II Probability & Measure

~ Richard Nickl

0. Introduction

1) Replace the Riemann integral by the more powerful Lebesgue integral.

A space of 'integrable functions' should be the topological completion of the normed space $(C[0,1], \|\cdot\|_1)$; $\|f\|_1 = \int_0^1 |f(x)|dx$,

in the sense that any $\{f_n : n \in \mathbb{N}\}$ in $C[0,1]$ that is Cauchy for $\|\cdot\|$ has a limit in the space $\overline{C[0,1]}$. It turns out $\neq$ {Riemann integrable} but equals the space $\mathcal{L}^1$ of Lebesgue-integrable functions.

$\mathcal{L}^2$ ... Hilbert space!

2) Axiomatic prob theory

↪ derive key results such as: LLN, CLT, ergodic thm

3) Problem of measure

Does every subset of $\mathbb{R}^d$ have a volume?

1. Measures

Def^n Let E be any set, let $\mathcal{P}(E)$ be the family of all subsets of E.
A <u>σ-algebra</u> $\mathcal{E}$ on E is a family of subsets of E such that

1) $\emptyset \in \mathcal{E}$
2) $A \in \mathcal{E} \Rightarrow A^c = E \setminus A \in \mathcal{E}$
3) $\forall n, A_n \in \mathcal{E} \Rightarrow \bigcup_n A_n \in \mathcal{E}$

Easy Since $\bigcap_n A_n = (\bigcup_n A_n^c)^c$, $\mathcal{E}$ closed under ctble intersection and also $B \setminus A = B \cap A^c \in \mathcal{E}$ whenever $A, B \in \mathcal{E}$

We call $(E, \mathcal{E})$ a <u>measurable space</u>, $A \in \mathcal{E}$ is a measurable set.

A <u>measure</u> μ on $(E, \mathcal{E})$ is a (non-negative) set function
$\mu : \mathcal{E} \to [0, \infty]$ such that

1) $\mu(\emptyset) = 0$
2) for $(A_n \in \mathcal{E}, n \in \mathbb{N})$ disjoint, $\mu(\bigcup_{n=1}^\infty A_n) = \sum_{n=1}^\infty \mu(A_n)$

"countable additivity"

We call $(E, \mathcal{E}, \mu)$ a <u>measure space</u>.

<u>Remark</u> When E is countable then any measure μ on $\mathcal{E} = \mathcal{P}(E)$ necessarily satisfies $\mu(A) = \mu(\bigcup_{a \in A} \{a\}) = \sum_{a \in A} \mu(\{a\})$.

So μ corresponds to a 'mass function' $m : E \to [0,\infty]$, $x \mapsto \mu(\{x\})$

When E is not countable, we generally need to work with σ-algebras $\mathcal{E} \subsetneq \mathcal{P}(E)$ which cannot be described explicitly.

We define, for $\mathcal{A}$ any collection of subsets of E,

$$\sigma(\mathcal{A}) = \{ A \subseteq E : A \in \mathcal{E} \text{ for all } \sigma\text{-algebras } \mathcal{E} \supseteq \mathcal{A} \}$$

the σ-algebra <u>generated by</u> $\mathcal{A}$. It can be shown (Ex Sheet) that it's just $\bigcap_{\substack{\mathcal{E}\ \sigma\text{-algebra} \\ \mathcal{E} \supseteq \mathcal{A}}} \mathcal{E}$ and is a σ-algebra.

Choices for $\mathcal{A}$ will include

→ $\mathcal{A}$ is a <u>ring</u> of subsets of E if $\emptyset \in \mathcal{A}$, and if
$\forall A, B \in \mathcal{A} \Rightarrow \underset{B \setminus A}{\cancel{A^c}} \in \mathcal{A}$, $A \cup B \in \mathcal{A}$

→ $\mathcal{A}$ is an <u>algebra</u> if $\emptyset \in \mathcal{A}$,
$\forall A, B \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$, $A \cup B \in \mathcal{A}$

<u>Remark</u> Since $A \cap B = (A \cup B) \setminus (A \Delta B)$
where $A \Delta B = (A \setminus B) \cup (B \setminus A)$, we see that $A \cap B \in \mathcal{A}$ when A, B belong to a ring $\mathcal{A}$.

SET FUNCTIONS

For $\mathcal{A} \subseteq \mathcal{P}(E)$ s.t. $\emptyset \in \mathcal{A}$, we say that μ is a <u>set function</u> if
$\mu : \mathcal{A} \to [0, \infty]$, $\mu(\emptyset) = 0$

We say that μ is 1) <u>increasing</u> if $\forall A, B \in \mathcal{A}$ ~~disjoint~~ s.t. $A \subseteq B$ we have $\mu(A) \leq \mu(B)$

2) <u>additive</u> if $\forall A, B \in \mathcal{A}$ disjoint s.t. $A \cup B \in \mathcal{A}$ we have $\mu(A \cup B) = \mu(A) + \mu(B)$

3) <u>ctbly additive</u> if $\forall (A_n) \in \mathcal{A}$ disjoint s.t. $\bigcup_n A_n \in \mathcal{A}$ we have $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$

4) ctbly subadditive if $\forall (A_n) \in \mathcal{A}$ s.t. $\bigcup_n A_n \in \mathcal{A}$
we have $\mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$

Remark (One shows that) a countably additive set function on a σ-algebra is a measure. One shows that a measure satisfies 1, 2, 4 too.

Theorem (Caratheodory)
Let $\mathcal{A} \subseteq \mathcal{P}(E)$ be a ring of subsets of E and let
$\mu : A \longrightarrow [0;\infty]$ be a ctbly additive set function.
Then μ extends to a measure on $\sigma(A)$.

$\mathcal{A}$ ring $\Rightarrow$ $\emptyset \in \mathcal{A}$

$A, B \in \mathcal{A} \Rightarrow A \cup B,\ B \setminus A \in \mathcal{A}$

- μ is countably additive on $\mathcal{A}$ if $\forall\ A_n \in \mathcal{A}$ disjoint s.t. $\bigcup_n A_n \in \mathcal{A} \Rightarrow \mu(\bigcup_n A_n) = \sum_n \mu(A_n)$

"Disjointification Fact"

Let $A_n \in \mathcal{A}$, $\mathcal{A}$ a ring (or σ-algebra), consider $\bigcup_n A_n$.

Then write $\widetilde{A}_n = \bigcup_{j=1}^{n} A_j$, so $\widetilde{A}_n$ is increasing $(\widetilde{A}_n \uparrow)$, and defining $\widetilde{B}_1, \ldots$ via $\widetilde{B}_n = \widetilde{A}_n \setminus \widetilde{A}_{n-1}$.

Then the $\widetilde{B}_n$ are all disjoint, and $\bigcup_n A_n = \bigcup_n \widetilde{B}_n$ note $\widetilde{B}_n \in \mathcal{A}$

Theorem $\mathcal{A}$ a ring on E, $\mu : \mathcal{A} \to [0, \infty]$ countably additive

- Then $\exists$ a measure μ^* on $\sigma(\mathcal{A})$ s.t. $\mu^*|_{\mathcal{A}} = \mu$.

Proof For $B \subseteq E$, define the outer measure

$$\mu^*(B) = \inf \left\{ \sum_n \mu(A_n) : A_n \in \mathcal{A},\ B \subseteq \bigcup_n A_n \right\}$$

set $\mu^*(B) = \infty$ when no such (A_n) exists.

Clearly $\mu^*(\emptyset) = 0$, and μ^* is increasing on $\mathcal{P}(E)$.

Let us say that $A \subseteq E$ is μ^*-measurable if $\forall\ B \subseteq E$, we have

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

$\leftarrow$ KEY IDEA

- and define

$$\mathcal{M} = \{ A \subseteq E \mid A \text{ is } \mu^*\text{-measurable} \},$$

the set of all μ^*-measurable subsets of E.

Plan Show $\mathcal{M}$ is a σ-algebra, $\mathcal{M} \supseteq \mathcal{A}$, μ^* is a measure on $\mathcal{M}$, $\mu^*|_{\mathcal{A}} = \mu$.

Step I (μ^* is countably sub-additive)

For any $B_n, B \subseteq E$, suppose $B \subseteq \bigcup_n B_n$, then

$$\boxed{\mu^*(B) \leq \sum_n \mu^*(B_n)} \qquad (+)$$

- To prove this, may assume $\mu^*(B_n) < \infty$ $\forall n$.

Thus $\forall n \in \mathbb{N}$, $\forall \varepsilon > 0$, $\exists A_{nm}$ s.t. $B_n \subseteq \bigcup_m A_{nm}$,
where $A_{nm} \in \mathcal{A}$ s.t. $\mu^*(B_n) + \frac{\varepsilon}{2^n} \geqslant \sum_m \mu(A_{nm})$

Also, $B \subseteq \bigcup_n B_n \subseteq \bigcup_n \bigcup_m A_{nm}$, and $\mu^* \uparrow$

$$\mu^*(B) \leq \mu^*\left(\bigcup_n \bigcup_m A_{nm}\right) \leq \sum_{n,m} \mu(A_{nm})$$
$$\leq \sum_n \mu^*(B_n) + \varepsilon \sum_n 2^{-n}$$
$$= \sum_n \mu^*(B_n) + \varepsilon$$

So (†) follows.

<u>STEP II</u> (μ^* extends μ)

<u>FACT (Ex Sheet)</u> A countably additive set function on a ring $\mathcal{A}$, then it is also additive, increasing and countably subadditive.

Now write $A \in \mathcal{A}$ as $A = \bigcup_n (A \cap A_n)$ for any seq $A_n \in \mathcal{A}$ s.t. $A \subseteq \bigcup_n A_n$.

By countable subadditivity, and $A \cap A_n \subseteq A_n$, we have

$$\mu(A) = \mu\left(\bigcup_n (A \cap A_n)\right) \overset{\text{inc}}{\leq} \sum_n \mu(A_n)$$

so that taking the infimum over all A_n we see that

$$\mu(A) \leq \mu^*(A), \quad \forall A \in \mathcal{A}$$

It is obvious that $\mu^*(A) \leq \mu(A)$, since we can take

$$A_1 = A, \; A_n = \emptyset \text{ for } n > 1$$

<u>Step III</u> ($\mathcal{M} \supseteq \mathcal{A}$)

Let $A \in \mathcal{A}$, $B \subseteq E$, then we need to check that

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

By countable subadditivity (†), since $B = (B \cap A) \cup (B \cap A^c)$,

$$\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

Need to show $\geqslant$ to prove the assertion.

For this, may assume $\mu^*(B) < \infty$. Hence as before, let $\varepsilon > 0$,

$$\exists A_n \in \mathcal{A} \text{ s.t. } B \subset \bigcup_n A_n \text{ and s.t. } \mu^*(B) + \varepsilon \geqslant \sum_n \mu(A_n)$$

Now, $B \cap A \subseteq \bigcup_n \underbrace{(A_n \cap A)}_{\text{ells of ring}}$, $B \cap A^c \subseteq \bigcup_n \underbrace{(A_n \cap A^c)}_{\text{ells of ring}}$

and so by (†),

$$\mu^*(B\cap A)+\mu^*(B\cap A^c)$$
$$\le \sum_n \mu(A_n\cap A)+\sum_n \mu(A_n\cap A^c)$$
$$\underset{\text{additivity of }\mu}{=} \sum_n \mu(A_n) \le \mu^*(B)+\varepsilon$$

So since ε arbitrary, $\mu^*(B\cap A)+\mu^*(B\cap A^c)\le\mu^*(B)$.

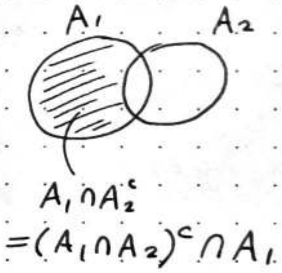
$M\supseteq A$ follows

<u>Step IV</u> ($\mathcal{M}$ is an algebra)

Clearly $\emptyset\in A\subseteq \mathcal{M}$. Likewise $\mathcal{M}$ closed under complements. (symmetry)

Now suppose $A_1, A_2\in\mathcal{M}$, $B\subseteq E$.

$$\mu^*(B)=\mu^*(B\cap A_1)+\mu^*(B\cap A_1^c) \qquad (A_1\in\mathcal{M})$$
$$(A_2\in\mathcal{M})$$
$$=\mu^*(B\cap A_1\cap A_2)+\mu^*(B\cap A_1\cap A_2^c)+\mu^*(B\cap A_1^c)$$
$$=\mu^*(B\cap A_1\cap A_2)+\mu^*(B\cap(A_1\cap A_2)^c\cap A_1)$$
$$+\mu^*(B\cap(A_1\cap A_2)^c\cap A_1^c)$$
$$=\mu^*(B\cap A_1\cap A_2)+\mu^*(B\cap(A_1\cap A_2)^c)$$



Thus $A_1\cap A_2\in\mathcal{M}$.

Since $A_1\cup A_2=(A_1^c\cap A_2^c)^c$, have $\mathcal{M}$ closed under unions.

<u>Step VI</u> ($\mathcal{M}$ is a σ-algebra, $\mu^*|_{\mathcal{M}}$ is a measure)

It suffices to prove (using disjointification) that $\forall A_n\in\mathcal{M}$ disjoint, $A=\bigcup_n A_n\in\mathcal{M}$, $\mu^*(A)=\sum_n\mu^*(A_n)$

Now for any $B\subseteq E$ we write as in V,

$$\mu^*(B)=\mu^*(B\cap A_1)+\mu^*(B\cap A_1^c)$$
$$=\mu^*(B\cap A_1)+\mu^*(B\cap \underbrace{A_1^c\cap A_2}_{=A_2})+\mu^*(B\cap A_1^c\cap A_2^c)$$
$$=\mu^*(B\cap A_1)+\mu^*(B\cap A_2)+\mu^*(B\cap A_1^c\cap A_2^c)$$
$$\vdots$$
$$=\sum_{n=1}^N \mu^*(B\cap A_n)+\mu^*(B\cap A_1^c\cap\cdots\cap A_N^c)$$

Now, $A\supseteq\bigcup_{n=1}^N A_n$ $\forall N$, so $A^c\subseteq\bigcap_{n=1}^N A_n^c$

So $\mu^*(B) \geqslant \sum_{n=1}^{N} \mu^*(B\cap A_n) + \mu^*(B\cap A^c)$ $(\mu^* \uparrow)$

Now let $N \to \infty$ to deduce

$$\mu^*(B) \geqslant \sum_n \mu^*(B\cap A_n) + \mu^*(B\cap A^c)$$
$$\underset{(+)}{\geqslant} \mu^*(B\cap \underbrace{\textstyle\bigcup_n A_n}_{A}) + \mu^*(B\cap A^c)$$

Conversely, $B \subseteq (B\cap A) \cup (B\cap A^c)$ gives other direction.

So $\mu^*(B) = \mu^*(B\cap A) + \mu^*(B\cap A^c)$ and $\mathcal{M}$ is a σ-algebra.

Now invoke above inequality with $B = A$,

$$\mu^*(A) \geqslant \sum_n \mu^*(\underbrace{A\cap A_n}_{A_n}) + \cancel{\mu^*(A\cap A^c)}^{\to \text{zero}}$$
$$= \sum_n \mu^*(A_n)$$

Again $(+)$ gives converse $\mu^*(A) \leq \sum_n \mu^*(A_n)$.

So indeed μ^* is countably additive so a measure on $\mathcal{M}$. $\square$

<u>Next topic</u>: check when μ^* is a unique extension

L3.1

(Uniqueness of extension)

A family $\mathcal{A}$ of subsets of E is called

1) a <u>π-system</u> if $\emptyset \in \mathcal{A}$ and $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$

2) a <u>d-system</u> if $E \in \mathcal{A}$ and $\forall A, B \in \mathcal{A}$ s.t. $A \subseteq B$,

and all $A_n \in \mathcal{A}$, $A_n \uparrow$ (increasing), then $B \setminus A \in \mathcal{A}$, $\bigcup_n A_n \in \mathcal{A}$

One shows (Ex Sheet) that if $\mathcal{A}$ is a d-system <u>and</u> a π-system then $\mathcal{A}$ is a σ-algebra.

<u>Lemma</u> (Dynkin) Let $\mathcal{A}$ be a π-system. Then any d-system containing $\mathcal{A}$ also contains $\sigma(\mathcal{A})$.

<u>Theorem</u> (uniqueness of extension)

Let μ_1, μ_2 be measures on $(E, \mathcal{E})$ s.t.

$\mu_1(E) = \mu_2(E) < \infty$. Suppose $\mu_1 = \mu_2$ on $\mathcal{A}$,

where $\mathcal{A}$ is some π-system generating $\mathcal{E}$ (i.e. $\mathcal{E} \overset{\text{equality?}}{\subseteq} \sigma(\mathcal{A})$)

Then $\mu_1 = \mu_2$ on $\mathcal{E}$.

<u>Proof</u> (Thm) Define the family

$$\mathcal{D} = \{ A \in \mathcal{E} : \mu_1(A) = \mu_2(A) \} \text{ of } \overset{\text{meas.}}{\text{subsets}} \text{ of } E$$

where μ_1, μ_2 agree. Clearly $\mathcal{A} \subseteq \mathcal{D}$; we will $\mathcal{D} \overset{\text{show}}{\supseteq} \sigma(\mathcal{A}) \supseteq \mathcal{E}$.

Clearly, $E \in \mathcal{D}$. Also, if $A, B \in \mathcal{E}$, $A \subseteq B$, then

$$\mu_i(A) + \mu_i(B \setminus A) = \mu_i(B) \leq \mu_i(E) < \infty \quad \text{for } i = 1, 2$$

So if $A, B \in \underset{A \subseteq B}{\mathcal{D}}$, then $B \setminus A \in \mathcal{D}$.

Next if $A_n \uparrow A$, where $A_n \in \mathcal{D}$ and disjoint? (wlog) then

$$\mu_i(A) = \mu_i\left(\bigcup_n \tilde{A}_n\right) = \lim_N \sum_{n=1}^N \mu_i(\tilde{A}_n) = \lim_N \mu_i\left(\bigcup_{n \leq N} \tilde{A}_n\right) = \lim_{n \to \infty} \mu_i(A_n),$$

so indeed $A = \bigcup_n A_n$ will lie in $\mathcal{D}$ too.

Thus $\mathcal{D}$ is a d-system containing the π-system $\mathcal{A}$.

So by Dynkin's lemma, $\mathcal{D} \supseteq \sigma(\mathcal{A}) \supseteq \mathcal{E}$, and $\mu_1 = \mu_2$ $\square$

Proof of Lemma

Def $\mathcal{D} = \bigcap_{\substack{\mathcal{D}_\alpha \text{ d-systems} \\ \mathcal{D}_\alpha \supseteq \mathcal{A}}} \mathcal{D}_\alpha$, the intersection of all d-systems containing $\mathcal{A}$,

which is again a d-system.

Plan show that $\mathcal{D}$ is a π-system

Define a new collection of sets

$$\mathcal{D}' = \{ B \in \mathcal{D} : B \cap A \in \mathcal{D} \ \forall A \in \mathcal{A} \}$$

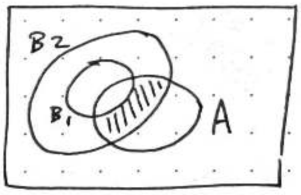
Clearly, $\mathcal{A} \subseteq \mathcal{D}'$ since $\mathcal{A}$ is a π-system.

Further, $\mathcal{D}'$ is a d-system: indeed

$$E \cap A = A \in \mathcal{A} \subseteq \mathcal{D} \quad \text{so} \quad E \in \mathcal{D}'$$

If $B_1, B_2 \in \mathcal{D}'$ s.t. $B_1 \subseteq B_2$ then $\forall A \in \mathcal{A}$

$$(B_2 \setminus B_1) \cap A = \underbrace{\underbrace{(B_2 \cap A)}_{\in \mathcal{D}} \setminus \underbrace{(B_1 \cap A)}_{\in \mathcal{D}}}_{\in \mathcal{D}}$$



So $B_2 \setminus B_1 \in \mathcal{D}'$.

Finally, if $B_n \uparrow B$ in $\mathcal{D}'$, so $B \in \mathcal{D}$, then

for any $A \in \mathcal{A}$ we have $\underbrace{(B_n \cap A)}_{\in \mathcal{D}} \uparrow (B \cap A) \in \mathcal{D}$

And so in fact $B \in \mathcal{D}'$.

Conclude that $\mathcal{D}'$ is a d-system containing $\mathcal{A}$, hence $\mathcal{D} \subseteq \mathcal{D}'$.

But $\mathcal{D}' \subseteq \mathcal{D}$ so in fact $\mathcal{D}' = \mathcal{D}$.

Now define a new system

$$\mathcal{D}'' = \{ B \in \mathcal{D} : B \cap A \in \mathcal{D} \ \forall A \in \mathcal{D} \}$$

Now from the above, $\mathcal{A} \subseteq \mathcal{D}''$.

Repeating the previous steps one shows that $\mathcal{D}'' = \mathcal{D}$.

So $\mathcal{D}$ is a π-system, and a d-system so a σ-algebra containing $\mathcal{A}$ $\square$

<u>Remark</u> A measure μ on $(E, \mathcal{E})$ is finite if $\mu(E) < \infty$.

The uniqueness theorem applies to "σ-finite" measures μ (i.e. μ s.t. $\exists\, E_n \in \mathcal{E}$ s.t. $E = \bigcup_n E_n$ and $\mu(E_n) < \infty \ \forall n$),

as we shall see later (for Lebesgue and other measures).

L4.1

Borel sets

Let E be a topological space; the σ-algebra generated by the open sets, $\mathcal{A} = \{A \subseteq E : A \text{ open}\}$

is called the Borel σ-algebra of E, we write $\mathcal{B}(E) = \sigma(\mathcal{A})$,

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

A measure μ on $(E, \mathcal{B}(E))$ is called a Borel measure.

If $\mu(K) < \infty$ for all compact subsets $K \subseteq E$ (check $K \in \mathcal{B}(E)$ ✓ Hausdorff)

then μ is a Radon measure.

CONSTRUCTION OF LEBESGUE MEASURE

Of key importance is the Borel measure μ on $\mathbb{R}$ that assigns to each interval its length, $\mu((a,b)) = b-a$, $a<b$ in $\mathbb{R}$.

Theorem (Lebesgue, 1902)

There exists a unique Borel measure μ on $\mathbb{R}$ s.t. $\forall a,b \in \mathbb{R}, a<b$, $\mu((a,b)) = b-a$.

