Probability and Measure

§1 Measures

1.1 Definitions

Definition Let E be a set.
A σ-algebra $\mathcal{E}$ on E is a set of subsets of E such that
$\emptyset \in \mathcal{E}$
$A \in \mathcal{E} \Rightarrow A^c \in \mathcal{E}$
$(A_n : n \in \mathbb{N})$ is a sequence in $\mathcal{E}$ $\Rightarrow$ $\bigcup_n A_n \in \mathcal{E}$

examples $\mathcal{E} = \{\emptyset, E\}$
$\mathcal{E} = 2^E$, the set of all subsets of E
$\mathcal{E} = \{\emptyset, A, A^c, E\}$, where $A \subseteq E$

Note $\mathcal{E}$ is 'closed under countable set operations'
$\bigcap_n A_n = (\bigcup_n A_n^c)^c$
$A \cup B = A \cup B \cup \emptyset \cup \emptyset \cup \dots$

Definitions $(E, \mathcal{E})$ is a measurable space
Given $(E, \mathcal{E})$, each $A \in \mathcal{E}$ is a measurable set

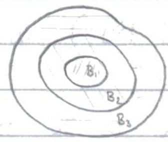
A measure μ on $(E, \mathcal{E})$ is a function $\mu : \mathcal{E} \to [0, \infty]$ such that
$\mu(\emptyset) = 0$
μ is countably additive
ie for any sequence $(A_n : n \in \mathbb{N})$ of disjoint elements of $\mathcal{E}$
$$\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$$

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

Note Let $A_1, \dots, A_n$ be disjoint
Then $\mu(A_1 \cup \dots \cup A_n) = \mu(A_1 \cup \dots \cup A_n \cup \emptyset \cup \emptyset \cup \dots)$
$= \sum_{i=1}^n \mu(A_i) + \sum \mu(\emptyset)$
$= \sum_{i=1}^n \mu(A_i)$

Claim: Suppose $(A_n : n \in \mathbb{N})$, $A \in \mathcal{E}$, with $A_n \uparrow A$
(ie $A_n \subseteq A_{n+1}$, $\bigcup_n A_n = A$)
Then $\mu(A_n) \uparrow \mu(A)$

Proof: Let $B_1 = A_1$
$B_n = A_n \setminus A_{n-1}$, $n \geq 2$
Then $(B_n : n \in \mathbb{N})$ are disjoint
$B_1 \cup \dots \cup B_n = A_n$
$\bigcup_n B_n = A_n$



$\mu(A_n) = \mu(\bigcup_{i=1}^n B_i)$
$= \sum_{i=1}^n \mu(B_i)$
$\uparrow \sum_{i=1}^\infty \mu(B_i)$
$= \mu(\bigcup_n B_n)$
$= \mu(A)$

1.2 Discrete measure theory

Definition

Let E be a countable set
$\mathcal{E}$ the set of all subsets of E
A mass function is a function $m : E \to [0, \infty]$

Consider two enumerations of E
$E = \{x_1, x_2, x_3, \dots\}$
$= \{y_1, y_2, y_3, \dots\}$
Note that $\forall\ n\ \exists\ t$ such that $\{x_1, \dots, x_n\} \subseteq \{y_1, \dots, y_t\}$
$\Rightarrow \sum_{i=1}^{\infty} m(x_i) \leq \sum_{i=1}^{\infty} m(y_i)$
By symmetry, $\sum_{i=1}^{\infty} m(x_i) \geq \sum_{i=1}^{\infty} m(y_i)$
$\Rightarrow \sum_{i=1}^{\infty} m(x_i) = \sum_{i=1}^{\infty} m(y_i)$
ie $\sum_{x} m(x)$ is well-defined

If μ is a measure on $(E, \mathcal{E})$, $A \subseteq E$,
then, by countable additivity, $\mu(A) = \sum_{x \in A} \mu(\{x\})$
So there is a one-to-one correspondence between measures and mass functions,
given by $m(x) = \mu(\{x\})$
$\mu(A) = \sum_{x \in A} m(x)$

1.3 Generated σ-algebras

Definition

Let $\mathcal{A}$ be a set of subsets of E
The σ-algebra generated by $\mathcal{A}$
is $\sigma(\mathcal{A}) = \{A \subseteq E : A \in \mathcal{E}\ \forall\ \sigma\text{-algebras } \mathcal{E} \text{ containing } \mathcal{A}\}$

$\sigma(\mathcal{A})$ is the smallest σ-algebra containing $\mathcal{A}$

1.4 π-systems and d-systems

Definitions

Let $\mathcal{A}$ be a set of subsets of E

$\mathcal{A}$ is a π-system if
$\emptyset \in \mathcal{A}$
$A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$

$\mathcal{A}$ is a d-system if
$E \in \mathcal{A}$
$A, B \in \mathcal{A}$ with $A \subseteq B \Rightarrow B \setminus A \in \mathcal{A}$
$(A_n : n \in \mathbb{N})$ is an increasing sequence in $\mathcal{A} \Rightarrow \bigcup_n A_n \in \mathcal{A}$

If $\mathcal{A}$ is both a π-system and a d-system
then $\mathcal{A}$ is a σ-algebra

Probability and Measure

Lemma 1.4.1 Dynkin's π-system lemma

Let $\mathcal{A}$ be a π-system

Then any d-system containing $\mathcal{A}$ also contains the σ-algebra generated by $\mathcal{A}$

Proof Let $\mathcal{D}$ be the intersection of all d-systems containing $\mathcal{A}$

Then $\mathcal{D}$ is itself a d-system; it is the smallest d-system containing $\mathcal{A}$

We will show that $\mathcal{D}$ is also a π-system and hence a σ-algebra, thus proving the lemma

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

Then $\mathcal{A}$ is a π-system $\Rightarrow \mathcal{A} \subseteq \mathcal{D}'$

Also, $\mathcal{D}'$ is a d-system, since

$E \in \mathcal{D}'$

Suppose $B_1, B_2 \in \mathcal{D}'$ with $B_1 \subseteq B_2$, and $A \in \mathcal{A}$

then $(B_2 \setminus B_1) \cap A = (B_2 \cap A) \setminus (B_1 \cap A) \in \mathcal{D}$, since $\mathcal{D}$ is a d-system

$\Rightarrow B_2 \setminus B_1 \in \mathcal{D}'$

Suppose $(B_n : n \in \mathbb{N})$ is an increasing sequence in $\mathcal{D}'$ with $B_n \uparrow B$, and $A \in \mathcal{A}$

then $B_n \cap A \uparrow B \cap A$

$$B \cap A = \left(\bigcup_n B_n \right) \cap A = \bigcup_n (B_n \cap A) \in \mathcal{D}$$

$\Rightarrow B \in \mathcal{D}'$

$\mathcal{D}'$ is a d-system, $\mathcal{A} \subseteq \mathcal{D}' \subseteq \mathcal{D}$, and $\mathcal{D}$ is the smallest d-system containing $\mathcal{A}$

