

Lecture 5: Measure Functions

Simple function (simple random variable)

Let $(\Omega, \mathcal{F}, P)$ be a probability space i.e. $P(\Omega) = 1$

Simple random variable $X: \Omega \rightarrow \mathbb{R}$

$\mu: \mathcal{F} \rightarrow \mathbb{R}^+$
 $\hookrightarrow X(\omega)$ only takes a finite # of values $\{x_1, \dots, x_p\}$
 also, the set $\{\omega \in \Omega: X(\omega) = x_i\} \in \mathcal{F}$

Let $\{A_i\}_{i=1}^p, A_i \in \mathcal{F}$ be a disjoint partition of Ω ,
 $\bigcup_{i=1}^p A_i = \Omega, A_i \cap A_j = \emptyset$ if $i \neq j$

$$X(\omega) = \sum_{i=1}^p x_i \mathbb{1}[\omega \in A_i]$$

We can then say that the probability that X is equal to x_i is $P(X = x_i) = P(\{\omega \in \Omega: X(\omega) = x_i\}) = P(A_i)$

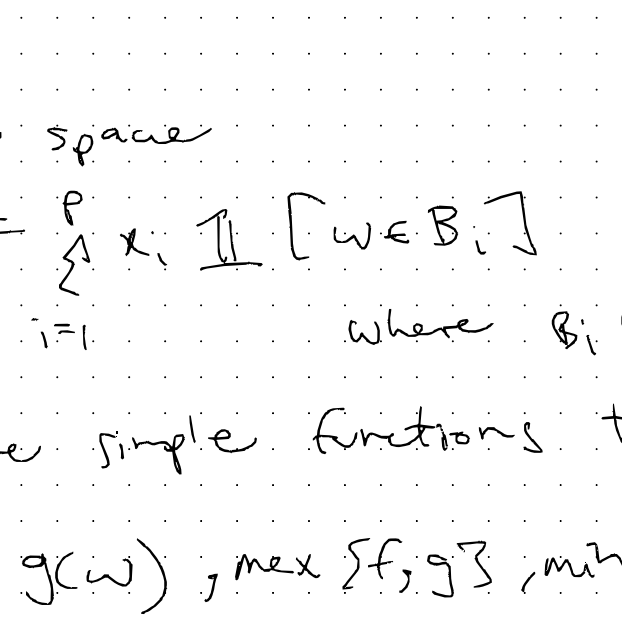
We can also define the expectation of X to be

$$EX = \sum_{i=1}^p x_i P(X = x_i)$$

Example: $\Omega = (0, 1]$

$$x_i = \frac{i-1}{4}$$

$$EX = 0.375$$



Exercise: if $f, g: \Omega \rightarrow \mathbb{R}$ are simple functions then
 so are $f+g, f \cdot g = f(\omega)g(\omega), \max\{f, g\}, \min\{f, g\}$
 are also simple functions

We define the integral of a simple function to be

$$\int f d\mu := \sum_{i=1}^p x_i \mu(B_i)$$

Exercise: For $f, g: \Omega \rightarrow \mathbb{R}^+$, simple functions, show that the integral is a linear operation

- $\int (f+g) d\mu = \int f d\mu + \int g d\mu$
- $\int cf d\mu = c \int f d\mu$ for some scalar $c > 0$

Measurable Functions and Random Variables

We need two measurable spaces

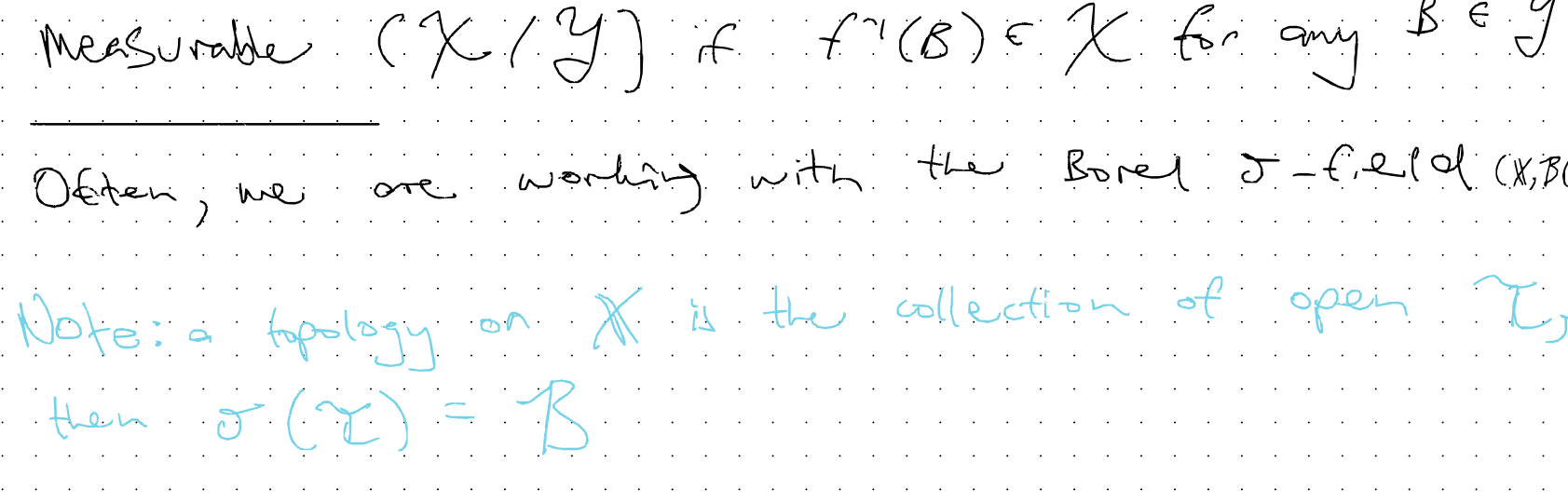
$$(\mathcal{X}, \mathcal{X}), (\mathcal{Y}, \mathcal{Y})$$

bold $\mathcal{X}$ script $\mathcal{X}$

Say $f: \mathcal{X} \rightarrow \mathcal{Y}$ and $A \in \mathcal{X}$ and $B \in \mathcal{Y}$

$$f(A) = \{y \in \mathcal{Y}: \exists x \in A \text{ s.t. } f(x) = y\}$$

$$f^{-1}(B) = \{x \in \mathcal{X}: \exists y \in B \text{ s.t. } f(x) = y\}$$



Definition: a function $f: \mathcal{X} \rightarrow \mathcal{Y}$ is said to be measurable $(\mathcal{X}/\mathcal{Y})$ if $f^{-1}(B) \in \mathcal{X}$ for any $B \in \mathcal{Y}$

Often, we are working with the Borel σ -field $\mathcal{B}(\mathbb{R})$

Note: a topology on $\mathcal{X}$ is the collection of open $\mathcal{U}$, then $\sigma(\mathcal{U}) = \mathcal{B}$

Secondly, $(\mathcal{Y}, \mathcal{Y})$ is often $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}^+))$$

$\downarrow$ in this case, we'd say that f is Borel measurable

Similarly, we can replace $\mathcal{B}(\mathbb{R})$ with $\mathcal{M}_2(\mathbb{R})$ to get Lebesgue measurable functions.

Good facts about Measurable Functions

1. Inverse images of set functions preserve set operations

i.e. for $f: \mathcal{X} \rightarrow \mathcal{Y}$ and $A, A_i \in \mathcal{Y}, i \in I$ then

$$f^{-1}\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f^{-1}(A_i) \quad \leftarrow \text{do unions}$$

$$f^{-1}(\mathcal{Y} \setminus A) = \mathcal{X} \setminus f^{-1}(A) \quad \leftarrow \text{do complements}$$

Hint: if I want to show that sets A, B are equal, then show that both $A \subseteq B$ and $B \subseteq A$ hold

Important: checking whether or not f is measurable can be done for $A \in \mathcal{A}$ where $\sigma(\mathcal{A}) = \mathcal{Y}$ instead of $\mathcal{U}$ sets in $\mathcal{Y}$.

e.g. let $A_t = (-\infty, t]$ $\forall t \in \mathbb{R}$ these sets generate $\mathcal{B}(\mathbb{R})$

$\Rightarrow f$ is measurable as long as $\{x: f(x) \leq t\}$ are measurable

2. for any $A \in \mathcal{X}$, the indicator function $f(x) = \mathbb{1}[x \in A]$ is measurable. The σ -field generated by f^{-1} is simply

$$\{\emptyset, A, A^c, \mathcal{X}\} \subset \mathcal{X}$$

3. for measurable functions $f, g: \mathcal{X} \rightarrow \mathbb{R}$ then $f+g$ and $f \cdot g$ are both measurable (talk about it later with product measures)

4. for measurable functions $\{f_i\}_{i=1}^\infty, \mathcal{X} \rightarrow \mathbb{R}$ then the following are also measurable:

$$\bullet \sup_i f_i \quad \bullet \inf_i f_i \quad \bullet \limsup_i f_i \quad \bullet \liminf_i f_i$$

and also $\lim_{i \rightarrow \infty} f_i$ if it exist

Proof sketch: In set notation $\{x: \sup_i f_i \leq t\} = \bigcap_{i=1}^\infty \{x: f_i(x) \leq t\}$

a countable intersections of measurable sets is measurable

• for $\inf$, do the same thing noting that $\{x: \inf_i f_i \leq t\} = \bigcup_{i=1}^\infty \{x: f_i(x) \leq t\}$

$$= \bigcup_{i=1}^\infty \{x: f_i(x) \leq t\}$$

$$\bullet \limsup_i f_i = \inf_i \sup_{j \geq i} f_j$$

$$\bullet \liminf_i f_i = \sup_i \inf_{j \geq i} f_j$$

• if $\lim_i f_i$ exists, then $\limsup_i f_i = \lim_i f_i = \liminf_i f_i \quad \square$

5. let $f: \mathcal{X} \rightarrow \mathbb{R}$ be continuous, then it is $\checkmark$ Borel measurable

Proof: A function between two topological is said to be continuous if the inverse image of an open set is open

i.e. $u \in \mathbb{R}$ is open then $f^{-1}(u)$ is open in $\mathcal{X}$

$\therefore f^{-1}(u)$ is measurable

Since the open sets generate $\mathcal{B}(\mathbb{R})$ we get that f is measurable $\square$

Note that $\sigma(\{f^{-1}(u): u \in \mathbb{R}, \text{open}\}) \subseteq \mathcal{B}(\mathcal{X})$

6. Given a collection of functions $f_i: \mathcal{X} \rightarrow \mathcal{Y}$

we can make them measurable by constructing

$(\mathcal{X}, \mathcal{X})$ where $\sigma(\{f_i\}_{i \in I}) \subseteq \mathcal{X}$, where this is the sigma-field constructed from the sets $f_i^{-1}(B)$

$\forall i, \forall B \in \mathcal{Y}$

Definition: (Almost everywhere / Almost surely)

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space

For $f, g: \Omega \rightarrow \mathbb{R}$, we say that $f = g$ a.e

when the set $N = \{\omega: f(\omega) \neq g(\omega)\}$

$$\mu(N) = 0$$

In probability, we say a.s. or w.p. 1

for "with probability 1"

Example: Two functions being equal almost everywhere

Let $((0, 1], \mathcal{B}, \lambda)$ be a measure space. let $f(t) = 0 \quad \forall t \in (0, 1]$.

$$\text{let } g(t) = \begin{cases} 0 & \forall t \in (0, 1] \setminus \mathbb{Q} \\ 1 & \forall t \in (0, 1] \cap \mathbb{Q} \end{cases}$$



Claim: $f = g$ a.e.

that is, $\lambda((0, 1] \cap \mathbb{Q}) = 0$

Show $\lambda(\mathbb{Q}) = 0$. Try enumerating $\{p_m\}_{m=1}^\infty$

to be the rational #'s and surrounding each by an interval $(p_m - \epsilon 2^{-m-1}, p_m + \epsilon 2^{-m-1})$

for any $\epsilon > 0$

Intro: Monotone Convergence, Fatou's Lemma, Lebesgue integration for Dominated convergence measurable functions

We will be considering measurable functions $f: \Omega \rightarrow [-\infty, \infty]$ where $(\Omega, \mathcal{F})$ is a measurable space.

We include $\pm\infty$ so we can consider sets like $f^{-1}(\infty) \subseteq \Omega$

Notation: $f_i \uparrow f, i \rightarrow \infty$ this means $f_i(\omega) \uparrow f(\omega)$

Given Ω and $f_i(\omega) \leq f_{i+1}(\omega)$

Theorem: Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{A}$ a π -system s.t. $\sigma(\mathcal{A}) = \mathcal{F}$ *converging from the space and $\mathcal{A}$ a π -system s.t. $\sigma(\mathcal{A}) = \mathcal{F}$ within up. important idea*
Let V be a linear space that contains

- (i) all indicators $\mathbb{1}_A$ and $\mathbb{1}_B$ for each $A \in \mathcal{A}$
- (ii) all functions f such that there exists a sequence $f_i \in V$ such that $f_i \uparrow f$

Then, V contains all measurable functions.

Proof: $\mathbb{1}_A \in V$ Let $\mathcal{L} = \{B \in \mathcal{F} : \mathbb{1}_B \in V\}$

Claim is that $\mathcal{L}$ is a σ -system *(skipping the details)*

$$\therefore \mathcal{A} \subseteq \mathcal{L} \Rightarrow \sigma(\mathcal{A}) = \mathcal{F} \subseteq \mathcal{L}$$

$$\therefore \mathcal{L} = \mathcal{F}$$

$$\therefore \mathbb{1}_B \in V \quad \forall B \in \mathcal{F}$$

Next, let f be a non-negative measurable function. we can define a sequence

$$f_i = 2^{-i} \lfloor 2^i f \rfloor$$

f_i are simple functions

$\hookrightarrow$ linear combination of Indicator Functions

$$\therefore f_i \in V$$

Furthermore, $f_i \uparrow f \therefore f \in V$.

Lastly, for a general measurable f we write $f = f^+ - f^-$ where f^+, f^- are non-negative measurable functions

$$f^+(\omega) = \begin{cases} f(\omega) & \text{if } f(\omega) \geq 0 \\ 0 & \text{else} \end{cases}$$

$$f^-(\omega) = \begin{cases} -f(\omega) & \text{if } f(\omega) \leq 0 \\ 0 & \text{else} \end{cases}$$

$$\therefore f \in V$$

Definition: (Integral of a measurable function)

For a non-neg measurable $f: \Omega \rightarrow [0, \infty]$ we define *($\Omega, \mathcal{F}, \mu$) is a measure space*

$$\int f d\mu = \sup \left[\sum_{i \in I} \inf_{\omega \in A_i} f(\omega) \mu(A_i) \right]$$

where the sup is taken over all finite partitions $\{A_i\}_{i \in I}$ of Ω

Inside [...] we have $\times$ Simple function!

$\therefore \int f d\mu$ is a sup over all simple g s.t. $0 \leq g \leq f$

For general functions f , we again write $f = f^+ - f^-$ and

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu$$

(dealing with ∞)

$$\text{In this course, } \bullet \cdot 0 \times \infty = 0$$

$$\bullet \cdot \infty \times \infty = \infty \text{ for any } \infty > 0$$

e.g. if $f = 0$ on A where $\mu(A) = \infty$ we still want

$$\int f d\mu = 0$$

Also, $\infty - \infty$ is undefined if both $\int f^+ d\mu$ and $\int f^- d\mu$ are finite we say that f is integrable

$$\text{(sometimes written as } f \in L^1(\Omega, \mathcal{F}, \mu) \text{)}$$

(pre-cursor) " L^1 spaces"

Theorem: (Monotone Convergence for simple functions)

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f \geq 0$ be measurable

and $f_n \geq 0$ then $\mathbb{N}$ be measurable simple functions s.t. $f_n \uparrow f$

Then, $\int f_n d\mu \uparrow \int f d\mu$.