Proof Consider the ring $\mathcal{A}$ of finite unions of disjoint intervals:

$$A = (a_1, b_1] \cup \cdots \cup (a_n, b_n]$$ ring ✓

This ring generates $\mathcal{B}$ (first it generates $\sigma(\{(a,b) : a<b\})$, see Ex Sheet; second for $a,b \in \mathbb{Q}$ provide basis for topology on $\mathbb{R}$)

For such $A \in \mathcal{A}$, define a set function

$$\mu(A) = \sum_{i=1}^{n} (b_i - a_i),$$ clearly is additive, subadditive,

and also well-defined: suppose A can be written as two distinct disjoint unions of intervals, say $A = \bigcup_j C_j$, $A = \bigcup_k D_k$,

then each $C_j = \bigcup_k (C_j \cap D_k)$, $D_k = \bigcup_j (D_k \cap C_j)$

and obtain $\mu(A) = \sum_j \mu(C_j) = \sum_{j,k} \mu(C_j \cap D_k) = \sum_k \mu(D_k)$

By virtue of Caratheodory's theorem, the result will follow if we can show that μ is countably additive on $\mathcal{A}$.

L 4.2

FACT (Ex Sheet) A finite-valued $\swarrow$ additive set function on a ring $\mathcal{A}$ is countably additive iff for any decreasing $A_n \downarrow$ with $\bigcap_n A_n = \emptyset$, $\mu(A_n) \longrightarrow 0$ as $n \to \infty$.

We will argue by contradiction; suppose $B_n \in \mathcal{A}$ is s.t. $B_n \downarrow \emptyset$ but $\mu(B_n) \not\to 0$. Then $\exists \varepsilon > 0$ s.t. for infinitely many $n \in \mathbb{N}$,

$$\mu(B_n) \geqslant 2\varepsilon > 0.$$

Now we approximate each B_n from within by $C_n \in \mathcal{A}$ s.t. $\overline{C_n} \subseteq B_n$ and $\mu(B_n \setminus C_n) \leqslant \varepsilon 2^{-n}$.

(take unions of intervals $(a_{n_i}, b_{n_i}]$ to $(a_{n_i} + \frac{\varepsilon 2^{-n}}{N_n}, b_{n_i}]$ where N_n is no. of ints in B_n)

Now, $\mu(B_n \setminus (C_1 \cap \cdots \cap C_n)) \leqslant \mu((B_1 \setminus C_1) \cup \cdots \cup (B_n \setminus C_n))$

$$\leqslant \sum_{i \leq n} \mu(B_i \setminus C_i) \leqslant \varepsilon \sum_{i \leq n} 2^{-i} \leq \varepsilon$$

Since $\mu(B_n) \geqslant 2\varepsilon$, "before removal",

we must have $\mu(C_1 \cap \cdots \cap C_n) \geqslant \varepsilon$, in particular the union of intervals $C_1 \cap \cdots \cap C_n$ is <u>non-empty</u>.

Likewise $\overline{C_1} \cap \cdots \cap \overline{C_n} = K_n$ is non-empty, (oh.... shit.... here it comes...) compact, so by compactness properties $\bigcap_n K_n \neq \emptyset$.

But $\bigcap_n K_n \subseteq \bigcap_n B_n = \emptyset$, $\times\!\!\!\times$

All in all, μ <u>is</u> countably additive and extends to a Borel measure.

$\mu((b, a)) = a - b$

Remains to prove uniqueness.

Let λ be any other measure on $\mathcal{B}$ s.t. $\lambda((a, b]) = b - a$

For any $n \in \mathbb{N}$, consider new measures

$$\mu_n = (A) = \mu(A \cap (n, n+1]), \quad \lambda_n = \lambda(A \cap (n, n+1]),$$

which are finite measures that coincide on $\mathcal{A}$, which forms a π-system generating σ-algebra. Hence by the uniqueness theorem applied to the <u>finite</u> measures λ_n, μ_n we must have that they coincide on $\mathcal{B}$.

But for any $A \in \mathcal{B}$,

$$\mu(A) = \sum_{n \in \mathbb{Z}} \mu(A \cap (n, n+1]) = \sum_{n \in \mathbb{Z}} \mu_n(A) = \sum_{n \in \mathbb{Z}} \lambda_n(A) = \cdots = \lambda(A). \quad \square$$

☺

Remark

1) A set $A \in \mathcal{B}$ is called a Lebesgue-null-set if $\mu(A) = 0$.

Since $\{x\} = \bigcap_n (x - \frac{1}{n}, x]$, but since μ is a measure we also have $\mu(\{x\}) = \lim_{n\to\infty} \mu((x-\frac{1}{n}, x]) = \lim_{n\to\infty} \frac{1}{n} = 0$, hence any singleton $\{x\}$, $x \in \mathbb{R}$ has Lebesgue measure zero. In fact for any countable subset of $\mathbb{R}$, eg $\mathbb{Q}$, we have $\mu(\mathbb{Q}) = \mu(\bigcup_{q\in\mathbb{Q}} \{q\}) = \sum_{q\in\mathbb{Q}} \underbrace{\mu(\{q\})}_{=0} = 0$

2) The measure μ is defined on $\mathcal{B}$, but in fact also on the class $\mathcal{M}$ of (outer) Lebesgue-measurable sets. One shows (Ex Sheet) that

$$\mathcal{M} = \{ A \cup N : A \in \mathcal{B},\ N \subseteq B \text{ for some } B \in \mathcal{B} \text{ s.t. } \mu(B) = 0 \}$$

so $\mathcal{M}$ contains any subset of a Borel null-set.

3) If we define the translation by x of $B \in \mathcal{B}$ to equal

$$B + x = \{ b + x : b \in B \} \qquad \text{(lies in } \mathcal{B}\text{)}$$

then $\mu_x(B) = \mu(B + x)$ equals $\mu(B)$

since $\mu_x((a,b]) = (b+x) - (a+x) = b - a$

we have $\mu_x = \mu$ by the uniqueness theorem.

So $\mu_x = \mu$ on $\mathcal{B}(\mathbb{R})$ but also on $\mathcal{B}((0,1])$ if "$+x$" is mod 1.

We say that Lebesgue measure μ is translation invariant.

Non-measurable sets

One may ask: $\mathcal{B} \subsetneq \mathcal{P}(\mathbb{R})$, $\mathcal{M} \subsetneq \mathcal{P}(\mathbb{R})$?

Similarly $\mathcal{B}((0,1]) \subsetneq \mathcal{P}((0,1])$?

For $x, y \in (0,1]$, let us write $x \sim y$ if $x - y \in \mathbb{Q}$, where "$-$" is mod 1 and $\mathbb{Q} = \mathbb{Q} \cap (0,1]$.

Now (using AoC) we select exactly one element from each equivalence cass $[x] = \{y \in (0,1] \text{ s.t. } y \sim x\}$ and let $S \subseteq (0,1]$ denote the resulting set.

Define further $S + q = \{ s + q \text{ mod } 1,\ s \in S\}$, so

$$(0,1] = \bigcup_{q\in\mathbb{Q}} (S + q) \ ; \text{ a disjoint union of cosets.}$$

L5.2

Now by translation invariance of $\mu|_{(0,1]} = \mu$,

$\mu(S) = \mu(S+q)$ for each q and hence

$\mu((0,1]) = \sum_{q \in \mathbb{Q}} \mu(S+q) = 1$ $\therefore$ either 0 or ∞, no

Therefore S cannot lie in $\mathcal{B}$, or in $\mathcal{M}$.

Is there another way to extend μ to $\mathcal{P}((0,1])$?

Theorem (Banach–Kuratowski)

Assume the continuum hypothesis. Then there exists no measure μ on $\mathcal{P}((0,1])$ s.t. $\mu((0,1]) = 1$ and $\mu(\{x\}) = 0 \ \forall x \in (0,1]$.

(see RM Dudley, Real Analysis & Prob p526)

PROBABILIY MEASURES $\mu(\Omega) = 1$

Call $(E, \mathcal{E}, \mu)$ a probability space.

We write $(\Omega, \mathcal{A}, \mathbb{P})$ for it, where

Ω ... outcome / sample space

$\mathcal{A}$ ... events (σ-algebra)

$\mathbb{P}$ ... prob measure

Axioms of probability (Kolmogorov, 1933)

1) $\mathbb{P}(\Omega) = 1$ $(\Leftrightarrow \mu(E) = 1, \mu(\phi) = 0)$

2) $\mathbb{P}(E) \geq 0 \quad \forall E \in \mathcal{A}$

3) $\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n)$ for $A_n \in \mathcal{A}$ disjoint

Let us say events $(A_i : i \in I)$ are independent if $\forall J \subseteq I$ finite

$$\mathbb{P}\Big(\bigcap_{i \in J} A_i\Big) = \prod_{i \in J} \mathbb{P}(A_i)$$

Sub-σ-algebras $\mathcal{A}_i \subseteq F$, $(\mathcal{A}_i : i \in I)$ are said to independent if $\forall J \subseteq I$ finite, the events $(A_i, i \in I)$ for $A_i \in \mathcal{A}_i$ are independent

$\therefore$

L5.3

<u>Thm</u> Let $\mathcal{A}_1, \mathcal{A}_2$ be π-systems of subsets in $\mathcal{F}$.

Suppose $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$ for all $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$.

Then $\sigma(\mathcal{A}_1)$, $\sigma(\mathcal{A}_2)$ are independent.

L6.1

Thm Let $\mathcal{A}_1, \mathcal{A}_2$ be π-systems of subsets in $\mathcal{F}$.

Suppose $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2) \quad \forall \; A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$

Then $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2)$ are independent

Proof Fix $A_1 \in \mathcal{A}_1$, define two measures

$\mu(A) = \mathbb{P}(A_1 \cap A)$, $\nu(A) = \mathbb{P}(A_1)\mathbb{P}(A)$;

which are finite measures s.t. $\mu(\Omega) = \mathbb{P}(A_1) = \nu(\Omega)$.

They coincide on the π-system $\mathcal{A}_2 \Rightarrow$ hence also on $\sigma(\mathcal{A}_2)$

Repeat the argument fixing $A_2 \in \mathcal{A}_2$. [not quite but ok] $\square$

In proving 'asymptotic' results in probability, it is useful to consider events (or sets) $\limsup A_n = \bigcap_n \bigcup_{m \geq n} A_m = \{A_n \text{ i.o.}\}$ ↑ infinitely often

$\liminf A_n = \bigcup_n \bigcap_{m \geq n} A_m = \{A_n \text{ eventually}\}$

Lemma (1st Borel-Cantelli lemma)

If $\sum_n \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(A_n \text{ i.o.}) = 0$

Pf We can write, for any n

$$\begin{aligned} \mathbb{P}(A_n \text{ i.o.}) &= \mathbb{P}\left(\bigcap_n \bigcup_{m \geq n} A_m\right) \\ &\leq \mathbb{P}\left(\bigcup_{m \geq n} A_m\right) \\ &\leq \sum_{m \geq n} \mathbb{P}(A_m) \longrightarrow 0 \text{ as } n \to \infty \end{aligned}$$

$\square$

2nd Borel-Cantelli Lemma

Suppose $(A_n : n \in \mathbb{N}) \subseteq \mathcal{F}$ are independent

If $\sum_n \mathbb{P}(A_n) = \infty$ then $\mathbb{P}(A_n \text{ i.o.}) = 1$

Proof We use $1 - a \leq e^{-a}$ for $a = a_n = \mathbb{P}(A_n)$

Then $\forall n$, $N \geq n$, by independence of A_n's (hence A_n^c's)

$$\mathbb{P}\left(\bigcap_{m=n}^{N} A_m^c\right) = \prod_{m=n}^{N} (1 - a_m) \leq \prod_{m=n}^{N} e^{-a_m} = \exp\left(-\sum_{m=n}^{N} a_m\right)$$

So necesarilly $\to 0$ as $N \to \infty$ and since $\bigcap_{m=n}^{N} A_m^c \downarrow \bigcap_{m=n}^{\infty} A_m^c$

we will deduce $\mathbb{P}\left(\bigcap_{m \geq n} A_m^c\right) = 0 \;\; \forall n$.

Conclude that $\mathbb{P}(A_n \text{ i.o.}) = 1 - \cancel{\lim} \, \mathbb{P}\left(\bigcup_n \bigcap_{m \geq n} A_m^c\right) = 1$ $\square$

L6.2

2. MEASURABLE FUNCTIONS & RANDOM VARIABLES

Measurable functions

Let $(E, \mathcal{E})$, $(G, \mathcal{G})$ be measurable spaces. A function

$$f : E \to G$$

is called <u>measurable</u> (or $\mathcal{E}$-$\mathcal{G}$ meas) if the inverse image (preimage)

$$f^{-1}(A) \in \mathcal{E} \quad \forall A \in \mathcal{G}$$

<u>Recall</u> $f^{-1}(A) = \{x \in E : f(x) \in A\}$

When the range space $(G, \mathcal{G}) = (\mathbb{R}, \mathcal{B})$, we simply say that $f : E \to \mathbb{R}$ is ~~Borel~~ measurable. If E is a topological space, $\mathcal{E} = \mathcal{B}(E)$, we say that $f : E \to \mathbb{R}$ is <u>Borel measurable</u>.

Note that <u>inverse images</u> preserve set operations

$$f^{-1}\Big(\bigcup_{i \in I} A_i\Big) = \bigcup_{i \in I} f^{-1}(A_i)$$

$$f^{-1}(\underset{\text{aah!}}{\mathcal{G}} \setminus A) = E \setminus f^{-1}(A)$$

so we deduce that

$$\{ f^{-1}(A) : A \in \mathcal{G} \} \text{ is itself a } \sigma\text{-algebra.}$$

Likewise $\{ A \in \mathcal{G} : f^{-1}(A) \in \mathcal{E} \}$ is a σ-algebra as well

In particular, to check measurability, the following is useful:

If $\mathcal{G} = \sigma(\mathcal{A})$ and $f^{-1}(A) \in \mathcal{E}$ $\forall A \in \mathcal{A}$ then

$\{A : f^{-1}(A) \in \mathcal{E}\}$ is a σ-algebra containing $\mathcal{A}$, so $\mathcal{G}$,

so f is measurable.

In particular, wen $(G, \mathcal{G}) = (\mathbb{R}, \mathcal{B})$, then the intervals

$$\{(-\infty, y] : y \in \mathbb{R}\} \text{ generate } \mathcal{B}$$

and so if $\{x \in E : f(x) \leq y\} \in \mathcal{E}$ $\forall y \in \mathbb{R}$

then f is 'Borel'-measurable.

If E is a topological space, and $f : E \to \mathbb{R}$ is continuous,

then $f^{-1}(U)$ is open in E whenever U is open in $\mathbb{R}$.

Since open sets generate $\mathcal{B}$, we see that f is measurable (Borel)

L6.3

If $A \in \mathcal{E}$ is measurable, then the indicator function

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{o/w} \end{cases}$$

is measurable, clearly.

One shows also (ExSheet) that $f_1 + f_2$, $f_1 \cdot f_2$, $\inf_n f_n$, $\sup_n f_n$,

$$\liminf_n f_n, \quad \limsup_n f_n$$

are all measurable when the f_n are.

Also $\{x \in E : f_n(x) \text{ converges as } n \to \infty\}$ is measurable.

<u>Finally</u> Given any maps $f_i : E \to G$,

we can make them all $\mathcal{E}-\mathcal{G}$ measurable by setting

$$\mathcal{E} = \sigma(f_i^{-1}(A) : A \in \mathcal{G},\ i \in I),$$

<u>the</u> σ-algebra generated by $(f_i)_{i \in I}$

<u>Theorem</u> (monotone classes)

Let $(E, \mathcal{E})$ be a measurable space, and consider $\mathcal{A}$ a π-system generating $\mathcal{E}$.

Let $\mathcal{V}$ be a vector space of bounded maps $f : E \to \mathbb{R}$

(i.e. $\sup_{x \in E} f(x) < \infty$ for each $f \in \mathcal{V}$) s.t.

1) $1 = (1_E) \in \mathcal{V}$ and $1_A \in \mathcal{V}$ $\forall A \in \mathcal{A}$

2) whenever $f_n \in \mathcal{V}$, f bounded s.t. $0 \le f_n \uparrow f$ pointwise
$\Rightarrow f \in \mathcal{V}$

Then $\mathcal{V}$ contains all bounded measurable <u>functions</u>.

<u>Proof</u> Define $\mathcal{D} = \{A \in \mathcal{E} : 1_A \in \mathcal{V}\}$,

which contains E, $1_{B \setminus A} = 1_B - 1_A \in \mathcal{V}$ (vector space) euh... $B \supset A$ ✓

and $0 \le 1_{A_n} \uparrow 1_A \in \mathcal{V}$ and so $A_n \uparrow A \Rightarrow A \in \mathcal{D}$, so $\mathcal{D}$ is a <u>d-system</u>.

It contains A (π-system), so by Dynkin's lemma,

$$\mathcal{D} \supseteq \sigma(A) \underset{\text{gen}}{\supseteq} \underset{\mathcal{E}}{\cancel{E}} \supseteq \mathcal{D} \quad \Rightarrow \quad \mathcal{E} = \mathcal{D}$$

L6.4

Now let $f: E \to [0,\infty)$ be any bounded non-negative measurable function.

We define a piecewise-constant approximation as follows

$$f_n(x) = 2^{-n} \sum_{j=0}^{n2^n} j\, 1_{A_{j,n}}$$

where $A_{j,n} = \{ x \in E : \frac{j}{2^n} < f(x) \leq \frac{j+1}{2^n} \}$ for $j = 0, 1, \dots, n2^n - 1$

$= f^{-1}\left(\left(\frac{j}{2^n}, \frac{j+1}{2^n} \right] \right) \in \mathcal{E}$ {j=0?}

$A_{n2^n,n} = \{ x \in E : f(x) > n \}$

$= f^{-1}((n,\infty)) \in \mathcal{E}$

since f is measurable.

In particular $f_n \in \mathcal{V}$.

Clearly $0 \leq f_n \leq n$, so for n large enough, we then must have $f_n(x) \leq f(x) \leq f_n(x) + \frac{1}{2^n}$.

So $f_n \uparrow f$ pointwise. Therefore by hypothesis, $f \in \mathcal{V}$.

For an arbitrary $f: E \to \mathbb{R}$ bounded measurable, we can write

$f = f^+ - f^-$ and apply the above to f^+, f^- □

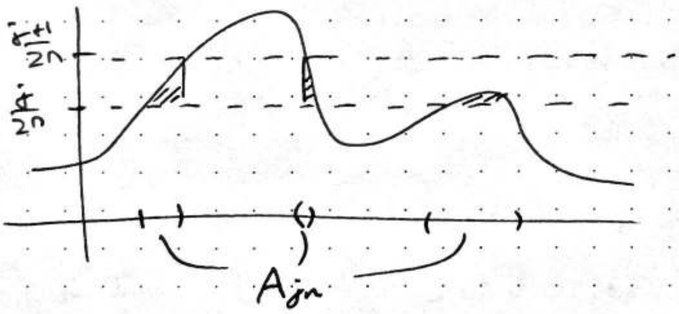


Image Measures 

L7.1

Image Measures

Let $(E, \mathcal{E})$ and $(G, \mathcal{G})$ measurable spaces,

$f : E \to G$ measurable map,

and let μ be a measure on $\mathcal{E}$. Then we can define a measure ν on $\mathcal{G}$ by $\nu(A) = \mu(f^{-1}(A))$, $A \in \mathcal{G}$,

which indeed is a measure on $\mathcal{G}$ (Ex Sheet)

It is called the image measure $\nu \equiv \mu \circ f^{-1}$

Lemma Let $g : \mathbb{R} \to \mathbb{R}$ be non-constant, right continuous, non-decreasing (↑).

Define $I = (g(-\infty), g(\infty))$ and a map $f : I \to \mathbb{R}$

$$f(x) = \inf\{y \in \mathbb{R} : g(y) \geq x\},$$

the smallest $y \in \mathbb{R}$ s.t. $g(y) \geq x$. Then f is left-continuous, non-decreasing and $\ulcorner f(x) \leq y \iff x \leq g(y) \lrcorner$

Proof For $x \in I$, define

$$J_x = \{y \in \mathbb{R} ; x \leq g(y)\}$$

is non-empty (since $g(y) > g(-\infty)$, $x > g(-\infty)$). wat (& bdd below)

So the infimum $f(x) = \inf J_x$ exists in $\mathbb{R}$.

If $y \in J_x$, $y' \geq y$ then $x \leq g(y) \leq g(y') \Rightarrow y' \in J_x$.

Also, for any sequence $y_n \in J_x$ s.t. $y_n \downarrow y$, by right-continuity,

$g(y_n) \downarrow g(y)$ and $x \leq g(y)$, i.e. $y \in J_x$.

So $J_x = [f(x), \infty)$ and $f(x) \leq y \iff y \in J_x \iff x \leq g(y)$ ⚬

Next note $x' \leq x \Rightarrow J_{x'} \supseteq J_x \Rightarrow f(x') \leq f(x)$.

Finally, if $x_n \uparrow x$ then $J_x = \bigcap_n J_{x_n}$ ✓ and by def^n of inf,

$f(x_n) \to f(x)$ as $n \to \infty$ □

$$\left\{ \bigcap_n [a_n, \infty) = [\sup_n a_n ; \infty) \right\}$$

L 7.2 (cadlag)

Theorem Let $g : \mathbb{R} \to \mathbb{R}$ be as in the previous lemma.

Then $\exists!$ Radon measure μ_g on $\mathbb{R}$ (on $\mathcal{B}(\mathbb{R})$) s.t.

$$\mu_g((a,b]) = g(b) - g(a) \quad \forall a,b \in \mathbb{R}, a<b$$

Moreover, any Radon measure on $\mathbb{R}$ can be represented in this way.

Remark We call μ_g the Lebesgue-Stieltjes measure with distr f^n g

Proof Let f, I be as in the lemma, and μ the Lebesgue measure on $\mathbb{R}$.

The $\{x \in I : f(x) \le z\} \underset{\text{lemma}}{=} \{x \in I : x \le g(z)\}$

$$= (g(-\infty), g(z)] \in \mathcal{B}(I)$$

$\forall z$, and since $((-\infty, z] : z \in \mathbb{R})$ generate $\mathcal{B}(\mathbb{R})$, we deduce that f is meas^{ble}.

So the image measure $\mu_g = \mu \circ f^{-1}$ exists on $\mathcal{B}(\mathbb{R})$

We also have

$$\mu_g((a,b]) = \mu(x : f(x) > a, f(x) \le b)$$

$$\underset{\text{lemma}}{=} \mu((g(a), g(b)])$$

$$= g(b) - g(a) \quad ☺$$

By the uniqueness argument for μ, we see that μ_g is the unique such measure, and it is Radon since any compact $K \in \mathcal{B}(\mathbb{R})$ necessarily satisfies $\mu_g(K) \le \mu_g([-n,n]) \le \mu_g((a,b]) < \infty$

Conversely, let ν be any Radon measure, and define

$$g(y) = \begin{cases} \nu((0,y]) & \text{for } y \ge 0 \\ -\nu((y,0]) & \text{for } y < 0 \end{cases}$$

which is clearly increasing and satisfies $(a<0<b)$

$$\nu((a,b]) = \nu((a,0]) + \nu((0,b])$$

$$= g(b) - g(a)$$; with a similar calculation for

$0<a<b$, $a<b<0$.

