) - B(s)] = 0 \\
&= \mathrm{Var}(B(s)) = s
\end{aligned}$$

And that gives the proof. □

Definition 208 (Second definition of Brownian Motion)**.** A Brownian Motion is a stochastic process that satisfies:

1. For any $k > 1$: $(t_1; \ldots; t_k)$ then:
   $(B_{t_1}; \ldots; B_{t_k})$ is a k-dimensional Gaussian random vector with mean zero and covariance $\mathrm{Covar}(B_{t_i}; B_{t_j}) = \min(t_t; t_j)$
2. continuous in time (same as First definition)

We have some properties of BM:

Theorem 209 (some properties of BM)**.** *For a BM $B(t;s)$, we have:*

1. ***Time scaling invariance of BM:*** *for any fixed $s > 0$:*

$$\{B_{st}; t \geq 0\} \overset{d}{=} \{\sqrt{s}B_t; t \geq 0\}$$

   Observe that $B(st) \sim N(0;st)$ and $\sqrt{s}B_t \sim \sqrt{s}N(0;t) \sim N(0;st)$

2. ***translation invariance:*** *With starting of zero*

$$B(t;\omega) - B(s;\omega) \sim B(t-s;\omega)$$

Proof. The translation invariance property is true because $B(t;\omega) - B(s;\omega) = N(0;t - s)$. □

17 Day 17: More properties of Brownian Motion

17.1 Properties of Brownian Motion

Remark 210. A Brownian motion satisfies first definition is called a **standard Brownian Motion**

Theorem 211 (scaling invariance)**.** *Let B be a standard BM. For any $a > 0$, we have $\frac{1}{a}B(a^2t)$ is also a standard BM.*

Proof. We checked all 4 parts in the definition. We denote $X(t) = \frac{1}{a}B(a^2t)$.

1. $X(0) = \frac{1}{a}B(a^2 0) = \frac{1}{a}B(0) = 0$
2. $X(t)$ is continuous since $B(t)$ is.

3.

$$\begin{aligned} X(t) - X(s) &= \frac{1}{a}\left[B(a^2 t) - B(a^2 s)\right] \\ &\overset{d}{=} \frac{1}{a} N(0, a^2 t - a^2 s) \\ &\overset{d}{=} N(0, t - s), \text{ since bring the constant term } \frac{1}{a} \text{ inside will be } \frac{1}{a^2} \end{aligned}$$

4. For any $t_0 = 0 < t_1 < \dots < t_n$: the same for:

$$a t_0^2 = 0 < a^2 t_1 < \dots < a^2 t_n$$

So we get $B(a^2 t_{i+1}) - B(a^2 t_i)$ are independent.

This completed the proof. □

Theorem 212 (Translation Invariance - Markov Property). *For a fixed $s > 0$, if B is a standard BM, so is $\{B(t+s) - B(s); t \geq 0\}$.*

Proof. Show all 4 properties. □

Theorem 213 (Regularity of Brownian Motion). *For any BM B, we have:*

1. *B is a locally α-Holder continous with $\alpha < \frac{1}{2}$:*
   For a fixed t, there exists a Holder constant $M < \infty$ and a $\delta > 0$ such that if $|t - s| < \delta$, then:

$$|B(t;\omega) - B(s;\omega)| \leq M|t - s|^{\alpha}$$

2. *BM B is nowhere differentiable. Moreover, for any $t \in [0, 1]$:*

$$\limsup_{h \to 0} \frac{B(t+h) - B(t)}{h} = \infty$$

or

$$\liminf_{h \to 0} \frac{B(t+h) - B(t)}{h} = -\infty$$

or both.

Proof. Suppose it's not true. Then there exists a constant M and for some t_o:

$$\sup_{h \in [0,1]} \left|\frac{B(t_0+h) - B(t_0)}{h}\right| \leq M$$

We will show that this cannot happen, in terms of probability, we would like to show:

$$\mathbf{P}(\omega : \exists t_0 : \sup_{h \in [0,1]} \left|\frac{B(t_0+h) - B(t_0)}{h}\right| \leq M) = 0$$

Let's begin proving this. WLOG we can assume that $t_0 \in [\frac{k-1}{2^n}, \frac{k}{2^n}]$. Then:

$$\begin{aligned} |B(\frac{k+j}{2^n}) - B(\frac{k+j-1}{2^n})| &= |B(\frac{k+j}{2^n}) - B(t_0) + B(t_0) - B(\frac{k+j-1}{2^n})| \\ &\leq \left|B(\frac{k+j}{2^n}) - B(t_0)\right| + \left|B(t_0) - B(\frac{k+j-1}{2^n})\right| \\ &\leq M\frac{j+1}{2^n} + M\frac{2j+1}{2^n} = M\left(\frac{2j+1}{2^n}\right) \text{ ,for } j = 1, 2, 3. \end{aligned}$$

We define:

$$\Omega_{n;k} := \left\{ ! : \left| B(\frac{k+j}{2^n}) \quad B(\frac{k+j \quad 1}{2^n}) \right| \quad M\left(\frac{2j+1}{2^n}\right) \text{ for } j = 1; 2; 3 \right\}$$

It's suffice to show that

$$\bigcup_{k=1}^{2^n \quad 3} \Omega_{n;k} \text{ cannot happen for infinitely many n}$$

because:

$$\left\{ ! : 9t_0; 9M : \sup_{h 2 [0;1]} \left| \frac{B(t_0+h) \quad B(t_0)}{h} \right| \quad M \right\} \quad \bigcup_{k=1}^{2^n \quad 3} \Omega_{n;k}$$

and therefore taking probability:

$$\mathbf{P}\left(! : 9t_0; 9M : \sup_{h 2 [0;1]} \left| \frac{B(t_0+h) \quad B(t_0)}{h} \right| \quad M \right) \quad \mathbf{P}\left(\bigcup_{k=1}^{2^n \quad 3} \Omega_{n;k} \text{ happens for infinitely many n} \right)$$