(Note: if $f_i \rightarrow f$ then we may or may not have $\int f_i d\mu \rightarrow \int f d\mu$)

Proof: Let g be any simple measurable function s.t. $0 \leq g \leq f$

Since $f_n \uparrow f$ and f is integrable Then $\int f_n d\mu \uparrow c \in [0, \infty]$ *(c is finite)*

Goal is to show that $\int g d\mu \leq c$

First write $g = \sum_{i \in I} \alpha_i \mathbb{1}_{A_i}$ where A_i are a disjoint partition of Ω

and similarly $f_n = \sum_{i \in J} b_i \mathbb{1}_{B_i}$

$$\therefore f_n = \sum_{i \in I} f_n \mathbb{1}_{A_i} = \sum_{i \in I} b_{ij} \mathbb{1}_{A_i \cap B_j} \quad \text{and} \quad \int f_n d\mu = \sum_{i \in I} \int_{A_i} f_n d\mu$$

Hence we need to show that for each $i \in I$ that

$$\lim_{n \rightarrow \infty} \int_{A_i} f_n d\mu \geq \alpha_i \mu(A_i) \iff$$

First, if $\alpha_i = 0$ then we're done!

If $\alpha_i > 0$ then we can divide by α_i and wlog we take

$$\alpha_i = 1 \text{ and consider } g = \mathbb{1}_{A_i} \iff$$

Now, for any $\varepsilon > 0$, $C_n = \{x \in A_i : f_n(x) > 1 - \varepsilon\}$

Then $C_n \uparrow A_i$ which means $C_1 \subseteq C_2 \subseteq \dots \subseteq A_i$, $\bigcup_{i=1}^{\infty} C_n = A_i$ *increasing sequence of simple functions*

By countable additivity of μ

$$\mu(C_{n+1}) = \mu(C_1) + \sum_{m=1}^n \mu(C_{m+1} \setminus C_m)$$

$$\text{and } \mu(C_n) \uparrow \mu(A_i)$$

Since $\int_{C_n} f_n d\mu \geq (1 - \varepsilon) \mu(C_n)$ we have that

$$c \geq (1 - \varepsilon) \mu(A_i)$$

$$\geq (1 - \varepsilon) \int g d\mu$$

$\therefore$ take $\varepsilon \rightarrow 0$ to get $\int g d\mu \leq c$

as this holds for any simple $g \leq f$ we have that $\int f d\mu \leq c$

$$\text{But also } \int f_n d\mu \leq \int f d\mu \quad \therefore \int f d\mu = c$$

$\square$

Theorem: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f, g: \Omega \rightarrow [-\infty, \infty]$ be measurable functions, and $f = g$ a.e. Then f is integrable *"extended and lin"*

iff g is integrable and if this is the case then $\int f d\mu = \int g d\mu$

Proof: Let $f = g$ on $\Omega \setminus N$ where $\mu(N) = 0$ for any measurable

$h: \Omega \rightarrow [-\infty, \infty]$ we can consider $\int_{\Omega} h d\mu$ and $\int_{\Omega \setminus N} h d\mu$

$\rightarrow$ True if h is an indicator function because $\mu(A) = \mu(A \setminus N)$

$\rightarrow \therefore$ also true for h being a

simple function

$\rightarrow$ from previous theorem, take non-neg

simple $\uparrow$ non-neg measurable

$\rightarrow$ both write $\int h d\mu = \int h^+ d\mu - \int h^- d\mu$

$$\therefore \int_{\Omega} f d\mu = \int_{\Omega} f^+ d\mu - \int_{\Omega} f^- d\mu = \int_{\Omega} g^+ d\mu - \int_{\Omega} g^- d\mu = \int_{\Omega} g d\mu \quad \square$$

Three Important Convergence Theorems.

(for X_i a sequence of measurable random variables then if $X_i(\omega) \rightarrow X(\omega)$ does $\mathbb{E} X_i \rightarrow \mathbb{E} X$?)

$\hookrightarrow$ not necessarily.

Theorem: (Monotone Convergence)

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\{f_i\}_{i=1}^{\infty}$ be measurable functions from $\Omega \rightarrow [-\infty, \infty]$ s.t. $f_i \uparrow f$ μ -a.e. and $\int f_i d\mu > -\infty$ *monotone conv for simple fns*

Then $\int f_i d\mu \uparrow \int f d\mu$

Proof: We didn't assume that f is measurable! but it is!

$\Rightarrow$ For $c \in \mathbb{R}$, consider the sets $(c, \infty]$ which generate the Borel σ -field $\mathcal{B}$

$$\text{Since } f_i \uparrow f, f^{-1}((c, \infty]) = \bigcup_{i=1}^{\infty} f_i^{-1}((c, \infty])$$

$$\text{and } f_i^{-1}((c, \infty]) \in \mathcal{F} \therefore f \text{ is measurable}$$

Next, assume $f_i \geq 0$ and for each f_i we take simple functions

g_{ij} s.t. $g_{ij} \uparrow f_i$ as $j \rightarrow \infty$.

$\therefore$ by a previous theorem: $\int g_{ij} d\mu \uparrow \int f_i d\mu$

Furthermore, define $g_i^* = \max\{g_{i1}, \dots, g_{in}\}$

$$f_i \quad g_{i1} \quad g_{i2} \quad \dots \quad g_{in}^* \quad \text{also is a simple}$$

and $g_i^* \uparrow f$ so once again $\int g_i^* d\mu \uparrow \int f d\mu$. But $g_i^* \leq f_i$

(by construction) $\therefore \int g_i^* d\mu \leq \int f_i d\mu \leq \int f d\mu \Rightarrow \int f_i d\mu \uparrow \int f d\mu$

Done for non-neg functions ($f_i \geq 0$)

Now, assume that $f \leq 0$, $\therefore f_i \uparrow f$ implies $-f_i \downarrow -f$

$$\text{Let } h = -f \text{ and } h_i = -f_i$$

$$\text{Now, } 0 \leq \int h_i d\mu \leq \int h d\mu$$

Next, we note that $0 \leq h_i - h_j \uparrow h_i - h$

$$\therefore \text{from above } \int (h_i - h_j) d\mu \uparrow \int (h_i - h) d\mu$$

by subtraction *(Note all integrals are finite)*

$$\int h_i d\mu \downarrow \int h d\mu \text{ and } \int f_i d\mu \uparrow \int f d\mu$$

for a general $f = f^+ - f^-$, we have $f_i^+ \uparrow f^+$ and $f_i^- \downarrow f^-$

and by assumption $\int f^- d\mu < \infty$

$$\therefore \text{from above } \int f_i^+ d\mu \uparrow \int f^+ d\mu$$

$$\int f_i^- d\mu \downarrow \int f^- d\mu$$

$$\therefore \int f_i d\mu \uparrow \int f d\mu \quad \square$$

Note: we can also prove a monotone convergence theorem for $f_i \downarrow f$ s.t. $\int f_i d\mu < \infty$

finiteness condition

Theorem (Fatou's Lemma):

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space

let $\{f_i\}_{i=1}^{\infty}$ be non-neg measurable $\Omega \rightarrow [-\infty, \infty]$

Then,

$$\int \liminf f_i d\mu \leq \liminf \int f_i d\mu$$

Proof: recall $\liminf f_i = \sup_{i \rightarrow \infty} \inf_{j \geq i} f_j$

Hence, let g_j be the $\inf\{f_i : i \geq j\}$

$$\text{Then } g_j \uparrow \liminf f_i \text{ and } f_i \geq 0$$

$$\therefore \text{Monotone convergence says } \int g_j d\mu \uparrow \int \liminf f_i d\mu$$

By construction, $g_j \leq f_i$ for any $i \geq j$

$$\therefore \int g_j d\mu \leq \int f_i d\mu \quad \forall i \geq j$$

$$\therefore \int g_j d\mu \leq \inf_{i \geq j} \int f_i d\mu$$

take $j \rightarrow \infty$ to get $\lim_{j \rightarrow \infty} \int g_j d\mu = \int \liminf f_i d\mu \leq \liminf_{i \rightarrow \infty} \int f_i d\mu$

$\square$

Theorem (Dominated Convergence)

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space

Let $\{f_i\}_{i=1}^{\infty}$ and g be measurable functions and absolutely integrable $\int |g| d\mu < \infty$

If $|f_i| \leq g$ μ -a.e. and $f_i(\omega) \rightarrow f(\omega)$ for each $\omega \in \Omega$ then

f is absolutely integrable and $\int f_i d\mu \rightarrow \int f d\mu$

Proof: Let $f_i^{\wedge} = \inf\{f_j : j \geq i\}$

$$f_i^{\vee} = \sup\{f_j : j \geq i\}$$

Then $f_i^{\wedge} \leq f_i \leq f_i^{\vee}$ *important: upper bounds*

also $f_i^{\wedge} \uparrow f$ and $\int f_i^{\wedge} d\mu \geq -\int g d\mu > -\infty$ *multitasking monotone convergence*

$$\text{then } \int f_i^{\wedge} d\mu \uparrow \int f d\mu$$

Do the same for $f_i^{\vee}$ i.e. $f_i^{\vee} \downarrow f \Rightarrow \int f_i^{\vee} d\mu \downarrow \int f d\mu$

$$\text{and since } \int f_i^{\wedge} d\mu \leq \int f_i d\mu \leq \int f_i^{\vee} d\mu$$

$$\downarrow \int f d\mu \leq \int f d\mu \leq \int f d\mu \quad \square$$

Halfway Point, yay. one more lecture on Integration.

L^p spaces, law of large numbers, CLT, etc. incoming.

use measurable functions to define a measure
between two measurable spaces.

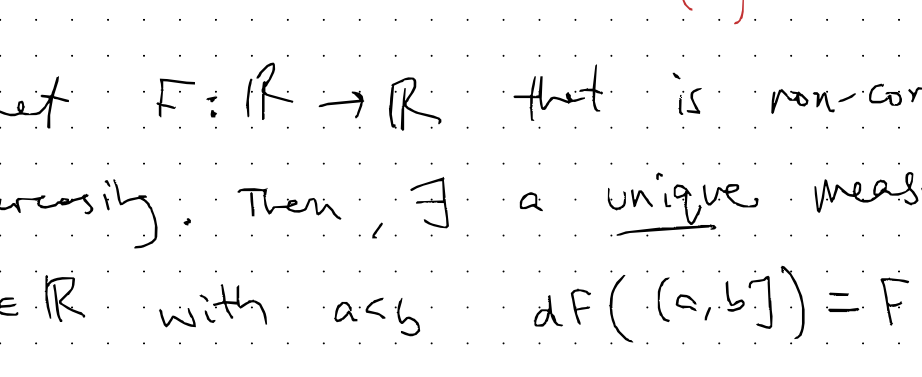
• let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be measurable spaces.

• let $\psi: X \rightarrow Y$ be measurable

• let μ be a measure on X then we can define

$$\nu = \mu \circ \psi^{-1} \quad \text{"a measure on } Y \text{"}$$

for some $B \in \mathcal{Y}$, $\nu(B) = \mu(\psi^{-1}(B))$



Theorem: let $F: \mathbb{R} \rightarrow \mathbb{R}$ that is non-constant right continuous and non-decreasing. Then, $\exists$ a unique measure dF on $\mathbb{R}$

s.t. $\forall a, b \in \mathbb{R}$ with $a < b$ $dF((a, b]) = F(b) - F(a)$

Proof: Let $F(\infty) = \lim_{x \rightarrow \infty} F(x)$ and $F(-\infty) = \lim_{x \rightarrow -\infty} F(x)$

Define an open interval $I = (F(-\infty), F(\infty))$

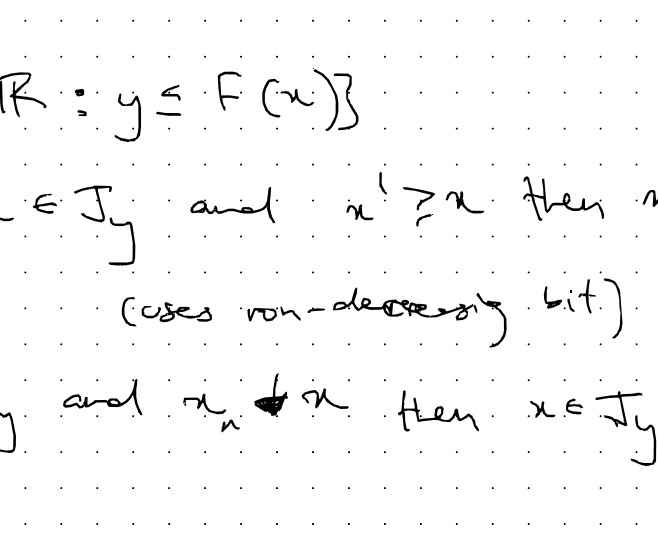
Define $g(y) = \inf \{x \in \mathbb{R} : y \leq F(x)\}$

We want to define dF to be $\lambda \circ g^{-1}$

show it makes sense

Claim: g is left continuous and non-decreasing and for $y \in I$,

$\lambda \in \mathbb{R}$, $g(y) \leq \lambda$ iff $y \leq F(\lambda)$



Fix a $y \in I$ and let $J_y = \{x \in \mathbb{R} : y \leq F(x)\}$

as F is non-decreasing if $x \in J_y$ and $x' \geq x$ then $x' \in J_y$ (uses non-decreasing bit)

as F is right-continuous if $x_n \in J_y$ and $x_n \downarrow x$ then $x \in J_y$

$\therefore J_y = [g(y), \infty)$

Furthermore, $g(y) \leq \lambda$ iff $y \leq F(\lambda)$

Also for $y \leq y'$ we have $J_{y'} \subseteq J_y$

$\therefore g(y) \leq g(y')$ So if $y_n \uparrow y$

Then $J_y = \bigcap_{n=1}^{\infty} J_{y_n}$ and thus $g(y_n) \rightarrow g(y)$

$\therefore g$ is cont. from the left

Next, g is Borel measurable! "inverse images preserve set operations"

$\therefore$ Defining $dF = \lambda \circ g^{-1}$ gives

$$dF((a, b]) = \lambda(\{y : g(y) > a, g(y) \leq b\})$$

$$= \lambda(F(a), F(b)]$$

$$= F(b) - F(a)$$

any other measure μ s.t. $\mu((a, b]) = F(b) - F(a)$ must coincide with dF on all Borel sets.

"turns out the L-S measure is related to random measure"

Def: (Random measure)

let $(\Omega, \mathcal{F}, \mu)$ be a measure space $\mathcal{F}$ is the Borel σ -field

μ is said to be random if $\mu(K)$ is finite for all compact K

$\in \mathcal{F}$

Note: dF is a random measure sequential etc.

• Every non-zero random measure on $\mathcal{B}(\mathbb{R})$ can be written as a Lebesgue-Stieltjes measure: $dF = \lambda \circ g^{-1}$ for some F .