Clearly g is right continuous since $A_n = (0, y_n] \downarrow (0, y]$

so $\nu(A_n) \longrightarrow \nu(A)$ by countable additivity ~

By uniqueness $\mu_g = \nu$. □

L7.3

Example The 'Dirac' point mass measure δ_x at $x \in \mathbb{R}$ is

$$\delta_x(A) = 1 \text{ iff } x \in A, \text{ else zero,}$$

is a Radon measure with df $g(t) = 1_{[x,\infty)}(t)$

2.3 RANDOM VARIABLES

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $(E, \mathcal{E})$ a measurable space. A measurable map $X : \Omega \to E$ is called a random variable (rv) in E. If $E = \mathbb{R}$, we just say r.v. The image measure on E obtained in this way $\mu_X = \mathbb{P} \circ X^{-1}$ is called the law or distribution of X.

When $E = \mathbb{R}$, $\mathcal{E} = \mathcal{B}$, this law is determined by the π-system $\{(-\infty, x] : x \in \mathbb{R}\}$ as

$$\begin{aligned} F_X(x) &= \mu_X((-\infty; x]) \\ &= \mathbb{P}(\omega \in \Omega : X(\omega) \leq x) \\ &=: \mathbb{P}(X \leq x), \end{aligned}$$

called the cumulative distribution function (cdf). By properties of measures, we see

- $\lim_{z \downarrow x} F(z) = F(x)$
- $\lim_{z \to \infty} F(z) = 1$
- $\lim_{z \to -\infty} F(z) = 0$, and $F \uparrow$

and we say that any F with the preceding properties is a cdf.

Indeed, if $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}((0,1))$, $\overset{\mathbb{P}}{\mu} = \mu|_{(0,1)}$ Lebesgue measure restricted to $(0,1)$, then $F = g$ in the lemma from earlier gives that

$$X(\omega) = \inf\{x : \omega \leq F(x)\}, \quad \omega \in (0,1)$$

is a random variable with distribution function

$$\begin{aligned} \mathbb{P}(\omega \in \Omega : X(\omega) \leq x) &= \mathbb{P}(\omega \in \Omega : \omega \leq F(x)) \\ &= F(x) - 0 = F(x) \end{aligned}$$

so any probability distribution can be realised by a prob dist f^n.

L8.1

We say that a countable family of random variables $(X_i : i \in I)$ is <u>independent</u> if the family of σ-algebras $(\sigma(X_i) : i \in I)$ is independent.

One can check (ExSheet) that this is equivalent to (for real rvs)

$$\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \cdots \mathbb{P}(X_n \le x_n)$$

$\forall x_i \in \mathbb{R}$, forall finite subsets $X_1, \dots, X_n$ of $(X_i : i \in I)$

To construct examples of such sequences of indep rvs, consider a prob space

$$(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{B}(0,1), \mu|_{(0,1)})$$

Recall that any $\omega \in \Omega = (0,1)$ has a binary expansion

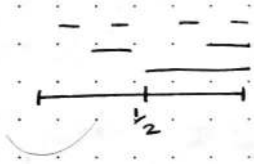
$$\omega = 0.\omega_1\omega_2\omega_3 \cdots, \quad \omega_i \in \{0,1\}$$
$$= 0 + \overset{1/2}{0} + \overset{1/4}{0} + \cdots$$

which is unique if we exclude ∞-many zeros

(in this case $\frac{1}{2} = 0 + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots$ ☺)

Hence define Rademacher rvs

$$R_n : (0,1) \to \{0,1\}$$
$$\omega \mapsto \omega_n$$



One sees $R_1 = 1_{(\frac{1}{2},1)}$

$$R_2 = 1_{(\frac{1}{4},\frac{1}{2}]} + 1_{(\frac{3}{4},1)}$$

$$R_3 = 1_{(\frac{1}{8},\frac{1}{4}]} + 1_{(\frac{3}{8},\frac{1}{2}]} + 1_{(\frac{5}{8},\frac{3}{4}]} + 1_{(\frac{7}{8},1)}$$

$\vdots$ "percollate"

all the R_n are indeed measurable,

$$\mathbb{P}(R_n = 1) = \mathbb{P}(R_n = 0) = \frac{1}{2}$$

Notice next that, for $x_i \in \{0,1\}^n$

$$\mathbb{P}(R_1 = x_1, \dots, R_n = x_n) = 2^{-n}$$

and also $\mathbb{P}(R_1 = x_1) \cdots \mathbb{P}(R_n = x_n) = (\frac{1}{2})^n = 2^{-n}$

Hence the R_n are all independent and identically distributed (iid) rvs

Thus we have constructed an infinite sequence of iid rvs s.t.

$$\mathbb{P}(R_n = 1) = \mathbb{P}(R_n = 0) = \frac{1}{2}$$

Now take a bijection $m : \mathbb{N}^2 \longleftrightarrow \mathbb{N}$ (e.g. $m = 2^{k-1}(2n-1)$)

and define new variables

$$Y_{nk} = R_{m(n,k)} \qquad (\text{iid})$$

$$Y_n = \sum_k 2^{-k} Y_{nk} \quad , \qquad \leftarrow \text{measurable}$$

which are also all independent (the series converges $\forall \omega \in \Omega$ since $|Y_{nk}(\omega)| \leq 1$)

Now consider intervals $\left(\frac{i}{2^m}, \frac{i+1}{2^m}\right)$ with dyadic rational endpoints and observe

$$\mathbb{P}\left(\frac{i}{2^m} < Y_n \leq \frac{i+1}{2^m} \right)$$

$$= \mathbb{P}\left(\frac{i}{2^m} \leq \sum_k 2^{-k} Y_{nk} \leq \frac{i+1}{2^m} \right)$$

$$= \mathbb{P}\left(Y_{n1} = y_1, \dots, Y_{nm} = y_m \right) = 2^{-m} \qquad \text{sth sth a.s.}$$

where $(y_1, \dots, y_m) \in \{0,1\}^m$ corresponds to i, m

Since these intervals form a π-system generating $\mathcal{B}((0,1))$, we see $\mathcal{L}(Y_n) = \mu|_{(0,1)}$ $\forall n$, and hence the Y_n form an infinite sequence of iid $U(0,1)$ variables.

<u>Prop^n</u> Let $(F_n : n \in \mathbb{N})$ be any sequence of probability dist^n functions on $\mathbb{R}$. Then there exist indep rvs $(X_n : n \in \mathbb{N})$ defined on $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{B}, \mu|_{(0,1)})$ s.t. their prob distributions $F_{X_n} = F_n$ $\forall n \in \mathbb{N}$

<u>Proof</u> Take $Y_n \overset{\text{iid}}{\sim} U(0,1)$, set $\mathcal{G}_n(y) = \inf\{x : y \leq F_n(x)\}$, and $X_n = \mathcal{G}_n(Y_n)$, a <u>r.v.</u> with distribution

$$\mathbb{P}(X_n \leq x) = \mathbb{P}(\mathcal{G}_n(Y_n) \leq x)$$

$$= \mathbb{P}(Y_n \leq F_n(x)) \qquad (\text{gen inverses})$$

$$= F_n(x)$$

□

§2.5 Convergence of meas f^ns and ms

Let $(E, \mathcal{E}, \mu)$ be a measure space.

- A property defining $A \in \mathcal{E}$ is said to hold almost everywhere (ae) if $\mu(A^c) = 0$.

When $(E, \mathcal{E}, \mu) = (\Omega, \mathcal{F}, \mathbb{P})$ we say $\mathbb{P}$-almost surely (a.s.)

A sequence of measurable $f_n : (E, \mathcal{E}, \mu) \to \mathbb{R}$ is said to converge μ-a.e. to $f : E \to \mathbb{R}$ if $\mu(x \in E : f_n(x) \underset{n\to\infty}{\not\to} f(x)) = 0$

Next, if f_n, f are measurable maps on E, we say that $f_n \underset{n\to\infty}{\longrightarrow} f$ in μ-measure if $\forall \varepsilon > 0$,

$$\mu(x \in E : |f_n(x) - f(x)| > \varepsilon) \longrightarrow 0 \quad \text{as } n \to \infty$$

- For r.v.s X_n, X (so that $\mu(E) = P(\Omega) = 1$) we say

$X_n \to X$ almost surely $X_n \xrightarrow{as} X$,

$X_n \to X$ in probability $X_n \xrightarrow{P} X$

respectively.

For ms (on $\mathbb{R}$), we say

$X_n \longrightarrow X$ in distribution, $X_n \xrightarrow{d} X$

if the distribution functions

$$F_{X_n}(t) \longrightarrow F_X(t) \quad \forall t \in \mathbb{R}$$

- where F_X is continuous.

Theorem Let $(f_n : n \in \mathbb{N})$ be a sequence of meas f^n on $(E, \mathcal{E}, \mu)$

1) if $\mu(E) < \infty$, if $f_n \xrightarrow{a.e.} 0 \Rightarrow f_n \longrightarrow 0$ in measure
2) if $f_n \to 0$ in measure, then $f_{n_k} \longrightarrow 0$ a.e. along a subsequence n_k

Proof 1) Let $\varepsilon > 0$.

$$\mu(|f_n| \le \varepsilon) \geq \mu(\underbrace{\bigcap_{m \geq n} \{|f_m| \le \varepsilon\}}_{A_m})$$

$$\uparrow \mu\left(\bigcup_n \bigcap_{m \geq n} A_m\right) = \mu(|f_n| \le \varepsilon \text{ eventually})$$

$$\geq \mu(f_n \longrightarrow 0 \text{ as } n \to \infty) = \mu(E) < \infty$$

- Taking complements, $\overline{\lim_n} \, \mu(|f_n| > \varepsilon) = 0$ and so $f_n \longrightarrow 0$ in measure.

L8.4

b) Choose $\varepsilon = \frac{1}{k}$, $k \in \mathbb{N}$, select a subsequence n_k s.t.

$$\sum_k \mu\left(|f_{n_k}| > \frac{1}{k}\right) < \infty$$

By the first Borel-Cantelli lemma,

$$\mu\left(|f_{n_k}| > \frac{1}{k} \text{ i.o.}\right) = 0$$

$$\Rightarrow f_{n_k} \longrightarrow 0 \text{ a.e.}$$

$\square$

L9.1

Remark 1) The requirement $\mu(E) < \infty$ is necessary in the previous theorem:

$$f_n = 1_{(n,\infty)} \longrightarrow 0 \text{ a.e. as } n \to \infty,$$

but $\mu(x : |f_n(x)| > \varepsilon) = \mu((n,\infty)) = \infty$ (μ Lebesgue)

2) Take independent $(A_n : n \in \mathbb{N})$ s.t. $\mathbb{P}(A_n) = \frac{1}{n}$ $(\to 0)$

Then $\sum_n \mathbb{P}(A_n) = \infty$, so by Borel-Cantelli

$\mathbb{P}(1_{A_n} = 1 \text{ i.o.}) = 1$, so

$1_{A_n} \not\to 0$ $\mathbb{P}$-a.s.

In contrast, $\mathbb{P}(1_{A_n} > \varepsilon) = \mathbb{P}(A_n) = \frac{1}{n} \longrightarrow 0$, so

$1_{A_n} \longrightarrow 0$ in $\mathbb{P}$ as $n \to \infty$.

3) One shows further that $X_n \xrightarrow{\mathbb{P}} X$ then $X_n \xrightarrow{d} X$

Moreover, if $X_n \xrightarrow{d} X$ then $\exists\, \tilde{X}_n, \tilde{X} : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ s.t.

$\mathcal{L}aw(\tilde{X}_n) = \mathcal{L}aw(X_n)$, $\mathcal{L}aw(\tilde{X}) = \mathcal{L}aw(X)$

and $\tilde{X}_n \to \tilde{X}$ a.s. as $n \to \infty$.

Example Consider $(X_n : n \in \mathbb{N})$ iid with cdf $\mathbb{P}(X_n \le x) = 1 - e^{-x}$

Consider $A_n = \{X_n \geq \alpha \log n\}$, $\alpha > 0$

so that $\mathbb{P}(A_n) = e^{-\alpha \log n} = n^{-\alpha}$, hence

$\sum_n \mathbb{P}(A_n) < \infty \iff \alpha > 1$, so

by the Borel-Cantelli lemmas,

$\mathbb{P}\left(\frac{X_n}{\log n} \geq 1 \text{ i.o.}\right) = 1$

$\mathbb{P}\left(\frac{X_n}{\log n} \geq 1 + \varepsilon \text{ i.o.}\right) = 0$ for any $\varepsilon > 0$

hence $\limsup_n \frac{X_n}{\log n} = 1$ a.s.

TAIL EVENTS

For $(X_n : n \in \mathbb{N})$ any sequence of rvs, set

$\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \dots)$

$\mathcal{T} = \bigcap_n \mathcal{T}_n$,

where $\mathcal{T}$ is called the tail-σ-algebra

L9.2

<u>Theorem</u> (Kolmogorov)

Let $(X_n : n \in \mathbb{N})$ be independent rvs

If $A \in \mathcal{T}$ then $\mathbb{P}(A) \in \{0,1\}$

If $Y : (\Omega, \mathcal{T}, \mathbb{P}) \longrightarrow \mathbb{R}$ is measurable, then Y is a.s. constant.

<u>Remark</u> "0-1-law"

<u>Proof</u> The σ-algebra $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ is generated by the π-system

$$A = \{X_1 \le x_1, \dots, X_n \le x_n\}, \quad x_i \in \mathbb{R}$$

while $\mathcal{T}_n$ is generated by the π-system of sets

$$B = \{X_{n+1} \le x_{n+1}, \dots, X_{n+k} \le x_{n+k}\}, \quad k \in \mathbb{N}$$

By independence, for any such A, B,

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

so by a theorem from earlier, the σ-algebras $\mathcal{F}_n, \mathcal{T}_n$ are independent.

Then $\bigcup_n \mathcal{F}_n$ is another π-system generating

$$\mathcal{F}_\infty = \sigma(X_1, X_2, \dots) = \sigma(X_n : n \in \mathbb{N}),$$

so we deduce, again by the theorem from earlier, that $\mathcal{F}_\infty, \mathcal{T}$ are indep.

But $\mathcal{T} \subseteq \mathcal{T}_n \subseteq \mathcal{F}_\infty$, *wheeze*, so if $A \in \mathcal{T}$, A is independent from itself and $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$.

Finally, if Y is $\mathcal{T}$-measurable, the

$$Y^{-1}((-\infty, y]) = \{Y \le y\} \in \mathcal{T}$$

has probability 1 or 0, so $Y = c$, [a.s.]

$$c = \inf\{y : F(y) = 1\}$$

$\square$

L9.3

3. Integration

For $(E, \mathcal{E}, \mu)$ a measure space, and f a measurable function on E, we will define its integral

$$\mu(f) = \int_E f \, d\mu = \int_E f(x) \, d\mu(x) = \int_E f(x) \mu(dx)$$

whenever $f \geq 0$, or when f is <u>integrable</u>.

Equally we will define the expectation of rvs X as

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

To start, let us say a function $f : E \to \mathbb{R}$ is simple if it is of the form $f = \sum_{k=1}^{m} a_k 1_{A_k}$, $a_k \geq 0$, $A_k \in \mathcal{E}$, $m \in \mathbb{N}$.

For such f, we set its integral to be

$$\mu(f) = \sum_{k=1}^{m} a_k \mu(A_k),$$

which is well-defined (easy) and one further shows

1) $\mu(\alpha f) = \alpha \mu(f)$, $\mu(f+g) = \mu(f) + \mu(g)$, f, g simple, $\alpha \geq 0$
2) $f \leq g \Rightarrow \mu(f) \leq \mu(g)$
3) $f = 0$ a.e. $\Rightarrow \mu(f) = 0$

For a general measurable, non-negative

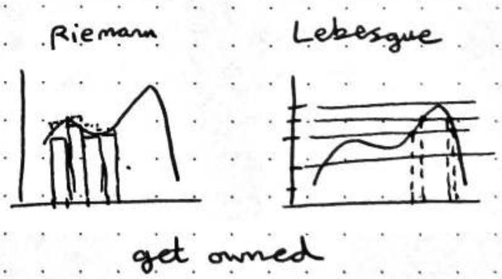
$$f : E \to [0, \infty]$$

we set $\mu(f) = \sup \{ \mu(g) : g \text{ simple}, g \leq f \}$

<u>Remark</u> We will see the sup can be realised for a particular g simple chosen as

$$f_n = 2^{-n} \lfloor 2^n f \rfloor \wedge n$$
$$= 2^{-n} \sum_{j=0}^{n 2^n - 1} j \, 1_{A_{n_j}}$$

$$A_{n_j} = f^{-1}\left(\frac{j}{2^n}, \frac{j+1}{2^n} \right]$$



Now if $0 \leq f \leq g \Rightarrow \mu(f) \leq \mu(g)$.

If $f : E \to [-\infty, \infty]$ is measurable, then set

$$f^+ = f \vee 0, \quad f^- = (-f) \vee 0,$$

the positive, negative parts

Then $f = f^+ - f^-$, $|f| = f^+ + f^-$

If $\mu(|f|) < \infty$, then we say f is integrable, then we define

$$\mu(f) = \mu(f^+) - \mu(f^-)$$

Clearly, $|\mu(f)| \leq \mu(|f|)$ $\forall$ integrable f

<u>Theorem</u> (Monotone convergence theorem)

Let f_n, f be measurable $(E, \mathcal{E}, \mu) \to [0, \infty]$,

suppose $f_n \uparrow f$ (pointwise). Then $\mu(f_n) \uparrow \mu(f)$

<u>Remark</u> $x_n \uparrow x$ if $x_n \leq x_{n+1}$, $x_n \to x$

For (meas) functions $f_n \uparrow f$ if $f_n(x) \uparrow f(x)$ for each x.

Theorem (Monotone convergence)

Let $f_n, f : (E, \mathcal{E}, \mu) \to [0,\infty]$ be measurable.

Suppose $f_n \uparrow f$ (pointwise). Then $\mu(f_n) \uparrow \mu(f)$.

Proof Recall $\mu(f) = \sup\{\mu(g) : g \le f,\ g \text{ simple}\}$

$M = \sup_n \mu(f_n)$, so $\mu(f_n) \uparrow M$ (since (f_n) monotone increasing)

and since $f_n \le f \ \forall n$, $\mu(f_n) \le \mu(f) \ \forall n \Rightarrow M \le \mu(f)$

We will show that $\mu(f) \le M$ via $\mu(g) \le M \ \forall g$ simple, $g \le f$

WLOG, take $g = \sum_{k=1}^{m} a_k 1_{A_k}$, A_k disjoint

Define a new simple function

$$g_n = (\underbrace{2^{-n}\lfloor 2^n f_n \rfloor}_{\overline{f_n}})^{n \vee ?} \overset{\text{minimum}}{\wedge} g \quad , \text{ so that } \overline{f_n} \le f_n$$

implies $g_n \le f_n$. Since $f_n \uparrow f$, $\Rightarrow \overline{f_n} \overset{? \text{ LMAO}}{\uparrow} f$ and since $f \ge g$,

we see $g_n = \overline{f_n} \wedge g \uparrow f \wedge g = g$

Fix $\varepsilon > 0$ and define sets

$$A_k(n) = \{x \in A_k \mid g_n(x) \ge (1-\varepsilon)a_k\}$$

and note $g = a_k$ on A_k. Since $g_n \uparrow g$, we must have $A_k(n) \uparrow A_n$ $(n\to\infty)$

so since μ is a measure, $\mu(A_k(n)) \uparrow \mu(A_k)$

Also, $1_{A_k} g_n \ge (1-\varepsilon) a_k 1_{A_k(n)}$ (on E)

so since $\mu(\cdot)$ is increasing,

$$\mu(1_{A_k} g_n) \ge (1-\varepsilon) a_k \mu(A_k(n))$$

Finally since the A_k are disjoint, we write

$$g_n = \left(\sum_{k=1}^{m} 1_{A_k}\right) g_n$$

Now using additivity of simple function integrals,

$$\mu(g_n) = \sum_{k=1}^{m} \mu(1_{A_k} g_n) \ge \sum_{k=1}^{m} (1-\varepsilon) a_k \mu(A_k(n))$$

$$\underset{n\to\infty}{\uparrow} (1-\varepsilon) \sum_{k=1}^{m} a_k \mu(A_k) = (1-\varepsilon)\mu(g)$$

So we see $\mu(g_n) \le \mu(f_n) \le M$

and overall, $\mu(g) \le \frac{M}{1-\varepsilon}$ and since ε arbitrary, result follows $\square$

Some remarks

1) Another way of writing the last result is

$$\lim_n \int f_n \, d\mu = \lim_n \mu(f_n) \stackrel{!}{=} \mu(f) = \mu(\lim_n f_n) = \int \lim_n f_n \, d\mu$$

as long as $0 \leq f_n \uparrow f$ pointwise.

Also, for $g_n \geq 0$ measurable f^ns, we have

$$\sum_n \mu(g_n) = \mu\left(\sum_n g_n\right)$$

2) The requirement $f_n \uparrow f$ pointwise can be weakened to $f_n \uparrow f$ a.e.

3) If $(f_n : \mathbb{N})$, f are measurable, $f_n \uparrow f$, $\mu(f_1) > -\infty$,

then $\mu(f_n) \uparrow \mu(f)$ as $n \to \infty$ (not required that $f_n \geq 0$)

The condition $\mu(f_1) > -\infty$ is necessary as the example

$f_n = -1_{(n,\infty)}$ shows ($\uparrow 0$, $\mu(f_n) = -\infty \ \forall n$)

4) A symmetric version where $f_n \downarrow f$; $\mu(f_1) < \infty$ implies $\mu(f_n) \downarrow f(f)$ also exists.

Theorem Let $f, g \geq 0$ measurable functions. Then

a) $\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)$, $\alpha, \beta \in \mathbb{R}_{\geq 0}$ (additivity)

b) $g \leq f \Rightarrow \mu(g) \leq \mu(f)$

c) $\mu(f) = 0 \Leftrightarrow f = 0$ a.e.

Proof a) Take simple functions $f_n = (2^{-n} \lfloor 2^n f \rfloor) \wedge n$

$$g_n = (2^{-n} \lfloor 2^n g \rfloor) \wedge n$$

so $0 \leq f_n \uparrow f$, $0 \leq g_n \uparrow g$,

and also $0 \leq \alpha f_n + \beta g_n \uparrow \alpha f + \beta g$, as $n \to \infty$.