In summary, we wish to show:

$$\mathbf{P}\left(\bigcup_{k=1}^{2^n \quad 3} \Omega_{n;k} \text{ i.o} \right) = 0$$

We will now begin to estimate:

$$\begin{aligned}
\mathbf{P}(\Omega_{n;k}) &= \prod_{j=1}^{3} \mathbf{P}\left(! : \left| B(\frac{k+j}{2^n}) \quad B(\frac{k+j \quad 1}{2^n}) \right| \quad M\left(\frac{2j+1}{2^n}\right) \right) \\
&\quad \text{because of independence of } B \\
&= \prod_{j=1}^{3} \mathbf{P}\left(\left| N(0; \frac{1}{2^n}) \right| \quad M\left(\frac{2j+1}{2^n}\right) \right) \\
&\quad \text{use definition 1, part 3} \\
&\quad \left[\mathbf{P}\left(jB(2 \quad ^n)j \quad \frac{7M}{2^n} \right) \right]^3 \\
&= \left[\mathbf{P}\left(p\frac{1}{2 \quad n} jB(2 \quad ^n)j \quad p\frac{1}{2 \quad n} \frac{7M}{2^n} \right) \right]^3 ; \text{ use scale invariance} \\
&= \left[\mathbf{P}\left(jB(1)j \quad \frac{7M}{2^{\frac{n}{2}}} \right) \right]^3 \\
&= \left[\int_{\frac{7M}{2^{\frac{n}{2}}}}^{\frac{7M}{2^{\frac{n}{2}}}} p\frac{1}{2} e^{\quad \frac{z^2}{2}} dz \right]^3 \\
&\quad \left[\int_{\frac{7M}{2^{\frac{n}{2}}}}^{\frac{7M}{2^{\frac{n}{2}}}} p\frac{1}{2} \frac{1}{2} dz \right]^3 = \left[\frac{7M}{2^{\frac{n}{2}}} \right]^3 = \frac{(7M)^3}{2^{\frac{3n}{2}}}
\end{aligned}$$

Therefore we take union:

$$\mathbf{P}(\bigcup_{k=0}^{2^n \; 3} \Omega_{n,k}) \quad \sum_{k=0}^{2^n \; 3} \frac{(7M)^3}{2^{\frac{3n}{2}}} \quad 2^n \frac{(7M)^3}{2^{\frac{3n}{2}}} = \frac{(7M)^3}{2^{\frac{n}{2}}}$$

And then we take sum:

$$\sum_n \mathbf{P}(\bigcup_{k=0}^{2^n \; 3} \Omega_{n,k}) \quad \sum_n \frac{(7M)^3}{2^{\frac{n}{2}}} < \infty$$

Hence by Borel-Cantelli Lemma:

$$\mathbf{P}(\bigcup_{k=0}^{2^n \; 3} \Omega_{n,k} \text{ i.o.}) = 0$$

. And that gives the desired result:

$$\mathbf{P}\left(\omega : \exists t_0, \exists M : \sup_{h \in [0,1]} \left| \frac{B(t_0 + h) \quad B(t_0)}{h} \right| \quad M\right) = 0$$

And concluded the proof. □

17.2 Conditional Probability

Definition 214 (conditional probability of event A given event B)**.** The conditional probability of event A given event B occurred is denoted as $\mathbf{P}(A|B)$ and defined as:

$$\mathbf{P}(A|B) := \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

Example 215. Dice rolling, one time $\Omega = \{1, \ldots, 6\}$, $F = 2$, $(\{i\}) = \frac{1}{6}$. Let A be the event that 2 is an outcome. Let B be the event that the outcome is even. Then:

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(\{2\})}{\mathbf{P}(\{2, 4, 6\})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Remark 216. If A and B are independent, then $\mathbf{P}(A|B) = \mathbf{P}(A)$, since $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.

Definition 217 (Conditional Expectation of a random variable)**.** The Conditional Expectation of a random variable $X(\omega)$ given an event B is:

$$\mathbf{E}[X|B] := \int X(\omega) d\mathbf{P}(\omega|B)$$

Example 218. Dice rolling 2 times, the space of events is $\Omega = \{(1, 1), \ldots\}$ $F = 2$, and $\mathbf{P}((i, j)) = \frac{1}{36}$. let define the random variable of summation of 2 outcomes $X(\omega_1, \omega_2) = \omega_1 + \omega_2$, and let B be the event that the first outcome is a 2, i.e., $\omega_1 = 2$. Then:

$$\begin{aligned}
\mathbf{E}[X|B] &= \sum_{\omega_1=1}^{6} \sum_{\omega_2=1}^{6} (\omega_1 + \omega_2) \mathbf{P}(\omega_1 + \omega_2 | B) \\
&= \sum_{\omega_1=2} \sum_{\omega_2=1}^{6} (\omega_1 + \omega_2) \mathbf{P}(\omega_1 + \omega_2 | \omega_1 = 2) \\
&= (2.6 + 1 + 2 + \ldots + 6) \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{12 + 21}{6} = 5.5 = 2 + 3.5
\end{aligned}$$

The sum was $(2+1)+(2+2)+\dots(2+6) = 2.6+(1+2+\dots+6) = 12+21$. And the probability was $\mathbf{P}((\omega_1+\omega_2) \cap (\omega_1 = 6)) = \frac{1}{6}$ (there are only 1 sum, given that we know the first ω_1 is 6) and $\mathbf{P}(\omega_1 = 6) = \frac{1}{6}$.

We can define random variable using conditional expectation with respect to either a random variable OR a sigma-algebra.