$\rightarrow$ if μ is random on the Borel sets of $\mathbb{R} : \mathcal{B}(\mathbb{R})$, then define

$$F(x) = \begin{cases} \mu([0, x]) & \text{if } x \geq 0 \\ \mu((x, 0]) & \text{if } x < 0 \end{cases}$$

$$\therefore F(b) - F(a) = \mu((a, b]) \text{ for } a < b$$

$$\therefore \mu = dF \text{ by uniqueness}$$

Product Measure (take 2) Using the integral

Given $\mathcal{X}, \mathcal{Y}$ σ -fields we define $\mathcal{X} \times \mathcal{Y}$ to be the

σ -field generated by the rectangles $A \times B$ where $A \in \mathcal{X}$,

$B \in \mathcal{Y}$. We denote the set of all rectangles to be $\mathcal{R}$.

$$\sigma(\mathcal{R}) =: \mathcal{X} \times \mathcal{Y}$$

sigma field

The big theorem we want to prove:

Theorem (Existence + Uniqueness of Product measure)

let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be σ -finite measure spaces.

let π be a set function on $\mathcal{X} \times \mathcal{Y}$ s.t. for $A \in \mathcal{X}$ and

$$B \in \mathcal{Y} \quad \pi(A \times B) = \mu(A)\nu(B)$$

Then, π extends uniquely to a measure on $(X \times Y, \mathcal{X} \times \mathcal{Y})$

$$\text{s.t. for any } E \in \mathcal{X} \times \mathcal{Y} \quad \pi(E) = \iint \mathbb{1}_E(x, y) d\mu d\nu$$

$$= \int \int \mathbb{1}_E(x, y) d\nu(y) d\mu(x)$$

can't extend to general measures. (the extension won't be unique)

Proof:

(1) do it for finite measures

(2) extend to σ -finite

Fact 1: On $\mathcal{R}$, set function π is countably additive

we can extend $\mathcal{R}$ to a field $\mathcal{A}$ by including countable

semi-ring unions of rectangles.

We need a theorem similar in style to Dynkin π - λ

Def: (Monotone class) similar to Dynkin- π Theorem

A collection $\mathcal{M}$ of subsets of Ω is **monotone** if

(1) for $\{A_i\}_{i=1}^{\infty}$ s.t. $A_i \in \mathcal{M}$ and $A_i \uparrow A = \bigcup_{i=1}^{\infty} A_i$ then $A \in \mathcal{M}$

"closed under increasing sequences" of sets

(2) for $\{A_i\}_{i=1}^{\infty}$, $A_i \in \mathcal{M}$ and $A_i \downarrow A = \bigcap_{i=1}^{\infty} A_i$ then $A \in \mathcal{M}$

Note: if a field $\mathcal{A}$ is also Monotone then it is a σ -field

Recall $\mathcal{A}$ is a field if

- $\emptyset, \Omega \in \mathcal{A}$

- if $B, A \in \mathcal{A}$ then $B \cap A \in \mathcal{A}$

- if $B, A \in \mathcal{A}$ then $B \cup A \in \mathcal{A}$ "finite unions"

replace: $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ "complements"

$\hookrightarrow$ true because $B \cap A = B \cap A^c$

Theorem (Monotone class Theorem)

Let $\mathcal{A} \subseteq \mathcal{A}(\Omega)$ and $\mathcal{M}$ be a monotone class

s.t. $\mathcal{A} \subseteq \mathcal{M}$, then $\sigma(\mathcal{A}) \subseteq \mathcal{M}$.

Recall that Dynkin- π -Lemma says:

$$(\pi \text{ sys}) \subseteq (\lambda \text{ sys})$$

$$\sigma(\pi \text{ sys}) \subseteq (\lambda \text{ sys})$$

Proof: define $\mathcal{M}(\mathcal{A})$ to be the minimal

monotone class that contains $\mathcal{A}$.

- Goal is to show $\sigma(\mathcal{A}) \subseteq \mathcal{M}(\mathcal{A})$

Smallest sigma field $\mathcal{A}$

Smallest monotone class containing $\mathcal{A}$

the intersection of all monotone classes containing $\mathcal{A}$

"could start with a field"

- Method: $\mathcal{M}(\mathcal{A})$ is also a field and thus a σ -field and thus contains $\sigma(\mathcal{A})$

Step 1: show $\mathcal{M}(\mathcal{A})$ is closed under complements

$$\text{let } \mathcal{F} = \{A \in \mathcal{M}(\mathcal{A}) : A^c \in \mathcal{M}(\mathcal{A})\}$$

$\therefore \mathcal{A} \subseteq \mathcal{F}$ since $\mathcal{A}$ is a field

Next, for $\{A_i\}_{i=1}^{\infty}$ s.t. $A_i \in \mathcal{F}$ and $A_i \uparrow A$

$$\text{then } A_i^c \in \mathcal{F} \text{ and } A_i^c \downarrow A^c = \bigcap_{i=1}^{\infty} A_i^c = \left(\bigcup_{i=1}^{\infty} A_i\right)^c$$

$$\therefore \bigcup_{i=1}^{\infty} A_i = A \in \mathcal{F}$$

$\therefore \mathcal{F}$ is monotone and $\mathcal{F} \subseteq \mathcal{M}(\mathcal{A})$

Do the same for $A_i \downarrow A$ to get that $\mathcal{F} = \mathcal{M}(\mathcal{A})$

(by minimality of $\mathcal{M}(\mathcal{A})$)

Therefore $\mathcal{M}(\mathcal{A})$ is closed under complements.

Step 2: Show that $\mathcal{M}(\mathcal{A})$ is closed under finite unions

$$\text{Define } \mathcal{G}_1 = \{A \in \mathcal{M}(\mathcal{A}) : A \cup B \in \mathcal{M}(\mathcal{A}), \forall B \in \mathcal{A}\}$$

all the elements of the monotone class.

As with $\mathcal{F}$, note that $\mathcal{A} \subseteq \mathcal{G}_1$

$\mathcal{G}_1$ is monotone as $\left(\bigcup_{i=1}^{\infty} A_i\right) \cup B = \bigcup_{i=1}^{\infty} (A_i \cup B) \in \mathcal{M}(\mathcal{A})$

$$\text{and } \left(\bigcap_{i=1}^{\infty} A_i\right) \cup B = \bigcap_{i=1}^{\infty} (A_i \cup B) \in \mathcal{M}(\mathcal{A})$$

$\therefore$ by minimality $\mathcal{G}_1$ is monotone

$$\text{Hence } \mathcal{G}_1 = \mathcal{M}(\mathcal{A})$$

$\therefore \mathcal{M}(\mathcal{A})$ is closed under unions with $\mathcal{A}$

$$\text{Define } \mathcal{G}_2 = \{B \in \mathcal{M}(\mathcal{A}) : A \cup B \in \mathcal{M}(\mathcal{A}), \forall A \in \mathcal{M}(\mathcal{A})\}$$

$\mathcal{A} \subseteq \mathcal{G}_2$ since if $B \in \mathcal{A}$ then $A \cup B \in \mathcal{M}(\mathcal{A})$

for any $A \in \mathcal{M}(\mathcal{A})$ from $\mathcal{G}_1$

$\mathcal{G}_2$ is also monotone

$$\text{Hence } \mathcal{G}_2 = \mathcal{M}(\mathcal{A}) \therefore \mathcal{M}(\mathcal{A}) \text{ closed under finite unions.}$$

$\therefore \mathcal{M}(\mathcal{A})$ is a field $\Rightarrow$ also a σ -field.

$$\text{Thus } \sigma(\mathcal{A}) \subseteq \mathcal{M}(\mathcal{A})$$

(Final lecture, in Chapter 2)

Theorem (Existence and uniqueness of Product Measure)

Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be σ -finite measure spaces. Let π be a set function on $X \times Y$, s.t. for $A \in \mathcal{X}$ and $B \in \mathcal{Y}$ then $\pi(A \times B) = \mu(A) \nu(B)$. Then π extends uniquely to a measure on $(X \times Y, \mathcal{X} \times \mathcal{Y})$ s.t. for any $E \in \mathcal{X} \times \mathcal{Y}$

$$\pi(E) = \iint \mathbb{1}_E(x, y) d\mu(x) d\nu(y) = \iint \mathbb{1}_E(x, y) d\nu(y) d\mu(x)$$

Definition: Let $\mathcal{R}$ collection of rectangles i.e. $R \in \mathcal{R}, R = A \times B$ some $A \in \mathcal{X}, B \in \mathcal{Y}$

Claim: π is countably additive on $\mathcal{R}$

Note: including finite unions of rectangles takes $\mathcal{R}$ to a field $\mathcal{A}$

Theorem (Monotone Class Theorem) *very similar to Dynkin π - λ lemma*
Let $\mathcal{A}$ be a field and $\mathcal{M}$ be monotone s.t. $\mathcal{A} \subset \mathcal{M}$. Then $\sigma(\mathcal{A}) \subset \mathcal{M}$

Lemma: Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be finite measure spaces. Let $\mathcal{F} = \{E \in \mathcal{X} \times \mathcal{Y} : \iint \mathbb{1}_E(x, y) d\mu(x) d\nu(y) = \iint \mathbb{1}_E(x, y) d\nu(y) d\mu(x)\}$

Then $X \times Y \in \mathcal{F}$

Proof: Let $E = A \times B, A \in \mathcal{X}, B \in \mathcal{Y}$

$$\begin{aligned} \iint \mathbb{1}_E(x, y) d\mu(x) d\nu(y) &= \mu(A) \int \mathbb{1}_B(y) d\nu(y) \\ &= \mu(A) \nu(B) \\ &= \nu(B) \mu(A) = \nu(B) \int \mathbb{1}_A(x) d\mu(x) \\ &= \int \int \mathbb{1}_E(x, y) d\nu(y) d\mu(x) \end{aligned}$$

Next, for disjoint $R_1, R_2 \in \mathcal{R}$

$$\mathbb{1}_{R_1 \cup R_2} = \mathbb{1}_{R_1} + \mathbb{1}_{R_2}$$

$\therefore \mathcal{F}$ contains finite disjoint unions of $\mathcal{R}$

$\therefore \mathcal{A} \subset \mathcal{F}$ (see Dudley prop 3.2.3)

Next, consider $\{E_i\}_{i=1}^\infty, E_i \in \mathcal{F}$

Then if $E_i \uparrow E$ then monotone convergence says

$$\iint \mathbb{1}_{E_i}(x, y) d\mu(x) d\nu(y) \uparrow \iint \mathbb{1}_E(x, y) d\mu(x) d\nu(y)$$

$$\text{and } \iint \mathbb{1}_{E_i}(x, y) d\nu(y) d\mu(x) \uparrow \iint \mathbb{1}_E(x, y) d\nu(y) d\mu(x)$$

$\therefore E \in \mathcal{F}$ and same holds for $E_i \downarrow E$

$\therefore \mathcal{F}$ is a monotone class and $\mathcal{A} \subset \mathcal{F}$

$\therefore X \times Y = \sigma(\mathcal{A}) \subset \mathcal{F}$

Proof (of Product Measure)

First we consider μ, ν to be finite measures and

$\pi(A \times B) = \mu(A) \nu(B)$. Then we extend to a set function

$$\pi(E) := \iint \mathbb{1}_E(x, y) d\mu(x) d\nu(y) \text{ any } E \in \mathcal{X} \times \mathcal{Y}$$

(Carleson says that definition makes sense)

Integrals are linear $\Rightarrow \pi$ is finitely additive

Apply monotone convergence:

$\Rightarrow \pi$ is countably additive

$\therefore \pi$ is a measure on $X \times Y$

To show that π is unique, let ρ be some other set function s.t. $\rho(A \times B) = \mu(A) \nu(B)$

$$\text{Let } \mathcal{M} = \{E \in \mathcal{X} \times \mathcal{Y} : \pi(E) = \rho(E)\}$$

$$\rightarrow \text{for } E = \bigcup_{i=1}^\infty E_i, \text{ write } E = \bigcup_{i=1}^\infty D_i$$

$D_i = E_i$ and $D_i = E_i \setminus \bigcup_{j=1}^{i-1} E_j$ are disjoint. So, by countable additivity $\pi(E) = \rho(E)$

$\rightarrow$ Do same idea for $E_i \downarrow E$

$\therefore$ By the monotone class theorem $X \times Y \in \mathcal{M} \therefore \pi$ is unique

on $X \times Y$ (for finite measures)

uniqueness!

Now let μ and ν be σ -finite measures and $\{A_i\}_{i=1}^\infty$ and $\{B_j\}_{j=1}^\infty$ be disjoint partitions of X and Y respectively

$$\text{s.t. } \mu(A_i) < \infty, \nu(B_j) < \infty$$

Then for any $E \in \mathcal{X} \times \mathcal{Y}$ we define $E_{i,j} = E \cap (A_i \times B_j)$

From the finite measure case

$$\begin{aligned} \iint \mathbb{1}_E(x, y) d\mu(x) d\nu(y) &= \iint \mathbb{1}_{E_{i,j}}(x, y) d\mu(x) d\nu(y) \end{aligned}$$

Now sum over all i, j and apply monotone convergence

$$\begin{aligned} \pi(E) &= \iint \mathbb{1}_E(x, y) d\mu(x) d\nu(y) \\ &= \iint \mathbb{1}_E(x, y) d\nu(y) d\mu(x) \end{aligned}$$

Furthermore, monotone convergence implies that π is countably additive.

$\therefore \pi$ is a measure on $X \times Y$

for any other ρ s.t. $\rho(A \times B) = \mu(A) \nu(B)$

Then countable additivity + uniqueness shown for finiteness

$$\pi(E) = \sum_{i,j} \pi(E_{i,j}) = \sum_{i,j} \rho(E_{i,j}) = \rho(E)$$

Fubini-Tonelli

Theorem: let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be σ -finite measure spaces.

Let $f: X \times Y \rightarrow \mathbb{R}$ be $X \times Y$ measurable s.t. either (1) $f \geq 0$ (non-negative) or (2) $\iint |f| d(\mu \times \nu) < \infty$ (absolutely integrable)

$$\text{Then } \int f d(\mu \times \nu) = \iint f(x, y) d\mu(x) d\nu(y) = \iint f(x, y) d\nu(y) d\mu(x)$$

Also $\int f(x, y) d\mu(x)$ is Y -measurable

$\int f(x, y) d\nu(y)$ is X -measurable

Proof: We have this for indicator functions from above

$\therefore$ We have the result for simple functions as integrals are linear

Then applying Monotone Convergence to simple functions

$\Rightarrow$ non-negative measurable functions

Instead assume $\iint |f| d(\mu \times \nu) < \infty$ then $f = f^+ - f^-$ and the above argument holds for f^+ and f^- i.e.

$$\int f^+(x, y) d\mu(x) < \infty \text{ v.a.e and similarly } \int f^-(x, y) d\nu(y) < \infty \text{ } \mu\text{-a.e and similarly for } f^-$$

$$\therefore \int |f|(x, y) d\nu(y) = \int f^+(x, y) d\nu(y) + \int f^-(x, y) d\nu(y)$$

$$\therefore \int f(x, y) d\nu(y) = \int f^+(x, y) d\nu(y) - \int f^-(x, y) d\nu(y) \text{ } \mu\text{-a.e}$$

$$\therefore \iint f(x, y) d\nu(y) d\mu(x) = \iint f^+(x, y) d\nu(y) d\mu(x) - \iint f^-(x, y) d\nu(y) d\mu(x)$$

Do same swapping μ and ν

By induction, we can extend the above theorems to a finite product of N measure spaces (σ -finite)

Infinite Products of Probability Spaces

Definition: Let $(\mathcal{J}_n, \mathcal{F}_n, P_n)_{n \in \mathbb{N}}$ be a seq. of prob. spaces.