By the monotone convergence theorem we see

$$\mu(f_n) \uparrow \mu(f), \quad \mu(g_n) \uparrow \mu(g), \quad \mu(\alpha f_n + \beta g_n) \uparrow \mu(\alpha f + \beta g)$$

Since $\mu(\cdot)$ additive on the simple functions,

$\mu(\alpha f_n + \beta g_n) = \alpha \mu(f_n) + \beta \mu(g_n)$ and done by above

somewhat subtle?

b) Blindingly obvious

c) If $f = 0$ a.e., then $f_n = 0 \ \forall n$ a.e. $\Rightarrow \mu(f_n) = 0 \uparrow f(\mu) = 0$

Conversely, if $\mu(f) = 0$, then $0 \leq \mu(f_n) \leq \mu(f) = 0 \Rightarrow \mu(f_n) = 0 \ \forall n$ $\Big\{$ $\therefore f_n = 0$ a.e. $\forall n$; $\therefore f = 0$ a.e. (pointwise limit) $\square$

L10.3

<u>Theorem</u> Let $f, g : E \to \mathbb{R}$ be μ-integrable;

then 1) $\forall \alpha, \beta \in \mathbb{R}$, $\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)$ (additivity)

2) $\mu(g) \leq \mu(f)$ whenever $g \leq f$

3) $f = 0$ a.e. $\Rightarrow$ $\mu(f) = 0$

<u>Proof</u> 1) For $\alpha \geq 0$, αf is also integrable, and

$$\mu(\alpha f) = \mu(\alpha f^+) - \mu(\alpha f^-)$$
$$= \alpha \mu(f^+) - \alpha \mu(f^-) \quad \text{(previous thm)}$$
$$= \alpha \mu(f)$$

and combine this with $\mu(-f) = -\mu(f)$ to get $\mu(\alpha f) = \alpha \mu(f)$ $\forall \alpha \in \mathbb{R}$

Now set $h = f + g$ can be written

$$h^+ + f^- + g^- = h^- + f^+ + g^+ \quad \text{(all non-negative)}$$

$\therefore$ $\mu(h^+) + \mu(f^-) + \mu(g^-) = \mu(h^-) + \mu(f^+) + \mu(g^+)$ by previous theorem

$\therefore$ $\mu(h) = \mu(f) + \mu(g)$ ☺

2) If $f \leq g \Rightarrow \mu(g) - \mu(f) = \mu(\underbrace{g - f}_{\geq 0}) \geq 0$

So $\mu(f) \leq \mu(g)$.

3) If $f = 0$ a.e. $\Rightarrow$ $f^+ = f^- = 0$ a.e

$\Rightarrow \mu(f^+) = \mu(f^-) = 0$

$\Rightarrow \mu(f) = 0$ $\square$

<u>Remark about 3)</u>

We cannot expect the converse to hold, but we can check that $f = 0$ a.e. by checking i) $\mu(|f|) = 0 \Rightarrow |f| = 0$ a.e. $\Rightarrow f = 0$ a.e.

ii) $\mu(f 1_A) = 0$ $\forall A \in \mathcal{A}$, where $\mathcal{A}$ π-system generating $\mathcal{E}$, with $E \in \mathcal{A}$

(uniqueness theorem for measures)

Lemma (Fatou)

Let $f_n : (E, \mathcal{E}, \mu) \to [0, \infty)$ be measurable.

Then $\mu(\liminf_n f_n) \leq \liminf_n \mu(f_n)$

Remark Recall that $(x_n \in \mathbb{R})$

$$\liminf_n x_n \equiv \underline{\lim}\, x_n = \sup_n \inf_{m \geq n} x_m$$

$$\limsup_n x_n \equiv \overline{\lim}\, x_n = \inf_m \sup_{n \geq m} x_n$$

Note: $\underline{\lim}\, f_n$, $\overline{\lim}\, f_n$ are measurable

$\overline{\lim}\, x_n = \underline{\lim}\, x_n = x \iff \lim_n x_n = x$

If $f_n \to f$ pointwise, then Fatou's Lemma says

$$\mu(f) \leq \underline{\lim}\, \mu(f_n)$$

Proof For any $k \geq n$, $\inf_{m \geq n} f_m \leq f_k$

So

$$\mu(\inf_{m \geq n} f_m) \leq \mu(f_k) \quad \forall k \geq n$$

$$\therefore \mu(\inf_{m \geq n} f_m) \leq \inf_{k \geq n} \mu(f_k) \leq \sup_n \inf_{k \geq n} \mu(f_k) = \underline{\lim}\, \mu(f_k)$$

Also, the f^ns $\inf_{m \geq n} f_m \nearrow \sup_n \inf_{m \geq n} f_m$ as $n \to \infty$

So by monotone convergence theorem,

$$\mu(\inf_{m \geq n} f_m) \nearrow \mu(\underline{\inf}\, f_n)$$

The result $\mu(\underline{\lim}\, f_n) \leq \underline{\lim}\, \mu(f_n)$ follows $\square$

Theorem (Dominated convergence)

Let $f_n, f : (E, \mathcal{E}, \mu) \to \mathbb{R}$ be measurable f^ns,

and suppose $f_n \to f$ pointwise on E,

and that $|f_n| \leq g$ $\forall n$, where g is μ-integrable

Then f_n, f are integrable, and

$$\mu(f_n) \to \mu(f) \quad \text{as } n \to \infty.$$

Proof Taking limits in $|f_n| \leq g$, deduce $|f| \leq g$,

so $\mu(|f|) \leq \mu(g) < \infty \Rightarrow f$ (and f_n) is integrable.

Now write

$$0 \le g \pm f_n \xrightarrow[n\to\infty]{} g \pm f = \underline{\lim}\,(g \pm f_n)$$

So applying Fatou's lemma twice

$$\mu(g) + \mu(f) = \mu(g+f) = \mu\left(\underline{\lim}\,(g+f_n)\right)$$
$$\underset{\text{Fatou}}{\le} \underline{\lim}\ \mu(g+f_n) = \underline{\lim}\left(\mu(g) + \mu(f_n)\right)$$
$$= \mu(g) + \underline{\lim}\ \mu(f_n)$$

$$\mu(g) - \mu(f) = \mu\left(\underline{\lim}\,(g-f_n)\right) \overset{\text{Fatou}}{\le} \underline{\lim}\left(\mu(g) - \mu(f_n)\right)$$
$$\le \mu(g) - \overline{\lim}\ \mu(f_n)$$

So collecting these inequalities, obtain

$$\overline{\lim}\ \mu(f_n) \le \mu(f) \le \underline{\lim}\ \mu(f_n) \le \overline{\lim}\ \mu(f_n)$$

$$\therefore \lim \mu(f_n) = \mu(f)$$ □

An immediate application: He's wearing air pods!

On $E = [0,1]$ (or $\mu(E) < \infty$),

suppose $f_n \to f$ pointwise and

$$\sup_n \|f_n\|_\infty = g < \infty$$

Then $\mu(g) < \infty$, so by the Dominated Convergence Theorem,

$$\mu(f_n) \longrightarrow \mu(f)$$

<u>Theorem</u> (Diff under an integral)

Let $U \subseteq \mathbb{R}$ be open and let $f : U \times E \to \mathbb{R}$

satisfying
1) $x \mapsto f(t,x)$ is integrable $\forall\, t \in U$
2) $t \mapsto f(t,x)$ is diff'ble $\forall\, x \in E$
3) $\left|\frac{\partial f}{\partial t}(t,x)\right| \le g(x) \quad \forall t, x$

s.t. g is integrable

Then $x \mapsto \frac{\partial f}{\partial t}(t,x)$ is integrable $\forall t$, if $F(U) \to \mathbb{R}$,

$$F(t) = \int_E f(t,x)\,d\mu(x) \quad \text{then} \quad \frac{dF}{dt} = \int_E \frac{\partial f}{\partial t}(t,x)\,d\mu(x)$$

Proof For $h_n \to 0$ set

$$g_n(x) = \frac{f(t+h_n, x) - f(t,x)}{h_n} - \frac{\partial f}{\partial t}(t,x)$$

⌈well defined?⌋ t fixed!

$\to 0$ as $n \to \infty$ (pointwise in x)

By the Mean Value Thm,

$$|g_n(x)| = \left| \frac{\partial f}{\partial t}(\tilde{t}_n, x) - \frac{\partial f}{\partial t}(t,x) \right| \leq 2\, g(x) \quad ; \mu\text{-integrable}$$

So by the dominated conv theorem,

$$\mu(g_n) \to \mu(\lim g_n) = 0$$

In other words,

$$\int_E \frac{f(t+h_n, x) - f(t,x)}{h_n}\, d\mu(x)$$

$$= \frac{F(t+h_n) - F(t)}{h_n} \longrightarrow \int_E \frac{\partial f}{\partial t}(t,x)\, d\mu(x) = \frac{dF}{dt} \quad \square$$

A few more remarks about the Riemann integral

1) One shows that if $f : [0,1] \to \mathbb{R}$ is Riemann-integrable, then it defines a Lebesgue integrable function on $([0,1], \mathcal{E}^{\mu}, \mu)$

$\mathcal{E}^{\mu} = \mathcal{M}$ Lebesgue-measurable sets

μ: Lebesgue measure on $[0,1]$

and $\mu(f) = \int_0^1 f(x)\,dx$ [But $\exists$ f Riemann integrable that is not Borel-measurable]

But note that $1_{\mathbb{Q}}$ is not Riemann-integrable, but $\mu(1_{\mathbb{Q}}) = \mu(\mathbb{Q}) = 0$ so $1_{\mathbb{Q}}$ is Lebesgue-integrable.

2) For continuous functions on $[0,1] \to \mathbb{R}$, the fundamental thm of calculus for the Lebesgue integral is proved in the same way as for the Riemann integral.

In fact, Lebesgue proved a stronger/ more general assertion to the effect that for $f:[0,1] \to \mathbb{R}$ Lebesgue integrable, if $F(x) = \int_0^x f(y)\, d\mu(y)$, then

$$\lim_{h \to \infty} \frac{F(x+h) - F(x)}{h} = \lim_{h \to 0} \frac{1}{h} \int_x^{x+h} f(y)\, d\mu(y) = f(x) \quad \mu\text{-a.e. } x$$

("Lebesgue's differentiation theorem") <<look it up in a book>>

L12.1

Transformations of integrals

Proposition Let $\phi : [a,b] \to \mathbb{R}$ be continuously diff, strictly increasing. Then for any Borel meas $g : [\phi(a), \phi(b)] \to [0,\infty]$, we have

$$\int_{\phi(a)}^{\phi(b)} g(y)\, dy = \int_a^b g(\phi(x))\, \phi'(x)\, dx$$

Proof (sketch, left as exercise)

First check indicators of intervals $g = 1_{(\phi(a_i), \phi(b_i))}$, $a \le a_i < b_i \le b$ in which case the FTC gives the result.

Then apply an approximation argument with the monotone class theorem. $\square$

More generally, if $f : (E, \mathcal{E}) \to (G, \mathcal{G})$ is measurable, and $\nu = \mu \circ f^{-1}$ is the image measure on $\mathcal{G}$, then for any map $g : \mathcal{G} \to [0,\infty]$, then $\mu \circ f^{-1}(g) = \int_G g\, d\nu \overset{!}{=} \int_E g(f(x))\, d\mu(x) = \int_E g \circ f\, d\mu = \mu(g \circ f)$, see Ex Sheet. In particular if $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ has law μ_X, and if $g : \mathbb{R} \to \mathbb{R}$ is Borel, then

$$\mathbb{E}\, g(X) = \int g\, d\mu_X = \int_\Omega g(X(\omega))\, d\mathbb{P}(\omega)$$

Prop Let $f : (E, \mathcal{E}, \mu) \to \mathbb{R}$ be measurable and define

$$\nu_f(A) = \mu(f 1_A), \quad A \in \mathcal{E}$$

Then ~~is~~ ν_f defines a new measure on $\mathcal{E}$ and

$$\nu_f(g) = \int g\, d\nu_f = \int_E g f\, d\mu$$

for any $g : (E, \mathcal{E}) \to [0,\infty]$ meas.

Proof Exercise. $\square$

We call f the density of ν_f wrt μ.

In particular, if ~~$\mu = \mu_X$~~ ν_f is the law of a random variable on $\mathbb{R}$, then we say that f is the prob density function of ν_f. Expectations are then often written as $\mathbb{E}\, g(X) = \int_{\mathbb{R}} g(x) f(x)\, d\mu(x)$

L12.2

PRODUCT MEASURES & FUBINI'S THEOREM

Let $(E_1, \mathcal{E}_1, \mu_1)$, $(E_2, \mathcal{E}_2, \mu_2)$ be <u>finite</u> measure spaces ($\mu_i(E_i) < \infty$)

We define 'rectangles'

$$A_1 \times A_2 = \{ (a_1, a_2) \in E_1 \times E_2 : a_1 \in A_1 ; a_2 \in A_2 \}$$

and the product space $E = E_1 \times E_2$, on which the family of subsets

$$\mathcal{A} = \{ A_1 \times A_2 : A_1 \in \mathcal{E}_1 , A_2 \in \mathcal{E}_2 \}$$

defines a π-system. We define the <u>product σ-algebra</u>

$$\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2 = \sigma(\mathcal{A})$$

We need two lemmas:

$E_2 \to E_1 \times E_2$, $x_2 \mapsto (x_1, x_2)$

<u>Lemma</u> Let $f : E \to \mathbb{R}$ be $\mathcal{E}$-measurable

Then for any $x_1 \in E_1$, the map

$$x_2 \longmapsto f(x_1, x_2) \quad \text{is } \mathcal{E}_2\text{-measurable}$$

<u>Proof</u> We use the monotone class theorem with

$\mathcal{A}$ the π-system generating $\mathcal{E}$ and

vector space $\mathcal{V} = \{ f : E \to \mathbb{R} \text{ bounded, } \mathcal{E}\text{-meas s.t. Lemma holds} \}$

Then clearly $1_A \in \mathcal{V} \;\; \forall A \in \mathcal{A}$ since $1_A = 1_{A_1}(x_1) 1_{A_2}(x_2) = c \, 1_{A_2}(x_2)$ is $\mathcal{E}_2$-measurable.

Then $f_n \in \mathcal{V}$ s.t. $0 \leq f_n \nearrow f$ then

$$f(x_1, \cdot) = \lim_n f_n(x_1, \cdot)$$

is also $\mathcal{E}_2$-measurable as the limit of $\mathcal{E}_2$-meas f^ns, hence by the monotone class theorem $\mathcal{V}$ contains all bounded meas f^ns

The general meas f can be approx by $f_n = (-n) \vee f \wedge n \nearrow f$, so satisfies the condition of the lemma too. $\square$

L12.3

<u>Lemma</u> Let $f : (E, \mathcal{E}) \to [-\infty, \infty]$ be bounded or non-negative, and measurable. Then define

$$f_1(x_1) = \int_{E_2} f(x_1, x_2)\, d\mu_2(x_2)$$

If f is bounded then $f_1 : E_1 \to \mathbb{R}$ is bounded, $\mathcal{E}_1$ measurable

If f is non-negative then $f_1 : E_1 \to \mathbb{R}$ is non-negative, $\mathcal{E}_1$-meas.

<u>Proof</u> Note that f_1 is well-defined as an integral by the previous Lemma.

We apply again a monotone class argument with π-system $\mathcal{A}$ and vector space

$$\mathcal{V} = \{ f : E \to \mathbb{R},\ \text{bounded meas., lemma is true} \}$$

To see $1_A \in \mathcal{V}$ for $A \in \mathcal{A}$, notice that for $1_{A_1 \times A_2}$, we have

$$f_1(x_1) = 1_{A_1}(x_1)\, \mu_2(A_2) \quad \text{which is } \geq 0;\ \underset{\mu_2(E_2) < \infty}{\text{bounded}},\ \mathcal{E}_1\text{-meas}$$

Next by the monotone convergence theorem (x_1 fixed), (applied to μ_2-integral)

$$\int_{E_2} f(x_1, x_2)\, d\mu_2(x_2) = \lim_n \underbrace{\int_{E_2} f_n(x_1, x_2)\, d\mu_2(x_2)}_{\text{are } \mathcal{E}_1\text{-meas}}$$

f monotone $f_n \nearrow$

so the LHS is $\mathcal{E}_1$-meas as the limit of $\mathcal{E}_1$-meas f's.

Clearly $f_1 \geq 0$ if all $f_n \nearrow$ are non-negative, and if f_n are bounded the limit is bounded by $\|f\|_\infty \mu_2(E_2) < \infty$.

So $\mathcal{V}$ contains all bounded measurable f's by the monotone class theorem $\square$

<u>Theorem</u> (Product measure)

There exists a unique $\mu = \mu_1 \otimes \mu_2$ measure on $(E, \mathcal{E})$ s.t.

$$\mu(A_1 \times A_2) = \mu_1(A_1)\, \mu_2(A_2) \quad (*)$$

$\forall A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2$.

<u>Proof</u> Since $\mathcal{A}$ is a π-system generating $\mathcal{E} = \sigma(\mathcal{A})$, if such a measure exists, it is necessarily unique. To prove existence, define

$$\mu(A) = \int_{E_1} \left(\int_{E_2} 1_A(x_1, x_2)\, d\mu_2(x_2) \right) d\mu_1(x_1) \quad \forall A \in \mathcal{E}.$$

This is well-defined by the previous lemmas.

Clearly $\mu(\emptyset) = 0$, and it satisfies $(*)$ since for $A = A_1 \times A_2$,

$$\mu(A) = \int_{E_1} 1_{A_1}(x_1) \left(\int_{E_2} 1_{A_2}(x_2)\, d\mu_2(x_2) \right) d\mu_1(x_1) = \mu_1(A_1)\, \mu_2(A_2)$$

To check countable additivity, let $A_n \in \mathcal{E}$ be disjoint, so

$$1_{\cup_n A_n} = \sum_n 1_{A_n} = \lim_{N\to\infty} \sum_{n\leq N} 1_{A_n} \quad \text{(monotone limit)}, \text{ so}$$

$$\mu(\cup_n A_n) = \int_{E_1} \left(\int_{E_2} \sum_n 1_{A_n}(x_1,x_2)\, d\mu_2(x_2) \right) d\mu_1(x_1)$$

$$= \int_{E_1} \sum_n \left(\int_{E_2} 1_{A_n}(x_1,x_2)\, d\mu_2(x_2) \right) d\mu_1(x_1) \quad \Big\downarrow \text{ monotone convergence}$$

$$= \sum_n \int_{E_1} \left(\int_{E_2} 1_{A_n}(x_1,x_2)\, d\mu_2(x_2) \right) d\mu_1(x_1)$$

$$= \sum_n \mu(A_n) \quad \text{, so } \mu \text{ is indeed a measure} \qquad \square$$

<u>Remark</u> The construction is symmetric in $1 \leftrightarrow 2$

<u>Remark</u> It is clear that on $\overline{\mathcal{E}} = \mathcal{E}_2 \otimes \mathcal{E}_1$, we can define $\overline{\mu} = \mu_2 \otimes \mu_1$ and if $\overline{f} : E_2 \times E_1 \longrightarrow [0;\infty]$ is defined as

$$\overline{f}(x_2, x_1) = f(x_1; x_2), \text{ then}$$

$\overline{\mu}(\overline{f}) = \mu(f)$, simply because

$\mu_1(A_1)\mu_2(A_2) = \mu_2(A_2)\mu_1(A_1)$ on π-system $\mathcal{A}$.

<u>Theorem</u> (Fubini)

Consider a product space $(E, \mathcal{E}; \mu) = (E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2, \mu_1 \otimes \mu_2)$ $\nwarrow \mu_1, \mu_2$ finite

a) Let $f : (E, \mathcal{E}; \mu) \longrightarrow [0;\infty]$. Then

$$\mu(f) \overset{(+)}{=} \int_{E_1} \left(\int_{E_2} f(x_1, x_2)\, d\mu_2(x_2) \right) d\mu_1(x_1)$$
$$\overset{\text{symm}}{=} \int_{E_2} \left(\int_{E_1} f(x_1, x_2)\, d\mu_1(x_1) \right) d\mu_2(x_2)$$

b) Let $f : (E, \mathcal{E}, \mu) \longrightarrow \mathbb{R}$ be μ-integrable

Set $A_1 = \{x \in E_1 : \int_{E_2} |f(x_1, x_2)|\, d\mu_2(x_2) < \infty\}$

and $f_1(x_1) = \int_{E_2} f(x_1, x_2)\, d\mu(x_2)$ for $x_1 \in A_1$

$= 0$ otherwise

Then $\mu_1(A_1^c) = 0$ and f_1 is μ_1-integrable and

$\mu_1(f_1) = \mu(f)$

<u>Proof</u> a) By the definition of the product measure μ in the previous proof, we know that the result is true (i.e. (+) holds) for $f = 1_{A_1 \times A_2}$, so by uniqueness$^{(?)}$, this extends to all $A \in \mathcal{E}$.

By linearity of the $\mu; \mu_1, \mu_2$ integrals, (+) extends to all simple functions

For general f take a simple approximation $f_n = \{2^{-n} \lfloor 2^n f \rfloor\} \wedge n$, so $0 \le f_n \nearrow f$; so (+) holds for f_n and by monotone converge theorem, we see $\mu(f_n) \nearrow \mu(f)$, and also $\forall x_1 \in E_1$,

$$0 \le \int_{E_2} f_n(x_1; x_2)\, d\mu_2(x_2) \nearrow \int_{E_2} f(x_1; x_2)\, d\mu_2(x_2)$$

so overall $\mu(f) = \lim\limits_n \mu(f_n) \overset{(+)}{=} \lim\limits_n \int_{E_1} \left(\int_{E_2} f_n(x_1; x_2)\, d\mu_2(x_2) \right) d\mu_1(x_1)$

$$\overset{\text{MCT}}{=} \int_{E_1} \left(\int_{E_2} f(x_1, x_2)\, d\mu_2(x_2) \right) d\mu_1(x_1)$$

which proves a)

b) $f: E \to \mathbb{R}$ is s.t. $\mu(|f|) < \infty$

The map $x_1 \mapsto \int_{E_2} |f(x_1, x_2)| \, d\mu_2(x_2)$

$$E_1 \to [0; \infty]$$

is $\mathcal{E}_1$-measurable by the lemma from last lecture, and by a) the μ_1-integral of this map satisfies

$$\int_{E_1} \left(\int_{E_2} |f(x_1, x_2)| \, d\mu_2(x_2) \right) d\mu_1(x_1) = \mu(|f|) < \infty$$

In particular, $\mu_1(E_1 \setminus A_1) = 0$

So the map f_1 is μ_1-integrable. Now we can decompose

$$f_1^{(\pm)} = \int_{E_2} f^{\pm}(x_1, x_2) \, d\mu_2(x_2)$$

and $f_1 = (f_1^{(+)} - f_1^{(-)}) 1_{A_1}$, by linearity of μ_2-integral

Then using a) we deduce

$$\begin{aligned} \mu(f) &= \mu(f^+) - \mu(f^-) \\ &= \mu_1(f^{(+)}) - \mu_1(f^{(-)}) \\ &= \mu_1(f_1) \end{aligned}$$

$\square$

Remarks 1) Clearly $\mu_1(f_1) = \mu_1(f_1 1_{A_1})$

2) The theorem extends to σ-finite measures (such as Lebesgue-measure on $\mathbb{N}$)

3) By a π-system argument, check

$$(\mathcal{E}_1 \otimes \mathcal{E}_2) \otimes \mathcal{E}_3 = \mathcal{E}_1 \otimes (\mathcal{E}_2 \otimes \mathcal{E}_3)$$

so by induction we can extend the previous arguments to any $n \in \mathbb{N}$

In particular, on $(\mathbb{R}^n, \mathcal{B}^{\otimes n})$, $\mathcal{B}^{\otimes n} = \bigotimes_{i=1}^{n} \mathcal{B}(\mathbb{R})$

one defines the n-dimensional Lebesgue measure $\mu^n = \bigotimes_{i=1}^{n} \mu$,

where μ is standard Lebesgue measure on $\mathbb{R}$. The resulting integral is written as

$$\begin{aligned} \int_{\mathbb{R}^n} f \, d\mu^n &= \int_{\mathbb{R}^n} f(x) \, dx \\ &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \dots, x_n) \, d\mu_1(x_1) \cdots d\mu_n(x_n) \end{aligned}$$

for $f: (\mathbb{R}^n, \mathcal{B}^{\otimes n}) \to \mathbb{R}$.