$\Rightarrow \mathcal{D} = \mathcal{D}'$

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

Then $\mathcal{D} = \mathcal{D}' \Rightarrow \mathcal{A} \subseteq \mathcal{D}''$

As before $\mathcal{D}''$ is a d-system

$\Rightarrow \mathcal{D} = \mathcal{D}''$

$\Rightarrow \mathcal{D}$ is a π-system $\square$

1.5 Set functions and properties

Definitions Let $\mathcal{A}$ be any set of subsets of E containing the empty set $\emptyset$

A set function is a function $\mu : \mathcal{A} \to [0, \infty]$ with $\mu(\emptyset) = 0$

Let μ be a set function

μ is increasing if $\forall A, B \in \mathcal{A}$ with $A \subseteq B$, $\mu(A) \leq \mu(B)$

μ is additive if $\forall$ disjoint sets $A, B \in \mathcal{A}$ with $A \cup B \in \mathcal{A}$, $\mu(A \cup B) = \mu(A) + \mu(B)$

μ is countably additive if $\forall$ sequences of disjoint sets $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$ with $\bigcup_n A_n \in \mathcal{A}$, $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$

μ is countably subadditive if $\forall$ sequences $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$ with $\bigcup_n A_n \in \mathcal{A}$, $\mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$

1·6 Construction of measures

Definitions Let $\mathcal{A}$ be a set of subsets of E

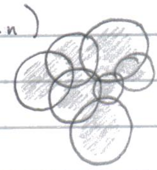
$\mathcal{A}$ is a ring on E if $\emptyset \in \mathcal{A}$
and $A, B \in \mathcal{A} \Rightarrow B \setminus A \in \mathcal{A}$
$A \cup B \in \mathcal{A}$

$\mathcal{A}$ is an algebra on E if $\emptyset \in \mathcal{A}$
and $A, B \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
$A \cup B \in \mathcal{A}$

Note $A \cap B = (A \cup B) \setminus ((A \setminus B) \cup (B \setminus A))$
$B \setminus A = (A \cup B^c)^c$
So σ-algebra $\Rightarrow$ algebra $\Rightarrow$ ring

Theorem 1·6·1 Carathéodory's extension theorem
Let $\mathcal{A}$ be a ring of subsets of E
let $\mu : \mathcal{A} \to [0, \infty]$ be a countably additive set function
Then μ extends to a measure on the σ-algebra generated by $\mathcal{A}$

Proof For any $B \subseteq E$, define the outer measure $\mu^*(B) = \inf \sum_n \mu(A_n)$
where the infimum is taken over all sequences $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$
such that $B \subseteq \bigcup_n A_n$
and is taken to be ∞ if there is no such sequence



Note that μ^* is increasing
and $\mu^*(\emptyset) = 0$
$A \subseteq E$ is μ^*-measurable if, $\forall\, B \subseteq E$,
$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$
Let $\mathcal{M} = \{A \subseteq E : A \text{ is } \mu^*\text{-measurable}\}$
We will show that $\mathcal{M}$ is a σ-algebra containing $\mathcal{A}$
and that μ^* restricts to a measure on $\mathcal{M}$, extending μ.
This will prove the theorem.

Step 1 μ^* is countably subadditive on 2^E

Suppose $(B_n : n \in \mathbb{N})$ is a sequence of subsets of E
$B \subseteq \bigcup_n B_n$
It is sufficient to consider the case where $\mu^*(B_n) < \infty \;\; \forall\, n$
Given $\varepsilon > 0$, for each $n \in \mathbb{N}$ $\exists$ sequences $(A_{nm} : m \in \mathbb{N})$ in $\mathcal{A}$
such that $B_n \subseteq \bigcup_m A_{nm}$
$$\mu^*(B_n) + \frac{\varepsilon}{2^n} \geq \sum_m \mu(A_{nm})$$
Now $B \subseteq \bigcup_n B_n \subseteq \bigcup_n \bigcup_m A_{nm}$
$$\Rightarrow \mu^*(B) \leq \sum_n \sum_m \mu(A_{nm}) \leq \sum_n \mu^*(B_n) + \varepsilon$$
ε was arbitrary $\Rightarrow \mu^*(B) \leq \sum_n \mu^*(B_n)$

Probability and Measure

Step 2 μ^* extends μ

$\mathcal{A}$ is a ring, μ is countably additive $\Rightarrow$ μ is countably subadditive and increasing

So, for $A \in \mathcal{A}$ and any sequence $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$ with $A \subseteq \bigcup_n A_n$,
$A = \bigcup_n (A \cap A_n) \Rightarrow \mu(A) \leq \sum_n \mu(A \cap A_n) \leq \sum_n \mu(A_n)$
Taking the infimum over all such sequences, $\mu(A) \leq \mu^*(A)$
On the other hand, clearly $\mu^*(A) \leq \mu(A)$ for $A \in \mathcal{A}$
$\Rightarrow \mu = \mu^*$ on $\mathcal{A}$

Step 3 $\mathcal{M}$ contains $\mathcal{A}$

Let $A \in \mathcal{A}$ and $B \subseteq E$.
By subadditivity of μ^*, $\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c)$
To show $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$, consider the case $\mu^*(B) < \infty$.
Given $\varepsilon > 0$, $\exists$ a sequence $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$
such that $B \subseteq \bigcup_n A_n$
$\mu^*(B) + \varepsilon \geq \sum_n \mu(A_n)$
Then $B \cap A \subseteq \bigcup_n (A_n \cap A)$ $\qquad A_n \cap A \in \mathcal{A}$
$B \cap A^c \subseteq \bigcup_n (A_n \cap A^c)$ $\qquad A_n \cap A^c \in \mathcal{A}$

$$\begin{aligned}
\Rightarrow \mu^*(B \cap A) + \mu^*(B \cap A^c) &\leq \sum_n \mu(A_n \cap A) + \sum_n \mu(A_n \cap A^c) \\
&= \sum_n (\mu(A_n \cap A) + \mu(A_n \cap A^c)) \\
&= \sum_n \mu(A_n) \qquad \mu \text{ is countably additive} \\
&\leq \mu^*(B) + \varepsilon
\end{aligned}$$

ε was arbitrary $\Rightarrow \mu^*(B \cap A) + \mu^*(B \cap A^c) \leq \mu^*(B)$
$\Rightarrow \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$
$\Rightarrow A \in \mathcal{M}$

Step 4 $\mathcal{M}$ is an algebra

Clearly $E \in \mathcal{M}$
$A \in \mathcal{M} \Rightarrow A^c \in \mathcal{M}$
Suppose $A_1, A_2 \in \mathcal{M}$ and $B \subseteq E$.

$$\begin{aligned}
\text{Then } \mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\
&= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c) \\
&= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1) \\
&\quad + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c) \qquad (A_i \cap A_j)^c = A_i^c \cup A_j^c \\
&= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap (A_1 \cap A_2)^c)
\end{aligned}$$

$\Rightarrow A_1 \cap A_2 \in \mathcal{M}$

Step 5 $\mathcal{M}$ is a σ-algebra
μ^* restricts to a measure on $\mathcal{M}$

$\mathcal{M}$ is an algebra, so it is sufficcent to show that for any sequence of disjoint sets $(A_n : n \in \mathbb{N})$ in $\mathcal{M}$, and for $A = \bigcup_n A_n$, we have $A \in \mathcal{M}$

$$\mu^*(A) = \sum_n \mu^*(A_n)$$

Take any $B \subseteq E$.