Example 219. Simple partition of $\Omega = \bigcup_{i=1} A_i$ where the A_i are pairwise disjoint. Consider $\mathcal{A} = \{A_i\}$. Then we define as new random variable as:

$$\mathbf{E}[X|\mathcal{A}] = \sum_i \mathbf{E}[X|A_i]\mathbf{1}_{A_i}(\omega)$$

Note that if $\omega \in A_j$ then $\mathbf{E}[X|\mathcal{A}] = \mathbf{E}[X|A_j]$.

Example 220. Consider $A_i = \{w_1 = i :$ first dice roll is $i\}$. We have union of A_i ($i = 1$ to $i = 6$) is Ω and that the A_i are pairwise disjoint. Denote $\mathcal{A} = \{A_1, \dots, A_6\}$. Then:

$$\begin{aligned}\mathbf{E}[X|\mathcal{A}] &= \sum_{i=1}^{6} \mathbf{E}[X|A_i]\mathbf{1}_{A_i}(\omega) \\ &= \mathbf{E}[X|A_1]\mathbf{1}_{A_1}(\omega) + \mathbf{E}[X|A_2]\mathbf{1}_{A_2}(\omega) + \mathbf{E}[X|A_3]\mathbf{1}_{A_3}(\omega) + \dots \\ &= 4.5\mathbf{1}_{\omega_1=1}(\omega) + 5.5\mathbf{1}_{\omega_1=2}(\omega) + 6.5\mathbf{1}_{\omega_1=3}(\omega) + \dots\end{aligned}$$

The $4.5, 5.5, 6.5, \dots$ come from earlier example.

Example 221. Define $Y(\omega_1, \omega_2) = \omega_1$: the outcome I get in the first round. Then:

$$\mathbf{E}[X|Y] = Y(\omega) + 3.5$$

Definition 222. Consider a random variable X on $(\Omega, \mathcal{F}, \mathbf{P})$ and a sub sigma algebra $\mathcal{A} \subset \mathcal{F}$. Then the $Y := \mathbf{E}[X|\mathcal{A}]$ which can be defined as either the integral $\int_A Y d\mathbf{P} = \int_A X d\mathbf{P}$ if $A \in \mathcal{A}$ or equal to a random variable that is measurable on $\mathcal{A}$.

Example 223. If the joint pdf of random variables X and Y exists as $f_{X,Y}(x,y)$, and their own pdf is $f_X(x)$ and $f_Y(y)$. Then :

$$d\mathbf{P}(X|Y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

18 Day 18: Markov Process

18.1 Markov Process Definition

Intuitively, the **Markov Process** or **Markov Chain** is a stochastic process that is memoriless,i.e., the future of a stochastic process (which could depend on the past and present in general) only depends on the present.

Definition 224 (Markov Process - Markov Chain)**.** A Markov Process is a stochastic process $X(t, \omega)$ with $t \in T$ on a state space S (that is, $X(t, \omega) : \Omega \to S$). We may also assume state $S = \{1, 2, \dots\}$ countable for simplicity. The temporal space is T:

1. $T = \{0, 1, 2, \dots, n, \dots \infty\}$ is discrete: called **Discrete Markov Chain**

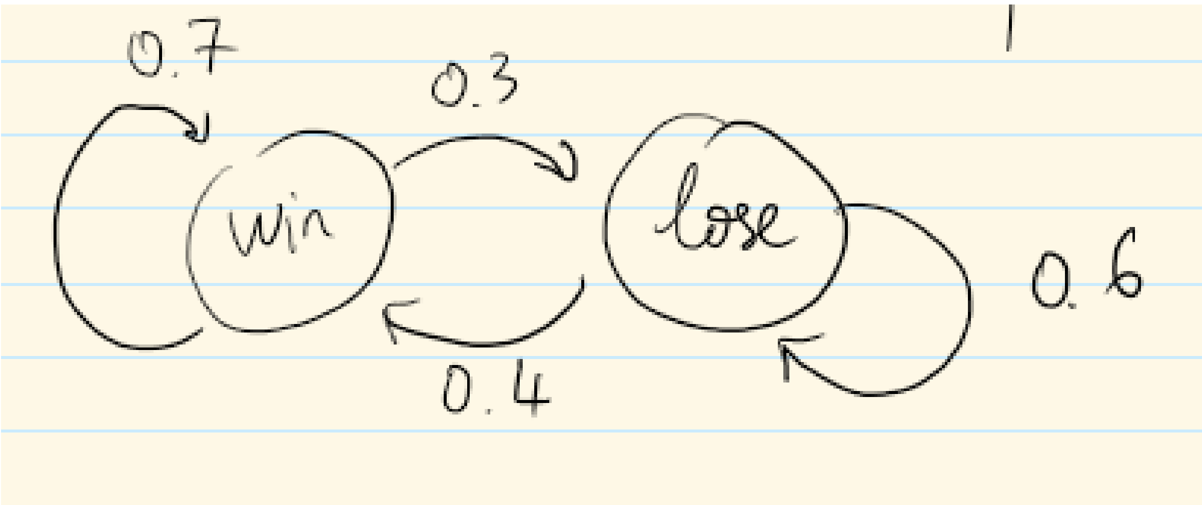


Figure 25: Markov Gambling example

2. When $T = [0, \infty)$ continuous, we called this a **Continuous Markov Chain**.

In addition, it must satisfied the Markov property. For simplicity, we will work with only the Markov property in discrete time case:

$$\mathbf{P}(X(n+1) = i_{n+1} | X(n) = i_n, \dots, X(i) = i_1) = \mathbf{P}(X(n+1) = i_{n+1} | X(n) = i_n)$$

where the $(i_1, \dots,) \in S$ That is, the next step state only depends on the current state.