L13.3

Product prob spaces & independence

Prop Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a prob space, and let $(E_i, \mathcal{E}_i)$; $i=1,\dots,n$ be measurable spaces, with product $(E, \mathcal{E}) = (\overset{n}{\underset{i=1}{\times}} E_i, \bigotimes_{i=1}^{n} \mathcal{E}_i)$

Consider a measurable map

$$X(\omega) = (X_1(\omega), \dots, X_n(\omega)) \; ; \; X : \Omega \to E$$

TFAE a) $X_1, \dots, X_n$ are independent

b) $\mu_X = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$

c) $\forall f_1, \dots, f_k$ bounded $\mathcal{E}_i$-meas on E_i, $i=1,\dots,n$

then $\mathbb{E}\left(\prod_{k=1}^{n} f_k(X_k)\right) = \prod_{k=1}^{n} \mathbb{E}(f_k(X_k))$

Proof Set $\nu = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$, which on the π-system

$$\mathcal{A} = \{ \overset{n}{\underset{k=1}{\times}} A_k : A_k \in \mathcal{E}_k \} \quad \text{generating } \mathcal{E}$$

satisfies $\forall A \in \mathcal{A}$,

$$\begin{aligned} \mu_X(A) &= \mathbb{P}(X_k \in A_k \; \forall k) \\ &= \mathbb{P}(\cap_k \{X_k \in A_k\}) \\ &\downarrow \text{indep} \\ &= \prod_k \mathbb{P}(X_k \in A_k) \\ &= \prod_k \mu_{X_k}(A_k) \\ &= \nu(A) \end{aligned}$$

which by uniqueness theorem yields $\mu_X = \nu$. So a) $\Rightarrow$ b).

Next suppose b) holds. Use Fubini's theorem n-times to see

$$\begin{aligned} \mathbb{E}\left(\prod_{k=1}^{n} f_k(X_k)\right) &= \int_E \prod_{k=1}^{n} f_k(x_k) \, d\mu_X(x) \\ &\downarrow \text{by a)} \\ &= \int_E \prod_{k=1}^{n} f_k(x_k) \, d\mu_{X_k}(x_k) \\ &\downarrow \text{Fubini} \\ &= \prod_{k=1}^{n} \int_{E_k} f_k(x_k) \, d\mu_{X_k}(x_k) \\ &= \prod_{k=1}^{n} \mathbb{E}(f_k(X_k)) \end{aligned}$$

So b) $\Rightarrow$ c)

Finally, c) $\Rightarrow$ a) by taking $f_k = 1_{A_k}$, $A_k \in \mathcal{E}_k$, $k=1,\dots,n$ $\square$

One can in fact construct infinite product prob spaces

Suppose $(\Omega_i, \mathcal{F}_i, \nu_i)_{i=1}^{\infty}$ and set $\Omega = \prod_{i=1}^{\infty} \Omega_i$

On Ω we define cylinder subsets

$$\mathcal{C} = \{ C \subseteq \Omega : C = A \times \prod_{i>n} \Omega_i ,\ A \in \bigotimes_{i=1}^{n} \mathcal{F}_i ,\ n \in \mathbb{N} \}$$

which is a π-system generating the 'cylindrical' σ-algebra $\mathcal{F} = \sigma(\mathcal{C})$

Then $\exists!$ ν on $\mathcal{F}$ s.t.

$$\nu(C) = (\nu_1 \otimes \cdots \otimes \nu_n)(A)$$

extending a sequence of prob measures $\nu_1, \ldots$ to $\sigma(\mathcal{C})$. (Kolmogorov's extension thm)

Specifically, we can in this way realise an ∞ seq of independent rvs with Law ν_i $i = 1, \ldots$ by just taking coordinate projection $X_n(\omega) = \omega_n$, $\omega \in \Omega$.

L^p- spaces, norms, inequalities

Let $(E, \mathcal{E}, \mu)$ be a measure space, and let $p \in [1, \infty]$.

Then we define spaces

$$L^p(E, \mathcal{E}; \mu) = L^p(\mu) = L^p$$

$$= \{ f \text{ measurable}, \; \|f\|_p < \infty \}$$

where $\|f\|_p = \left(\int_E |f(x)|^p \, d\mu(x) \right)^{1/p}$, $p < \infty$

$$\|f\|_\infty = \operatorname*{ess\,sup}_{x \in E} |f(x)| = \inf\{ \lambda \in \mathbb{R} : f \leq \lambda \text{ a.e.} \}$$

Clearly if $\mu(E) < \infty$ then

$\|f\|_p \leq \|f\|_\infty (\mu(E))^{1/p}$, so in this case

$L^\infty \subseteq L^p \; \forall p$.

We say $f_n \to f$ in L^p as $n \to \infty$ if $\|f_n - f\|_p \to 0$ as $n \to \infty$.

Markov's (Chebyshev's) inequality

Prop Let $f \geq 0$ meas. Then $\forall \lambda \geq 0$,

$$\mu(x \in E : f(x) \geq \lambda) = \mu(f \geq \lambda) \leq \left(\frac{\lambda}{\mu(f)} \right)^{-1}$$

Proof Integrate wrt μ the inequality

$\lambda 1_{\{f \geq \lambda\}} \leq f$ on E. $\square$

So if $g \in L^p(\mu)$, then applying Markov's ineq to $f = |g|^p$ gives

$$\mu(|g| > \lambda) = \mu(|g|^p > \lambda^p)$$

$$\leq \frac{\|g\|_p^p}{\lambda^p} = O(\lambda^{-p}) \quad \text{as } \lambda \to \infty$$

Jensen's inequality

Let $I \subseteq \mathbb{R}$ be an interval. A map $c : I \to \mathbb{R}$ is convex if

$$\forall x, y \in I, \; t \in [0,1], \quad c(tx + (1-t)y) \leq t\,c(x) + (1-t)\,c(y)$$

~~Ex~~ Since c is continuous on int I, one shows that c is Borel

One also shows that a convex function (equivalently) satisfies

$$\forall x < t < y \text{ in } I, \quad \frac{c(t) - c(x)}{t - x} \leq \frac{c(y) - c(t)}{y - t} \quad (0)$$

<u>Lemma</u> For $c: I \to \mathbb{R}$ convex and any $m \in \text{int } I$, there exist $a, b \in \mathbb{R}$ s.t. $c(x) \geq ax + b \quad \forall x \in I$, with equality at $x = m$

<u>Proof</u> Applying (o) with $t = m$ we set

$$a \doteq \sup_{x<m} \left\{ \frac{c(m) - c(x)}{m - x} : x \in I \right\}$$

which exists by (o). Now for $y > m$, by (o) we have

$$a \leq \frac{c(y) - c(m)}{y - m} \quad \text{or} \quad c(y) \geq a(y-m) + c(m)$$
$$= ay + b \quad \text{for } b = c(m) - am$$

For $x < m$, we know

$$\frac{c(m) - c(x)}{m - x} \leq a \quad \text{so} \quad c(x) \geq a(x-m) + c(m)$$
$$= ax + b$$

Also, $c(m) = am + b$, so we are done. □

<u>Theorem</u> (Jensen's inequality)

Let X be an integrable r.v. ($\mathbb{E}|X| < \infty$) taking values in $I \subset \mathbb{R}$.

Let $c: I \to \mathbb{R}$ be convex. Then $\mathbb{E}\, c(X)$ is well-defined, and

$$\mathbb{E}\, c(X) \geq c(\mathbb{E} X)$$

<u>Proof</u> Assume first $m = \mathbb{E} X \in \text{int } I$.

Now by the previous Lemma we see

$$c(X) \geq aX + b \quad \text{and} \quad \mathbb{E}\, c^-(X) \leq |a|\, \mathbb{E}|X| + |b| < \infty$$

since X integrable

$\therefore \mathbb{E}\, c(X) = \mathbb{E}\, c^+(X) - \mathbb{E}\, c^-(X)$ is well defined in $(-\infty, \infty]$

Now integrating the preceding ineq gives

$$\mathbb{E}\, c(X) \geq a\, \mathbb{E} X + b = am + b = c(m) = c(\mathbb{E} X) \quad \text{as desired}$$

If $m = \mathbb{E} X$ lies at $\inf I$ or $\sup I$ then $X = \text{const.}$ a.s. and so

$$\mathbb{E}\, c(X) = c(\underset{X}{\overset{\shortparallel}{\mathbb{E} X}})$$ □

A direct implication, with convex function $c(x) = |x|^{q/p}$, $1 \le p < q < \infty$,

is $\|X\|_p = (\mathbb{E}|X|^p)^{1/p} = (c(\mathbb{E}|X^p))^{1/q}$

$$\underset{\text{Jensen}}{\le} (\mathbb{E}(c(|X|^p)))^{1/q}$$

$$= (\mathbb{E}|X|^q)^{1/q} = \|X\|_q \quad \text{, so } L^q(\mathbb{P}) \subseteq L^p(\mathbb{P})$$

for all $p \le q$.

Hölder and Minkowski's inequality

<u>Theorem</u> (Hölder, C-S for $p=q=2$)

Let $p, q \in [1, \infty]$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$ ("conjugate")

Then $\mu(|fg|) \le \|f\|_p \|g\|_q \quad \forall f, g$ meas

<u>Proof</u> If $\|f\|_p = \infty$ or $\|g\|_q = \infty$ ✓

and if $f=0$ or $g=0$ a.e. ✓

Then $\|f\|_p \neq 0$ as o/w $|f|^p = 0$ a.e. $\Rightarrow f = 0$ a.e.

We can thus divide desired ineq by $\|f\|_p$ and assume wlog $\|f\|_p = 1$.

In particular,
$$\mathbb{P}(A) = \int_A |f|^p \, d\mu \quad , \quad A \in \mathcal{E}$$

defines a p.m. on $(E, \mathcal{E})$ with density $|f|^p$.

So we can write, noting also that $q(p-1) = p$ for conjugate indices,

$$\mu(|fg|) = \int_E |fg| \, d\mu$$

$$= \int_E |g| \frac{|f|^p}{|f|^{p-1}} 1_{\{|f|>0\}} \, d\mu$$

$$= \mathbb{E}\left[\frac{|g|}{|f|^{p-1}} 1_{\{|f|>0\}} \right]$$

$$\underset{L^q(\mathbb{P}) \subseteq L^1(\mathbb{P})}{\le} \left(\mathbb{E}\left[\frac{|g|^q}{|f|^{q(p-1)}} 1_{\{|f|>0\}} \right] \right)^{1/q}$$

$$= \left(\int_E |g|^q \frac{|f|^p}{|f|^p} 1_{\{|f|>0\}} \, d\mu \right)^{1/q}$$

$$\le \mu(|g|^q)^{1/q} = \|g\|_q \|f\|_p$$

□

Theorem (Minkowski)

Let $p \in [1, \infty]$, f, g measurable. Then

$$\|f+g\|_p \le \|f\|_p + \|g\|_q \quad (\Delta\text{-ineq for } L^p\text{- "norms"})$$

Proof $p = 1, \infty$ are clear. Also if $\|f\|_p = \infty$ or $\|g\|_p = \infty$ the result is also clear. So assume o/w.

Using the basic (pointwise) inequality

$$|f+g|^p \le 2^p(|f|^p + |g|^p) \quad \text{we see}$$

$$\mu(|f+g|^p) \le 2^p(\|f\|_p^p + \|g\|_p^p) < \infty, \text{ so } f+g \in L^p$$

If $\|f+g\|_p = 0$, nothing to prove.

So assume >0, in particular

$$\mu(|f+g|^p) = \|f+g\|_p^p > 0, \text{ and so}$$

for conjugate p, q;

$$\| |f+g|^{p-1} \|_q = (\mu(|f+g|^p))^{1-\frac{1}{p}}$$

Now we use Hölder's inequality

$$\|f+g\|_p^p = \int_E |f+g|^{p-1} |f+g| \, d\mu \le \int_E |f+g|^{p-1} |f| \, d\mu + \int_E |f+g|^{p-1} |g| \, d\mu$$

$$\overset{\text{Hölder}}{\le} \left(\int_E |f+g|^{q(p-1)} d\mu\right)^{1/q} \left(\int |f|^p d\mu\right)^{1/p}$$

$$+ \left(\int_E |f+g|^{q(p-1)} d\mu\right)^{1/q} \left(\int |g|^p d\mu\right)^{1/p}$$

$$= (\|f\|_p + \|g\|_p)(\|f+g\|_p^{p/q})$$

so dividing by $\|f+g\|_p^{p/q} > 0$ we obtain

$$\|f+g\|_p^{p-p/q} \le \|f\|_p + \|g\|_p$$

which proves the theorem since $p(1-\frac{1}{q}) = 1$. $\square$

By linearity of $\int \cdot \, d\mu$ and Minkowski's inequality, the $\|\cdot\|_p$ - functionals satisfy 1) + 2), but not 3), since $\|f\|_p = 0 \Rightarrow |f|^p = 0$ a.e.

$\Rightarrow f = 0$ a.e.,

with the qualification 'a.e.'.

Define equivalence classes $[f] = \{ h \text{ meas.}; \text{ s.t. } h = f \text{ a.e} \}$, we can define

$$\mathcal{L}^p(\mu) = L^p(\mu) / \{ h \text{ meas.}, h = 0 \ \mu\text{.a.e.} \}$$
$$= \{ [f] : f \in L^p \}$$

which equipped with the $\|\cdot\|_p$ - norm is a normed vector space ($[0] = 0$).

Since $\|[f]\|_p = \|f\|_p$ this distinction of spaces is often notationally suppressed.

A normed space $(V, \|\cdot\|)$ is called <u>complete</u> if every Cauchy sequence (v_n) in V convergences to some $v \in V$. More precisely, recall that v_n is Cauchy for $\|\cdot\|$ if $\forall \varepsilon > 0, \exists N \in \mathbb{N}$ s.t. $\forall m, n \geq N$,

$$\| v_n - v_m \| \leq \varepsilon ,$$

and in a complete space, $\exists v \in V$ s.t. $v_n \to v$.

<u>Terminology</u> a complete normed linear space is called a <u>Banach</u> space

<u>Theorem</u> ($\mathcal{L}^p$ is a Banach space)

Let $1 \leq p \leq \infty$, and let $(f_n : n \in \mathbb{N})$ be s.t.

$$\| f_n - f_m \|_p \to 0 \text{ as } m, n \to \infty \quad (\text{Cauchy for } L^p\text{-norm})$$

Then $\exists f \in L^p$ s.t. $\| f_n - f \|_p \to 0$ as $n \to \infty$.

<u>Proof</u> ($p = \infty$ exercise)

By hypothesis we can find a subsequence n_k ($\varepsilon = 2^{-k}$) s.t.

$$\| f_{n_{k+1}} - f_{n_k} \|_p \leq 2^{-k} \quad \forall k \in \mathbb{N},$$

so $$S = \sum_{k=1}^{\infty} \| f_{n_{k+1}} - f_{n_k} \|_p < \infty$$

By Minkowski's inequality, for any finite integer K,

$$\Big\| \sum_{k=1}^{K} | f_{n_{k+1}} - f_{n_k} | \Big\|_p \leq \sum_{k=1}^{K} \| f_{n_{k+1}} - f_{n_k} \|_p \leq S$$

So by the monotone convergence theorem,

$$\Big\| \sum_{k=1}^{\infty} | f_{n_{k+1}} - f_{n_k} | \Big\|_p \leq S .$$

So in particular

$$\sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| < \infty \quad \text{a.e.}\ , \quad \text{hence } \forall x \in A, \quad \mu(E \setminus A) = 0,$$

the real series $\sum_{k=1}^{K} (f_{n_{k+1}}(x) - f_{n_k}(x)) = f_{n_{K+1}}(x) - f_{n_0}(x)$

so by the completeness of $\mathbb{R}$, $f_{n_{K+1}}(x) \longrightarrow \lim_k f_{n_k}(x)$ as $K \to \infty$

for $x \in A$. Define

$$f(x) = \begin{cases} \lim_k f_{n_k}(x), & \text{for } x \in A \\ 0 & \text{o/w} \end{cases} \quad \text{(meas)}$$

and by what precedes we have shown that

$$f_{n_k} \longrightarrow f \quad \text{a.e.}$$

Now let $\varepsilon > 0$ be fixed and choose N large enough s.t.

$$\|f_n - f_m\|_p^p < \varepsilon \qquad \forall\, m, n \geq N$$

Then $\|f_n - f\|_p^p = \mu(|f_n - f|^p)$

$$= \mu\left(\underline{\lim}_K |f_n - f_{n_k}|^p\right)$$

$$\overset{\text{Fatou}}{\leq} \underline{\lim}_K\, \mu(|f_n - f_{n_k}|^p) \overset{\text{Cauchy prop}}{\leq} \varepsilon$$

So in particular, $f \in L^p$,

$$\|f\|_p \leq \|f - f_N\|_p + \|f_N\|_p < \infty \quad (N \text{ fixed})$$

and $f_n \longrightarrow f$ in L^p. $\square$

<u>Remark</u> If $f_n \in C([0,1])$ or f_n is simple, and Cauchy for $\|\cdot\|_1$ (μ Lebesgue measure) on $(0,1)$, then $\exists\, f \in L^1(\mu)$ s.t. $f_n \to f$ in $L^1(\mu)$. One further shows that $C([0,1])$ or $V = \{f \text{ simple}\}$ are <u>dense</u> in $L^1(\mu)$ for $\|\cdot\|_1$-norm ($\forall f \in L^1(\mu)\ \exists\, v_f \in C([0,1])$, $v_f \in V$ resp s.t. $\|f - v_f\| < \varepsilon$, ε arbitrary). In this sense, $L^1(\mu)$ is the <u>completion</u> of $C([0,1])$ or V for the $\|\cdot\|_1$-norm. (Same for L^p, $p < \infty$)

L15.3

$L^2(\mu)$ as a Hilbert Space

A symmetric bilinear map DX

$$(v,w) \mapsto \langle v,w \rangle \; ; \; V \times V \longrightarrow \mathbb{R}$$

on a vector space V is called an inner product if

$\langle v,v \rangle \geq 0$, $= 0$ iff $v = 0$

Then $\|v\| = \sqrt{\langle v,v \rangle}$ is a norm on V

Then $\mathcal{L}^2(\mu)$ is an inner product space for

$\langle f,g \rangle = \int_E fg \, d\mu$, inducing the norm $\|\cdot\|_2$

A complete inner product space (such as $\mathcal{L}^2(\mu)$) is called a <u>Hilbert space</u>

Basic geometric properties are

1) (Pythagoras) $\|f+g\|_2^2 = \|f\|_2^2 + 2\langle f,g \rangle + \|g\|_2^2$

We say f is orthogonal to g in $\mathcal{L}^2$ if $\langle f,g \rangle = \int_E fg \, d\mu = 0$

2) (parallelogram identity)

$$\|f+g\|_2^2 + \|f-g\|_2^2 = 2\|f\|_2^2 + 2\|g\|^2$$

~~Theorem~~ For $V \subseteq L^2$, we define the orthogonal complement

$$V^\perp = \{ f \in L^2 : \langle f,v \rangle = 0 \; \forall v \in V \}$$

A subset V of L^2 (or $\mathcal{L}^2$) is closed if for any $f_n \in V$ s.t. $f_n \to f$ in $\mathcal{L}^2$ we know

$$f \in V \; (f = v \text{ a.e.}, v \in V)$$

<u>Theorem</u> (orthogonal projection)

Let $V \subseteq L^2$ be a closed vector subspace (of L^2)

Then for each $f \in L^2$ there exists a (a.e. unique) orthogonal decomposition

$$f = v + u \;, \quad v \in V, \; u \in V^\perp$$

s.t. $\|f - v\|_2 \overset{(*)}{\leq} \|f - g\|_2 \;\; \forall g \in V$, with equality iff $v = g$ a.e.

[We call v the orthogonal projection of f onto V]

<u>Proof</u> $\|\cdot\| = \|\cdot\|_2$ for this proof

Define a minimal distance $d(f,V) = \inf_{g \in V} \|f - g\|$

and take a sequence $g_n \in V$ s.t. $\|f - g_n\| \longrightarrow d(f,V)$.

By the ▱-law, $m, n \in \mathbb{N}$

$$2\|f-g_n\|^2 + 2\|f-g_m\|^2 = \underbrace{\|2(f - \underbrace{\tfrac{g_m+g_n}{2}}_{\in V})\|^2}_{\geq 4\, d(f,V)^2} + \|g_n - g_m\|^2$$

so rearranging,

$$0 \leq \|g_n - g_m\|^2 \leq 2\left(\|f-g_n\|^2 + \|f-g_m\|^2\right) - 4\, d(f,V)^2$$

$$\longrightarrow 0 \text{ as } n,m \longrightarrow \infty$$

So we conclude that g_n is Cauchy for $\|\cdot\|$, hence $\to g$ in L^2, and since V was closed, $g \in V$ too

Then $\|f-g\| = \lim \|f-g_n\| = d(f,V)$ as desired,

so g indeed minimises the distance and $(+)$ holds (a.e. shenanigans).

In particular the map

$$F : t \mapsto \|f - (g+th)\|^2, \quad t \in \mathbb{R},\ h \in V$$

is minimised at $t=0$, and

$$F(t) = \|f-g\|^2 - 2t\langle f-g, h\rangle + t^2\|h\|^2$$

$$F'(t) = -2\langle f-g, h\rangle + 2t\,\|h\|^2$$

$$F''(t) = 2\|h\|^2 \geq 0$$

so necessarily we must have $0 = F'(0) = -2\langle f-g, h\rangle \quad \forall h \in V$

This means $\langle f-g, h\rangle = 0 \ \forall h \in V$, so $u = f-g \in V^\perp$

and we have our decomposition $f = u+v$.