Then $\mu^*(B) = \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c)$

$= \mu^*(B \cap A_1) + \mu^*(B \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c)$

A_n are disjoint $\Rightarrow A_i \cap A_j^c = A_i$

$\dots = \sum_{i=1}^n \mu^*(B \cap A_i) + \mu^*(B \cap A_1^c \cap \dots \cap A_n^c)$

Note that $A \supseteq \bigcup_{i=1}^n A_i \Rightarrow A^c \subseteq \bigcap_{i=1}^n A_i^c$

$\Rightarrow \mu^*(B \cap A^c) \leq \mu^*(B \cap A_1^c \cap \dots \cap A_n^c) \quad \forall n$

Hence, on letting $n \to \infty$, $\mu^*(B) \geq \sum_{n=1}^\infty \mu^*(B \cap A_n) + \mu^*(B \cap A^c)$

$\geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$ by countable subadditivity

Also, by countable subadditivity, $\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c)$

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

$\Rightarrow A \in \mathcal{M}$

Taking $B = A \Rightarrow \mu^*(A) = \sum_{n=1}^\infty \mu^*(A_n)$ $\square$

Lecture 3

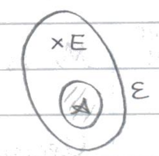
1.7 Uniqueness of measures

Theorem 1.7.1 Uniqueness of extension

Let μ_1, μ_2 be measures on $(E, \mathcal{E})$ with $\mu_1(E) = \mu_2(E) < \infty$

Suppose that $\mu_1 = \mu_2$ on $\mathcal{A}$, for some π-system $\mathcal{A}$ generating $\mathcal{E}$ $[\mu_1(A) = \mu_2(A) \ \forall A \in \mathcal{A}]$

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



Proof Consider $\mathcal{D} = \{A \in \mathcal{E} : \mu_1(A) = \mu_2(A)\}$

By the hypothesis, $E \in \mathcal{D}$ and $\mathcal{A} \subseteq \mathcal{D}$

For $A, B \in \mathcal{E}$ with $A \subseteq B$,

we have $\mu_1(A) + \mu_1(B \setminus A) = \mu_1(B) < \infty$

$\mu_2(A) + \mu_2(B \setminus A) = \mu_2(B) < \infty$

so if $A, B \in \mathcal{D}$

then $\mu_1(A) = \mu_2(A)$, $\mu_1(B) = \mu_2(B)$

$\Rightarrow \mu_1(B \setminus A) = \mu_2(B \setminus A)$

$\Rightarrow B \setminus A \in \mathcal{D}$

$E = \{0\}$
$\mathcal{E} = \{\emptyset, \{0\}\}$
$\mathcal{A} = \{\emptyset\}$
$\mu_k(\emptyset) = 0$
$\mu_k(\{0\}) = k$, $k = 1, 2$

If $(A_n : n \in \mathbb{N})$ is a sequence in $\mathcal{D}$ with $A_n \uparrow A$

then $\mu_1(A) = \lim_{n \to \infty} \mu_1(A_n)$

$= \lim_{n \to \infty} \mu_2(A_n)$

$= \mu_2(A)$

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

So by Dynkin's lemma, $\mathcal{D} = \mathcal{E}$ $\square$

Probability and Measure

1·8 Borel sets and measures

Definition Let E be a topological space
The Borel σ-algebra of E is the σ-algebra generated by the set of open sets in E

Notation $\mathcal{B}(E)$ is the Borel σ-algebra of E
$\mathcal{B}$ is the Borel σ-algebra of $\mathbb{R}$

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

A Radon measure on E is a Borel measure μ on E
such that $\mu(K) < \infty$ $\forall$ compact sets K

examples On $(\mathbb{R}, \mathcal{B})$, $\mu(A) = \text{card}(A)$ is a Borel measure

Consider $\mathcal{A} = \{A \in \mathcal{B} : \mu(A^c) < \infty\} \cup \{\emptyset\}$
Then $\mathbb{R} \in \mathcal{A}$
Note that $\mu = 2\mu$ on $\mathcal{A}$
So $\mu_1(E) < \infty$ is a necessary condition in Theorem 1·7·1

1·9 Probability measures, finite and σ-finite measures

Definitions Let $(E, \mathcal{E}, \mu)$ be a measure space
if $\mu(E) = 1$ then μ is a probability measure
$(E, \mathcal{E}, \mu)$ is a probability space

Notation $(\Omega, \mathcal{F}, \mathbb{P})$ is often used to denote a probability space

Definitions If $\mu(E) < \infty$ then μ is a finite measure

if $\exists$ a sequence of sets $(E_n : n \in \mathbb{N})$ in $\mathcal{E}$ with $\mu(E_n) < \infty$ $\forall n$
and $\bigcup_n E_n = E$

then μ is a σ-finite measure

example On $\mathcal{A}$, $\mu(A) = \text{card}(A)$ is a σ-finite measure if $\mathcal{A}$ is countable

1.10 Lebesgue measure

Theorem 1.10.1 $\exists$ a unique Borel measure μ on $\mathbb{R}$ such that, $\forall a, b \in \mathbb{R}$ with $a < b$, $\mu((a,b]) = b - a$

Definition The Lebesgue measure on $\mathbb{R}$ is $\mu((a,b]) = b - a$
A Lebesgue measure is an example of a Haar measure

Proof Existence
Consider the ring $\mathcal{A}$ of finite unions of disjoint intervals of the form
$A = (a_1, b_1] \cup \dots \cup (a_n, b_n]$, $\quad a_1 \le b_1 \le a_2 \le \dots \le b_n$
Note that $\mathcal{A}$ generates $\mathcal{B}$
Define for $A \in \mathcal{A}$, $\quad \mu(A) = \sum_{i=1}^{n} (b_i - a_i)$
Note that the presentation of A is not unique,
as $(a,b] \cup (b,c] = (a,c]$ whenever $a < b < c$
Nevertheless, μ is well-defined and additive.
We will show that μ is countably additive on $\mathcal{A}$,
from which the existence of a Borel measure extending μ
follows by Carathéodory's extension theorem
By additivity of μ, it suffices to show that
if $A \in \mathcal{A}$, and $(A_n : n \in \mathbb{N})$ is an increasing sequence in $\mathcal{A}$ with $A_n \uparrow A$,
then $\mu(A_n) \to \mu(A)$