Definition 225 (Markov property in discrete time)**.** For easier notation, we rewrite the Markov property in discrete time here, consider $X(t, \omega) := X_n(\omega)$ with $n \in T$:

$$\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1} \dots, X_0 = i_0) = \mathbf{P}(X_{n+1} = j | X_n = i)$$

The next time step state only depends on the current state.

Example 226. Gambling: $S = \{\text{win} = 1, \ \text{lose} = 2\}$

1. If I win, I have probability of win again is 0.7 and 0.3 of losing.

2. IF I lose, I have probability of win again is 0.4 and 0.6 of losing.

See figure (25) for clarification. Now

$$\begin{aligned} \mathbf{P}(X_{n+1} = j | X_n = i) &= \mathbf{P}(X_{n+1} = j | X_n = i) \\ &= \{p_{i,j}\} \text{ where } i, j \in S = \{1, 2\} \end{aligned}$$

We have this matrix:

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

We call P the transition matrix of this discrete Markov Chain.
Assume initially I have 0.5 change to win or lose: $= X_0 = [0.5, 0.5]$ Then: $X_1 = X_0 P$ and so $X_n = X_0 P^n$. Does the limit exits? Is there

$$\lim_{n \to \infty} X_n \stackrel{?}{=} X$$

In this example, numerically, yes, $X = [0.5714, 0.4786]$.

Example 227 (Random Walk in one dimension)**.** $X_0 = 0$, and $_k$ is either 1 step right or 1 step left, with $\mathbf{P}(\ _k = 1) = \mathbf{P}(\ = \ \ 1) = \frac{1}{2}$. Then $X_n = X_0 + \sum_{k=1}^{n} \ _k$. We have the following:

1. $p_{k;k+1} = \mathbf{P}(X_{n+1} = k + 1 j X_n = k) = \mathbf{P}(\ _{k+1} = 1) = \frac{1}{2}$
2. $p_{k;k+1} = j\mathbf{P}(\ _{k+1} = \ \ 1) = \frac{1}{2}$
3. $p_{i;j} = 0$, otherwise

Example 228 (Ehrenfest Chain)**.** Consider r balls on 2 sides. Each step, we choose 1 ball and move it to the other side. Let X_n denote the number of balls on the left side. Then

1. $p_{k;k+1} = \frac{r \ k}{r}$
2. $p_{k;k+1} = \frac{k}{r}$
3. $p_{i;j} = 0$ otherwise

18.2 Recurrent and Transience

Definition 229 (Recurrent and Transience)**.** Let X_n be a discrete Markov process that started from $X_0 = i$. Then

1. We say that state i is a **recurrent** if

$$\mathbf{P}(X_n(!) = i; \text{i.o.} j X_0 = i) = 1$$

2. We say that state i is **transient** if

$$\mathbf{P}(X_n(!) = i; \text{i.o.} j X_0 = i) = 0$$

Remark 230. We will always assume $X_0 = i$ from now on and just write $\mathbf{P}(\ ) = \mathbf{P}(\ j X_0 = i)$

Definition 231 (passage time)**.** We have the following definition:

1. T_i is called the **first passage time** to state i:

$$T_i := \inf f n \quad 1 : X_n(!) = i g$$

   Note that inf is taking on n.

2. is called the **r-th passage time to state i**: $T_i^{(0)} = 0$ and

$$T_i^{(r)} := \inf f n \quad (T_i^{(n \ \ 1)} + 1) : X_n(!) = i g$$

3. We called the **length of r-th visit to i**:

$$S_i^{(r)} = \begin{cases} T_i^{(r)} \quad T_i^{(r \ \ 1)} & \text{if } T_i^{(r)} < 1 \\ 1 & \text{otherwise} \end{cases}$$

Definition 232 (number of visit to state i)**.** The **number of visit to state i** is defined as:

$$V_i := \sum_{n=0}^{\infty} \mathbf{1}_{X_n(\omega)=i}$$

Remark 233. From the number of visit definition we have:

1. State i is **recurrent** if $\mathbf{P}(V_i = \infty) = 1$
2. State i is **transient** if $\mathbf{P}(V_i = \infty) = 0$

Definition 234 (return probability to i)**.** f_i is the return probability to i and is defined as:

$$f_i = \mathbf{P}(T_i < \infty) = \mathbf{P}(T_i < \infty | X_0(\omega) = i)$$

Remark 235. From the definition of return probability to i, we get:

1. $f_i = 0$ if and only if X_n is recurrent.
2. $f_i < 1$ if and only if X_n is transient

Definition 236 (returning back to i after n steps)**.** We define:

$$p_{i,i}^{(n)} := \mathbf{P}_i(X_n(\omega) = i) = \mathbf{P}(X_n(\omega) = i | X_0 = i)$$

as the returning back to i after n steps.

Theorem 237. *Given the definition of returning back to i after n steps above, we have:*

1. *State i is recurrent if* $\sum_{n=1}^{\infty} p_{i,i}^{(n)} = \infty$
2. *State i is transient if* $\sum_{n=1}^{\infty} p_{i,i}^{(n)} < \infty$

Proof. Consider V_i the number of visit to i. Then:

$$\begin{aligned}
\mathbf{P}_i(V_i > k) &= \mathbf{P}_i(T_i^{(k)} < \infty) \\
&= \mathbf{P}_i(S_i^{(k)} < \infty | T_i^{(k-1)} < \infty)\mathbf{P}(T_i^{(k-1)} < \infty) \\
&= \mathbf{P}_i(S_i^{(k)} < \infty | T_i^{(k-1)} < \infty)\mathbf{P}_i(S_i^{(k-1)} < \infty | T_i^{(k-2)} < \infty) \\
&= \dots = \text{ repeat this process of Markov chain} \\
&= \left[\mathbf{P}_i(T_i^1 < \infty)\right]^k = f_i^k
\end{aligned}$$