Then, $\mathcal{J} = \bigotimes_{n=1}^\infty \mathcal{J}_n$ consists of sequences like

$$\{\omega_i\}_{i=1}^\infty \text{ s.t. } \omega_i \in \mathcal{J}_i$$

"little omega" Then, let $\mathcal{R}$ be the collection of finite dimensional rectangles in $\mathcal{J}$. i.e. $R \in \mathcal{R}$ then $R = \bigotimes_{n=1}^\infty A_n$

for $A_n \in \mathcal{F}_n$ s.t. $\exists N$ s.t. $\forall n > N, A_n = \mathcal{J}_n$

We can show that $\mathcal{R}$ is a semi-ring

Denote $\mathcal{G}$ to be the field generated by $\mathcal{R}$

Next, we define P is a set function on $\mathcal{R}$ s.t. $P(R) = \prod_{n=1}^\infty P_n(A_n)$

as eventually $P(A_n) = P(\mathcal{J}_n) = 1, n > N$

Next, define $\mathcal{F}$ to be a σ -field on $\mathcal{J}$ s.t. for projections $\omega_n: \mathcal{J} \rightarrow \mathcal{J}_n$ where $\omega_n((\omega_1, \dots)) = \omega_n$

$\mathcal{F}$ is the smallest σ -field s.t. ω_n are measurable

$\forall n \in \mathbb{N}$

$$(\mathcal{J}, \mathcal{F}) \rightarrow (\mathcal{J}_n, \mathcal{F}_n)$$

i.e. $\mathcal{F}$ contains sets $\omega_n^{-1}(A_n), A_n \in \mathcal{F}_n$

Definition: $(X \times Y, \mathcal{X} \times \mathcal{Y}, \pi)$

and $A \in \mathcal{X} \times \mathcal{Y}, x \in X$

$$\text{then } A_x := \{y \in Y : (x, y) \in A\}$$

Note, $A_x \in \mathcal{Y}$ for any $x \in X$

Theorem ($\exists!$ ∞ -prod-prob-measures)

The set function P extends uniquely from $\mathcal{R}$ to a probability measure on $\mathcal{F}$

Lemma: Let μ be a finitely additive set function on a field $\mathcal{S}$. Then μ is countably additive $\Leftrightarrow$

for any seq $A_i \downarrow \emptyset$ with $A_i \in \mathcal{S}$ then $\mu(A_i) \rightarrow 0$

Proof: we want to extend P from $\mathcal{R}$ to $\mathcal{S}$ and show it makes sense. It can be written as

• disjoint union of $\mathcal{R}$'s in $\mathcal{R}$.

• for any $S \in \mathcal{S}$

$$S = \bigcup_{i=1}^k R_i = \bigcup_{i=1}^k \bigotimes_{n=1}^\infty A_{i,n}$$

where $A_{i,n} \in \mathcal{F}_n$ and all but a finite $\#$ of $A_{i,n} = \mathcal{J}_n$

$\therefore$ we can treat $P(S)$ as a finite product measure on $\mathcal{J}_1 \times \mathcal{J}_2 \times \dots \times \mathcal{J}_N$ and countably additive on $\mathcal{S}$.

By Carathéodory, if P is countably additive on $\mathcal{S}$, then it extends to a measure on $\sigma(\mathcal{S}) = \mathcal{F}$

$\therefore$ We use above lemma to show that for some $\{A_i\}_{i=1}^\infty$ decreasing s.t. $\bigcap_{i=1}^\infty A_i = \emptyset$ then $\sum_{i=1}^\infty P(A_i) > 1$

Let $P^{(n)}$ be P on $\mathcal{S}$. For $n \geq 1$ we define $\mathcal{S}^{(n)} = \bigotimes_{m=n}^\infty \mathcal{J}_m$

$\mathcal{S}^{(n)}$ disjoint union rectangles from $\mathcal{J}^{(n)}$ "infinite tail prod"

(a) $P^{(n)}$ is P restricted to $\mathcal{S}^{(n)}$

For any $E \in \mathcal{S}^{(n)}$ and $(x_1, \dots, x_n) \in \mathcal{J}_1 \times \dots \times \mathcal{J}_n$ we

define $\mathcal{F}^{(n)}(x_1, \dots, x_n) = \{ \bigotimes_{m=n}^\infty A_m \in \mathcal{S}^{(n)} : \{x_m\}_{m=n}^\infty \in E \}$

For any $E \in \mathcal{S}, \exists N > 0$ s.t. $E = F \times \bigotimes_{n=N}^\infty \mathcal{J}_n$

$$\therefore F = \bigcup_{i=1}^k F_i = \bigcup_{i=1}^k \bigotimes_{n=1}^N A_{i,n} \text{ where } A_{i,n} \in \mathcal{F}_n$$

$$F_i \in \mathcal{F}_N$$

For any choice of $M \in \mathbb{N}$ and $(x_1, \dots, x_M) \in$

$$E^{(M)}(x_1, \dots, x_M) = G \times \mathcal{J}^{(M)} \text{ for } G = \bigcup_{i=1}^k \bigotimes_{n=M+1}^\infty A_{i,n}$$

$$\therefore E^{(M)}(x_1, \dots, x_M) \in \mathcal{S}^{(M)} \therefore P^{(M)} \text{ is defined on } \mathcal{S}^{(M)}$$

via Fubini-Tonelli Theorem $P(E) = \int \mathbb{1}_E dP = \int \mathbb{1}_E dP_1 \times \dots \times dP_N \times dP^{(N)}$

$$= \int P^{(N)}(E^{(N)}(x_1, \dots, x_N)) dP_1 \times \dots \times dP_N$$

Returning to the A_i from above $\{A_i\}_{i=1}^\infty$ decreasing s.t.

$$P(A_i) > \frac{\epsilon}{2} > 0$$

$$\text{Define } F_i := \{x_1 \in \mathcal{J}_1 : P^{(1)}(A_i^{(1)}(x_1)) > \frac{\epsilon}{2}\}$$

Set of x_1 's s.t. tail seq have measure $> \frac{\epsilon}{2}$

$$\mathcal{J}_1, \mathcal{J}_2, \mathcal{J}_3, \dots$$

$$F_i, A_i^{(1)}(x_1)$$

Using the Fubini-Tonelli Integral formula with $E = A_i$

$$E \subset P(A_i) = \int P^{(1)}(A_i^{(1)}(x_1)) dP_1(x_1)$$

$$= \int_{F_i} P^{(1)}(A_i^{(1)}(x_1)) dP_1(x_1) + \int_{\mathcal{J}_1 \setminus F_i} \dots dP_1(x_1)$$

$$\leq P_1(F_i) + \frac{\epsilon}{2}$$

$\therefore P_1(F_i) > \frac{\epsilon}{2}$ A_i 's are decreasing so F_i are decreasing

and P_1 is countably additive prob measure

on $(\mathcal{J}_1, \mathcal{F}_1)$

$$P_1(\bigcap_{i=1}^\infty F_i) = \int \mathbb{1}_{\bigcap_{i=1}^\infty F_i} dP_1 = \int \inf_i \mathbb{1}_{F_i} dP_1$$

$$= \inf_i \int \mathbb{1}_{F_i} dP_1 = \inf_i P_1(F_i) > \frac{\epsilon}{2}$$

$$\therefore \bigcap_{i=1}^\infty F_i \neq \emptyset$$

Next, fix $y_1 \in \bigcap_{i=1}^\infty F_i$ and define

$$G_i := \{x_2 \in \mathcal{J}_2 : P^{(2)}(A_i^{(2)}(y_1, x_2)) > \frac{\epsilon}{4}\}$$

By same argument as above

$$P_2(G_i) > \frac{\epsilon}{4} \Rightarrow \bigcap_i G_i \neq \emptyset$$

so pick $y_2 \in \bigcap_i G_i$

Continue via induction to construct a sequence

$$\{y_i\}_{i=1}^\infty, y_i \in \mathcal{J}_i \text{ s.t. } P(A_i^{(i)}(y_1, \dots, y_n)) > \frac{\epsilon}{2^n}$$

$\forall i$ show that $\{y_i\}_{i=1}^\infty \in A_i \forall i$

For each j , select $n_j \in \mathbb{N}$ s.t. $(x_1, \dots, x_{n_j}) \in \mathcal{J}_1 \times \dots \times \mathcal{J}_{n_j}$

s.t. $A_j^{(n_j)}(x_1, \dots, x_{n_j}) = \mathcal{J}_1 \times \dots \times \mathcal{J}_{n_j}$

denote as $A_j \in \mathcal{S}$

Thus for n large enough

$$A_j^{(n)}(y_1, \dots, y_n) = \mathcal{J}_1 \times \dots \times \mathcal{J}_n$$

Since $(y_1, \dots, y_n) \in A_j \forall j, \bigcap_{j=1}^\infty A_j \neq \emptyset$

$\therefore P$ is countably additive on $\mathcal{S}$ and can be uniquely extended to all of $\mathcal{F}$

$\square$

L^p spaces are Banach spaces

(p)th moment of ω on $\mathbb{R}^d$

Define: (L^p space)

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f: \Omega \rightarrow [-\infty, \infty]$

f measurable, then we say that $f \in L^p(\Omega, \mathcal{F}, \mu)$

for $1 \leq p < \infty$ if $\int |f|^p d\mu < \infty$

and for $p = \infty$, $f \in L^\infty(\Omega, \mathcal{F}, \mu)$ if

$$\inf \{ t \in [-\infty, \infty] : |f| \leq t \text{ } \mu.a.e \} < \infty$$

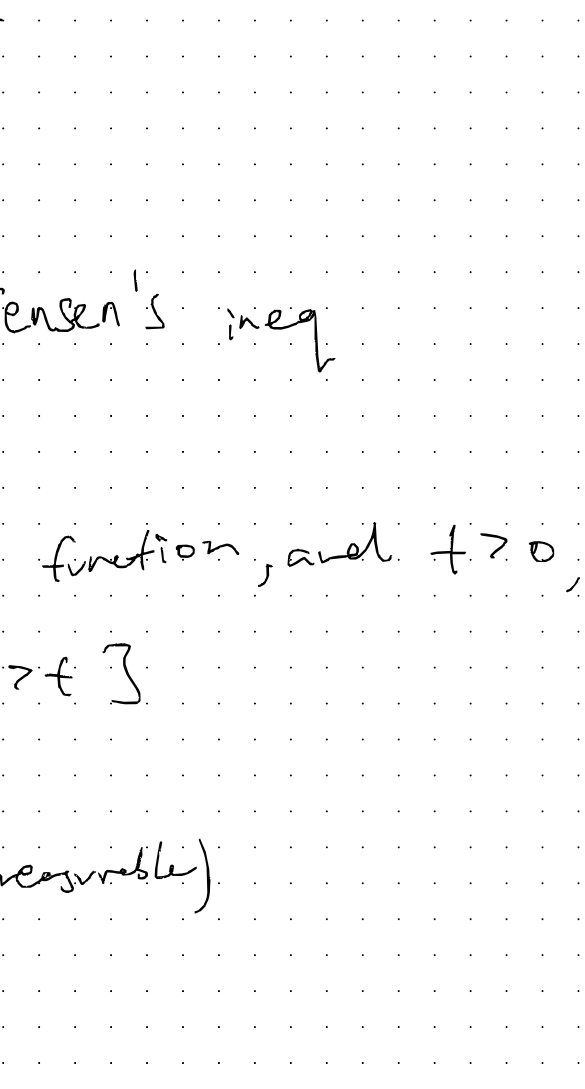
L^p spaces are normed spaces

$$\|f\|_p = \left[\int |f|^p d\mu \right]^{\frac{1}{p}}, \quad 1 \leq p \leq \infty$$

"ball is no longer convex for $p < 1$ "

$$\|f\|_\infty = \inf \{ t \in [-\infty, \infty] : |f| \leq t \text{ } \mu.a.e \}$$

essential supremum.



Def: conjugate pairs

for $1 < p, q < \infty$, we say p, q

are conjugate pairs if $\frac{1}{p} + \frac{1}{q} = 1$

Also, 1 and ∞ are conjugates.

Markov / Chebyshev / Chernoff ineq. and Jensen's ineq

Theorem: (Markov's Inequality)

Let f be a non-negative measurable function, and $t > 0$,

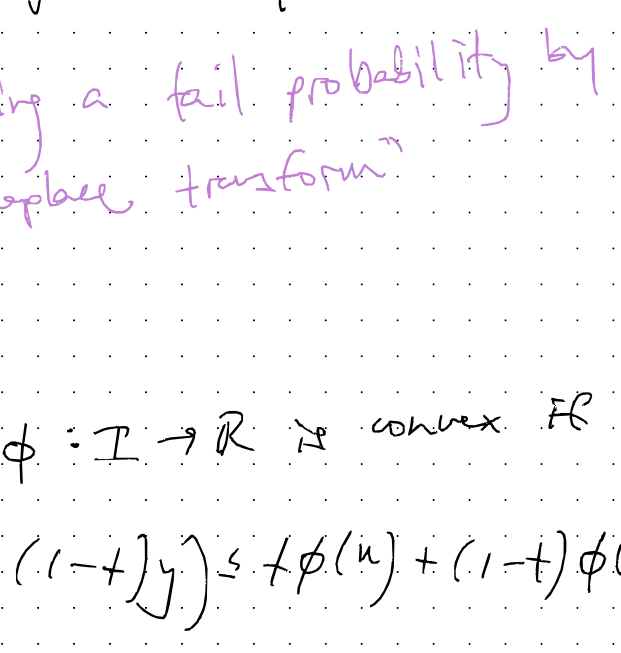
$$\mu(\{f > t\}) \leq \frac{1}{t} \int f d\mu$$

"subset of omega"
 element of σ -algebra (because it's measurable)

Proof: Note that $t \cdot \mathbb{1}_{\{f > t\}} \leq f$,

then by monotonicity of the integral

$$t \mu(\{f > t\}) = \int t \cdot \mathbb{1}_{\{f > t\}} d\mu \leq \int f d\mu$$



$$\Rightarrow \mu(\{f > t\}) \leq \frac{1}{t} \int f d\mu$$

Useful ways to use Markov's inequality:

(1) Chebyshev's Ineq

for f measurable, and $m \in \mathbb{R}$ (not necessarily > 0)

$$\mu(\{f - m > t\}) \leq \frac{1}{t^2} \int (f - m)^2 d\mu$$

$P(|X - EX| > t) \leq \frac{\text{Var}(X)}{t^2}$
 tail probability
 inverse quadratic

(2) Chernoff's Ineq

for f measurable, $\eta \in \mathbb{R}$

$$\mu(\{f > t\}) \leq e^{-\eta t} \int e^{\eta f} d\mu$$

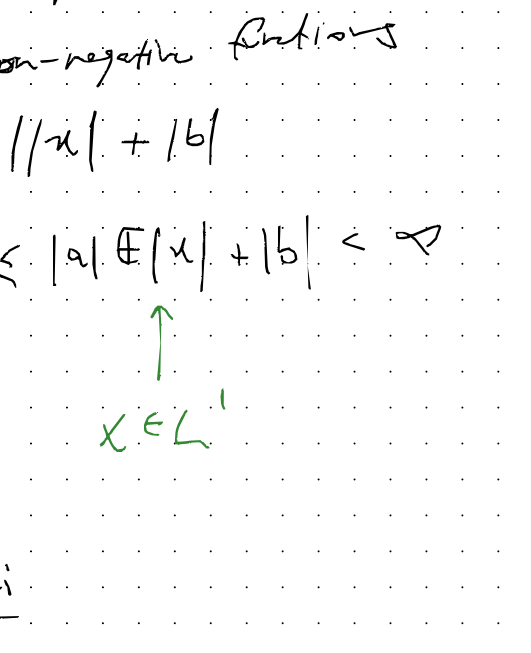
padding out with?