Set $B_n = A \setminus A_n$
then $B_n \in \mathcal{A}$, and $B_n \downarrow \emptyset$
By additivity again, it suffices to show that $\mu(B_n) \to 0$
Suppose that $\exists \ \varepsilon > 0$, and a subsequence B_{n_j} of B_n, such that $\mu(B_{n_j}) \ge 2\varepsilon \ \forall j$
For each n_j, $\exists \ C_{n_j} \in \mathcal{A}$ with $\overline{C}_{n_j} \subseteq B_{n_j}$
and $\mu(B_{n_j} \setminus C_{n_j}) \le \frac{\varepsilon}{2^j}$
Then
$$\mu(B_{n_j} \setminus (C_{n_1} \cap \dots \cap C_{n_j})) \le \mu((B_{n_1} \setminus C_{n_1}) \cup \dots \cup (B_{n_j} \setminus C_{n_j})) \le \sum_{j \in \mathbb{N}} \frac{\varepsilon}{2^j} = \varepsilon$$

$$\mu(B_{n_j}) \ge 2\varepsilon \Rightarrow \mu(C_{n_1} \cap \dots \cap C_{n_j}) \ge \varepsilon$$
$$\Rightarrow C_{n_1} \cap \dots \cap C_{n_j} \ne \emptyset$$
$$\Rightarrow K_j = \overline{C}_{n_1} \cap \dots \cap \overline{C}_{n_j} \ne \emptyset$$
$(K_{n_j} : j \in \mathbb{N})$ is a decreasing sequence of bounded non-empty closed sets in $\mathbb{R}$, and $\mathbb{R}$ is complete
$$\Rightarrow \emptyset \ne \bigcap_j K_j \subseteq \bigcap_j B_{n_j} \quad \Rightarrow\Leftarrow \quad B_n \downarrow \emptyset$$

Lecture 4

Uniqueness
Let λ be any measure on $\mathcal{B}$ with $\lambda((a,b]) = b - a \quad \forall \ a < b$
Fix n and consider $\mu_n(A) = \mu((n, n+1] \cap A)$
$\lambda_n(A) = \lambda((n, n+1] \cap A)$

Then μ_n and λ_n are probability measures on $\mathcal{B}$
and $\mu_n = \lambda_n$ on the π-system of intervals of the form $(a,b]$ which generates $\mathcal{B}$.
So, by Theorem 1.7.1, $\mu_n = \lambda_n$ on $\mathcal{B}$.
Hence, $\forall \ A \in \mathcal{B}$, $\quad \mu(A) = \sum_n \mu_n(A) = \sum_n \lambda_n(A) = \lambda(A)$ $\square$

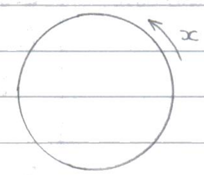
Probability and Measure

Definition A measure μ is translation invariant if $\mu(B+x) = \mu(B)$
where $B + x = \{ b + x : b \in B \}$

Let μ be a Lebesgue measure
Define $\mu_x(B) = \mu(B+x)$
Then $\mu_x((a,b]) = (b+x) - (a+x)$
$= b - a$

$\Rightarrow \mu_x = \mu$
$\Rightarrow$ the Lebesgue measure is translation invariant

Let μ be the Lebesgue measure restricted to $\mathcal{B}((0,1])$
so $B + x = \{ b + x \pmod 1 : b \in \mathcal{B}((0,1]) \}$
$\subseteq (0,1]$
Then μ is, again, a translation invariant



If we inspect the proof of Caratheodory's Extension Theorem, and consider its application in Theorem 1.10.1, we see we have constructed not only a Borel measure μ but also an extension of μ to the set of outer measurable sets $\mathcal{M}$. In this context, the extension is also called Lebesgue measure and $\mathcal{M}$ is called the Lebesgue σ-algebra. In fact, the Lebesgue σ-algebra can be identified as the set of all sets of the form $A \cup N$, where $A \in \mathcal{B}$ and $N \subseteq B$ for some $B \in \mathcal{B}$ with $\mu(B) = 0$.
Moreover $\mu(A \cup N) = \mu(A)$ in this case.

1.11 Existense of a non-Lebesgue-measurable subset of $\mathbb{R}$

For $x, y \in [0,1)$, write $x \sim y$ if $x - y \in \mathbb{Q}$
Then $\sim$ is an equivalence relation.
Using the Axiom of Choice, we can find a subset S of $[0,1)$ containing exactly one representative of each equivalence class.
We will show that S cannot be Lebesgue measurable.

Set $Q = \mathbb{Q} \cap [0,1)$
For each $q \in Q$, define $S + q = \{ s + q \pmod 1 : s \in S \}$
Then the sets $S + q$ are all disjoint
and $[0,1) = \bigcup_{q \in Q} (S+q)$
On the other hand, the Lebesgue σ-algebra and Lebesgue measure on $(0,1]$ are translation invariant for addition modulo 1
Hence, if S is Lebesgue measurable
then so is $S+q$, with $\mu(S+q) = \mu(S)$
But then $1 = \mu([0,1))$
$= \sum_{q \in Q} \mu(S+q)$
$= \sum_{q \in Q} \mu(S)$
which is impossible.

$\mu(S+q) = 0 \Rightarrow \sum_{q \in Q} \mu(S) = 0$
$\mu(S+q) > 0 \Rightarrow \sum_{q \in Q} \mu(S) = \infty$

Hence S is not Lebesgue measurable.

1.12 Independence

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ provides a model for an experiment whose outcome is subject to chance, according to the following interpretation:

Ω is the set of all possible outcomes
$\mathcal{F}$ is the set of observable sets of outcomes, or events,
$\mathbb{P}(A)$ is the probability of event A

Definition Let I be a countable set
A family $(A_i : i \in I)$ of events is independent
if, $\forall$ finite subsets $J \subseteq I$, $\quad \mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$

Lecture 5

A family $(\mathcal{A}_i : i \in I)$ of sub-σ-algebras of $\mathcal{F}$ is independent if the family $(A_i : i \in I)$ is independent whenever $A_i \in \mathcal{A}_i \ \forall i$

Theorem 1.12.1 Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be π-systems contained in $\mathcal{F}$
and suppose that $\quad \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$
whenever $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$.
Then $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are independent.

Proof Fix $A_1 \in \mathcal{A}_1$ and define for $A \in \mathcal{F}$, $\quad \mu(A) = \mathbb{P}(A_1 \cap A)$
$\nu(A) = \mathbb{P}(A_1)\mathbb{P}(A_2)$

Then μ and ν are measures which agree on the π-system $\mathcal{A}_2$
with $\mu(\Omega) = \nu(\Omega)$
$= \mathbb{P}(A_1)$
$< \infty$

So, by uniqueness of extension, $\forall A_2 \in \sigma(\mathcal{A}_2)$, $\quad \mathbb{P}(A_1 \cap A_2) = \mu(A_2)$
$= \nu(A_2)$
$= \mathbb{P}(A_1)\mathbb{P}(A_2)$

Now fix $A_2 \in \mathcal{A}_2$ and define for $A \in \mathcal{F}$, $\quad \mu'(A) = \mathbb{P}(A \cap A_2)$
$\nu'(A) = \mathbb{P}(A)\mathbb{P}(A_2)$

As before, $\forall A_1 \in \sigma(\mathcal{A}_1)$, $\quad \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$ $\quad \square$

1.13 Borel-Cantelli Lemmas

Let $(A_n : n \in \mathbb{N})$ be a sequence of events.
Then $\quad \limsup A_n = \bigcap_n \bigcup_{m \geq n} A_m$
$\liminf A_n = \bigcup_n \bigcap_{m \geq n} A_m$

Notation $\limsup A_n$ can be written $\{A_n \text{ infinitely often}\}$ or A_n i.o.
because $\omega \in \limsup A_n \Leftrightarrow \omega \in A_n$ for infinitely many A_n
$\liminf A_n$ can be written $\{A_n \text{ eventually}\}$ or A_n ev.