If state i is recurrent, then $f_i = 1$, so

$$\mathbf{P}_i(V_i = \infty) = \lim_{k \to \infty} \mathbf{P}_i(V_i > k) = \lim_{k \to \infty} f_i^k = 1$$

If state i is transient, then $f_i < 1$, so:

$$\mathbf{P}_i(V_i = \infty) = \lim_{k \to \infty} \mathbf{P}_i(V_i > k) = \lim_{k \to \infty} f_i^k = 0$$

If state i is either recurrent or transient:

$$
\begin{aligned}
\mathbf{E}_i[V_i] &= \mathbf{E}_i[\sum_{n=1}^{\infty} \mathbf{1}_{\{X_n = i\}}] \\
&= \sum_{n=1}^{\infty} \mathbf{P}_i(X_n = i) \\
&= \sum_{n=1}^{\infty} p_{i,i}^{(n)} \\
&= \sum_{n=1}^{\infty} \mathbf{P}_i(V_i > k) = \sum_{n=1}^{\infty} f_i^k
\end{aligned}
$$

1. If recurrent, then $f_i = 1$, so $sum_{n=1}^{\infty}\mathbf{P}_i(V_i > k) = \mathbf{E}_i[V_i] = \sum_{n=1}^{\infty} f_i^k = \infty$
2. if transient, then $f_i < 1$, so $sum_{n=1}^{\infty}\mathbf{P}_i(V_i > k) = \mathbf{E}_i[V_i] = \sum_{n=1}^{\infty} f_i^k < \infty$

□

Example 238. Consider random walk in 1 dimension with:$X_n = 0$, the ξ_k is either 1 step right or 1 step left, with $\mathbf{P}(\xi_k = 1) = p$, and $\mathbf{P}(\xi = -1) = 1 - p$. Also

$$X_n = X_0 + \sum_{k=1}^{n} \xi_k$$

Theorem 239. *For 1 dimension random walk:*

1. *If $p = \frac{1}{2}$, then all states are recurrent.*
2. *If $p \neq \frac{1}{2}$, then all states are transient.*

Proof. WLOG, we estimate for $i = 0$ and returning to zero.

1. When $p \neq \frac{1}{2}$:
$$\mathbf{E}[\xi_k] = 1.p + (-1)(1 - p) = 2p - 1 \neq 0$$
Then by Strong LLN:
$$\lim_{n \to \infty} \mathbf{E}[\frac{X_n}{n}] = \lim_{n \to \infty} \mathbf{E}[\frac{\xi_1 + \cdots \xi_n}{n}] \overset{as}{SLLN} 2p - 1 \neq 0$$
So then $X_n \overset{as}{\to} \infty$, and therefore $\mathbf{P}(X_n = 0 \text{ i.o.}) = 0$, and therefore state $i = 0$ is transient.
2. When $p = \frac{1}{2}$: So $\mathbf{E}[\xi_k] = 0$. Then:
$$p_{0,0}^{(n)} = \begin{cases} 0 & \text{if } n = 2k+1 \\ \left(\frac{1}{2}\right)^k \binom{2k}{k} & \text{if } n = 2k \end{cases}$$

Recall Stirlin Formula $n! \approx \sqrt{2\pi n}(\frac{n}{e})^n$. Apply this when $n = 2k$ we get:

$$p_{0,0}^{(n)} = \frac{1}{2^k}\frac{(2k)!}{k!k!} \approx \frac{1}{2^{2k}}\frac{\sqrt{2\pi 2k}}{\sqrt{2\pi k}\sqrt{2\pi k}}\frac{(2k)^{2k}}{k^k k^k}\frac{e^k e^k}{e^{2k}}$$
$$\approx \frac{1}{\sqrt{\pi k}}$$

Then:

$$\sum_{n=1}^{\infty} p_{0,0}^{(n)} \geq \sum_{k=1}^{\infty} p_{0,0}^{(k)} = \sum_{k=1}^{\infty} \frac{1}{\sqrt{\pi k}} = \infty$$

By previous theorem, we concluded that 0 is recurrent.

This completed the proof. □

Remark 240. Random Walk in d-dimension. X_n is d-dimensional lattice $\mathbb{Z}^d$. And $X_n = X_0 + \sum_{k=1}^n \xi_k$ Where $\mathbf{P}(\xi_k = e_j) = \mathbf{P}(\xi_k = -e_j) = \frac{1}{2d}$, where the e_j is the standard basic vector all 0 except at j coordinate. We get:

$$p_{k,k+e_1} = p_{k,k-e_1} = \cdots = p_{k,k+e_d} = p_{k,k-e_d} = \frac{1}{2d}$$

If the random walk is 1 or 2 dimension, all states are recurrent. If random walk is in 3 or higher dimension, all states are transient.

18.3 Filtration

Intuitively, if we have a discrete time Stochastic Process, then this is a sigma algebra $\sigma(X_1, \ldots, \xi_n)$: considering upto n-th time step.
If this is a continuous time stochastic process, we introduce the idea of **Filtration**.