"bounding a tail probability by the Laplace transform"

Definition: (Convex function on $\mathbb{R}$)

Let $I \subseteq \mathbb{R}$ be an interval, a function $\phi: I \rightarrow \mathbb{R}$ is convex if

$$\forall t \in [0, 1] \text{ and all } x, y \in I \quad \phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y)$$



"secant line is always on top"

Theorem (Jensen's Ineq)

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

an integrable random variable (i.e. $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$)

s.t. $X: \Omega \rightarrow I \subseteq \mathbb{R}$, then for any convex $\phi: I \rightarrow \mathbb{R}$, we

$$\text{have } \phi\left(\int X d\mu\right) \leq \int \phi(X) d\mu$$

$$\text{i.e. } \phi(\mathbb{E}X) \leq \mathbb{E}[\phi(X)]$$

Proof: For some $c \in I$, if $X = c$ p.a.e. then we're done

else, let $m = \mathbb{E}X$ is in the interior of I

$$\text{i.e. } \exists \varepsilon > 0 \text{ s.t. } (m - \varepsilon, m + \varepsilon) \subseteq I$$

Then, we can choose $a, b \in \mathbb{R}$ s.t. $\phi(x) \geq ax + b \quad \forall x \in I$

and $\phi(m) = am + b$

Then $\phi(X) \geq aX + b$ (x)

$$\therefore \phi(\mathbb{E}X) = am + b = \mathbb{E}[am + b] \leq \mathbb{E}[\phi(X)]$$

Lastly, $\mathbb{E}[\phi(X)]$ well defined? i.e. $\neq \infty - \infty$

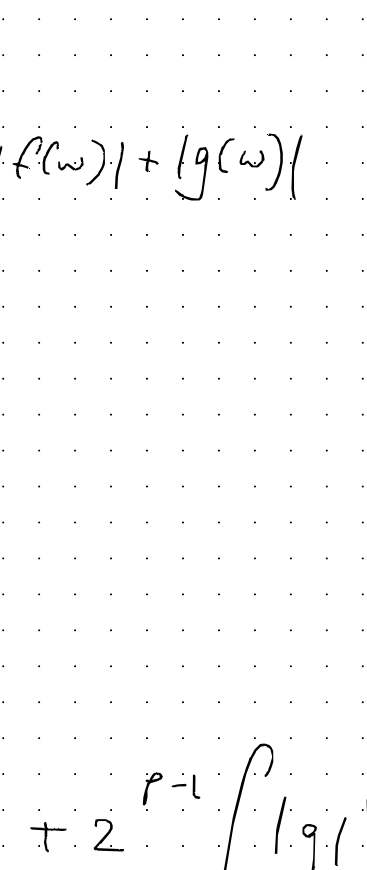
write $\phi = \phi^+ - \phi^-$ where ϕ^+ is convex.

with non-negative Radicals

and $\phi^-(x) \leq |a||x| + |b|$

$$\text{hence, } \mathbb{E}[\phi^-(X)] \leq |a|\mathbb{E}|X| + |b| < \infty$$

$X \in L^1$



Hölder + Minkowski

Theorem: (Hölder) Let $p, q \in [1, \infty]$ conjugate pairs f, g measurable

$$\text{Then } \|fg\|_1 \leq \|f\|_p \|g\|_q$$

Proof: If either $\|f\|_p = 0, \infty$ or $\|g\|_q = 0, \infty$ then done.

Hence, let $0 < \|f\|_p < \infty$ then we can "normalise" wlog $\|f\|_p = 1$

Thus, we can define a probability measure

$$\mathbb{P}(A) := \int_A f d\mu$$

($1 < p, q < \infty$)

$$\|fg\|_1 = \int |fg| d\mu = \int \frac{|g|}{|f|^{p-1}} \mathbb{1}_{|f| > 0} |f|^p d\mu$$

$$\leq \left[\int \frac{|g|^q}{|f|^{q(p-1)}} \mathbb{1}_{|f| > 0} |f|^p d\mu \right]^{\frac{1}{q}}$$

$$\leq \left[\int |g|^q d\mu \right]^{\frac{1}{q}} = \|g\|_q$$

$$\leq \left[\int \phi \left(\frac{|g|}{|f|^{p-1}} \mathbb{1}_{|f| > 0} \right) |f|^p d\mu \right]^{\frac{1}{q}}$$

$$\leq \left[\int \phi \left(\frac{|g|}{|f|^{p-1}} \mathbb{1}_{|f| > 0} \right) |f|^p d\mu \right]^{\frac{1}{q}}$$

Note: $p = q = 2$, we have Cauchy Schwarz Ineq
 " L^2 is a Hilbert Space"

Theorem: (Minkowski's Ineq)

Let $p \in [1, \infty]$, f, g measurable then $\|f+g\|_p \leq \|f\|_p + \|g\|_p$

Proof: If $\|f\|_p = \infty$ or $\|g\|_p = \infty$ then done

• if $\|f+g\|_p = 0$ done

• if $p = 1$ then use $|(f+g)(\omega)| \leq |f(\omega)| + |g(\omega)|$ done

Otherwise, $|f+g|^p = 2^p \left| \frac{f+g}{2} \right|^p$

$$\leq 2^p \left(\frac{1}{2} |f|^p + \frac{1}{2} |g|^p \right)$$

$$= 2^{p-1} (|f|^p + |g|^p)$$

$$\therefore \int |f+g|^p d\mu \leq 2^{p-1} \int |f|^p d\mu + 2^{p-1} \int |g|^p d\mu < \infty$$

$$\therefore f+g \in L^p(\Omega, \mathcal{F}, \mu)$$

Assuming $\|f+g\|_p > 0$ and $p > 1$ and p, q conjugates then

$$\| |f+g|^{p-1} \|_q = \left[\int |f+g|^{(p-1)q} d\mu \right]^{\frac{1}{q}}$$

$$= \left[\int |f+g|^p d\mu \right]^{\frac{p-1}{p}}$$

$$= \|f+g\|_p^{p-1}$$

$$\therefore \|f+g\|_p^p = \int |f+g|^p d\mu \leq \int |f| |f+g|^{p-1} d\mu + \int |g| |f+g|^{p-1} d\mu$$

$$\|fg\|_p^{p-1} \leq \|f\|_p \| |f+g|^{p-1} \|_q + \|g\|_q \| |f+g|^{p-1} \|_q$$

$$= \|f\|_p \|f+g\|_p^{p-1} + \|g\|_q \|f+g\|_p^{p-1}$$

$$\|f+g\|_p^{p-1}$$

Theorem (Approx in L^p space)

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $\mathcal{A}$ is a

π -system such that $\sigma(\mathcal{A}) = \mathcal{F}$ and $\mu(A) < \infty \quad \forall A \in \mathcal{A}$

and $\exists A_i \in \mathcal{A}$ s.t. $A_i \uparrow \Omega$

$$\text{Let } V_0 := \left\{ \sum_{i=1}^n a_i \mathbb{1}_{A_i} : a_i \in \mathbb{R}, A_i \in \mathcal{A}, n \in \mathbb{N} \right\}$$

For $p \in [1, \infty)$, $V_0 \subset L^p$ and $\forall f \in L^p, \forall \varepsilon > 0 \exists v \in V_0$

$$\text{s.t. } \|f - v\|_p < \varepsilon$$

Proof: for any $A \in \mathcal{A}$, $\|\mathbb{1}_A\|_p = \left(\int \mathbb{1}_A^p d\mu \right)^{\frac{1}{p}} = \mu(A)^{\frac{1}{p}} < \infty$

"you can approximate in L^p with simple functions"

$\therefore \mathbb{1}_A \in L^p$ for any $A \in \mathcal{A}$ as L^p is linear, $V_0 \subset L^p$

Next, let $V \subset L^p$ be all $f \in L^p$ that can be approximated as above by some $v \in V_0$

let $f, g \in V$ be approximated by v_f, v_g respectively;

then, by Minkowski's Ineq:

$$\|(f+g) - (v_f + v_g)\|_p \leq \|f - v_f\|_p + \|g - v_g\|_p < 2\varepsilon$$

$\therefore V$ is also a linear space

• Next, assume $\Omega \in \mathcal{A}$, i.e. $\mu(\Omega) < \infty$

let $\mathcal{L} = \{ B \subset \Omega : \mathbb{1}_B \in V \}$

• We have $\mathcal{A} \subset \mathcal{L}$ and $\therefore \Omega \in \mathcal{L}$

• for $A, B \in \mathcal{L}$ s.t. $A \subset B$ then $\mathbb{1}_{B \setminus A} = \mathbb{1}_B - \mathbb{1}_A \in V$

$$\therefore B \setminus A \in \mathcal{L}$$

• for $\{A_i\}_{i=1}^\infty$ pairwise disjoint, $A_i \in \mathcal{L}$

$$\text{let } A = \bigcup_{i=1}^\infty A_i, \quad B_j = \bigcup_{i=1}^j A_i$$

$$\text{Then } B_j \uparrow A \text{ and } \|\mathbb{1}_A - \mathbb{1}_{B_j}\|_p = \mu(A \setminus B_j)^{\frac{1}{p}} \rightarrow 0$$

as $j \rightarrow \infty$

approx $\mathbb{1}_A$ by $\mathbb{1}_{B_j}$

$$\therefore A \in \mathcal{L}$$

$\Rightarrow \mathcal{L}$ is a λ -system

$\Rightarrow$ Dynkin π - λ says $\mathcal{F} = \sigma(\mathcal{A}) \subseteq \mathcal{L}$

$$\therefore \mathbb{1}_B \in V \quad \forall B \in \mathcal{F}$$

Now for any non-negative $f \in L^p$, we construct simple

$$f_n = \min \{ n, 2^{-n} \lfloor 2^n f \rfloor \}$$

similar to the simple functions with simple L^p

s.t. $f_n \uparrow f$ then $\|f - f_n\|_p \rightarrow 0$

$$\text{and } \|f - f_n\|_p \leq \|f\|_p \leftarrow \text{bound } \|f - f_n\|_p$$

$$\therefore \text{By Dominated Convergence } \|f - f_n\|_p = \left[\int |f - f_n|^p d\mu \right]^{\frac{1}{p}} \rightarrow 0$$

$\therefore f \in V$ and by linearity $V = L^p$

Lastly, for general $\Omega \notin \mathcal{A}$, by assumption $A_i \uparrow \Omega$

$$\therefore \text{for any } f \in L^p, f \mathbb{1}_{A_i} \in V$$

and $\|f - f \mathbb{1}_{A_i}\|_p \rightarrow 0$ pointwise and $\|f - f \mathbb{1}_{A_i}\|_p \leq \|f\|_p$

$$\therefore \|f - f \mathbb{1}_{A_i}\|_p \rightarrow 0 \text{ by Dominated Convergence}$$

$$\therefore f \in V$$

$$\therefore f = f^+ - f^-$$

Lecture 10: Convergence of Measure

Weak convergence of measure

$(\mathcal{X}, \mathcal{F})$ be a measurable space $\{P_i\}_{i=1}^\infty$ a seq of prob measures
 s.t. $P_i: \mathcal{X} \rightarrow \mathbb{R}^+$

Q: what does $P_i \rightarrow P$ mean?

e.g. Maybe we want $P_i(A) \rightarrow P(A) \quad \forall A \in \mathcal{F}$?

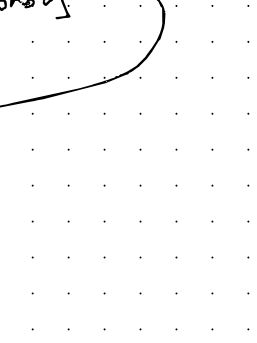
Defn: (Weak Conv) "softs"
 Let S be a metric space $\mathcal{F}$ be the Borel σ -field on S
 Then, we write $P_i \Rightarrow P$ for weak convergence if $\int f dP_i \rightarrow \int f dP$
 $\forall f \in C_b(\mathbb{R})$, all continuous bounded functions $S \rightarrow \mathbb{R}$
 (metric d. Define open sets (balls))

Note: Weak convergence is dependent on the metric space S and the topology generated by the metric

$$d: S \times S \rightarrow \mathbb{R}^+$$

If $\{P_i\}$ converges then P_i is "close" to P for i large, more precisely, for a finite collection of $f_1, \dots, f_n \in C_b(\mathbb{R})$ then we define an ε -neighborhood, any $\varepsilon > 0$, of P to be all Q s.t.

$$\left| \int f_i dP - \int f_i dQ \right| < \varepsilon \quad \forall i = 1, \dots, n$$



portmanteau

synergy = synchronous + energy

Theorem: (Portmanteau)

For P and P_i prob measures on $(S, \mathcal{F})$, T.F.A.E.:

"The following are equivalent"

- $P_i = P$
- $\int f dP_i \rightarrow \int f dP \quad \forall$ bounded uniformly cont f
- $\limsup P_i(C) \leq P(C) \quad \forall$ closed C
- $\liminf P_i(U) \geq P(U) \quad \forall$ open U
- $\lim P_i(A) = P(A) \quad \forall A \in \mathcal{F}$ s.t. $P(\partial A) = 0$

$$\partial A = \overline{A} \cap A^c$$

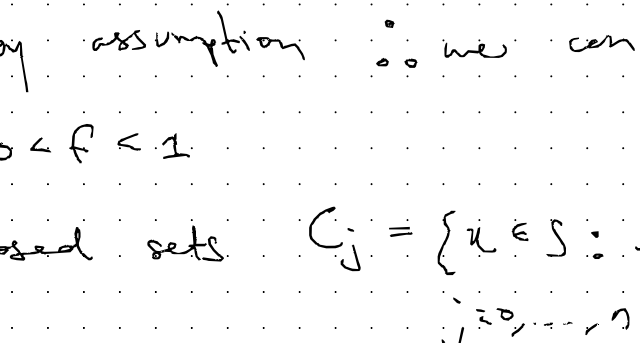
Proof: (1) $\Rightarrow$ (2) If convergence holds $\forall f \in C_b(\mathbb{R})$ then it holds for $f \in C_b(\mathbb{R})$ that are also uniformly cont

(2) $\Rightarrow$ (3). take C_δ expand it by some delta.

for any closed C and $\varepsilon > 0 \quad \exists \delta$ s.t. $C_\delta = \{x \in S : d(x, C) < \delta\}$
 $P(C_\delta) < P(C) + \varepsilon$ and $C_\delta \downarrow C$ as $\delta \downarrow 0$

Next define $f = 1$ on C and s.t. f is uniformly continuous
 $= 0$ on $S \setminus C_\delta$