Probability and Measure

Lemma 1.13.1 First Borel - Cantelli Lemma

If $\sum_n \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(A_n \text{ happens infinitely often}) = 0$

Proof

$$\begin{aligned}\mathbb{P}(A_n \text{ happens infinitely often}) &= \mathbb{P}\Big(\bigcap_n \bigcup_{m \geq n} A_m\Big) \\ &\leq \mathbb{P}\Big(\bigcup_{m \geq n} A_m\Big) \\ &\leq \sum_{m \geq n} \mathbb{P}(A_m) \\ &\to 0 \quad \text{as } n \to \infty \qquad \square\end{aligned}$$

Remark This argument is valid whether or not $\mathbb{P}$ is a probability measure.

Lemma 1.13.2 Second Borel - Cantelli Lemma

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

If $\sum_n \mathbb{P}(A_n) = \infty$ then $\mathbb{P}(A_n \text{ happens infinitely often}) = 1$

Proof

The events $(A_n : n \in \mathbb{N})$ are independent

$\Rightarrow$ the events $(A_n^c : n \in \mathbb{N})$ are independent

$$\begin{aligned}\forall n \in \mathbb{N}, \quad \mathbb{P}\Big(\bigcap_{m \geq n} A_m^c\Big) &\leq \mathbb{P}\Big(\bigcap_{m=n}^{N} A_m^c\Big) \\ &= \prod_{m=n}^{N} (1 - \mathbb{P}(A_m)) \\ &\leq e^{-\sum_{m=n}^{N} \mathbb{P}(A_m)} \qquad 1 - a \leq e^{-a} \\ &\to 0 \quad \text{as } N \to \infty\end{aligned}$$

$\Rightarrow \mathbb{P}\Big(\bigcap_{m \geq n} A_m^c\Big) = 0 \quad \forall\, n \in \mathbb{N}$

$$\begin{aligned}\mathbb{P}(A_n \text{ happens infinitely often}) &= 1 - \mathbb{P}\Big(\Big(\bigcap_n \bigcup_{m \geq n} A_m\Big)^c\Big) \\ &= 1 - \mathbb{P}\Big(\bigcup_n \bigcap_{m \geq n} A_m^c\Big) \\ &= 1 \qquad \square\end{aligned}$$

Remark Without independence the result is false

eg $A_n = A \quad \forall\, n$

where $\mathbb{P}(A) \in (0, 1)$

§2 Measurable Functions and Random Variables

2.1 Measurable functions

Definitions

The inverse image of A by f is $f^{-1}(A) = \{x \in E : f(x) \in A\}$

Let $(E, \mathcal{E})$ and $(G, \mathcal{G})$ be measurable spaces
A function $f : E \to G$ is measurable if $f^{-1}(A) \in \mathcal{E}$ whenever $A \in \mathcal{G}$

If $(G, \mathcal{G}) = (\mathbb{R}, \mathcal{B})$
then f is a measurable function on E

If $(G, \mathcal{G}) = ([0, \infty], \mathcal{B}([0, \infty]))$
then f is a non-negative measurable function on E.
This terminology is convenient, but it has the consequence that some non-negative measurable functions are not (real-valued) measurable functions.

If E is a topological space, $\mathcal{E} = \mathcal{B}(E)$
$(G, \mathcal{G}) = (\mathbb{R}, \mathcal{B})$
then f is a Borel function

For any function $f : E \to G$, the inverse image preserves set operations

$$f^{-1}\left(\bigcup_i A_i\right) = \{x \in E : f(x) \in A_i \text{ for some } i\} = \bigcup_i f^{-1}(A_i)$$

$$f^{-1}(G \setminus A) = \{x \in E : f(x) \in G,\ f(x) \notin A\} = f^{-1}(G) \setminus f^{-1}(A)$$

Therefore, the set $\{f^{-1}(A) : A \in \mathcal{G}\}$ is a σ-algebra on E
and $\{A \subseteq G : f^{-1}(A) \in \mathcal{E}\}$ is a σ-algebra on G
In particular, if $\mathcal{G} = \sigma(\mathcal{A})$ and $f^{-1}(A) \in \mathcal{E}$ whenever $A \in \mathcal{A}$
then $\{A : f^{-1}(A) \in \mathcal{E}\}$ is a σ-algebra containing $\mathcal{A}$ and hence $\mathcal{G}$
$\Rightarrow f$ is measurable

Proposition

Suppose $f : E \to G$
and $\mathcal{A}$ is some collection of subsets of G
Then $\sigma(f^{-1}(\mathcal{G})) = f^{-1}(\sigma(\mathcal{G}))$

Suppose $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B})$
Then $\mathcal{B} = \sigma(\{(-\infty, y] : y \in \mathbb{R}\})$
so f is Borel measurable $\Leftrightarrow f^{-1}((-\infty, x]) \in \mathcal{E} \quad \forall x \in \mathbb{R}$
where $f^{-1}((-\infty, x]) = \{t \in E : f(t) \leq x\}$

Probability and Measure

Let E be a topological space
$f : E \to \mathbb{R}$ be a continuous function
Then the open sets $U \subseteq \mathbb{R}$ generate $\mathcal{B}$
and U is open in $\mathbb{R} \Rightarrow f^{-1}(U)$ is open in E
Hence any continuous function is measurable

Definition For $A \subseteq E$, the indicator function 1_A of A is the function $1_A : E \to \{0, 1\}$

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

Alternative Notations
I_A
χ_A
'characteristic function of A'

Note The indicator function of any measurable set is a measurable function

Given any family of functions $f_i : E \to G$, $i \in I$,
we can make them all measurable by taking $\mathcal{E} = \sigma(f_i^{-1}(A) : A \in \mathcal{G},\ i \in I)$
Then $\mathcal{E}$ is the σ-algebra generated by $(f_i : i \in I)$

Note The composition of measurable functions is measurable

Proposition 2.1.1 Let $(f_n : n \in \mathbb{N})$ be a sequence of non-negative measurable functions on E.
Then the following functions are measurable

$f_1 + f_2$
$f_1 f_2$
$\inf_n f_n$
$\sup_n f_n$
$\liminf_n f_n$
$\limsup_n f_n$

The same conclusion holds for real-valued measurable functions provided the limit functions are also real-valued