Uniqueness, suppose $f = w+z$ for some $w \in V$, $z \in V^\perp$. Then

$$(v-w) + (u-z) = 0 \quad \text{and using Pythagoras}$$

$$0 = \|v-w\|^2 + \|u-z\|^2 \Rightarrow v = w,\ u = z. \qquad \square$$

When the Hilbert space is $L^2(\mathbb{P})$, then the <u>covariance</u> of r.v.s X, Y in $L^2(\mathbb{P})$ is $\quad \operatorname{cov}(X,Y) = \mathbb{E}((X-\mathbb{E}X)(Y-\mathbb{E}Y))$

$$= \langle X - \mathbb{E}X, Y - \mathbb{E}Y \rangle$$

$\operatorname{var}(X) = \operatorname{cov}(X,X)$; and we may declare

<u>$X \perp Y$ in</u> $L^2(\mathbb{P})$ if $\mathbb{E}X = \mathbb{E}Y = 0$, $\operatorname{cov}(X,Y) = 0$.

Next, if $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$ generated by $\{\mathcal{G}_i : i \in I\}$, I countable, $\mathcal{G}_i \in \mathcal{F}$, and $L^2_{(\star)}(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$; and one defines the conditional expectation of $X \in L^2(\Omega,\mathcal{F},\mathbb{P})$ given $\mathcal{G}$ as $\quad Y = \sum_{i\in I} \mathbb{E}(X|G_i)\, 1_{\mathcal{G}_i}$,

$$\mathbb{E}(X|\mathcal{G}_i) = \frac{\mathbb{E}(X 1_{\mathcal{G}_i})}{\mathbb{P}(\mathcal{G}_i)}, \quad \mathbb{P}(\mathcal{G}_i) > 0$$

One shows that the law of Y coincides with the law of the $L^2(\mathbb{P})$-<u>projection</u> of X onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$. (Dudley p350)

<u>Convergence in $L^1(\mathbb{P})$</u>

In what follows, $L^1 = L^1(\Omega, \mathcal{F}, \mathbb{P})$ for some prob space [could take $\mu(E) < \infty$]:

<u>Theorem</u> (Bounded convergence)

Let $(X_n : n \in \mathbb{N})$ s.t. $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$, and $|X_n| \leq c \quad \forall n \in \mathbb{N}$, $c < \infty$, then $X_n \xrightarrow[n\to\infty]{} X$ in L^1

<u>Proof</u> From earlier, we know that $X_{n_k} \xrightarrow{a.s.} X$ along a subsequence n_k

$|X| = \lim_{n\to\infty} |X_{n_k}| \leq c$, $\Rightarrow X \in L^1(\mathbb{P})$

Then

$$\mathbb{E}|X_n - X| = \mathbb{E}|X_n - X| 1_{\{|X_n - X| > \varepsilon/2\}} + \mathbb{E}|X_n - X| 1_{\{|X_n - X| \leq \varepsilon/2\}}$$

$$\leq 2c\, \mathbb{P}(|X_n - X| > \varepsilon/2) + \varepsilon/2 < \varepsilon$$

for $n = n(c)$ large enough by def^n of conv in $\mathbb{P}$. $\square$

Next, a lemma

<u>Lemma</u> Let $X \in L^1(\mathbb{P})$. Then

$$I_X(\delta) = \sup\{ \mathbb{E}|X| 1_A : A \in \mathcal{F}, \mathbb{P}(A) \leq \delta \}$$
$$\downarrow$$
$$0 \text{ as } \delta \to 0$$

pf suppose not, so $\forall \varepsilon > 0$, $\exists A_n \in \mathcal{F}$ $(\delta = 2^{-n})$ s.t. $\mathbb{P}(A_n) \leq 2^{-n}$ but $\mathbb{E}(|X| 1_{A_n}) \geq \varepsilon$

Then $\sum_n \mathbb{P}(A_n) < \infty$ so by Borel-Cantelli we see that

$\mathbb{P}(A_n \text{ i.o.}) = 0$.

Now use the dominated convergence theorem to deduce

$$\varepsilon \leq \mathbb{E}(|X| 1_{A_n}) \leq \mathbb{E}(|X| 1_{\bigcup_{m \geq n} A_m}) \longrightarrow \underset{\overset{\parallel}{0}}{\mathbb{E}(|X| 1_{\{A_n \text{ i.o.}\}})}$$

a contradiction. $\square$ #

The previous limit can be required to hold uniformly in a collection $\mathcal{X}$ of r.v.s in $L^1(\mathbb{P})$, specifically define

$$I_{\mathcal{X}}(\delta) = \sup \{ \mathbb{E}(|X| 1_A) : X \in \mathcal{X}, A \in \mathcal{F}; \mathbb{P}(A) \leq \delta \}$$

Clearly $I_{\mathcal{X}}(1) < \infty$ iff $\mathcal{X}$ is bounded in L^1 (i.e. $\sup_{X \in \mathcal{X}} \|X\|_1 < \infty$), and we say $\mathcal{X}$ is uniformly integrable (UI) if it is bounded in $L^1(\mathbb{P})$ and satisfies $I_{\mathcal{X}}(\delta) \to 0$ as $\delta \to 0$

Remark If $\mathcal{X}$ is bounded in $L^p(\mathbb{P})$ for $1 < p \leq \infty$, then by Hölder's inequality (conjugate q)

$$\mathbb{E}(|X| 1_A) \leq \underbrace{(\mathbb{E}|X|^p)^{1/p}}_{\leq C} (\underbrace{\mathbb{P}(A)}_{\leq \delta})^{1/q} \longrightarrow 0 \text{ as } \delta \to 0$$

so $\mathcal{X}$ is UI. (Not true for $p=1$; $\mathbb{P}$ is Lebesgue m on $(0,1)$, $\{X_n = n 1_{(0,1/n)}\}$ is bounded in L^1, but not UI.)

Lemma $\mathcal{X}$ is UI $\iff$ $\sup \{ \mathbb{E}|X| 1_{\{|X| > K\}} : X \in \mathcal{X} \} \to 0$ as $K \to \infty$

Proof "$\Rightarrow$" Choose δ such that $I_{\mathcal{X}}(\delta) < \varepsilon$; K large enough s.t. $I_{\mathcal{X}}(1) \leq K\delta$. By Markov's inequality,

$$\mathbb{P}(|X| > K) \leq \frac{\mathbb{E}|X|}{K} \leq \frac{I_{\mathcal{X}}(1)}{K} \leq \delta \quad \text{, hence}$$

$$\mathbb{E}|X| \underbrace{1_{\{|X|>K\}}}_{\substack{\text{event } A, \\ \mathbb{P}A \leq \delta}} \leq I_{\mathcal{X}}(\delta) < \varepsilon \quad \text{for such } K$$

"$\Leftarrow$" Conversely, take K s.t.

$$\sup_{X \in \mathcal{X}} \mathbb{E}(|X| 1_{\{|X|>K\}}) < \varepsilon/2$$

then $\mathcal{X}$ is bounded in $L^1(\mathbb{P})$ since

$$\mathbb{E}|X| \le \mathbb{E}(|X| 1_{\{|X| \le K\}}) + \mathbb{E}(|X| 1_{\{|X|>K\}})$$
$$\le K + \varepsilon/2 < \infty,$$

and if $A \in \mathcal{F}$, $\mathbb{P}(A) \le \delta$, then

$$\mathbb{E}|X| 1_A \le \mathbb{E}(|X| \underbrace{1_A}_{\le 1_E} 1_{\{|X|>K\}}) + \mathbb{E}(|X| 1_A 1_{\{|X| \le K\}})$$
$$< \varepsilon/2 + K\,\mathbb{P}(A)$$
$$\le \varepsilon/2 + K\delta < \varepsilon$$

for $\delta = \delta(K, \varepsilon)$ small enough. $\square$

<u>Theorem</u> Let $(X_n : n \in \mathbb{N})$, X be r.v.s in $(\Omega, \mathcal{F}, \mathbb{P})$. TFAE

(a) $X, X_n \in L^1(\mathbb{P})$ $\forall n$; $X_n \xrightarrow[n\to\infty]{} X$ in $L^1(\mathbb{P})$

(b) $\mathcal{X} = (X_n : n \in \mathbb{N})$ is UI and $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$

<u>Proof</u> "$a \Rightarrow b$" By Markov's inequality,

$$\mathbb{P}(|X_n - X| > \varepsilon) \le \frac{\mathbb{E}|X_n - X|}{\varepsilon} \to 0 \quad \text{as } n \to \infty,$$

so $X_n \to^{\mathbb{P}} X$. Since any finite collection of r.vs is UI, we have $\forall A \in \mathcal{F}$ s.t. $\mathbb{P}(A) \le \delta$ that

$$\mathbb{E}|X| 1_A < \varepsilon/2, \quad \mathbb{E}|X_n| 1_A < \varepsilon/2 \quad \forall n \le N, \quad N = N_\varepsilon \text{ chosen}$$

s.t. $\mathbb{E}|X_n - X| < \varepsilon/2$ $\forall n > N$.

Then $\mathbb{E}(|X_n| 1_A) \le \mathbb{E}(|X_n - X| 1_A) + \mathbb{E}(|X| 1_A)$

$$\le \varepsilon/2 + \varepsilon/2 \quad \text{for } n \ge N$$

so (X_n) is UI.

Next "$b \Rightarrow a$" From earlier we know $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X \Rightarrow X_{n_k} \xrightarrow{a.s.} X$ as $k \to \infty$, so by Fatou's lemma

$$\mathbb{E}|X| = \mathbb{E} \underline{\lim}_k |X_{n_k}| \le \underline{\lim}_k \mathbb{E}|X_{n_k}| \le I_{\mathcal{X}}(1) < \infty$$

So $X \in L^1(\mathbb{P})$. Now for K to be chosen

$$X_n^K = (-K) \vee X_n \wedge K, \quad X^K = (-K) \vee X_n \wedge K, \text{ so that}$$

$X_n^K \xrightarrow{\mathbb{P}} X^K$ as $n \to \infty$

since $\mathbb{P}(|X_n^k - X^k| > \epsilon) \leq \mathbb{P}(|X_n - X| > \epsilon) \to 0$

and by the lemma from earlier applied to $|X_n^k| \leq K = C$, we deduce

$$\mathbb{E}|X_n^k - X^k| \xrightarrow[n\to\infty]{} 0$$

Now write

$$\mathbb{E}|X_n - X| \leq \mathbb{E}|X_n - X_n^k| + \mathbb{E}|X_n^k - X^k| + \mathbb{E}|X^k - X|$$

$$\leq \mathbb{E}|X_n| 1_{\{|X_n| > k\}} + \underbrace{\varepsilon/3}_{\text{for } n(k) \text{ large enough}} + \mathbb{E}|X| 1_{\{|X| > k\}}$$

$$\overset{\text{UI}}{\underset{\text{lemma}}{<}} \varepsilon/3 + \varepsilon/3 + \varepsilon/3 \quad \text{for } k \text{ large enough, } n \text{ large enough}$$

So $X_n \to X$ in $L^1(\mathbb{P})$ as $n \to \infty$. $\square$

In this section, $L^p = L^p(\mathbb{R}^d)$, $1 \le p < \infty$, denote spaces of complex-valued integrable functions $f : \mathbb{R}^d \to \mathbb{C}$ s.t.

$$\|f\|_p^p = \int_{\mathbb{R}^d} |f(x)|^p \, dx < \infty,$$

dx is (product) Lebesgue-measure μ^d on $\bigotimes_{j=1}^{d} \mathcal{B}(\mathbb{R}) = \mathcal{B}$

The integral of a complex valued map $f = (\mathrm{Re} f) + i(\mathrm{Im} f)$ is defined as

$$\int f(x)\,dx = \int \mathrm{Re} f(x)\,dx + i \int \mathrm{Im}\, f(x)\,dx$$

where we assume $\int |f(x)|\,dx < \infty$. Then L^p is a complex vector space, satisfying $\left|\int f(x)\,dx\right| \le \int |f(x)|\,dx$.

On $\mathcal{L}^2(\mathbb{R}^d)$ the inner product $\langle \cdot, \cdot \rangle_{L^2}$ is now given by

$$\langle f, g \rangle_{L^2} = \int_{\mathbb{R}^d} f(x)\,\overline{g(x)}\,dx$$

Moreover, dx is <u>translation-invariant</u> (see Ex Sheet) in that

$$\int_{\mathbb{R}^d} f(x-y)\,dx = \int_{\mathbb{R}^d} f(x)\,dx = \int_{\mathbb{R}^d} f(-x)\,dx$$

(Fourier's idea: write $f = \sum_k c_k(f)\, e^{ikx}$, an 'orthogonal' decomposition of f in $L^2([0,2\pi])$)

A 'continuous' version of this idea is based on <u>Fourier transforms</u> $\hat{f} : \mathbb{R}^d \to \mathbb{C}$ of $f \in L^1(\mathbb{R}^d)$ defined as

$$\hat{f}(u) = \int_{\mathbb{R}^d} f(x)\, e^{i\langle u, x\rangle}\,dx \quad , \; u \in \mathbb{R}^d, \; \langle u, x\rangle = \sum_{i=1}^{d} u_i x_i$$

Clearly, since $|e^{i\langle u,x\rangle}| = 1$, we see $|\hat{f}(u)| \le \|f\|_1 \;\; \forall u$

Likewise, by the dominated convergence thm, if $u_n \to u$ in $\mathbb{R}^d$, then $\hat{f}(u_n) \to f(u)$, so $\hat{f} : \mathbb{R}^d \to \mathbb{C}$ is bdd continuous

<u>"TRAILER"</u> 1) Fourier Inversion, $f, \hat{f} \in L^1$ then $f(x) = \frac{1}{(2\pi)}^d \int_{\mathbb{R}^d} \hat{f}(u)\, e^{-i\langle x,u\rangle}\,du$

2) <u>Plancherel's theorem</u>, $f \in L^1 \cap L^2$, then

$$\|\hat{f}\|_{L^2} = (2\pi)^{d/2} \|f\|_{L^2}, \quad \langle f, g\rangle_{L^2} = (2\pi)^d \langle \hat{f}, \hat{g}\rangle_{L^2}$$

3) Proof of CLT (cheese lettuce tomato) from prob theory

Some more def^{n}s : For a prob. measure (or finite) μ we define the Fourier transform $\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} d\mu(x)$, $u \in \mathbb{R}^d$;

again $|\hat{\mu}(u)| \leq \mu(\mathbb{R}^d) = 1$

If $\mu = \mu_f$ arises from prob density f, then $\hat{\mu}_f = \hat{f}$

Related to this is the characteristic function of a r.v. X taking values in $\mathbb{R}^d$,

$$\phi_X(u) = \mathbb{E}(e^{i\langle u,X\rangle})$$
$$= \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} d\mu_X(x) = \hat{\mu}_X(u)$$

where μ_X is the law of the r.v. X.

A key concept in Fourier analysis is the notion of convolution of f^{n}s or measures. For $f \in L^p$, $1 \leq p < \infty$, and ν a prob measure on $\mathbb{R}^d$, set

$$f * \nu(x) = \int_{\mathbb{R}^d} f(x-y)\, d\nu(y) \quad , x \in \mathbb{R}^d$$
$$= \nu(f(x - \cdot))$$

By Jensen's inequality and Fubini's theorem ;

$$\|f*\nu\|_p^p = $$

$$\int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} |f(x-y)| \underset{\text{p.m.}}{\underset{\uparrow}{d\nu(y)}} \Big)^p dx$$

$$\overset{\text{Jensen}}{\leq} \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} |f(x-y)|^p d\nu(y) \Big) dx$$

$$= \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} |f(\underset{\text{drop } y}{\underset{\uparrow}{x-y}})|^p dx \Big) d\nu(y) = \|f\|_p^p \cdot 1$$

which implies $\|f*\nu\|_p \leq \|f\|_p$, in particular, $f*\nu \in L^p$ and hence is $< \infty$ a.e.

When ν has a density $g \in L^1$ so that $d\nu(x) = g(x)dx$, then we write $f*g$ for $f*\nu_g$ ($g \in L^1$)

For two prob measures μ, ν on $\mathbb{R}^d$ we define their <u>convolution</u> as the new measure

$$\mu*\nu(A) = \int_{\mathbb{R}^d\times\mathbb{R}^d} 1_A(x+y)\, d\mu(x)\, d\nu(y)$$
$$= \mu\otimes\nu(X+Y \in A), \quad A \in \mathcal{B}$$

So if μ has density f then by translation invariance of dx and Fubini's Theorem,

$$\mu * \nu(A) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1_A(x+y)\, f(x)\,dx\, d\nu(y)$$

$$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1_A(x)\, f(x-y)\, d\nu(y)\, dx$$

$$= \int_{\mathbb{R}^d} 1_A(x)\, f*\nu(x)\, dx \quad , A \in \mathcal{B}$$

so $\mu * \nu$ has density $f*\nu$ (in L^1)

A key observation is that, $f \in L^1$

$$\widehat{f*\nu}(u) = \int_{\mathbb{R}^d} f*\nu(x)\, e^{i\langle u,x\rangle}\, dx$$

$$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x-y)\, e^{i\langle u, x-y+y\rangle}\, d\nu(y)\, dx$$

$$\overset{\text{Fubini}}{=} \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} f(\underset{z}{x-y})\, e^{i\langle u, \underset{z}{x-y}\rangle}\, dx\right) e^{i\langle u,y\rangle}\, d\nu(y)$$

$$= \hat{f}(u) \underbrace{\int_{\mathbb{R}^d} e^{i\langle u,y\rangle}\, d\nu(y)}_{\hat{\nu}(u)} = \hat{f}(u)\cdot\hat{\nu}(u) \quad \forall u \in \mathbb{R}^d$$

Likewise, by independence of $X \sim \mu$, $Y \sim \nu$,

$$\widehat{\mu*\nu} = \mathbb{E}\left(e^{i\langle u, X+Y\rangle}\right)$$

$$\underset{\text{indep}}{=} \mathbb{E}\left(e^{i\langle u,X\rangle}\right)\mathbb{E}\left(e^{i\langle u,Y\rangle}\right)$$

$$= \hat{\mu}(u)\,\hat{\nu}(u) \quad , u \in \mathbb{R}^d$$

Fourier transforms of Gaussians

Recall that the Gaussian density of a $N(0,t)$ variable on $\mathbb{R}$ is

$$g_t(x) = \frac{1}{(2\pi t)^{1/2}}\, e^{-|x|^2/2t}, \quad x \in \mathbb{R}$$

In particular, for a $X \sim N(0,1)$ variable, the c.f.

$$\phi_X(u) = \mathbb{E}\left(e^{iuX}\right) \quad \text{satisfies}$$

L17.4

$$\frac{d}{du}\phi_X(u) = \frac{1}{\sqrt{2\pi}}\frac{d}{du}\int_{\mathbb{R}} e^{iux} e^{-x^2/2}\,dx$$

$$\overset{\text{diff under }\int}{=} \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{iux}\, ix\, e^{-x^2/2}\,dx$$

$$\overset{\text{by parts}}{=} 0 + i^2 u \int_{\mathbb{R}} e^{iux} e^{-x^2/2} \frac{1}{\sqrt{2\pi}}\,dx$$

$$= -u\,\phi_X(u)$$

So we obtain $\frac{d}{du}\left(e^{u^2/2}\phi_X(u)\right) = u e^{u^2/2}\phi_X(u) + e^{u^2/2}\left(-u\,\phi_X(u)\right)$

$$= 0 \quad \text{on } \mathbb{R}$$

So $\phi_X(u) = \phi_X(0) e^{-u^2/2} = e^{-u^2/2}$ ☺

So in $d=1$, $\hat{g}_1(u) = \sqrt{2\pi}\, g_1(u)$

We have shown in $d=1$ that

$$\hat{g}_1(u) = \sqrt{2\pi}\, g_1(u) \quad \forall u \in \mathbb{R}$$

Next let $Z_1, \dots, Z_d$ be $\overset{iid}{\sim} N(0,1)$, $Z = (Z_1, \dots, Z_d) \sim N(0, I_{\mathbb{R}^d})$

which has pdf

$$g_t(x) = \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|x|^2}{2t}\right), \quad x \in \mathbb{R}^d$$

then $\sqrt{t} Z \sim N(0, tI)$, $t>0$

$$\begin{aligned}
\hat{g}_t(u) &= \mathbb{E}\left(e^{i\langle u, \sqrt{t} Z\rangle}\right) \\
&= \mathbb{E}\left(\prod_{j=1}^d e^{i u_j \sqrt{t} Z_j}\right) \\
&\overset{ind}{=} \prod_{j=1}^d \mathbb{E}\left(e^{i \underbrace{u_j \sqrt{t}}_{\text{"}u\text{"}} Z_j}\right) \\
&= \prod_{j=1}^d e^{-u_j^2 t/2} \\
&= e^{-|u|^2 t/2}.
\end{aligned}$$

In other words, $\hat{g}_t(u) = (2\pi)^{d/2}\, t^{-d/2}\, g_{1/t}(u)$,

and Fourier-transforming again we see

$$\hat{\hat{g}}_t = (2\pi)^d g_t \quad \Rightarrow \quad g_t = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{g}(u)\, e^{-iux}\, du$$

Hence 'Fourier inversion' holds for (centred) Gaussians. (translation invariance)

A Gaussian convolution of $f \in L^1(\mathbb{R}^d)$ is

$$f * g_t = \int_{\mathbb{R}^d} f(\cdot - y)\, g_t(y)\, dy, \text{ where } g_t \text{ is a Gaussian}$$

pdf as above. As before, $f * g_t$ is bounded and cts on $\mathbb{R}^d$, and

$\| f * g_t \|_1 \leq \| f \|_1$ and also, since

$$| f * g_t(x) | \leq \frac{1}{(2\pi t)^{d/2}} \int | f(x-y) | \underbrace{e^{-|x|^2/2t}}_{\leq 1} dx = \frac{\|f\|_1}{(2\pi t)^{d/2}} \quad \forall x \in \mathbb{R}^d$$

Also $\widehat{f * g}(u) = \hat{f}(u)\, e^{-|u|^2 t/2}$, so

$$\| \widehat{f * g} \|_1 \leq \| f \|_1 \int_{\mathbb{R}^d} e^{-|u|^2 t/2}\, du = \left(\frac{2\pi}{t}\right)^{d/2} \| f \|_1$$

recall $\| \hat{f} \|_\infty \leq \| f \|_1$.

$$\| \widehat{f * g} \|_\infty \leq \| f \|_1$$

Remark If μ is a prob measure, then writing $g_t = g_{t/2} * g_{t/2}$

$$\mu * g_t = \underbrace{(\mu * g_{t/2})}_{\in L^1} * g_{t/2} \quad \text{is also a Gaussian conv}^{\underline{n}}$$

<u>Lemma</u> Fourier inversion holds for all Gaussian convolutions

<u>Proof</u> For $f \in L^1$, $t > 0$, use Fourier inversion for g_t and Fubini's theorem

$$(2\pi)^d f * g_t(x) = (2\pi)^d \int_{\mathbb{R}^d} f(x-y)\, g_t(y)\, dy$$

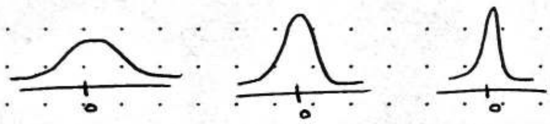
$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x-y)\, \hat{g}_t(u)\, e^{-i\langle u, \overset{-x+x}{y} \rangle}\, du\, dy$$

$$= \int_{\mathbb{R}^d} \underbrace{\left(\int_{\mathbb{R}^d} f(x-y)\, e^{i\langle u, x-y \rangle}\, dy \right)}_{\hat{f}(u)} \hat{g}_t(u)\, e^{-i\langle u, x \rangle}\, du$$

$$= \int_{\mathbb{R}^d} \hat{f}(u)\, \hat{g}_t(u)\, e^{-i\langle u, x \rangle}\, du$$

$$= \int_{\mathbb{R}^d} \widehat{f * g_t}(u)\, e^{-i\langle u, x \rangle}\, du \qquad \square$$

The next step uses that $f * g_t \approx f$ as $t \to 0$

( $\longrightarrow \delta_0$)

<u>Lemma</u> Let $f \in L^p$, $1 \leq p < \infty$. Then

$$\| f * g_t - f \|_p \longrightarrow 0 \quad \text{as } t \to 0$$

<u>Proof</u> $\forall \varepsilon > 0$, $\exists h \in C_c(\mathbb{R}^d)$ s.t. $\| f - h \|_p < \varepsilon/3$ (see ExSheet)

Then by linearity of $*$ we see

$$\| (f * g_t - h * g_t) \|_p = \| (f-h) * g_t \|_p \leq \| f - h \|_p$$

Next define $e(y) = \int_{\mathbb{R}^d \cap K_h} | h(x-y) - h(x) |^p\, dx$ which satisfies

$$0 \leq e(y) \leq 2^{p+1} \| h \|_p^p \quad \forall y \text{ ; which is } dx\text{-integrable on } K_h \text{ (bdd)}$$

Then since $h(x-y) - h(x) \longrightarrow 0$ as $y \to 0$,

the dominated converge theorem implies $e(y) \longrightarrow 0$ as $y \to 0$ $(\dagger)$

Next
$$\| h * g_t - h \|_p^p \overset{g_t\, \text{a pdf}}{=} \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \big(h(x-y) - h(x) \big)\, g_t(y)\, dy \right|^p dx$$

$$\overset{\text{Jensen}}{\leq} \int_{\mathbb{R}^d} e(y)\, g_t(y)\, dy$$

$$\overset{\frac{y}{\sqrt{t}} = z}{=} \int_{\mathbb{R}^d} e(\underbrace{\sqrt{t}\, z}_{\to 0})\, g_1(z)\, dz \longrightarrow 0 \quad \text{as } t \to 0$$

by bounded convergence

Summarising, by Minkowski's inequality

$$\|f*g_t - f\|_p \le \|f*g_t - h*g_t\|_p + \|h*g_t - h\|_p + \|h-f\|_p$$
$$\le 2\varepsilon/3 + \varepsilon/3$$

for t small enough. $\square$

<u>Theorem</u> (Fourier inversion) Suppose $f \in L^1(\mathbb{R}^d)$ and $\hat{f} \in L^1(\mathbb{R}^d)$ ($\in$ Fourier algebra). Then

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(u) \, e^{-i\langle u,x\rangle} \, du$$

for a.e. $x \in \mathbb{R}^d$.