(See Urysohn's lemma or Theorem 1.2 Billingsley)



Then, by (2) $P_i(C) \leq \int f dP_i \rightarrow \int f dP \leq P(C_\delta) < P(C) + \varepsilon$

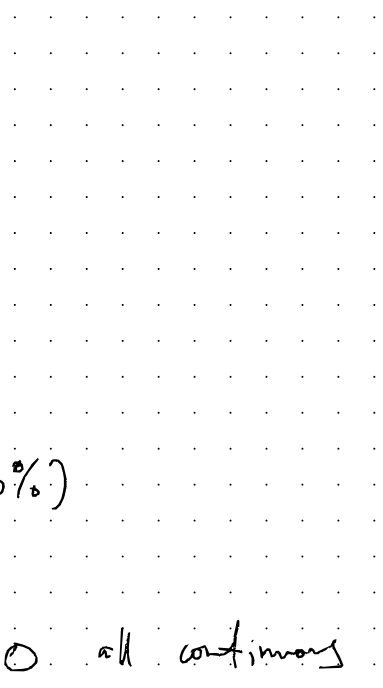
$\therefore$ take $\limsup$ and $\varepsilon \downarrow 0$

$$\limsup P_i(C) \leq P(C)$$

(3) $\Rightarrow$ (1) Let $f \in C_b(\mathbb{R})$. Goal is to show $\limsup \int f dP_i \leq \int f dP$ and similarly for $\liminf$

f is bounded by assumption $\therefore$ we can shift and scale it. wlog, assume $0 < f \leq 1$

Define nested closed sets $C_j = \{x \in S : f(x) \geq \frac{j}{n}\}$



$$\sum_{j=1}^n \frac{j-1}{n} P(C_{j-1} \setminus C_j) \leq \int f dP \leq \sum_{j=1}^n \frac{j}{n} P(C_j \setminus C_{j-1})$$

$$\text{also, } P(C_{j-1} \setminus C_j) = P(C_{j-1}) - P(C_j)$$

$$\therefore \frac{1}{n} \sum_{j=1}^n P(C_j) \leq \int f dP \leq \frac{1}{n} + \frac{1}{n} \sum_{j=1}^n P(C_j)$$

$$\therefore \limsup \int f dP_i \leq \frac{1}{n} + \frac{1}{n} \sum_{j=1}^n \limsup P_i(C_j)$$

$$\leq \frac{1}{n} + \frac{1}{n} \sum_{j=1}^n P(C_j)$$

$$\leq \frac{1}{n} + \int f dP$$

Take $n \rightarrow \infty$, $\limsup \int f dP_i \leq \int f dP$

Now, do the same thing but for $-f$ replacing f .

$$\Rightarrow \liminf \int f dP_i \geq \int f dP$$

$$\therefore \int f dP_i \rightarrow \int f dP \quad \square \text{ (60\%)}$$

Remark: - weak conv $P_i \rightarrow P, f \in C_b(\mathbb{R})$

- Radon metric: $\sup_f [\int f dP_i - \int f dP] \rightarrow 0$ all continuous $f: S \rightarrow [-1, 1]$

- if we take all f measurable $f: S \rightarrow [-1, 1]$, "Total Variation"

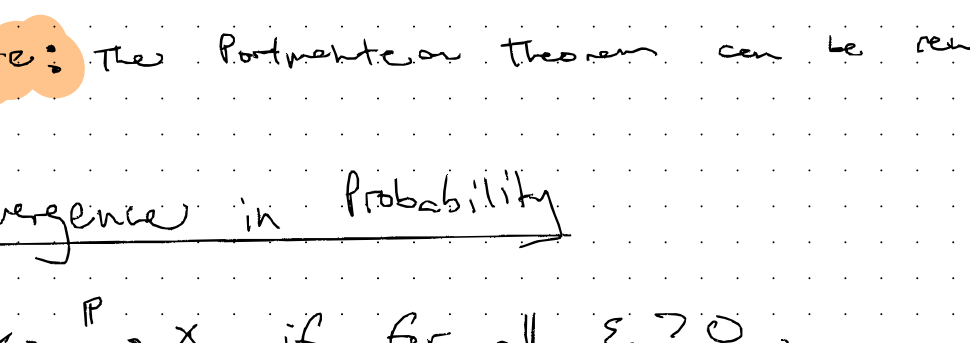
- if f are Lipschitz, constant 1 "how quickly does the convergence occur"
 conv in 1-wasserstein $\hookrightarrow$ useful in optimal transport theory of ML.

Conv of Random Variables

Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a Probability space $(S, \mathcal{F})$ be a metric space

Then a random variable $X: \mathcal{X} \rightarrow S$ defines a measure on $\mathcal{F}$

$$P(A) := \mu(X^{-1}(A)), \quad A \in \mathcal{F}$$



For a seq $\{X_i\}_{i=1}^\infty, X_i \xrightarrow{d} X$ (in distribution) if $P_i \rightarrow P$

For Notation: $E[X] = \int x(\omega) d\mu(\omega) = \int x dP(x)$

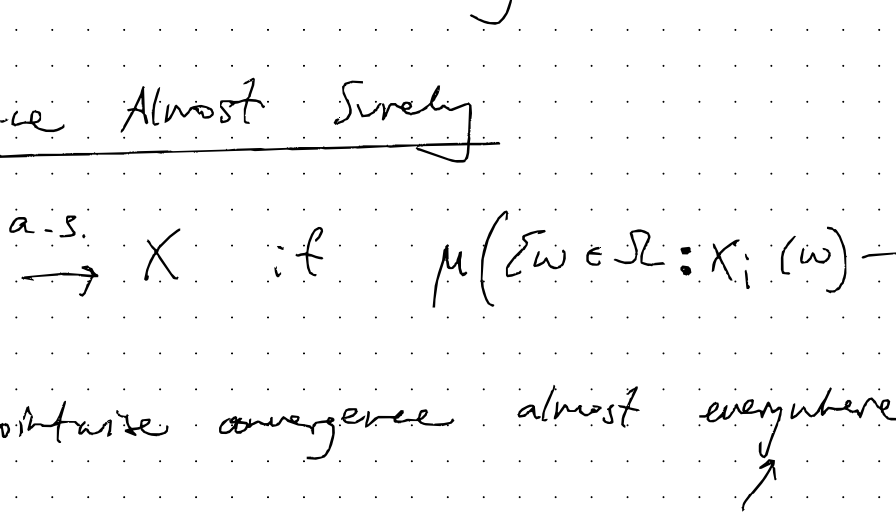
$$\text{also } P_i(A) := P_i(X_i \in A)$$

Note: The Portmanteau theorem can be rewritten for Random Variables.

Convergence in Probability

$X_i \xrightarrow{P} X$, if for all $\varepsilon > 0$,

$$\mu(\{\omega \in \mathcal{X} : d(X_i(\omega), X(\omega)) > \varepsilon\}) \rightarrow 0$$



In short, $P(d(X_i, X) > \varepsilon) \rightarrow 0$.

Convergence in Probability is closely related to d.

Convergence Almost Surely

$X_i \xrightarrow{a.s.} X$ if $\mu(\{\omega \in \mathcal{X} : X_i(\omega) \rightarrow X(\omega)\}) = 1$

i.e. pointwise convergence almost everywhere / surely
 analyst. probabilist

$$\mu(\{\omega \in \mathcal{X} : X_i(\omega) \not\rightarrow X(\omega)\}) = 0$$

Note: the metric d does not appear here

Convergence in L^p

$$X_i \xrightarrow{L^p} X, \quad E[d(X_i, X)^p] = \int d(X_i(\omega), X(\omega))^p d\mu(\omega) \rightarrow 0$$

function $\mathcal{X} \rightarrow \mathbb{R}^+$

if $\mathcal{F} = \mathbb{R}$ then $\int |X_i - X|^p d\mu \rightarrow 0$

convergence in the p th moment

Hierarchy of Convergence

- conv a.s. $\Rightarrow$ conv prob
- conv prob $\Rightarrow$ conv dist
- $1 \leq p \leq q \leq \infty, \text{ conv } L^p \Rightarrow \text{ conv } L^q$
- any $p \in [1, \infty], \text{ conv } L^p \Rightarrow \text{ conv prob}$

But convergence a.s. $\not\Rightarrow$ conv L^p

Borel-Cantelli Lemmas

Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a probability space $\{A_i\}_{i=1}^\infty, A_i \in \mathcal{F}$

then

$$\bullet \limsup_i A_i = \bigcap_{i=1}^\infty \bigcup_{j \geq i} A_j$$

$$\bullet \liminf_i A_i = \bigcup_{i=1}^\infty \bigcap_{j \geq i} A_j$$

Sometimes, we say A_i "infinitely often" or A_i i.o. for $\limsup_i A_i$ because if $\omega \in \limsup_i A_i$ Then $\forall N \in \mathbb{N} \quad \exists n > N$ s.t. $\omega \in A_n$

$\hookrightarrow$ We also say A_i "eventually", A_i ev. for $\liminf_i A_i$

because $\exists N \in \mathbb{N}$ s.t. $\forall n \geq N \quad \omega \in A_n$

Theorem (1st Borel-Cantelli)

Let $\{A_i\}_{i=1}^\infty$ with $A_i \in \mathcal{F}$ If $\sum_{i=1}^\infty \mu(A_i) < \infty$ then $\mu(\limsup_i A_i) = 0$

Proof: As the summation converges to a finite value, then $\sum_{i=n}^\infty \mu(A_i) \rightarrow 0$ as $n \rightarrow \infty$.

$$\therefore \mu(\limsup_i A_i) = \mu(\bigcap_{i=1}^\infty \bigcup_{j \geq i} A_j)$$

$$\leq \mu(\bigcup_{j=1}^\infty A_j)$$

$$\leq \sum_{j=1}^\infty \mu(A_j) \rightarrow 0 \text{ as } i \rightarrow \infty \quad \square$$

Theorem (2nd Borel-Cantelli Lemma)

Let $\{A_i\}_{i=1}^\infty$ be independent and $A_i \in \mathcal{F}$. Then if $\sum_{i=1}^\infty \mu(A_i) = \infty$

then $\mu(\limsup_i A_i) = 1$

Proof: Note that $1 - t \leq e^{-t}, \forall t \in \mathbb{R}$

One can check that if $\{A_i\}_{i=1}^\infty$ are independent then $\{A_i^c\}_{i=1}^\infty$ are independent

$$\bullet \text{ for any } i \in \mathbb{N} \text{ and } k \geq i \quad \mu\left(\bigcap_{j=i}^k A_j^c\right) = \prod_{j=i}^k [1 - \mu(A_j)]$$

$$\leq \exp\left[-\sum_{j=i}^k \mu(A_j)\right]$$

Take $K \rightarrow \infty$ and RHS $\rightarrow 0$

$$\therefore \mu\left(\bigcap_{j=i}^\infty A_j^c\right) = 0 \quad \forall i$$

$$\therefore \mu(\limsup_i A_i) = \mu\left(\bigcap_{i=1}^\infty \bigcup_{j \geq i} A_j\right)$$

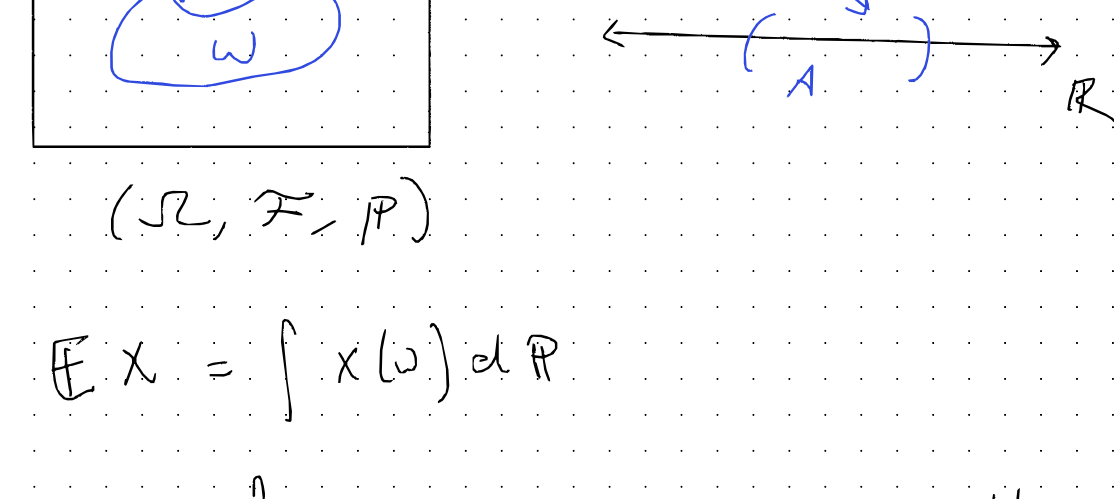
$$= 1 - \mu\left(\bigcup_{i=1}^\infty \bigcap_{j \geq i} A_j^c\right) = 1$$

$\uparrow$

$\square$

Let $\{X_i\}_{i=1}^{\infty}$ be random variables from $(\Omega, \mathcal{F}, \mathbb{P}) + (\mathbb{R}, \mathcal{B})$.
Hence, for any $A \in \mathcal{B}$

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



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

$S_n = \sum_{i=1}^n X_i$ is also a measurable random variable

Definition (Independence)

X and Y on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ but with possibly different $(\mathcal{X}, \mathcal{X})$ and $(\mathcal{Y}, \mathcal{Y})$.

"coordinates"

Then we say that X and Y are independent if

$$\mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B) \quad \forall A \in \mathcal{X} \text{ and } B \in \mathcal{Y}$$

This can be extended to **finite** collections $\{X_i\}_{i=1}^n$ with

$$\mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i), \quad \forall A_i$$

Also, $\{X_i\}_{i=1}^{\infty}$ is independent if every finite set of X_i 's is independent $\rightarrow$ similar to what we saw for σ -fields.

Note: since $\{X \in A\} = \{\omega \in \Omega : X(\omega) \in A\} = X^{-1}(A)$

Then X and Y are independent if and only if σ -fields $\sigma(X)$ and $\sigma(Y)$ are independent.

Identically Distributed "you want the RV's to have the same distribution"

$X: \Omega \rightarrow \mathbb{R}$ the distribution (or law) is the measure induced by X on $\mathbb{R}$ i.e. for $A \in \mathcal{B}$, this $\mathbb{P} \circ X^{-1}(A)$ we say that X and Y are identically distributed if measures $\mathbb{P} \circ X^{-1}$ and $\mathbb{P} \circ Y^{-1}$ coincide a.s.

Theorem (Weak law of large numbers)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. $\{X_i\}_{i=1}^{\infty}$ be random variables from $\Omega \rightarrow \mathbb{R}$ s.t. $\mathbb{E} X_i = c \in \mathbb{R}$ and $\mathbb{E} X_i^2 = 1 \quad \forall i = 1, \dots$

and $\mathbb{E}[(X_i - c)(X_j - c)] = 0 \quad \forall i \neq j$

Then $\frac{S_n}{n} \xrightarrow{\mathbb{P}} c$ in probability.

uncorrelated is a weaker condition than independence

Proof: wlog $c=0$, otherwise replace X_i with $X_i - c$.