Lecture 6

Theorem 2.1.2 Monotone Class Theorem

Let $(E, \mathcal{E})$ be a measurable space
let $\mathcal{A}$ be a π-system generating $\mathcal{E}$
Let $\mathcal{V}$ be a vector space of bounded measurable functions $f : E \to \mathbb{R}$ such that
i. $1 \in \mathcal{V}$ and $1_A \in \mathcal{V}$ $\forall A \in \mathcal{A}$
ii. if $f_n \in \mathcal{V}$ $\forall n$ and f is bounded with $0 \le f_n \uparrow f$
then $f \in \mathcal{V}$
Then $\mathcal{V}$ contains every bounded measurable function

Proof Consider $\mathcal{D} = \{ A \in \mathcal{E} : 1_A \in \mathcal{V} \}$

Clearly $\mathcal{D} \subseteq \mathcal{E}$

$\mathcal{D}$ is a d-system:

$$A, B \in \mathcal{D} \text{ with } A \subseteq B \Rightarrow 1_A, 1_B \in \mathcal{V}$$
$$\Rightarrow 1_{A \setminus B} = 1_A - 1_B \in \mathcal{V} \quad (\mathcal{V} \text{ is a vector space})$$
$$\Rightarrow A \setminus B \in \mathcal{D}$$

$$A_n \in \mathcal{D},\ A_n \uparrow A \Rightarrow 0 \leq 1_{A_n} \uparrow 1_A \leq 1$$
$$\Rightarrow 1_A \in \mathcal{V} \quad (\text{by ii})$$
$$\Rightarrow A \in \mathcal{D}$$

$\mathcal{A}$ is a π-system, so by Dynkin's Lemma, $\mathcal{D} \supseteq \sigma(\mathcal{A}) = \mathcal{E}$

$\Rightarrow \mathcal{D} = \mathcal{E}$

Since $\mathcal{V}$ is a vector space, it contains all finite linear combinations of indicator functions of measurable sets.

If f is a bounded and non-negative measurable function

then the functions $f_n = 2^{-n} \lfloor 2^n f \rfloor \wedge 1$ are a finite linear combination of indicators of sets in $\mathcal{E}$, so belong to $\mathcal{V}$

and $|f(x) - f_n(x)| \leq 2^{-n} \ \forall x \Rightarrow 0 \leq f_n \uparrow f$
$\Rightarrow f \in \mathcal{V}$

Finally, if f is any bounded, measurable function

then $f^+ = \max\{f, 0\}$

$f^- = \max\{-f, 0\}$ are bounded, non-negative measurable functions

$\Rightarrow f^+, f^- \in \mathcal{V}$

$f = f^+ - f^- \Rightarrow f \in \mathcal{V}$ $\square$

Remark
$$f_n = \frac{1}{2^n} \lfloor 2^n f \rfloor = \sum_{j \geq 0} \frac{j}{2^n} 1_{A_{nj}}, \quad \text{where } A_{nj} = \{x \in E : \frac{j}{2^n} < f(x) \leq \frac{j+1}{2^n}\} = f^{-1}\left(\left(\frac{j}{2^n}, \frac{j+1}{2^n}\right]\right), \text{ which is in } \mathcal{E}$$

2·2 Image measures

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

let μ be a measure on $\mathcal{E}$

Then any measurable function $f: E \to G$ induces an image measure $\nu = \mu \circ f^{-1}$ on $\mathcal{G}$, given by

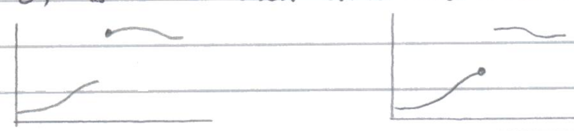
$$\nu(A) = \mu(f^{-1}(A))$$

where $A \in \mathcal{G}$

f^{-1} is the pre-image

We shall construct some new measures from Lebesgue measure in this way.

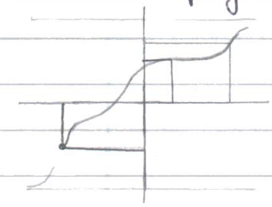
Definition f is right continuous at a if given ε, $\exists\ \delta > 0$ such that $a < x < a + \delta \Rightarrow |f(x) - f(a)| < \varepsilon$



right-continuous left-continuous

Probability and Measure

Lemma 2.2.1 Let $g : \mathbb{R} \to \mathbb{R}$ be non-constant, right-continuous and non-decreasing.
Set $g(\pm\infty) = \lim_{x \to \pm\infty} g(x)$
$I = (g(-\infty), g(\infty))$
Define $f : I \to \mathbb{R}$
by $f(x) = \inf \{ y \in \mathbb{R} : x \leq g(y) \}$ (f is the geometrised inverse of g)
Then f is left-continuous and non-decreasing.
Moreover, for $x \in I$ and $y \in \mathbb{R}$,
$$f(x) \leq y \iff x \leq g(y)$$



Proof Fix $x \in I$ and consider the set $J_x = \{ y \in \mathbb{R} : x \leq g(y) \}$
Note that $J_x \neq \emptyset$, $J_x \neq \mathbb{R}$
Since g is non-decreasing, if $y \in J_x$ and $y' \geq y$
then $y' \in J_x$
Since g is right-continuous, if $y_n \in J_x$ and $y_n \downarrow y$
then $x \leq g(y_n)$ $\forall n$ $\Rightarrow x \leq \lim_{n\to\infty} g(y_n) = g(y)$
$\Rightarrow y \in J_x$
Hence since g is non-constant, and $x \in I$, $J_x = [\inf J_x, \infty)$
$= [f(x), \infty)$

so $x \leq g(y) \iff f(x) \leq y$
f is non-decreasing: $x \leq x' \Rightarrow J_x \supseteq J_{x'}$
$\Rightarrow f(x) \leq f(x')$
f is left-continuous: $x_n \uparrow x \Rightarrow J_x = \bigcap_n J_{x_n}$
$\Rightarrow f(x_n) \to f(x)$ $\square$

Theorem 2.2.2 Let $g : \mathbb{R} \to \mathbb{R}$ be non-constant, right-continuous and non-decreasing.
Then $\exists$ a unique Radon measure dg on $\mathbb{R}$ such that, $\forall a, b \in \mathbb{R}$ with $a < b$,
$$dg((a,b]) = g(b) - g(a)$$
Moreover, we obtain in this way all non-zero Radon measures on $\mathbb{R}$.

The measure dg is called the Lebesgue-Stieltjes measure associated with distribution function g.

Proof Set $g(\pm\infty) = \lim_{x\to\pm\infty} g(x)$
$I = (g(-\infty), g(\infty))$
Define $f : I \to \mathbb{R}$
by $f(x) = \inf \{ y \in \mathbb{R} : x \leq g(y) \}$
Let μ be the Lebesgue measure on I
Then $f^{-1}((a,b]) = \{ x \in I : a < f(x) \leq b \}$
$= (g(a), g(b)]$
which is a Lebesgue measurable subset of $I = (g(-\infty), g(\infty))$
so f is Borel measurable
and the induced measure $dg = \mu \circ f^{-1}$ on $\mathbb{R}$ satisfies
$$dg((a,b]) = \mu(\{x : f(x) > a \text{ and } f(x) \leq b\})$$
$$= \mu((g(a), g(b)])$$
$$= g(b) - g(a)$$

The argument used for uniqueness of Lebesgue measure shows that there is at most one Borel measure with this property.