<u>Remark</u> 1) The identity holds everywhere for the unique cts $f \in [f]$

2) If $\hat{f} = 0$ a.e. $\Rightarrow f = 0$ a.e.

<u>Proof</u> Set $f_t = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \underbrace{\hat{f}(u) \, e^{-|u|^2 t/2}}_{=\widehat{f*g_t}(u)} e^{-i\langle u,x\rangle} \, du$

so $f_t = f * g_t$ so by the previous Lemma,

$$\|f_t - f\|_1 = \|f*g_t - f\|_1 \longrightarrow 0 \text{ as } t \to 0.$$

At the same time, since $\hat{f}(u) \, e^{-|u|^2 t/2} \longrightarrow \hat{f}(u)$ as $t \to 0$ and $\hat{f} \in L^1$, the dominated converge theorem implies

$$f_t \longrightarrow \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(u) \, e^{-i\langle x,u\rangle} \, du$$

But since L^1-conv $\Rightarrow f_{t_n} \to f$ a.e. along a subsequence, then by uniqueness of limits,

$$f = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(u) \, e^{-i\langle x,u\rangle} \, du \quad \text{a.e.} \quad \square$$

<u>Theorem</u> (Plancherel) Let $f \in L^1 \cap L^2$. Then

$$\|\hat{f}\|_{L^2} = (2\pi)^{d/2} \|f\|_{L^2}$$

<u>Proof</u> First assume $\hat{f} \in L^1$, so Fourier inversion applies

$$\Rightarrow f, \hat{f} \in L^\infty \Rightarrow f, \hat{f} \in L^2$$ We see

$$(x,u) \mapsto f(x) \, \hat{f}(u)$$ is $dx\,du$-integrable, so

$$(2\pi)^d \|f\|_{L^2}^2 = (2\pi)^d \int_{\mathbb{R}^d} f(x) \, \overline{f(x)} \, dx$$

$$\overset{\text{Fourier inversion}}{=} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \hat{f}(u) e^{-i\langle u,x\rangle} du \right) \overline{f(x)}\, dx$$

$$\overset{\text{Fubini}}{=} \int_{\mathbb{R}^d} \hat{f}(u) \overline{\int_{\mathbb{R}^d} f(x) e^{i\langle u,x\rangle} dx}\, du$$

$$= \int_{\mathbb{R}^d} \hat{f}(u) \overline{\hat{f}(u)}\, du = \|\hat{f}\|_{L^2}^2$$

To extend this to general $f \in L^1 \cap L^2$, take the Gaussian conv

$f_t = f * g_t$; then $f_t \longrightarrow f$ in L^2 by the Lemma from earlier,

so $\|f_t\|_2 \longrightarrow \|f\|_2$ as $t \to 0$

Likewise $|\hat{f}_t| = |\hat{f}| e^{-|u|^2 t/2} \nearrow |\hat{f}|$ as $t \to 0$

So by the monotone convergence theorem,

$$\|\hat{f}_t\|_2^2 \nearrow \|\hat{f}\|_2^2 \quad \text{as } t \to 0$$

Then $(2\pi)^d \|f\|_2^2 = \lim_{t\to 0} (2\pi)^d \|f_t\|_2^2$

$= \lim_{t \to 0} \|\hat{f}_t\|_2^2$ by Plancherel, since $\hat{f}_t \in L^1$

$= \|\hat{f}\|_2^2$. $\square$

Weak convergence and characteristic f^{n}'s

We say that Borel p.m. $(\mu_n : n \in \mathbb{N})$ on $\mathbb{R}^d$ <u>converge weakly</u> to a Borel p.m. μ on $\mathbb{R}^d$ if

$$\mu_n(f) = \int_{\mathbb{R}^d} f \, d\mu_n \longrightarrow \int_{\mathbb{R}^d} f \, d\mu$$

$\forall f : \mathbb{R}^d \to \mathbb{R}$ bounded and continuous.

If $X_n \sim \mu_n$, $X \sim \mu$ are r.v.s then we say $X_n \to X$ weakly.

A few facts are (see Ex Sheet)

1) If $\int f d\mu = \int f d\nu$ $\forall f : \mathbb{R}^d \to \mathbb{R}$ and μ, ν p.m $\Rightarrow$ $\mu = \nu$

2) If $d = 1$ then $\mu_{X_n} \to \mu_X$ weakly for r.v's X_n, X $\Leftrightarrow$ $X_n \xrightarrow[n\to\infty]{d} X$

3) In fact since μ (and then also the μ_n) is a Radon measure, it concentrates its mass on a compact subset of $\mathbb{R}^d$; and one then shows further that $\mu_n \to \mu$ weakly $\Leftrightarrow$ $\int f d\mu_n \to \int f d\mu$ $\forall f \in C_c^\infty(\mathbb{R}^d)$

infinitely diff maps of compact support

A famous theorem due to P. Lévy is that to establish weak convergence, it suffices to check pointwise conv of c.f.s $\phi_{X_n} \to \phi_X$.

<u>Theorem</u> 1) Let X be a r.v. in $\mathbb{R}^d$. Then the law μ_X of X is uniquely determined by $\phi_X \; (= \hat{\mu}_X)$ and if $\phi_X \in L^1$ then μ_X has a pdf $f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi_X(u) e^{-iux} du$

2) If $(X_n : n \in \mathbb{N})$, X are rvs in $\mathbb{R}^d$ s.t. $\phi_{X_n}(u) \xrightarrow{p.w.} \phi_X(u)$ as $n \to \infty$ then $X_n \to X$ $(\mu_{X_n} \to \mu_X)$ weakly.

<u>Proof</u> We can take $Z \sim N(0, I)$ with density g_1, independent of X (later also, X_n) (if necessary enrich the prob space to be $(\Omega, \mathcal{F}, \mathbb{P}) \times (\mathbb{R}, \mathcal{B}, g_1 dx)$ and note that all statements are about 'laws' only)

a) Then $\sqrt{t} Z$ has density g_t and $X + \sqrt{t} Z$ has density $f_t = \mu_X * g_t$ (a Gaussian convolution)

By Fourier inversion for Gaussian convolutions,

$$f_t(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi_X(u) \, e^{-|u|^2 t/2} \, e^{-i\langle u, x\rangle} du$$

and if we test against any $g : \mathbb{R}^d \to \mathbb{R}$ bdd cts

$$\int_{\mathbb{R}^d} g(x)\, f_t(x)\, dx = \mathbb{E}\, g(X + \sqrt{t}\, Z) \xrightarrow[\text{convergence}]{\text{bounded}} \mathbb{E}(g\, X) = \int_{\mathbb{R}^d} g(x)\, d\mu_X(x)$$

so ϕ_X determines $\int g \, d\mu$ $\forall g$ bdd and cts

so by the remark from earlier, also μ_X

Next if $\phi_X \in L^1$, then $|f_t| \leq \frac{1}{(2\pi)^d} \|\phi_X\|_1$

and applying the dom. conv theorem with $\leq |\phi_X|$, $f_t(x) \to f_X(x)$ p.w. as $t \to 0$

But $f_t = \mu_X * g_t \; \left(= \mathcal{L}(X + \sqrt{t}\, Z)\right)$

≥ 0 $\forall t$, so the limit $f_X \geq 0$ as well

And for $g \in C_c(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} g \, d\mu_X = \lim_{t \to 0} \int_{\mathbb{R}^d} g(x)\, f_t(x)\, dx = \int_{\mathbb{R}^d} g(x)\, f_X(x)\, dx$$

by dominated convergence, so f_X is the pdf of μ_X.

b) Now let X_n be s.t. $\phi_{X_n}(u) \to \phi_X(u)$; and let $g : \mathbb{R}^d \to \mathbb{R}$ be compactly supported and Lipschitz $\left(|g(x) - g(y)| \leq \|g\|_{Lip} |x - y| \;\; \forall x, y \in \mathbb{R}^d \right)$

Then let $\varepsilon > 0$; choose t s.t. $\sqrt{t}\, \|g\|_{Lip}\, \mathbb{E}|Z| \leq \varepsilon/3$

so that

$$\mathbb{E}\left| g(X_n + \sqrt{t}\, Z) - g(X_n) \right| \leq \sqrt{t}\, \|g_{Lip}\|\, \mathbb{E}|Z| \leq \varepsilon/3$$

$$\mathbb{E}\left| g(X + \sqrt{t}\, Z) - g(X) \right| \leq \cdots \leq \varepsilon/3 \text{ ; now}$$

$$\begin{aligned} \left| \int g\, d\mu_{X_n} - \int g\, d\mu_X \right| &= \left| \mathbb{E}\, g(X_n) - \mathbb{E}\, g(X) \right| \\ &\leq \left| \mathbb{E}\, g(X_n) - \mathbb{E}\, g(X_n + \sqrt{t}\, Z) \right| \\ &\quad + \left| \mathbb{E}\, g(X_n + \sqrt{t}\, Z) - \mathbb{E}\, g(X + \sqrt{t}\, Z) \right| \\ &\quad + \left| \mathbb{E}\, g(X + \sqrt{t}\, Z) - \mathbb{E}\, g(X) \right| \\ &\leq \tfrac{2}{3}\varepsilon + \text{middle term} \end{aligned}$$

, so suffices to show convergence to zero of middle term as $n \to \infty$

Indeed, $\mathbb{E}\, g(X_n + \sqrt{t}\, Z) = \int_{\mathbb{R}^d} g(x)\, \overbrace{f_{t,n}(x)}^{\mu_{X_n} * g_t}\, dx$

$$\underset{\substack{\text{Fourier}\\ \text{inv}^n}}{=} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} g(x) \int_{\mathbb{R}^d} \phi_{X_n}(u)\, e^{-|u|^2 t/2} e^{-i\langle x, u\rangle}\, du\, dx$$

but $|g(x)|\, e^{-|u|^2 t/2}$ is $du\,dx$ integrable

and $|\phi_{X_n}(u)| \leq 1$, $\phi_{X_n} \to \phi_X$ pointwise

so by the dominated convergence theorem,

$$\underset{n\to\infty}{\longrightarrow} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} g(x)\, \phi_X(u)\, e^{-|u|^2 t/2} e^{-i\langle x, u\rangle}\, du\, dx$$

$$= \int_{\mathbb{R}^d} g(x)\, f_t(x)\, dx = \mathbb{E}\, g(X + \sqrt{t} Z)$$

So overall, $\left| \int g\, d\mu_{X_n} - \int g\, d\mu_X \right| < \varepsilon$ for all n large enough

for any $g \in C_c^\infty$, so $\mu_{X_n} \to \mu_X$ weakly $\square$

Multivariate Gaussian distributions

Def A r.v. X in $\mathbb{R}^d$ is Gaussian if $\langle u, X\rangle = \sum_{i=1}^d u_i X_i$ is one-dimensional Gaussian $\forall\, u \in \mathbb{R}^d$.

(Remark A 1-d-Gaussian is a normal distribution with pdf $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ for some $\mu \in \mathbb{R}$, $\sigma > 0$. $[N(\mu, \sigma^2)$, $N(\mu, 0) \sim \delta_\mu]$

Theorem Let $X = (X_1, \dots, X_n)$ be Gaussian in $\mathbb{R}^n$

a) if A is a $n \times n$ matrix, $b \in \mathbb{R}^n$ then $AX + b$ is Gaussian in $\mathbb{R}^n$

b) $X \in L^2$, $\mu = \mathbb{E}X$ exists, and the covariance matrix
$V = \operatorname{var}(X) = (\operatorname{cov}(X_i, X_j)_{i,j=1}^n)$ exists

c) $\phi_X(u) = e^{i\langle u, \mu\rangle - \langle u, Vu\rangle/2}$ $\forall\, u \in \mathbb{R}^n$

d) if V^{-1} exists then X has a pdf on $\mathbb{R}^n$ given by

$$f_X(x) = (2\pi)^{-\frac{n}{2}} (\det V)^{-\frac{1}{2}} \exp\left\{ -\frac{\langle x-\mu, V^{-1}(x-\mu)\rangle}{2} \right\}$$

e) if $X = (\underset{\mathbb{R}^{n_1}}{X_{(1)}}, \underset{\mathbb{R}^{n_2}}{X_{(2)}})$ is s.t. $\operatorname{cov}(X_1, X_2) = 0$, then $X_{(1)}, X_{(2)}$ are independent

L19.4

Proof Use Ex Sheet + uniqueness of c.f.s
+ props of 1-dim Gaussians $\square$

L20.1 ~~ct et~~

Sums of independent rvs

$X_1, \dots, X_n$ iid rvs ; $\mathrm{Var}(X_i) = 1$; $\mathbb{E} X_i = 0$, then $\forall \varepsilon > 0$

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n X_i \right| > \varepsilon \right) \leq \frac{\frac{1}{n^2} \mathrm{Var}\left(\sum_{i=1}^n X_i \right)}{\varepsilon^2} = \frac{1}{n\varepsilon^2} \longrightarrow 0$$

So $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mathbb{E} X = 0$ (weak LLN)

Theorem Let $(X_n : n \in \mathbb{N})$ be independent; s.t.

$$\mathbb{E} X_n = \mu \ \forall n, \quad \mathbb{E} X_n^4 \leq M \ \forall n$$

Then if $S_n = \sum_{i=1}^n X_i$; then

$$\frac{S_n}{n} \longrightarrow \mu \ \text{a.s. as } n \to \infty$$

Proof We can assume $\mu = 0$; since $Y_n = X_n - \mu$; then

$$\mathbb{E} Y_n = 0 \ \forall n \ \text{ and } \ \mathbb{E} Y_n^4 \leq \mathbb{E}\left(2^4 (X_n^4 + \mu^4)\right) \leq 16 (M + \mu^4) = \tilde{M}$$

Now to prove the result, notice

$$X_n, X_n^2, X_n^3 \ \text{are integrable} \ \left(L^4(\mathbb{P}) \subseteq L^p(\mathbb{P}), \ 1 \leq p \leq 4 \right)$$

and by independence, $\mu = 0$;

$$\mathbb{E} X_i X_j^3 = \mathbb{E} X_i X_j X_k^2 = \mathbb{E} X_i X_j X_k X_l = 0$$

for distinct i, j, k, l. Thus

$$\mathbb{E}(S_n^4) = \mathbb{E}\left(\sum_{i \leq n} X_i^4 + 6 \sum_{i < j} X_i^2 X_j^2 \right)$$

Now $\mathbb{E} X_i^2 X_j^2 \stackrel{\text{ind}}{=} (\mathbb{E} X_i^2)(\mathbb{E} X_j^2) \stackrel{\text{C-S}}{\leq} \sqrt[2]{\mathbb{E} X_i^4} \sqrt[2]{\mathbb{E} X_j^4} \leq M$ so

$$\mathbb{E}(S_n^4) \leq nM + 3n(n-1)M \leq 3n^2 M$$

so

$$\mathbb{E}\left(\sum_n \left(\frac{S_n}{n} \right)^4 \right) \leq 3M \sum_n \frac{1}{n^2} < \infty$$

So in particular, $\sum_n \left(\frac{S_n}{n} \right)^4 < \infty$ a.s.

So $\frac{S_n}{n} \longrightarrow 0$ a.s. (woah!) $\square$

Qu's How fast is the convergence of $\frac{S_n}{n} \to \mathbb{E}X \ (=\mu)$

Theorem (Cheese Lettuce Tomato)

Let $X_1, \dots, X_n$ be iid r.v.s on $\mathbb{R}$ st.

$\mathbb{E}X_i = 0 \ \forall i, \quad \mathrm{Var}(X_i) = 1$

Then for $S_n = \sum_{i=1}^n X_i$ we have, for all $x \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} S_n \le x\right) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$$

$$= \mathbb{P}(Z \le x), \quad Z \sim N(0,1)$$

Proof Apply to theorem from last lecture, and recall that weak convergence of $\frac{1}{\sqrt{n}} S_n \to Z$ as $n \to \infty$ will imply the result $(d=1)$

If $\phi(u) = \mathbb{E}e^{iuX_1} \Rightarrow \phi(0) = 1$, but differentiating ϕ wrt u under the integral, using $\mathbb{E}|X_1| \le \sqrt{\mathbb{E}|X_1|^2} = 1$,

$$\phi'(u) = i\,\mathbb{E}X_1 e^{iuX_1} \Rightarrow \phi'(0) = i\,\mathbb{E}X_1 = 0$$

and likewise

$$\phi''(u) = i^2\, \mathbb{E}X_1^2 e^{iuX_1} \Rightarrow \phi''(0) = -1$$

So applying Taylor's theorem to ϕ we see that as $u \to 0$,

$$\phi(u) = 1 - \frac{u^2}{2} + o(u^2) \quad \star$$

Then if ϕ_n is the c.f. of $S_n/\sqrt{n}$, we see

$$\phi_n(u) = \mathbb{E}\left(e^{iu(X_1 + \dots + X_n)/\sqrt{n}}\right) \overset{\text{iid}}{=} \left(\phi\left(\frac{u}{\sqrt{n}}\right)\right)^n$$

$$\underset{\substack{n\to\infty \\ u \text{ fixed}}}{=} \left(1 - \frac{u^2}{2n} + o\left(\frac{u^2}{n}\right)\right)^n$$

Using that the complex logarithm satisfies $\log(1+z) = z + o(|z|)$

so $\log \phi_n(u) = n \log\left(1 - \frac{u^2}{2n} + o\left(\frac{u^2}{n}\right)\right) = -\frac{u^2}{2} + o(1)$ as $n \to \infty$

hence $\forall u \in \mathbb{R}$, $\phi_n(u) \to \underset{\mathbb{E}(e^{iuZ})}{\underset{\uparrow}{e^{-u^2/2}}}$ as $n \to \infty$. $\square$

Remarks 1) A d-dimensional version can be proved either in the same way (Dudley, p306), or also using the Cramér-Wold device: $X_n \xrightarrow{d} X$ in $\mathbb{R}^k$ $\Leftrightarrow \langle u, X_n \rangle \xrightarrow{d} \langle u, X \rangle$ in $\mathbb{R}$ $\forall u \in \mathbb{R}^k$

2) The theorem says

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X \right) \xrightarrow{d} N(0,1), \text{ non-deg}$$

but this limit does not hold a.s. / in $\mathbb{P}$

One proves the law of the iterated lograithm (LIL)

$$\overline{\lim_n} \frac{S_n}{\sqrt{2n \log\log n}} \overset{a.s.}{=} +1$$

$$(\underline{\lim}) \quad '' \quad \overset{a.s.}{=} -1$$

The following results are often useful in applications of the CLT

Proposition 1) (Continuous mapping) If $X_n \to X$ weekly in $\mathbb{R}^d$, X_n takes values in $\mathcal{X} \subseteq \mathbb{R}^d$; and $g: \mathcal{X} \to \mathbb{R}^k$ is continuous, then $g(X_n) \to g(X)$ weekly in $\mathbb{R}^k$.

2) (Slutsky's Lemma)

If $X_n \to X$ weakly in $\mathbb{R}^k$, $Y_n \xrightarrow{\mathbb{P}} c$, c a constant (a.s.) then

$$(X_n, Y_n) \longrightarrow (X, c) \text{ weakly (jointly) in } \mathbb{R}^k \times \mathbb{R}^k$$

In particular, $X_n + Y_n \to X_n + c$, $X_n Y_n \to Xc$

$$X_n / Y_n \to X/c \quad \text{if } c \neq 0$$

Proof 1) Use that $f \circ g : \cancel{\mathbb{R}^d} \to \mathbb{R}$ is bounded, continuous and so ($\mathcal{X} = \mathbb{R}^d$)

$$\int f \, d\mu_{g(X_n)} = \int f\circ g \, d\mu_{X_n} \longrightarrow \int f \circ g \, d\mu_X = \int f \, d\mu_{g(X)}$$

2) First we prove an auxiliary result that if

(0) $X_n \xrightarrow{w} X$ as $n \to \infty$ and $|X_n - Y_n| \xrightarrow{\mathbb{P}} 0$

then $Y_n \xrightarrow{w} X$ as $n \to \infty$.