Definition 241 (Filtration)**.** We have 2 definitions of Filtration:

1. $\{\mathcal{F}_t\}_{t\geq 0}$ is a filtration of $X(t)$ defined on $(\Omega, \mathcal{F}, P)$ if:
   $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s \leq t$, and $\mathcal{F}_t \subset \mathcal{F}$, and

$$\mathcal{F}_t := \sigma(\{X_s : 0 \leq s \leq t\})$$

2. Since BM is also a random variable, we have the definition of Filtration of BM:

$$\begin{aligned}
\mathcal{F}_t &= \sigma(\{B(s) : 0 \leq s \leq t\}) \\
&= \sigma(\{\omega \in \Omega : B(s,t) \in B \text{ for any Borel set } B \in \mathcal{B}(\mathbb{R}), 0 \leq s \leq t\}) \\
&= \sigma(\{\omega \in \Omega : B(t_j, \omega) \in B_j \text{ for any } \{t_j\}_{j=1}^n, 0 \leq t_j \leq t; \text{ for any } B_j \in \mathcal{B}(\mathbb{R})\})
\end{aligned}$$

Theorem 242 (Brownian Motion is a Markov process)**.** *For a fixed $s \geq 0$, $\{B(t+s) - B(s)\}_{t\geq 0}$ is a Brownian Motion starting at* 0.
In addition, the set $\{B(t+s) - B(s)\}$ is independent to $\mathcal{F}_t = \sigma(\{B(s) : 0 \leq s \leq t\})$

Proof. We try to show that $\sigma(\{B(t+s) - B(s); t \geq 0\})$ and $\sigma(\{B(\tau) : 0 \leq \tau \leq t\})$ are idnependent. The stratregy is to show that all generating sets are independent and they are π-system (recall, a π-system is when $A \in \Pi$, $B \in \Pi$, then $A \cap B \in \Pi$).

1. For any $t_1 \geq 0$: and for any $t_2 \leq s$. the $B(t_1+s) - B(s)$ and $B(t_2)$ are independent, since by definition of BM any increments are independent.

2. Show that $\{B(\tau) : 0 \leq \tau \leq t\}$ is a π-system. The:

$$A = \{B(t_j) \in B_j : 0 \leq t_j \leq s, B_j \in \mathcal{B}(\mathbb{R}), j = 1, \ldots, n\} \text{ is a } \pi\text{-system}$$

   And

$$B = \{B(\tau_i) \in E_i : 0 \leq \tau_i \leq s, E_i \in \mathcal{B}(\mathbb{R}), i = 1, \ldots, n\} \text{ is a } \pi\text{-system}$$

   Therefore A intersect B is also a finite projection generating set (and is a π-system).

□

19 Day 19: Martingale

19.1 More on Conditional Probability - with Filtration

Definition 243 (Conditional Expectation with sigma algebra)**.** Let X be a random variable on $(\Omega, \mathcal{F}, \mathbf{P})$. Let $\mathcal{F}_0$ be a sigma algebra that is $\mathcal{F}_0 \subseteq \mathcal{F}$. Then $Y := \mathbf{E}[X|\mathcal{F}_0]$ is a random variable:

1. Y is a $\mathcal{F}_0$-measurable

2.
$$\int_A Y d\mathbf{P} = \int_A X d\mathbf{P}, \text{ for any } A \in \mathcal{F}_0$$

   Equivalently:

$$\mathbf{E}[Y\mathbf{1}_A] = \mathbf{E}[X\mathbf{1}_A]$$

Example 244. Random variable X and $\mathcal{F}_0 = \mathcal{F}$. Then:

$$Y = \mathbf{E}[X|\mathcal{F}] \stackrel{as}{=} X$$

Example 245. X random variable, $\mathcal{F}_0 = \{\Omega, \emptyset\}$ Then:

$$Y = \mathbf{E}[X|\mathcal{F}_0] \stackrel{as}{=} \mathbf{E}[X]$$

We need to check condition 2 in the definition:

$$\int Y d\mathbf{P} = \int X d\mathbf{P} = \mathbf{E}[X]$$

Example 246. Let X random variable, and $\mathcal{F}_0$ is independent to X. That is, for any $B \in \mathcal{B}(\mathbb{R})$, $\{X \in B\}$ and $A \in \mathcal{F}_0$ is independent:

$$\mathbf{P}(\{X \in B\} \cap A) = \mathbf{P}(\{X \in B\})\mathbf{P}(A)$$

Let's check condition 2 in the definition, for any $A \in \mathcal{F}_0$:

$$\int_A Y d\mathbf{P} = \int_A X d\mathbf{P} = \int_A X\mathbf{1}_A d\mathbf{P} = \mathbf{E}[X]\mathbf{P}(A)$$

Therefore $Y \stackrel{as}{=} \mathbf{E}[X]$

Example 247. X random variable, $\mathcal{F}_0 = \{\emptyset, B, B^C, \Omega\}$ where B satisfying $0 < \mathbf{P}(B) < 1$. Then

$$Y(\omega) := \mathbf{E}[X|\mathcal{F}_0](\omega) = \begin{cases} \left(\int_B X d\mathbf{P}\right)\frac{1}{\mathbf{P}(B)} & \text{if } \omega \in B \\ \left(\int_{B^C} X d\mathbf{P}\right)\frac{1}{\mathbf{P}(B^C)} & \text{if } \omega \notin B \end{cases}$$

Check the condition 2 of definition:

$$\begin{aligned} \int_B Y d\mathbf{P} &= \int_B \int_B \frac{X d\mathbf{P}}{\mathbf{P}(B)} d\mathbf{P} \\ &= \int_B \frac{\mathbf{E}[X\mathbf{1}_B]}{\mathbf{P}(B)} d\mathbf{P} \\ &= \mathbf{E}[X\mathbf{1}_B]\frac{\mathbf{P}(B)}{\mathbf{P}(B)} = \mathbf{E}[X\mathbf{1}_B] \end{aligned}$$