Finally, if ν is any Radon measure on $\mathbb{R}$,

define $g : \mathbb{R} \to \mathbb{R}$

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

then g is right-continuous and non-decreasing

and $\nu((a, b]) = g(b) - g(a)$ whenever $a < b$

so $\nu = dg$ by uniqueness $\square$

Lecture 7

2.3 Random variables

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

let $(E, \mathcal{E})$ be a measurable space

A measurable function $X : \Omega \to E$ is a random variable in E or an E-valued random variable

It has the interpretation of a quantity, or state, determined by chance.

Where no space E is mentioned, it is assumed that X takes values in $\mathbb{R}$.

The image measure $\mu_X = \mathbb{P} \circ X^{-1}$ is the law or distribution of X

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

For real-valued random variables, μ_X is uniquely determined by its values on the π-system of intervals $((-\infty, x] : x \in \mathbb{R})$, given by

$$F_X(x) = \mu_X((-\infty, x])$$
$$= \mathbb{P}(X \leq x)$$

F_X is the distribution function of X

A distribution function is a function $F : \mathbb{R} \to [0, 1]$ such that F is increasing and right-continuous with

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

Let $\Omega = (0, 1)$

$\mathcal{F}$ is the Borel σ-algebra on Ω; $\mathcal{F} = \mathcal{B}((0,1))$

$\mathbb{P}$ is the restriction of Lebesgue measure on $\mathcal{F}$

Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space

Let F be any distribution function

Define $X : \Omega \to \mathbb{R}$

by $X(\omega) = \inf\{x : \omega \leq F(x)\}$

Then, by Lemma 2.2.1, X is a random variable and $X(\omega) \leq x \iff \omega \leq F(x)$

$\omega \in \Omega$, $x \in \mathbb{R}$

Probability and Measure

So $F_X(x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \le x\})$
$= \mathbb{P}(X \le x)$
$= \mathbb{P}((0, F(x)])$
$= F(x)$

X is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution function F.
Thus every distribution function is the distribution function of a random variable.

A countable family of random variables $(X_i : i \in I)$
is independent if the family of σ-algebras $(\sigma(X_i) : i \in I)$ is independent
$\sigma(X_i) = \{\{X_i \in A\} : A \in \mathcal{E}_i\}$
so each family of events $(\{X_i \in A\} : i \in I)$ is independent
where $A_i \in \mathcal{E}_i$

For a sequence $(X_n : n \in \mathbb{N})$ of real-valued random variables,
this is equivalent to the condition

$$\mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \ldots \mathbb{P}(X_n \le x_n) \quad \forall\, x_1, \ldots, x_n \in \mathbb{R},\ \forall\, n$$

A sequence of random variables $(X_n : n \ge 0)$ is often regarded as a process evolving in time.
The σ-algebra generated by $X_0, \ldots, X_n$
$\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$
contains those events depending (measurably) on $X_0, \ldots, X_n$
and represents what is known about the process by time n.

2.4 Rademacher functions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space such that
$\Omega = (0, 1)$
$\mathcal{F} = \mathcal{B}((0,1))$
$\mathbb{P}$ is the restriction of Lebesgue measure to $\mathcal{F}$

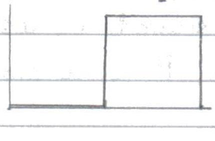
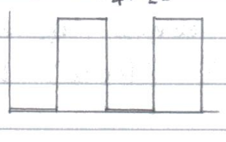
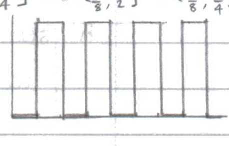
Provided that we forbid infinite sequences of 0's, each $\omega \in \Omega$ has a unique binary expansion $\omega = 0 \cdot \omega_1 \omega_2 \omega_3 \ldots$ (eg $0 \cdot 0 1111 \ldots = 0 \cdot 10000 \ldots$)

$$= \sum_{k=1}^{\infty} \omega_k 2^{-k}$$

Define the Rademacher functions as random variables $R_n : \Omega \to \{0, 1\}$
$R_n(\omega) = \omega_n$

$(R_n : n \in \mathbb{N})$ is a sequence of $\{0,1\}$-valued random variables on Ω

Then $R_1 = 1_{(\frac{1}{2}, 1]}$ $\quad R_2 = 1_{(\frac{1}{4}, \frac{1}{2}]} + 1_{(\frac{3}{4}, 1]}$ $\quad R_3 = 1_{(\frac{1}{8}, \frac{1}{4}]} + 1_{(\frac{3}{8}, \frac{1}{2}]} + 1_{(\frac{5}{8}, \frac{3}{4}]} + 1_{(\frac{7}{8}, 1]}$

etc

The random variables $R_1, R_2, \ldots$ are independent and Bernoulli
ie $\mathbb{P}(R_n = 0) = \mathbb{P}(R_n = 1) = \frac{1}{2}$

The strong law of large numbers (proved in §10) applies here to show that

$$\mathbb{P}\left(\left\{\omega \in (0,1) : \frac{|\{k \le n : \omega_k = 1\}|}{n} \to \frac{1}{2}\right\}\right) = \mathbb{P}\left(\frac{R_1 + \dots + R_n}{n} \to \frac{1}{2}\right)$$
$$= 1$$

This is called Borel's normal number theorem:
almost every point in $(0,1)$ is normal,
ie has 'equal' proportions of 0's and 1's in its binary expansion.

We now use a trick involving the Rademacher functions to construct on $\Omega = (0,1)$,
not just one random variable,
but an infinite sequence of independent random variables with given distribution functions

Proposition 2·4·1 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space of Lebesgue measure on the Borel subsets of $(0,1)$.
Let $(F_n : n \in \mathbb{N})$ be a sequence of distribution functions.
Then $\exists$ a sequence $(X_n : n \in \mathbb{N})$ of independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$
such that X_n has distribution function $F_{X_n} = F_n \quad \forall n$

Proof

Choose a bijection $m : \mathbb{N}^2 \to \mathbb{N}$
and set $Y_{k,n} = R_{m(k,n)}$, where R_m is the m^{th} Rademacher function.
Set $Y_n = \sum_{k=1}^{\infty} \frac{1}{2^k} Y_{k,n}$
Then $Y_1, Y_2, \dots$ are independent and, $\forall n$, for $\frac{i}{2^k} = 0 \cdot y_1 \dots y_k$, we have

$$\mathbb{P}\left(\frac{i}{2^k} < Y_n \le \frac{i+1}{2^k}\right) = \mathbb{P}(Y_{1,n} = y_1, \dots, Y_{k,n} = y_k)$$
$$= \frac{1}{2^k}$$

so $\mathbb{P}(Y_n \le x) = x \qquad \forall x \in [0,1]$
Set $G_n(y) = \inf\{x : y \le F_n(x)\}$
then, by Lemma 2·2·1, G_n is Borel and $G_n(y) \le x \iff y \le F_n(x)$
So, if we set $X_n = G_n(Y_n)$
then $X_1, X_2, \dots$ are independent random variables on Ω
and

$$\begin{aligned} F_{X_n}(x) &= \mathbb{P}(X_n \le x) \\ &= \mathbb{P}(G_n(Y_n) \le x) \\ &= \mathbb{P}(Y_n \le F_n(x)) \\ &= F_n(x) \end{aligned}$$