Then for any $t > 0$, Chebyshev says,

$$\mathbb{P}\left(\frac{|S_n|}{n} \geq t\right) \leq \frac{\mathbb{E} S_n^2}{t^2 n^2} = \frac{\sum_{i,j=1}^n \mathbb{E}[X_i X_j]}{t^2 n^2} = \frac{1}{nt^2} \rightarrow 0$$

as $n \rightarrow \infty$

□

In the next theorem, we require independent X_i 's we also need the variance

$$\text{Var}(X) = \int (X - \mathbb{E} X)^2 d\mathbb{P}(\omega)$$

Theorem (Strong law of Large Numbers)

Let $\{X_i\}_{i=1}^{\infty}$ be iid random variables from $\Omega \rightarrow \mathbb{R}$.

If $\mathbb{E}|X_i| < \infty$, then $\frac{S_n}{n} \xrightarrow{\text{a.s.}} c$ for $c = \mathbb{E} X_i$

Also, if $\mathbb{E}|X_i| = \infty$, then $\frac{S_n}{n}$ does not converge to any finite limit.

we assume that the second moment is finite

Proof: (2) Assume $n^{-1} S_n \rightarrow c \in \mathbb{R}$ but also $\mathbb{E}|X_i| = \infty$

Note that $\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} \rightarrow 0$

$$\frac{S_n}{n} - \frac{S_{n-1}}{n} \rightarrow c - c = 0$$

Since $\mathbb{E}|X_i| = \infty$ then $\sum_{n=0}^{\infty} \mathbb{P}(X_i > n) = \infty$, then Borel-Cantelli says that $|X_n| > n$ for i.o. Thus: $\mathbb{P}\left(\frac{S_n - S_{n-1}}{n} \rightarrow 0\right) = 0$

infinitely often

This average happens on a set of measure 0

$$\therefore n^{-1} S_n \not\rightarrow c \Rightarrow \mathbb{R}$$

(1) Assume $\mathbb{E} X_i = c \in \mathbb{R}$ wlog $X_i \geq 0$ or else write $X = X^+ - X^-$ and do everything for X^+, X^- .

Also, independence of X, Y implies independence of X^+ and Y^+

Also, F denotes the law of X i.e. $F(x) = \mathbb{P}(X \leq x)$

We define $Y_i = X_i \mathbb{1}_{X_i \leq i}$

X_i

$$T_n = \sum_{i=1}^n Y_i$$

"truncated versions"

"forcing finite variance"

For any $\delta > 1$ we can define a non-decreasing integer sequence $k_n = \lfloor \delta^n \rfloor$

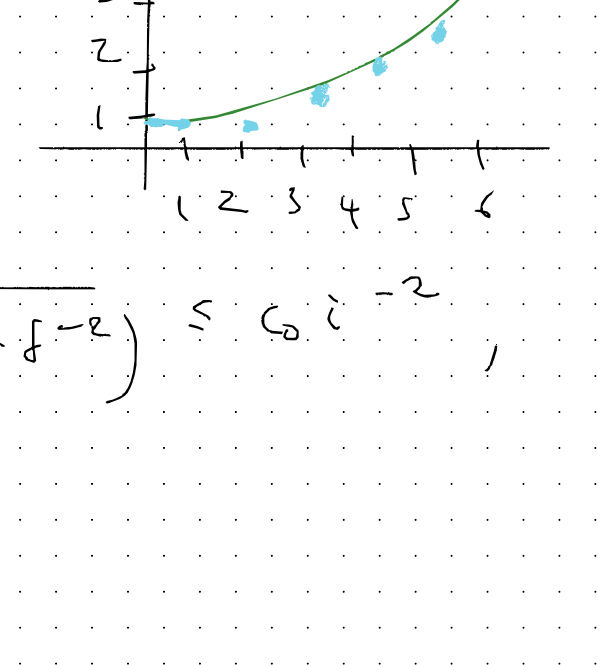
Then, $1 \leq k_n \leq \delta^n < k_{n+1} \leq \delta k_n$

$$k_n^{-2} \leq 4 \delta^{-2n}$$

$$\sum_{n=1}^{\infty} k_n^{-2} \mathbb{1}_{k_n \geq i} \leq 4 \sum_{n=1}^{\infty} \delta^{-2n} \mathbb{1}_{\delta^n \geq i} \leq \frac{4}{i^2(1-\delta^{-2})} \leq C_0 i^{-2}$$

for some $C_0 > 0$.

$$\sum_{i=k+1}^{\infty} i^{-2} < \int_k^{\infty} x^{-2} dx = \frac{1}{k}$$



Now, for any $t > 0, \exists c_1 > 0$ depending only on t and δ s.t.

"applying Chebyshev on infinite number of times"

$$\sum_{n=1}^{\infty} \mathbb{P}(|T_{k_n} - \mathbb{E} T_{k_n}| > t k_n) \leq c_1 \sum_{n=1}^{\infty} k_n^{-2} \text{Var}(T_{k_n})$$

$$= c_1 \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{i=1}^{k_n} \text{Var}(Y_i)$$

$$= c_1 \sum_{i=1}^{\infty} \left[\text{Var}(Y_i) \sum_{k_n \geq i} \frac{1}{k_n^2} \right]$$

$$\leq c_2 \sum_{i=1}^{\infty} i^{-2} \mathbb{E} Y_i^2$$

Note: $\text{Var}(Y_i) \leq \mathbb{E} Y_i^2$

$$= c_2 \sum_{i=1}^{\infty} i^{-2} \int_0^i x^2 dF(x)$$

$$= c_2 \sum_{i=1}^{\infty} i^{-2} \left\{ \sum_{k=0}^{i-1} \int_k^{k+1} x^2 dF(x) \right\}$$

$$\leq c_3 \sum_{k=0}^{\infty} \frac{1}{k+1} \int_k^{k+1} x^2 dF(x)$$

$k \leq x \leq k+1$

$$\leq c_3 \sum_{k=0}^{\infty} \int_k^{k+1} x dF(x)$$

$\therefore \frac{x}{k+1} \leq 1$

$$\leq c_3 \sum_{k=0}^{\infty} \mathbb{E} X_i < \infty$$

"increase at an exponential rate to allow for the convergence"

In the end, $\sum_{n=1}^{\infty} \mathbb{P}(|T_{k_n} - \mathbb{E} T_{k_n}| > t k_n) < \infty$

Borel-Cantelli (BC) says $\frac{1}{k_n} |T_{k_n} - \mathbb{E} T_{k_n}| \xrightarrow{\text{a.s.}} 0$

Since $\mathbb{E} Y_n \uparrow \mathbb{E} X_i$ we have that

group up convergence in probability to almost surely

$$k_n^{-1} \mathbb{E} T_{k_n} \uparrow \mathbb{E} X_i \text{ as } n \rightarrow \infty$$

$$\therefore k_n^{-1} T_{k_n} \xrightarrow{\text{a.s.}} \mathbb{E} X_i$$

To get back to S_n , note that for $X, \mathbb{E} X < \infty$

$$\iff \sum_{i=0}^{\infty} \mathbb{P}(X > i) < \infty$$

Consider sets like $I_k = \{k < X \leq k+1\}$

$$\text{Then } \sum_{i=1}^{\infty} \mathbb{P}(X_i \neq Y_i) = \sum_{i=1}^{\infty} \mathbb{P}(X_i > i) < \infty$$

BC says $\mathbb{P}(\limsup \{X_i \neq Y_i\}) = 0$

Hence, for i large enough $X_i = Y_i$ a.s.

? a function dependent on ω .

Let "large enough" mean all $i > m(\omega)$

(i.e. $X_i(\omega) = Y_i(\omega) \quad \forall i > m(\omega)$)

Furthermore, $k_n^{-1} S_{k_n} \rightarrow 0$ and $k_n^{-1} T_{k_n} \rightarrow 0$

as $n \rightarrow \infty$

$\therefore$ the "terms" where X_i and Y_i do not coincide become negligible as $n \rightarrow \infty$.

$\therefore k_n^{-1} S_{k_n} \xrightarrow{\text{a.s.}} \mathbb{E} X_i \quad \therefore$ we have almost sure (a.s.) convergence for a subsequence

Finally, $\frac{k_{n+1}}{k_n} \rightarrow \delta$ as $n \rightarrow \infty$

$$\therefore \exists n \text{ large enough s.t. } 1 \leq \frac{k_{n+1}}{k_n} < \delta^2$$

Thus, for $k_n < i < k_{n+1}$

$$\frac{S_{k_n}}{k_n} \leq \delta^2 \frac{S_i}{i} \leq \delta^n \frac{S_{k_{n+1}}}{k_{n+1}}$$

$$\delta^{-2} \mathbb{E} X_i \leq \liminf_{i \rightarrow \infty} \frac{S_i}{i} \leq \limsup_{i \rightarrow \infty} \frac{S_i}{i} \leq \delta^2 \mathbb{E} X_i$$

take $\delta \downarrow 1$

□

fails without finite first moment

Prohorov's Theorem

"Theorem about compactness"

Defn: (Uniform Tightness)

A collection of probability measures $\{\mu_i\}_{i \in \mathbb{I}}$ in a metric space is said to be "uniformly tight" if $\forall \varepsilon > 0 \exists K_\varepsilon$ compact s.t. $\mu_i(K_\varepsilon) > 1 - \varepsilon$ $\forall i$.

Bolzano-Weierstrass: every bounded sequence has a convergent subsequence

Theorem (Prohorov)

← couple different versions

"uniform tightness"

For sequence of prob measures $\{\mu_i\}_{i=1}^\infty$ if the seq is U.T. then it is relatively compact (sequentially compact) i.e. subseq $\mu_{i_k} \Rightarrow$ weakly conv. subsubseq $\mu_{i_k} \Rightarrow \mu$ for some μ depending on the subsequence.
 (possibly different μ)

Proposition:

If $\{\mu_i\}_{i=1}^\infty$ and μ are prob. measures s.t. $\forall \mu_{i_k} \Rightarrow \mu$
 (Same μ)

Proof:

Assume not, then $\exists f \in C_b$ s.t. $\int f d\mu_i \not\rightarrow \int f d\mu$.

$\exists \varepsilon > 0 \exists i_k$ s.t. $|\int f d\mu_{i_k} - \int f d\mu| > \varepsilon \forall k$

However, $\exists i_k$ s.t. $\mu_{i_k} \Rightarrow \mu$ contradiction. $\square$

Defn (Gaussian measure)

A Borel measure γ on $(\mathbb{R}, \mathcal{B})$ is said to be Gaussian with mean m and variance σ^2 if

$$\gamma([a, b]) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b \exp\left[-\frac{1}{2\sigma^2}(x-m)^2\right] d\lambda(x)$$

Also, for $\sigma = 0$, $\gamma = \delta_m$ (Dirac measure). We say γ is a degenerate Gaussian measure.

Defn (in $\mathbb{R}^d$)

γ on $(\mathbb{R}^d, \mathcal{B})$ is Gaussian if $\forall$ linear functions $f: \mathbb{R}^d \rightarrow \mathbb{R}$ the induced measure $\gamma \circ f^{-1}$ on $(\mathbb{R}, \mathcal{B})$ is Gaussian.

Any linear combination is Gaussian

Defn (R.V.)

A R.V. Z from $(\Omega, \mathcal{F}, \mu)$ to $(\mathbb{R}^d, \mathcal{B})$ is Gaussian if $\gamma := \mu \circ Z^{-1}$ is a Gaussian measure.

For vectors $u, v \in \mathbb{R}^d$, the inner product $\langle u, v \rangle = \sum_{i=1}^d u_i v_i$
 $|u|^2 = \langle u, u \rangle$

Characteristic function

A probability measure μ on $(\mathbb{R}^d, \mathcal{B})$ the characteristic function is $\tilde{\mu}: \mathbb{R}^d \rightarrow \mathbb{C}$

$$\tilde{\mu}(v) := \int \exp\{i \langle u, v \rangle\} d\mu(u)$$

↑
"up to scaling" $= \sqrt{1}$

If $\tilde{\mu}$ is integrable w.r.t λ on $\mathbb{R}^d$

$$p(u) = \frac{1}{(2\pi)^d} \int \tilde{\mu}(v) \exp\{-i \langle u, v \rangle\} d\lambda(v)$$

$p(u)$ is the probability density function for μ λ a.e.

$$\int f d\mu = \int f \cdot p d\lambda$$

Convolution

"convergence of characteristic functions"

For two measures μ and ν on $(\mathbb{R}^d, \mathcal{B})$ the convolution measure

$$(\mu * \nu)(B) := \int \nu(B-u) d\mu(u) \quad \forall B \in \mathcal{B}$$

$$B-u = \{y \in \mathbb{R}^d : y+u \in B\}$$

Note: the operator $*$ is both **associative** and **commutative**

$$\text{Also } (\overline{\mu * \nu}) = \tilde{\mu} \cdot \tilde{\nu}$$

↑ convolution ↑ product!

Also, for independent R.V.'s X, Y with measures μ, ν $X+Y$ has induced measure $\mu * \nu$

Theorem (uniqueness of $\sim$)

Let μ and ν be prob measures if $\tilde{\mu} = \tilde{\nu}$ then $\mu = \nu$.

Proof: Let γ_σ be a mean zero Gaussian measure on $\mathbb{R}^d$ with co-variance $\sigma^2 I_d$.

Denote $\mu^{(\sigma)} = \mu * \gamma_\sigma$ and same for $\nu^{(\sigma)}$.

It can be shown that the corresponding density functions for $\mu^{(\sigma)}$ and $\nu^{(\sigma)}$ are

$$p^{(\sigma)}(u) = \frac{1}{(2\pi)^d} \int \tilde{\mu}(t) \exp\{-i \langle u, t \rangle - \frac{1}{2} \sigma^2 |t|^2\} d\lambda(t)$$

$$\tilde{\nu}^{(\sigma)}(u) = \frac{1}{(2\pi)^d} \int \tilde{\nu}(t) \exp\{-i \langle u, t \rangle - \frac{1}{2} \sigma^2 |t|^2\} d\lambda(t)$$

$$\text{Since } \tilde{\mu} = \tilde{\nu} \text{ then } \mu^{(\sigma)} = \nu^{(\sigma)} \quad \forall \sigma > 0$$

Let X be a R.V. corresponding to μ and Z to γ_σ

Then $\mu^{(\sigma)}$ comes from $X + \sigma Z$
 (multivariate normal random variable)

$$\therefore X + \sigma Z \xrightarrow{\sigma \downarrow 0} X \quad \text{as } \sigma \downarrow 0$$

$\therefore$ conv in probability

$\therefore$ convergence in distribution $\mu^{(\sigma)} \Rightarrow \mu$ as $\sigma \downarrow 0$

$$\text{Since } \mu^{(\sigma)} \Rightarrow \mu \text{ and } \nu^{(\sigma)} \Rightarrow \nu$$

$$\mu = \nu$$

Multivariate

Theorem (CLT) (requires finite second moment)

Let $(\Omega, \mathcal{F}, \mu)$ be a prob space $\{X_n\}_{n=1}^\infty$ are iid random variables on $(\mathbb{R}^d, \mathcal{B})$ s.t. $\mathbb{E} X_1 = 0$, $\mathbb{E} |X_1|^2 < \infty$

Let $S_n = \sum_{j=1}^n X_j$, then $n^{-\frac{1}{2}} S_n \xrightarrow{d} Z$, where Z is Gaussian random variable with 0 mean and co-variance Σ with

$$\Sigma_{jk} = \mathbb{E}[X_{1j} X_{1k}]$$

Lemma: For a U.T. seq of prob measures μ_i on $\mathbb{R}^d$, if $\forall v \in \mathbb{R}^d$, $\tilde{\mu}_i(v) \rightarrow \tilde{\mu}(v)$, then $\mu_i \Rightarrow \mu$ where μ corresponds to $\tilde{\mu}$.
 (uniform tightness)

Proof: μ_i are U.T. $\therefore$ Prohorov says every subseq μ_{i_k} has convergent μ_{i_k} .