Example 248. X and Z are discrete, say X is $x_1, \ldots, x_n$ and Z is $z_1, \ldots, z_n$. Then:

$$Y = \mathbf{E}[X|\sigma(Z)] = \mathbf{E}[X|\sigma(\{z_1, \ldots, z_n\})]$$

Equivalently:

$$y_i = Y(\omega) = \mathbf{E}[X|Z = z_i]$$

for any $\omega \in G_i = \{Z = z_i\}$.
We now check condition 2 of definition:

$$\begin{aligned} \mathbf{E}[Y\mathbf{1}_{G_i}] &= y_i \mathbf{P}(G_i) \\ &= \mathbf{E}[X|G_i]\mathbf{P}(G_i) \\ &= \sum_{j=1}^{n} x_j \frac{\mathbf{P}(X = x_j, Z = z_j)}{\mathbf{P}(Z = z_j)} . \mathbf{P}(Z = z_j) \\ &= \sum_{j=1}^{n} x_j \mathbf{P}(X = x_j, Z = z_j) \\ &= \sum_{j=1}^{n} x_j \mathbf{1}_{G_j} = \mathbf{E}[X\mathbf{1}_{G_j}] \end{aligned}$$

Definition 249 (Filter of probability space)**.** Consider $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n \geq 1}, \mathbf{P})$ in the discrete case. Or $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbf{P})$ in the continuous case, in which $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space.

1. The $\{\mathcal{F}_n\}_{n \geq 1}$ is a sigma algebra and is a filtration such that: $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \mathcal{F}_\infty = \bigcup_{i=1}^{\infty} \mathcal{F}_i \subseteq \mathcal{F}$

2. in the continuous case, the $\{\mathcal{F}_t\}_{t \geq 0}$ is a filtration such that:

$$\mathcal{F}_s \subseteq \mathcal{F}_t$$

   for all $s \leq t$ and $\mathcal{F}_t \subseteq \mathcal{F}$.

Then the X_n is **adapted** to $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n \geq 1}, \mathbf{P})$ in the most natural way: the X_n is $\mathcal{F}_n$-measurable.
Note that another way to define filtration is $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$

19.2 Martingale

Definition 250 (Martingale)**.** A **Martingale** $\{M_n\}$ is a stochastic process satisfying:

1. adapted to $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n \geq 1}, \mathbf{P})$
2. $\mathbf{E}[|M_n|] < \infty$
3. $\mathbf{E}[M_{n+1}|\mathcal{F}_n] \stackrel{as}{=} M_n$

Remark 251. In the condition 3 above,

1. M_n is called a **sub-martingale** if:

$$E[M_{n+1}|\mathcal{F}_n] \geq M_n$$

2. M_n is called a **super-martingale** if:

$$E[M_{n+1}|\mathcal{F}_n] \leq M_n$$

Theorem 252 (Random Walk is a Martingale)**.** *Let ξ_k be iid and that $\mathbf{E}[\xi_k] = 0$. Set $X_0 = 0$ (or a constant), then: $X_n = X_0 + \sum_{i=1}^{n} \xi_i$ is a martingale.*

Proof. We check

1. $E[|x_n|] = 0 < \infty$
2. For this:

$$\begin{aligned}
\mathbf{E}[X_{n+1}|\mathcal{F}_n] &= \mathbf{E}[\xi_{n+1} + X_n|\mathcal{F}_n] \\
&= \mathbf{E}[\xi_{n+1}|\mathcal{F}_n] + \mathbf{E}[X_n|\mathcal{F}_n] \\
&= \mathbf{E}[\xi_{n+1}] + X_n \\
&\text{since } \xi_{n+1} \text{ is independent with } \mathcal{F}_n \text{, and } X_n \text{ is } \mathcal{F}_n\text{-measurable} \\
&= 0 + X_n = X_n \text{ since } \mathbf{E}[\xi_{n+1}] = 0
\end{aligned}$$

Therefore Random Walk is a martingale. □

Example 253. With $M_n = X_n^2 - n\sigma^2$ where $\mathbf{E}[\xi_k] = 0$ and $\mathrm{Var}(\xi_k) = \sigma^2 < \infty$:

$$\begin{aligned}
\mathbf{E}[M_{n+1}|\mathcal{F}_n] &= \mathbf{E}[X_{n+1}^2 - (n+1)\sigma^2|\mathcal{F}_n] \\
&= \mathbf{E}[\xi_{n+1}^2 - 2\xi_{n+1}X_n + X_n^2 - (n+1)\sigma^2|\mathcal{F}_n] \\
&= \mathbf{E}[\xi_{n+1}^2] + 2\mathbf{E}[\xi_{n+1}]\mathbf{E}[X_n|\mathcal{F}_n] + \mathbf{E}[X_n^2|\mathcal{F}_n] - (n+1)\sigma^2 \\
&= \sigma^2 + 0.X_n + X_n^2 - (n+1)\sigma^2 \\
&= X_n^2 - n\sigma^2 = M_n
\end{aligned}$$

So M_n is a martingale.

Remark 254. Following the above we could also see that X_n^2 is a sub-martingale.

Example 255 (Product Martingale)**.** We let $\mathbf{E}[\xi_k] = 1$ and $M_n = \prod_{i=1}^{n} \xi_i$. Then:

$$\begin{aligned}
\mathbf{E}[M_{n+1}|\mathcal{F}_n] &= \mathbf{E}[\xi_{n+1}M_n|\mathcal{F}_n] \\
&= \mathbf{E}[\xi_{n+1}|\mathcal{F}_n] + \mathbf{E}[M_n|\mathcal{F}_n] \\
&= \mathbf{E}[\xi_n]M_n = 1.M_n = M_n
\end{aligned}$$

Definition 256 (Stopping Time)**.** The **Stopping Time** is $N(\omega) \in \mathbb{N}$ such that the $\{N(\omega) = n\}$ is F_n -measurable.