For any f bounded and Lipschitz

$$|\mathbb{E}(f(Y_n)) - \mathbb{E}f(X)| \leq \underbrace{|\mathbb{E}f(X_n) - \mathbb{E}f(X)|}_{\to 0 \text{ as } n \to \infty \text{ since } X_n \xrightarrow{w} X} + |\mathbb{E}f(Y_n) - \mathbb{E}f(X_n)|$$

$$\leq \left| \mathbb{E}\left(f(Y_n) - f(X_n)\right) 1_{\{|X_n - Y_n| \leq \varepsilon/2\}} \right|$$

$$+ \left| \mathbb{E}\left(f(Y_n) - f(X_n)\right) 1_{\{|X_n - Y_n| > \varepsilon/2\}} \right|$$

$$\leq \|f\|_{Lip} \, \varepsilon/2 + \|f\|_\infty \, \mathbb{P}(|X_n - Y_n| > \varepsilon/2)$$

$\leq \varepsilon$ for n large enough ☺

Now Slutsky's lemma follows from (o) with

$$|(X_n, Y_n) - (X_n, c)| = |Y_n - c| \xrightarrow{\mathbb{P}} 0$$

Also $(X_n, c) \xrightarrow{w} (X, c)$ weekly since for $f: \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ bdd & continuous the map $x \mapsto f(x, c)$ is bdd & cts on $\mathbb{R}^k$

so $\mathbb{E} f(X_n, c) \longrightarrow \mathbb{E} f(X, c)$ by weak conv of $X_n \to X$ in $\mathbb{R}^k$

The last claim follows from continuity of $+, \cdot, /$ on $\mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ $\square$ and the continuous mapping theorem

§ Ergodic Theory

A dynamical system consists of a state space E and a transformation $\theta : E \to E$. If we apply θ n-times we write

$$\theta^n = \underbrace{\theta \circ \cdots \circ \theta}_{n\text{-times}}$$

In ergodic theory one studies the long-term 'statistical' behaviour of the orbits $\theta^{(n)}(x)$, $x \in E$, specifically we will consider averages

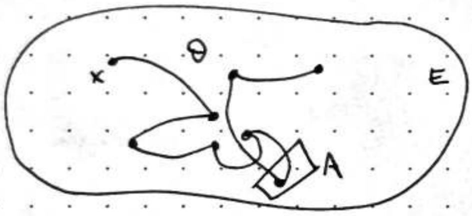
$$\frac{1}{n} \sum_{k=1}^{N} \theta^k(x) \quad , x \in E \text{ and } E \text{ endowed with some measure } \mu.$$

For instance, for $A \subseteq E$ we can ask whether

$$\frac{1}{N} \sum_{k=1}^{N} 1_A \circ \theta^k(x) = \frac{1}{N} \sum_{k=1}^{N} 1_{\{\theta^k(x) \in A\}}$$

stabilises as $N \to \infty$ at some "invariant" measure $\mu(A)$.

Ex (Boltzmann's ergodic hypothesis)



¿ no. of times the system visits A is approx $\approx vol(A)$ $(\mu(A))$?

Some definitions

Let $(E, \mathcal{E}, \mu)$ be a measure space.

A measurable map / transformation $\theta : E \to E$ is called measure-preserving (m.p.) if $\mu(\theta^{-1}(A)) = \mu(A)$ $\forall A \in \mathcal{E}$. In this case, if $f : E \to \mathbb{R}$ is measurable then necessarily $\int_E f \circ \theta \, d\mu = \int_E f \, d\mu$

A meas map f is called invariant (under θ) if $f = f \circ \theta$, and $A \in \mathcal{E}$ is called invariant if $\theta^{-1}(A) = A$.

The collection of invariant subsets of E, denoted $\mathcal{E}_\theta$, is a σ-algebra and $f : E \to \mathbb{R}$ is invariant $\iff$ f is $\mathcal{E}_\theta$-measurable.

Def^n We say that θ is ergodic (μ-ergodic) if $A \in \mathcal{E}_\theta$

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

Fact If θ is ergodic and f is invariant (for θ), then $f = c$ const. μ-a.e.

Some examples

1) On the n-torus $E = (0,1]^n$, $\mathcal{E} = \hat{\bigotimes}_{i \le n} \mathcal{B}((0,1])$, $\mu = \mu^n$ product Lebesgue measure, then the translation map

$$\theta_a(x_1, \dots, x_n) = (x_1 + a_1, \dots, x_n + a_n) \mod 1$$

is measure-preserving by translation invariance of μ; for $n = 1$, θ_a is ergodic $\iff$ a is irrational (see Ex Sheet)

2) On $E = (0,1]$ with μ = Lebesgue measure on $(0,1]$ the map $x \mapsto 2x \mod 1$ is m.p., ergodic for μ (Ex Sheet)

[It is also ergodic...]

3) Recall from earlier that on the infinite product space

$$E = \mathbb{R}^{\mathbb{N}} = \{x = (x_n : n \in \mathbb{N}) : x_n \in \mathbb{R}\}$$

we can consider the cylindrical σ-algebra $\sigma(\mathcal{C})$ generated by

$$\mathcal{C} = \{\prod_{n \in \mathbb{N}} A_n : A_n \in \mathcal{B}, A_n = \mathbb{R} \text{ for } n \text{ large enough}\}$$

One shows that

$$\sigma(\mathcal{C}) = \sigma(X_n : n \in \mathbb{N}), \quad X_n(x) = x_n$$

Now let m be any prob meas on $\mathbb{R}$; then we have shown earlier the existence of an infinite sequence $(Y_n : n \in \mathbb{N})$ iid of law m on $\mathbb{R}$; defined on $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{B}, \mu_{(0,1)})$ Lebesgue

which can be injected into $(\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}))$ by

$$Y : \Omega \to E, \quad Y(\omega) = (Y_n(\omega) : n \in \mathbb{N})$$

which is $\mathcal{F}$-$\sigma(\mathcal{C})$ measurable since for any cylinder $C = A_1 \times \dots \times A_N \times \dots$

$$Y^{-1}(C) = Y_1^{-1}(A_1) \cap \dots \cap Y_N^{-1}(A_N) \in \mathcal{F}$$

We define on $(\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}))$ the image measure

$$\mu = \mathbb{P} \circ Y^{-1}, \text{ which satisfies}$$

$$\mu\Big(\prod_{n \in \mathbb{N}} A_n\Big) = \prod_{n \in \mathbb{N}} \cancel{\mu}\, m(A_n) \quad \forall \text{ cylinders},$$

the unique 'product measure' on $\sigma(\mathcal{C})$.

One calls $(\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}), \mu) = (E, \mathcal{E}, \mu)$ the 'canonical model'
for an infinite sequence of iid variables.

On $(E, \mathcal{E}, \mu)$ we can define the (Bernoulli-) <u>shift map</u>

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

<u>Theorem</u> On $(\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}), \mu)$, the shift map θ is measure-preserving and ergodic

<u>Proof</u> For $A \in \mathcal{C}$ of the form

$A = \prod_{i \le n} (-\infty, x_i] \times \mathbb{R} \times \cdots$ we can check

$$\begin{aligned}
\mu(A) &= \mathbb{P}(Y_1 \le x_1, \dots, Y_N \le x_N) \\
&= m((-\infty, x_1]) \cdots m((-\infty, x_N]) \\
&= \mathbb{P}(Y_2 \le x_1, \dots, Y_{N+1} \le x_N) \\
&= \mu(\theta \circ 1_A) = \mu \circ \theta^{-1}(A)
\end{aligned}$$

So θ is m.p. (by Dynkin's Lemma)

To prove ergodicity, recall the tail-σ-algebra

$$\mathcal{T} = \bigcap_n \mathcal{T}_n \quad \text{where} \quad \mathcal{T}_n = \sigma(X_m : m \ge n+1)$$

Then for $A \in \mathcal{C}$, $A = \prod_k A_k$

$$\theta^{-n}(A) = \{X_{n+k} \in A_k \ \forall k\} \in \mathcal{T}_n$$

so since $\mathcal{T}_n$ is itself a σ-algebra, this extends to all $A \in \sigma(\mathcal{C})$.

Now suppose A is invariant, then

$$\theta^{-n}(A) = A \text{ for all } n, \quad \text{so } A \in \mathcal{T}_n \ \forall n, \text{ so } A \in \mathcal{T}$$

Thus $\mathcal{E}_\theta \subseteq \mathcal{T}$, and by Kolmogorov's 0-1 law,
we deduce

$$\mu(A) = 0 \text{ or } 1 \quad \text{for any invariant set}$$

$\therefore \theta$ is ergodic $\square$

next time: $\frac{1}{n} S_n(f) = \frac{1}{n} \sum_{m=1}^{n} \theta^m \circ f \longrightarrow$?

ergodic theorem

Ergodic theorems

Let $(E, \mathcal{E}, \mu)$ be a σ-finite measure space with

$\theta : E \to E$ a measure preserving transf $(\mu \circ \theta^{-1} = \mu)$

Set $S_0 = 0$, and for $f : E \to \mathbb{R}$ meas.,

$$S_n = S_n(f) = f + f \circ \theta + \cdots + f \circ \theta^{n-1}$$
$$= \sum_{k=0}^{n-1} f \circ \theta^k \qquad (\theta^0 = id)$$

Theorem (Birkhoff)

Let $f \in L^1(\mu)$. Then $\exists\ \bar{f} : (E, \mathcal{E}) \to \mathbb{R}$ θ-invariant s.t.

$$\mu(|\bar{f}|) \leq \mu(|f|) \quad \text{and} \quad \frac{S_n(f)}{n} \longrightarrow \bar{f} \quad \text{a.e.}$$

The proof is based on a key lemma

Lemma (maximal ergodic lemma)

For $f \in L^1(\mu)$, set $S^* = S^*(f) = \sup_{n \geq 0} S_n(f)$

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

Proof Define $S_n^* = \max_{0 \leq m \leq n} S_m$ $\quad (S_n^* \nearrow S^*)$

Then $S_m \leq S_n^*$ (on E) $\forall\, 0 \leq m \leq n$,

and $S_{m+1} = S_m \circ \theta + f$ (def)
$$\leq S_n^* \circ \theta + f \quad (+)$$

Now define $A_n = \{S_n^* > 0\}$ $\quad (\nearrow \{S^* > 0\})$.

On A_n, $S_n^* = \max_{1 \leq m \leq n} S_m$ (since $S_0 = 0$), so

$$S_n^* = \max_{1 \leq m \leq n} S_m \leq \max_{0 \leq m \leq n} S_{m+1} \overset{(+)}{\leq} S_n^* \circ \theta + f$$

Now integrating this inequality wrt $d\mu$,

$$\int_{A_n} S_n^* \, d\mu \leq \int_{A_n} (S_n^* \circ \theta) \, d\mu + \int_{A_n} f \, d\mu$$

On A_n^c we must have $S_n^* = 0 \leq S_n^* \circ \theta$ which implies

$$\int_E S_n^* \, d\mu \leq \underbrace{\int_E S_n^* \circ \theta \, d\mu}_{= \int_E S_n^* d\mu \text{ since } \theta \text{ is m.p.}} + \int_{A_n} f \, d\mu$$

Hence $\int_{A_n} f \, d\mu \geq 0$.

Since $f 1_{A_n} \longrightarrow f 1_{\{S^*>0\}}$ and $|f 1_{A_n}| \leq |f|$, $f \in L^1(\mu)$

we deduce from the dominated convergence theorem that

$$\int_{\{S^*>0\}} f \, d\mu = \lim_{n\to\infty} \int_{A_n} f \, d\mu \geq 0. \qquad \square$$

Now to prove Birkhoff's theorem, notice that

$$\limsup_n \frac{S_n(f)}{n} = \limsup_n \frac{S_n(f)\circ\theta}{n}$$
(inf) (inf)

$$\frac{S_{n+1}(f) - f}{n+1} \cdot \frac{n+1}{n}$$

are invariant functions and hence their inverse images

$$D = D_{a,b} = \left\{ \underline{\lim}_n \frac{S_n(f)}{n} < a < b < \overline{\lim}_n \frac{S_n(f)}{f} \right\}$$

are measurable and invariant ($\theta^{-1} D = D$).

Wlog we can assume $b > 0$. Then take $B \in \mathcal{E}$, $B \subseteq D$ s.t. $\mu(B) < \infty$ and set

$$g = f - b 1_B \in L^1(\mu)$$

Then

$$\begin{aligned} S_n(g) &= S_n(f) - b S_n(1_B) \\ &\geq S_n(f) - bn \quad (\text{since } 1_B + 1_B\circ\theta + \dots + 1_B\circ\theta^{n-1} \leq n) \\ (*) &> 0 \quad \text{on } D \text{ for some } n \in \mathbb{N} \quad (\text{def}^n \text{ of } \overline{\lim}) \end{aligned}$$

Now we apply the maximal ergodic lemma with $E = D$, $\mu = \mu|_D$ which is still θ-invariant:

$$\begin{aligned} \mu|_D(A) &= \mu(A \cap D) \overset{\text{inv}}{=} \mu(\theta^{-1}(A \cap D)) \\ &= \mu(\theta^{-1}(A) \cap \underbrace{\theta^{-1}(D)}_{D}) \\ &= \mu|_D(\theta^{-1}(A)) \qquad (\text{still mp}) \end{aligned}$$

And note that $\{S^* > 0\} \subseteq D$ as we are restricting, but by (*) get in fact $\{S^*(g) > 0\} \supseteq D$.

$$0 \leq \int_{\{S^*>0\}} g \, d\mu = \int_D g \, d\mu = \int_D f \, d\mu - b\mu(B)$$

So $b\mu(B) \leq \int_D f \, d\mu$.

By σ-finiteness, this inequality extends to D (take $B_n \uparrow D$, $\mu(B_n) < \infty$) and take limits $b\mu(D) = b \lim_n \mu(B_n) \leq \int_D f \, d\mu$ (1)

Repeating this argument with $-f, -a$ we deduce

$$(-a)\mu(D) \leq \int_D (-f) \, d\mu \qquad (2)$$

$$\therefore \ a\mu(D) \geq \int_D f \, d\mu$$

which combined give

$$b\mu(D) \leq \int_D f \, d\mu \leq a\mu(D)$$

but $a < b$, so necessarily $\mu(D) = 0$.

Now define
$$\Delta = \left\{ \underline{\lim} \frac{S_n(f)}{n} < \overline{\lim} \frac{S_n}{n} \right\} = \bigcup_{\substack{a,b \in \mathbb{Q} \\ a<b}} \left\{ \underline{\lim} \frac{S_n}{n} < a < b < \overline{\lim} \frac{S_n}{n} \right\}$$

Hence since $\mu(D_{a,b}) = 0$ by the above, $\mu(\Delta) = 0$.

On Δ^c, $\frac{S_n}{n}$ converges in $[-\infty, \infty]$ and, has invariant

$$\bar{f} = \begin{cases} \lim_n \frac{S_n}{n} & \text{on } \Delta^c, \\ 0 & \text{on } \Delta, \end{cases}$$

so that $\frac{S_n}{n} \to \bar{f}$ a.e. as $n \to \infty$.

Finally, since $\mu(|f \circ \theta^{n-1}|) = \mu(|f|)$ so

$$\mu(|S_n|) \leq n\,\mu(|f|) \quad \text{and thus}$$

$$\mu(|\bar{f}|) = \mu\left(\underline{\lim} \left|\frac{S_n}{n}\right|\right) \overset{\text{Fatou}}{\leq} \underline{\lim}\, \mu\left(\left|\frac{S_n}{n}\right|\right) \leq \mu(|f|) < \infty,$$

in particular $|\bar{f}| < \infty$ a.e. $\square$

L23.1

Theorem (von Neumann's L^p ergodic theorem)

Assume $\mu(E) < \infty$, $1 \le p < \infty$. Then $\forall f \in L^p(\mu)$,

$$\frac{S_n(f)}{n} = \frac{f + f\circ\theta + \cdots + f\circ\theta^{n-1}}{n} \longrightarrow \bar{f}$$

as $n \to \infty$ in $L^p(\mu)$.

Proof Since θ is m.p., so

$$\|f\circ\theta^i\|_p^p = \int_E |f|^p \circ \underbrace{\theta^i \, d\mu}_{= d\mu} = \|f\|_p^p \quad \forall i \geq 0$$

So by Minkowski's inequality, for any $f \in L^p(\mu)$,

$$\left\| \frac{S^n(f)}{n} \right\|_p \leqslant \frac{n}{n} \|f\|_p = \|f\|_p \quad (\dagger)$$

Now let $\varepsilon > 0$ be given, choose K large enough such that for

$$f_K = (-K) \vee f \wedge K,$$

$$\|f - f_K\|_p^p = \int_{\{|f|>K\}} (|f(x)| - K)^p \, d\mu(x) < (\varepsilon/3)^p$$

possible since $|f(x)|^p \mathbb{1}_{\{|f|>K\}} \searrow 0$ a.e. as $K \to \infty$;

and using dominated convergence.

Now f_K is bounded and hence in $L^1(\mu)$ (for $\mu(E) < \infty$); so by Birkhoff's theorem we know that

$$\frac{S_n(f_K)}{n} \xrightarrow[n\to\infty]{a.e.} \overline{f_K}, \quad \bar{f}_K \text{ invariant}$$

Since $\left| \frac{S_n(f_K)}{n} \right| \leqslant K$ for all n, bounded convergence implies that the last limit extends to L^p-convergence ($g_n \to g$ a.e., $|g_n| < K$ $\Rightarrow g_n \to g$ in $L^1(\mu)$, $\|g_n - g\|_p \leqslant K^{\frac{p-1}{p}} \|g_n - g\|_1 \to 0$)

so choose N large enough s.t. for $n \geqslant N$,

$$\left\| \frac{S_n(f_K)}{n} - \bar{f}_K \right\|_p < \varepsilon/3$$

Finally, for $\bar{f}$ the limit in Birkhoff's theorem of $\frac{S_n(f)}{f}$ (noting $L^p \subseteq L^1$)

$$\|\bar{f} - \bar{f}_K\|_p^p = \int \lim_{n\to\infty} \left| \frac{S_n(f)}{n} - \frac{S_n(f_K)}{n} \right|^p d\mu$$

$$\underset{\text{Fatou's lemma}}{\leq} \underline{\lim}_n \int \left| \frac{S_n(f-f_K)}{n} \right|^p d\mu$$

$$\underset{\text{For } f-f_K}{(+) \leq} \|f - f_K\|_p^p < (\varepsilon/3)^p \quad \text{by choice of } K$$

Collecting all terms, using Minkowski's inequality

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

$$\leq \underset{(+)}{\frac{\varepsilon}{3}} \quad \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \quad \text{for } n \text{ large enough}$$

$\square$

The strong law of large numbers

The previous ergodic theorems apply to the shift map

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

on the canonical p-space $(E, \mathcal{E}, \mu) = (\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}), \bigotimes_{i=1}^{\infty} m)$

where m is any p.m. on $\mathbb{R}$.

Theorem Let m be a p.m. on $\mathbb{R}$ s.t. $\int_{\mathbb{R}} |x| \, dm(x) < \infty$, $\int_{\mathbb{R}} x \, dm(x) = \nu$

Then

$$\mu\left(x \in \mathbb{R}^{\mathbb{N}} : \frac{x_1 + \dots + x_n}{n} \to \nu \text{ as } n \to \infty\right) = 1$$

Proof Choose $f(x) = x_1$, $f \in L^1(\mu)$,

so by the ergodic theorems

$$\frac{S_n(f)}{n} = \frac{f + f\circ\Theta + \dots + f\circ\Theta^{n-1}}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} \xrightarrow[L^1(\mu)]{a.e.} \bar{f}$$

where $\bar{f}$ is invariant, and = const a.e. since Θ was shown to be ergodic. Also

$$\text{const} = \mu(\bar{f}) \overset{L^1 \text{ limit}}{=} \lim_{n\to\infty} \mu\left(\frac{S_n(f)}{n}\right) = n \cdot \frac{\nu}{n} = \nu \quad \square$$

Theorem (Kolmogorov - Khinchine)

Let $(X_n : n \in \mathbb{N})$ be a sequence of iid r.v.'s defined on some p.space $(\Omega, \mathcal{F}, \mathbb{P})$ s.t. $\mathbb{E}|X_1| < \infty$, $\mathbb{E}X_1 = \nu$. Then if

$S_n = \sum_{i=1}^{n} X_i$, we have $\frac{S_n}{n} \xrightarrow{a.s.} \nu$ as $n \to \infty$

Proof Denote by m the law of X_1, and define

$$X : (\Omega; \mathcal{F}; \mathbb{P}) \longrightarrow (\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}))$$

by $X(\omega) = (X_1, X_2, \dots)$, which is measurable by arguments from earlier.

$$\mu = \mathbb{P} \circ X^{-1} = \bigotimes_{i=1}^{\infty} m$$

Using this injection

$$\mathbb{P}\left(\frac{S_n}{n} \to \nu \text{ as } n \to \infty\right)$$
$$= \mu\left(x \in \mathbb{R}^{\mathbb{N}} : \frac{x_1 + \dots + x_n}{n} \to \nu \text{ as } n \to \infty\right)$$
$$= 1 \text{ by the previous theorem.} \quad \square$$

To show that $\mathbb{E}|X_1| < \infty$ is necessary, we finally prove

Prop Let $X_1, \dots, X_n$ be iid s.t. $\mathbb{E}|X_1| = \infty$. Then a.s.,

$\frac{S_n}{n}$ does not converge to any finite limit

Proof Suppose to the contrary that $\frac{S_n}{n}$ converges a.s. to some limit.

Then $\frac{S_{n-1}}{n} = \frac{X_1 + \dots + X_{n-1}}{n-1} \cdot \frac{n-1}{n}$ converges to the same limit,

and hence
$$\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} \to 0 \text{ as } n \to \infty \text{ a.s.}$$

But for any non-negative r.v. Y (see ExSheet)

$$\mathbb{E}Y \leq \sum_n \mathbb{P}(Y > n), \quad \text{so for } Y = |X_1|,$$

we have
$$\infty = \mathbb{E}|X_1| \leq \sum_n \mathbb{P}(|X_1| > n)$$
$$\stackrel{iid}{=} \sum_n \mathbb{P}(|X_n| > n)$$

so by the second Borel-Cantelli Lemma,

$$\mathbb{P}\left(\frac{|X_n|}{n} > 1 \text{ i.o.}\right) = 1 \quad ※ \text{ to } \frac{X_n}{n} \to 0 \text{ a.s.} \quad \square$$