All these subsubsequences have a corresponding seq of characteristic functions that converge to $\tilde{\mu}$.

Thus by uniqueness all $\mu_{i_k} \Rightarrow \mu$ (weakly converge).

By the prop from above $\mu_i \Rightarrow \mu$ $\square$

Proof (CLT)

Let $(\Omega, \mathcal{F}, \mu)$ be a prob space $X_j: \Omega \rightarrow \mathbb{R}^d$ "characteristic function of a Gaussian measure"

R.V.'s X_j are mean zero and independent $\mathbb{E} \langle X_j, X_k \rangle = 0$ for $j \neq k$

Thus $\mathbb{E} |n^{-\frac{1}{2}} S_n|^2 = \frac{1}{n} \mathbb{E} \left[\sum_{j,k=1}^n \langle X_j, X_k \rangle \right]$

$$= \mathbb{E} |X_1|^2$$

For any $\varepsilon > 0 \exists M_\varepsilon > 0$ s.t. $\mathbb{E} |X_1|^2 < \varepsilon$
 (by assumption on the finiteness of the second moment)

Thus, by Chebyshev's Ineq:

$$\mathbb{P}(|n^{-\frac{1}{2}} S_n| > M_\varepsilon) < \varepsilon$$

"random variables induce measures"

$\therefore$ seq $n^{-\frac{1}{2}} S_n$ is U.T.

For $v \in \mathbb{R}^d$, R.V.'s $\langle v, X_j \rangle$ are iid real R.V.'s

with $\mathbb{E} \langle v, X_j \rangle = 0$, $\mathbb{E} \langle v, X_j \rangle^2 < \infty$.

$$\text{Let } h(v) := \mathbb{E} \exp\{i \langle v, X_j \rangle\}$$

$$h(0) = 1$$

$$\nabla h(0) = 0$$

$$\nabla^2 h(0) = -\mathbb{E} [X_j X_j^T] = -\Sigma$$

By Taylor's Theorem, $h(v) = 1 - \frac{1}{2} v^T \Sigma v + o(|v|^2)$
 (covariance matrix)
 (Euclidean Norm)

$\therefore$ for any fixed v , $\mathbb{E} \exp\{i \langle n^{-\frac{1}{2}} S_n, v \rangle\} = h(n^{-\frac{1}{2}} v)^n$
 (little o)
 (as $|v|^2 \rightarrow 0$)

$$= \left[1 - \frac{v^T \Sigma v}{2n} + o\left(\frac{|v|^2}{n}\right) \right]^n \xrightarrow{n \rightarrow \infty} \exp\left\{-\frac{1}{2} v^T \Sigma v\right\}$$

which is the characteristic function for Z .

by the above lemma.

$$n^{-\frac{1}{2}} S_n \xrightarrow{d} Z$$

 $\square$

Recap: L^p spaces, convergence, strong law, CLT.
now ergodic $\rightarrow$ stochastic processes & dynamic

"Ergodic if it explores the whole space"

Defn (Measure Preserving Map): "probability space not required"

let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space, a map/transformation $T: \mathcal{X} \rightarrow \mathcal{X}$ is called measure preserving if

$$\mu(T^{-1}(A)) = \mu(A) \quad \forall A \in \mathcal{F}$$

Definition (Invariance)

A set $A \in \mathcal{F}$ is T -invariant if $T(A) = A$

Note: The collection of all T -invariant sets forms a σ -field $\mathcal{F}_T$

A measurable function is T -invariant if $f = f \circ T$

Note f is T -invariant $\iff f = f(T(\dots))$

f is $\mathcal{F}_T$ -measurable

Defn (Ergodic)

A mapping T is ergodic if for any $A \in \mathcal{F}_T$ we have

$$\mu(A) = 0 \text{ or } \mu(A^c) = 0$$

Examples $([0,1], \mathcal{B}, \lambda)$

$\rightarrow$ Shift Map: $T(x) = x + a \pmod 1$

$$= \begin{cases} x+a, & x+a \leq 1 \\ x+a-1, & x+a > 1 \end{cases}$$

$\rightarrow$ Baker's Map: $T(x) = 2x - \lfloor 2x \rfloor$

"measure preserving map"

Fun Facts (things we won't prove)

(1) If f is integrable and T is measure-preserving

then $f \circ T$ is integrable and $\int f d\mu = \int f \circ T d\mu$

(2) If T is ergodic and f is T -invariant

then $f = c$ μ -a.e. some constant c

Ergodic Theorems

$(\mathcal{X}, \mathcal{F}, \mu)$ is a measure space

T is measure preserving

$f: \mathcal{X} \rightarrow \mathbb{R}$ is measurable

and we define

$$S_n = S_n(f) = f + f \circ T + f \circ T^2 + \dots + f \circ T^{n-1}$$

and $S_0 = 0$

Lemma (Maximal Ergodic Lemma)

let f be integrable and $S^* = \sup_{n \geq 0} S_n(f)$

Then $\int_{S^* > 0} f d\mu \geq 0$ $\leftarrow S^* > 0$

$\{ \omega \in \mathcal{X} : S^*(\omega) > 0 \}$

Proof:

let $S_n^* = \max_{0 \leq m \leq n} S_m(f)$

and $A_n = \{ \omega \in \mathcal{X} : S_n^* > 0 \}$. Then, for $1 \leq m \leq n$

$$S_m(f) = f + S_{m-1} \circ T \leq f + S_n^* \circ T$$

Furthermore, on A_n

$$S_n^* = \max_{0 \leq m \leq n} S_m(f) \leq f + S_n^* \circ T$$

And on A_n^c , $S_n^* = 0 \leq S_n^* \circ T$

$$\therefore \int_{\mathcal{X}} S_n^* d\mu \leq \int_{A_n} f d\mu + \int_{\mathcal{X}} S_n^* \circ T d\mu$$

Since S_n^* is integrable and T is measure preserving

$$\int S_n^* d\mu = \int S_n^* \circ T d\mu < \infty$$

$$\therefore 0 \leq \int_{A_n} f d\mu \text{ as } n \rightarrow \infty, A_n \uparrow \{S^* > 0\}$$

$$\text{Thus } \int_{S^* > 0} f d\mu = \lim_{n \rightarrow \infty} \int_{A_n} f d\mu \geq 0$$

by Dominated convergence with $|f|$.

$$\lim_{n \rightarrow \infty} \int_{\mathcal{X}} f \mathbb{1}_{A_n} d\mu = \int_{\mathcal{X}} f \mathbb{1}_{\{S^* > 0\}} d\mu$$

Theorem (Birkhoff's Ergodic Theorem) (almost everywhere)

let $(\mathcal{X}, \mathcal{F}, \mu)$ be a σ -finite measure space and

$f \in L^1(\mathcal{X}, \mathcal{F}, \mu)$. Then $\exists$ an invariant $\bar{f} \in L^1(\mathcal{X}, \mathcal{F}, \mu)$

$$\text{s.t. } \int |f| d\mu < \infty \text{ and } \frac{S_n(f)}{n} \rightarrow \bar{f} \quad \mu\text{-a.e. as } n \rightarrow \infty$$

Proof:

Both $\liminf_{n \rightarrow \infty} n^{-1} S_n(f)$ and $\limsup_{n \rightarrow \infty} n^{-1} S_n(f)$

are T -invariant.

$$\text{Indeed, } n^{-1} S_n(f) \circ T = \left[\frac{S_{n+1}(f) - f}{n} \right]$$

$$= \left(\frac{n+1}{n} \right) \frac{S_{n+1}(f)}{n+1} - \frac{f}{n}$$

$\therefore$ we define for $a < b$

$$D_{a,b} = \{ \omega \in \mathcal{X} : \liminf_{n \rightarrow \infty} n^{-1} S_n(f) < a < b < \limsup_{n \rightarrow \infty} n^{-1} S_n(f) \}$$

Note that $D_{a,b}$ is T -invariant

which, assume $b > 0$ else $a < 0$ and multiply by -1

For some $B \in \mathcal{F}$ s.t. $\mu(B) < \infty$ set $g = f - b \mathbb{1}_B$

$\therefore g$ is integrable and for each $\omega \in D_{a,b}$

$$\exists n \text{ s.t. } S_n(g)(\omega) \geq S_n(f)(\omega) - nb \geq 0$$

$\therefore S^*(g) > 0$ and maximal ergodic lemma says

$$0 \leq \int_{D_{a,b}} (f - b \mathbb{1}_B) d\mu = \int_{D_{a,b}} f d\mu - b \mu(B)$$

μ is σ -finite $\therefore \exists B_n \in \mathcal{F}$ s.t. $B_n \uparrow D_{a,b}$ and $\mu(B_n) < \infty$

$$\text{th. Thus } b \mu(D_{a,b}) = \lim_{n \rightarrow \infty} b \mu(B_n) \leq \int_{D_{a,b}} f d\mu$$

$$\therefore \mu(D_{a,b}) < \infty$$

Repeat the above for $-a$ and $-f$ to get

$$-a \mu(D_{a,b}) \leq \int_{D_{a,b}} -f d\mu$$

$$\therefore b \mu(D_{a,b}) \leq \int_{D_{a,b}} f d\mu \leq a \mu(D_{a,b})$$

$$\text{but } a < b$$

$$\therefore \mu(D_{a,b}) = 0$$

$$\text{Now, let } E = \{ \omega \in \mathcal{X} : \liminf_{n \rightarrow \infty} \frac{S_n(f)}{n} < \limsup_{n \rightarrow \infty} \frac{S_n(f)}{n} \}$$

E is T -invariant since T -invariant

$$E = \bigcup_{a < b} D_{a,b} \Rightarrow \mu(E) = 0$$

$\therefore n^{-1} S_n(f)$ converges in $[-\infty, \infty]$ on E^c

$$\therefore \text{Define } \bar{f} := \lim_{n \rightarrow \infty} \frac{S_n(f)}{n}, \quad \omega \in E^c$$

0, $\omega \in E$

$$\text{Finally, } \int |f \circ T^n| d\mu = \int |f| d\mu$$

$$\therefore \int |S_n(f)| d\mu \leq n \int |f| d\mu \quad \forall n$$

$$\text{By Fatou, } \int |f| d\mu = \int \liminf_{n \rightarrow \infty} |n^{-1} S_n(f)| d\mu$$

$$\leq \liminf_{n \rightarrow \infty} \int |n^{-1} S_n(f)| d\mu$$

$$\leq \int |f| d\mu$$

Theorem (von Neumann's Ergodic Theorem) (L^2)

let $\mu(\mathcal{X}) < \infty$ and $p \in [1, \infty)$. Then, $\forall f \in L^p(\mathcal{X}, \mathcal{F}, \mu)$

$$\exists \bar{f} \in L^p \text{ s.t. } n^{-1} S_n(f) \rightarrow \bar{f} \text{ in } L^p \quad \|n^{-1} S_n(f) - \bar{f}\|_p \rightarrow 0$$

Proof:

$$\text{Note } \|f \circ T\|_p^p = \int |f| \circ T^p d\mu = \int |f|^p d\mu$$

$$= \|f(T^n(\dots))\|_p^p$$

By Minkowski, $\|n^{-1} S_n(f)\|_p \leq \|f\|_p$ as $f \in L^p$, for any $\varepsilon > 0$,

we can choose $c > 0$ s.t. $\|f - g\| < \frac{\varepsilon}{3}$

$$g = \min[\max\{-c, f\}, c]$$

By Birkhoff's Theorem

$$\frac{S_n(g)}{n} \rightarrow \bar{g} \quad \mu\text{-a.e.}$$

because the measure space is finite, we can apply dominated convergence.

Note that $|n^{-1} S_n(g)| \leq c$

$\therefore$ by Dom. Conv $\exists N$ s.t. $\forall n > N$

$$\|n^{-1} S_n(g) - \bar{g}\|_p < \frac{\varepsilon}{3}$$

By Fatou,

$$\|f - \bar{g}\|_p^p = \int \liminf_{n \rightarrow \infty} |n^{-1} S_n(f - g)|^p d\mu$$

$$\leq \liminf_{n \rightarrow \infty} \int |n^{-1} S_n(f - g)|^p d\mu$$

$$= \|f - g\|_p^p < \left(\frac{\varepsilon}{3}\right)^p$$

$\therefore$ for $n > N$

$$\left\| \frac{S_n(f)}{n} - \bar{f} \right\|_p \leq \left\| \frac{S_n(f - g)}{n} \right\|_p + \left\| \frac{S_n(g)}{n} - \bar{g} \right\|_p + \|\bar{g} - \bar{f}\|_p$$

$$\leq \varepsilon$$

$\square$

Example Application $\Rightarrow$ Strong Law of Large Numbers (SLLN)

let $(\mathcal{X}, \mathcal{F}, P)$ be a probability space with iid real-valued random variables $\{X_i\}_{i=1}^\infty$ with distribution measure dF s.t.

$$dF(A) = \int \mathbb{1}_A dF$$

let $(S, \mathcal{P})$ with $S = \mathbb{R}^{\mathbb{N}}$ be a countable product of copies of $\mathbb{R}$ on $\mathcal{P}$ generated by the π -system

$$\mathcal{P} = \sigma(\mathcal{A}) \text{ where}$$

$$\mathcal{A} = \{ \prod_{n=1}^\infty A_n : A_n \in \mathcal{B} \quad \forall n, A_n = \mathbb{R} \}$$

essentially

let $X: \mathcal{X} \rightarrow \mathbb{R}^{\mathbb{N}}$ be $X(\omega) = (X_1(\omega), X_2(\omega), \dots)$

$$\mu(A) := P \circ X^{-1}(A) = \prod_{n=1}^\infty dF(A_n) \text{ where } A = \prod_{n=1}^\infty A_n$$

For a sequence $(x_1, x_2, \dots)$ define the shift map

$$T: \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}^{\mathbb{N}} \text{ s.t.}$$

$$T(x_1, x_2, \dots) = (x_2, x_3, \dots)$$

Proposition: T is measure preserving + ergodic

(look up Kolmogorov's zero-one law)

Theorem (SLLN)

let $\{X_i\}_{i=1}^\infty$ be iid r.v.s $\mathcal{X} \rightarrow \mathbb{R}$ if $E|X_i| < \infty$,

$$\text{then } \frac{S_n}{n} \xrightarrow{\text{a.s.}} c = E X_i$$

Proof:

Let $f: \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$

$$f(x_1, x_2, \dots) = x_1$$

$$\text{Then } S_n = f + f \circ T + \dots + f \circ T^{n-1} = x_1 + \dots + x_n$$

$\therefore$ By Birkhoff $\exists \bar{f} \in L^1$, invariant s.t. $\frac{S_n}{n} \rightarrow \bar{f}$ a.s.

Since T is ergodic then $\bar{f} = c$ a.s

Lastly, via von Neumann with $p=1$,

$$c = \int \bar{f} d\mu = \lim_{n \rightarrow \infty} \int n^{-1} S_n(f) d\mu = \int f d\mu$$

$$= E X_i$$

$\square$