Example 257. Couple possible stopping time:

1. $N(\omega) = 5$
2. $N(\omega) = \inf\{n : X_n \geq 0\}$
3. $N(\omega) = \inf\{n : X_n \geq 0\} \wedge 500$

20 Appendix

Lemma 258. Suppose $c_n \to c$. Then we always have:

$$(1 + \frac{c_n}{n})^n \to e^c$$

Proof. We use an identity:

$$w_1 w_2 \cdots w_n - z_1 z_2 \cdots z_n = (w_1 - z_1) z_2 \cdots z_n + w_1 (w_2 - z_2) z_3 \cdots z_n + \cdots + w_1 \cdots w_{n-1} (w_n - z_n)$$

The idea of the above identity is to $\pm w_1 z_2 \cdots z_n \pm w_1 w_2 z_3 \cdots z_n \pm \cdots$. Now we apply the identity to

$$w_j = 1 + \frac{c_n}{n} \text{ and } z_j = e^{\frac{c_n}{n}}$$

$$(1 + \frac{c_n}{n})^n - e^{\frac{c_n}{n} n} = (1 + \frac{c_n}{n} - e^{\frac{c_n}{n}}) e^{\frac{c_n}{n}(n-1)} + (1 + \frac{c_n}{n})(1 + \frac{c_n}{n} - e^{\frac{c_n}{n}}) e^{\frac{c_n}{n}(n-2)} \cdots +$$
$$= \sum_{j=1}^{n} (1 + \frac{c_n}{n})^{j-1} (1 + \frac{c_n}{n} - e^{\frac{c_n}{n}}) e^{\frac{c_n}{n}(n-j)}$$

Now in the above sum, we take a look at the term $(1 + \frac{c_n}{n})$. Since $c_n \to c$, there exists a constant M such that $|c_n| \leq M$ for all n(converging sequence is bounded). Therefore:

$$|1 + \frac{c_n}{n}|^{j-1} \leq (1 + \frac{M}{n})^{j-1} \leq (1 + \frac{M}{n})^j \leq e^{\frac{M}{n} j}$$
$$\text{the last inequality is true because } 1 + x \leq e^x, \forall x \in \mathbb{R}$$

Then we now have:

$$(1 + \frac{c_n}{n})^n - e^{\frac{c_n}{n} n} \leq \sum_{j=1}^{n} e^{\frac{M}{n} j} (1 + \frac{c_n}{n} - e^{\frac{c_n}{n}}) e^{\frac{M}{n}(n-j)}$$
$$= \sum_{j=1}^{n} (1 + \frac{c_n}{n} - e^{\frac{c_n}{n}}) e^{\frac{M}{n} n}$$

We take n large enough so that $|\frac{c_n}{n}| \leq 1$. We have a little claim: for any z with $|z| \leq 1$, we have:

$$|e^z - (1 + z)| = |1 + z + \frac{z^2}{2} + \cdots - (z + 1)| = |\sum_{k=2}^{\infty} \frac{z^k}{k!}|$$
$$\leq |z|^2 \sum_{k=1}^{\infty} \frac{1}{k!} \leq |z|^2$$

That is

$$je^z \quad (1+z)j \quad jzj^2$$

Applying that claim to $z = \frac{c_n}{n}$:

$$(1+\frac{c_n}{n})^n \quad e^{\frac{c_n}{n}n} \quad n\frac{jc_nj^2}{n^2}e^M = \frac{jc_nj^2}{n}e^M \quad {}^{n\,!\,1}\!\!! \;\; 0$$

This completed the lemma proof. □

Lemma 259. For any real number y, we always have:

$$je^{ity} \quad (1+iy+\frac{i^2y^2}{2})j \quad \min f\frac{jyj^3}{6}; jyj^2 g$$

Proof. We use Taylor's series integral form of a function f that is $(k+1)$ times differentiable:

$$f(x) = \sum_{j=0}^{k} \frac{f^{(j)}}{j!}(x \quad a)^j + \int_a^x \frac{f^{(k+1)}}{k!}(x \quad t)^k dt$$

We apply this for $f(y) = e^{iy}$ for 2 cases: $k = 2$ and $k = 1$.(and the point $a = 0$ in both cases):

1. Case $k = 2; a = 0$:

$$e^{iy} = 1 + iy + \frac{t^2}{2}y^2 + \int_0^y \frac{i^3e^{iy}}{3!}(y \quad t)^2 dt$$

Then substract and take absolute value gives:

$$\begin{aligned} je^{iy} \quad (1+iy+\frac{t^2}{2}y^2)j &= j\int_0^y \frac{i^3e^{iy}}{3!}(y \quad t)^2 dtj \\ & \int_0^y \frac{ji^3e^{iy}j}{6}(y \quad t)^2 dt \\ &= \frac{1}{6}\int_0^y (y \quad t)^2 dt \text{ ,since } ji^3e^{iy}j = 1 \\ &= \frac{y^3}{6} \end{aligned}$$

2. Case $k = 1; a = 0$:

$$e^{iy} = 1 + iy + \int_0^y \frac{i^2e^{iy}}{2!}(y \quad t)dt$$

Then:

$$\begin{aligned} je^{iy} \quad (1+iy)j &= j\int_0^y \frac{i^2e^{iy}}{2!}(y \quad t)dtj \\ & \int_0^y \frac{ji^2e^{iy}j}{2}jy \quad tjdt \\ &= \int_0^y (y \quad t)dt, \text{ since } ji^2e^{iy}j = 1 \\ &= \frac{y^2}{2} \end{aligned}$$

Therefore:

$$je^{iy} \quad (1+iy+\frac{i^2y^2}{2})j \quad je^{iy} \quad (1+iy)j+j\frac{i^2y^2}{2}j=\frac{y^2}{2}+\frac{y^2}{2}=y^2$$

So the quantity $je^{iy} \quad (1+iy+\frac{i^2y^2}{2})j$ must be less than or equal to the minimum of either $\frac{y^3}{6}$ or y^2.

□

References

[1] R. Durrett. (2010) , *Probability: Theory and examples*, Cambridge University Press.