$\square$

Lecture 8

2·5 Convergence of measurable functions and random variables

Let $(E, \mathcal{E}, \mu)$ be a measure space.
A set $A \in \mathcal{E}$ is sometimes defined by a property shared by its elements
ie $A = \{x \in E : P(x) \text{ holds}\}$, where P is a property

A property holds almost everywhere (or a.e.) if $\mu(A^c) = 0$

A property holds almost surely (or a.s.) if $\mu(A^c) = 0$
and $(E, \mathcal{E}, \mu)$ is a probability space
ie $\mu(E) = 1$

Probability and Measure

For a sequence of measurable functions $(f_n : n \in \mathbb{N})$, f a measurable function,
f_n converges to f almost everywhere if $\mu(\{x \in E : f_n(x) \not\to f(x)\}) = 0$
(or almost surely if $\mu(E) = 1$)

f_n converges to f in measure (or in probability if $\mu(E) = 1$)
if $\mu(\{x \in E : |f_n(x) - f(x)| > \varepsilon\}) \to 0$ $\forall\, \varepsilon > 0$

For a sequence of (real-valued) random variables $(X_n : n \in \mathbb{N})$
X_n converges to X in distribution if $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$
at all points $x \in \mathbb{R}$ where F_X is continuous.

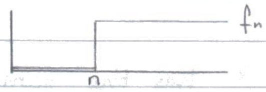
Note that this does not require the random variables to be defined on the same probability space.

examples 1. $E = \mathbb{R}$

$f_n(x) = 1_{\{x \geq n\}}$

$f(x) = 0$

$f_n \to f$ almost everywhere



2. $E = (0,1)$

$f(x) = 0$

$f_n \to f$ in measure

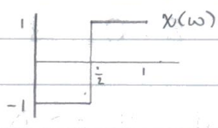
etc

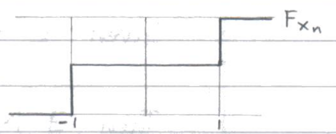
3. $E = (0,1)$

$X_n(\omega) = (-1)^n \chi(\omega)$

$X(\omega) = \chi(\omega)$

$X_n \to X$ in distribution





Theorem 2·5·1 Let $(f_n : n \in \mathbb{N})$ be a sequence of measurable functions

i. Suppose $\mu(E) < \infty$
If $f_n \to f$ almost everywhere, then $f_n \to f$ in measure

ii. If $f_n \to f$ in measure, then $f_{n_k} \to f$ almost everywhere for some subsequence (n_k)

Proof Reduce to $f_n - f$

i. Suppose $f_n \to 0$ almost everywhere, and $\mu(E) < \infty$.

Given $\varepsilon > 0$,

$$
\begin{aligned}
\mu(|f_n| \leq \varepsilon) &\geq \mu\Big(\bigcap_{m \geq n} \{|f_m| \leq \varepsilon\}\Big) \\
&\uparrow \mu\Big(\bigcup_n \bigcap_{m \geq n} \{|f_m| \leq \varepsilon\}\Big) \\
&= \mu(|f_n| \leq \varepsilon \text{ eventually}) \\
&\geq \mu(f_n \to 0) \\
&= \mu(E)
\end{aligned}
$$

$$\mu(|f_n| > \varepsilon) + \mu(|f_n| \leq \varepsilon) = \mu(E) < \infty$$

$$\mu(|f_n| \leq \varepsilon) \uparrow \mu(E)$$

$$\Rightarrow \mu(|f_n| > \varepsilon) \to 0$$

$f_n \to 0$ in measure

ii. Suppose $f_n \to 0$ in measure

Then $\exists$ a subsequence (n_k) such that $\mu(|f_{n_k}| > \frac{1}{k}) \leq \frac{1}{k^2}$

$\sum_k \mu(|f_{n_k}| > \frac{1}{k}) \leq \frac{\pi^2}{6} < \infty$

By the first Borel-Cantelli lemma, $\mu(|f_{n_k}| > \frac{1}{k} \text{ infinitely often}) = 0$

$\mu(|f_{n_k}| \not\to 0) = 0$

$\Rightarrow f_{n_k} \to 0$ almost everywhere $\square$

Theorem 2·5·2 Let X and $(X_n : n \in \mathbb{N})$ be real-valued random variables

i. If X and $(X_n : n \in \mathbb{N})$ are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $X_n \to X$ in probability

then $X_n \to X$ in distribution

ii. If $X_n \to X$ in distribution

then $\exists$ random variables $\tilde{X}$ and $(\tilde{X}_n : n \in \mathbb{N})$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$

such that $\tilde{X}$ has the same distribution as X ie $\mu_{\tilde{X}} = \mu_X$

$\tilde{X}_n$ has the same distribution as X_n $\forall n$ $\mu_{\tilde{X}_n} = \mu_{X_n}$

and $X_n \to X$ almost surely

Proof Let S be the subset of $\mathbb{R}$ where F_X is continuous

i. Suppose $X_n \to X$ in probability

Given $x \in S$ and $\varepsilon > 0$, $\exists\ \delta > 0$ such that $F_X(x - \delta) \geq F_X(x) - \frac{\varepsilon}{2}$

and $F_X(x + \delta) \leq F_X(x) + \frac{\varepsilon}{2}$

Then $\exists N$ such that, $\forall n \geq N$, $\mathbb{P}(|X_n - X| > \delta) \leq \frac{\varepsilon}{2}$

$\Rightarrow F_{X_n}(x) \leq \mathbb{P}(X \leq x + \delta) + \mathbb{P}(|X_n - X| > \delta)$

$\leq F_X(x) + \varepsilon$

and $F_{X_n}(x) \geq \mathbb{P}(X \leq x - \delta) \cancel{+} - \mathbb{P}(|X_n - X| > \delta)$

$\geq F_X(x) - \varepsilon$

ε was arbitrary $\Rightarrow F_{X_n}(x) \to F_X(x)$

ii. Suppose $X_n \to X$ in distribution

Take the probability space $(\Omega, \mathcal{F}, \mathbb{P})$

where $\Omega = (0,1)$

$\mathcal{F} = \mathcal{B}((0,1))$

$\mathbb{P}$ is the Lebesgue measure

For $\omega \in (0,1)$, define $\tilde{X}_n(\omega) = \inf\{x \in \mathbb{R} : \omega \leq F_{X_n}(x)\}$

$\tilde{X}(\omega) = \inf\{x \in \mathbb{R} : \omega \leq F_X(x)\}$

$\omega \leq F_X(x) \Leftrightarrow \tilde{X}(\omega) \leq x$

Then $\tilde{X}$ has the same distribution as X ie $\mu_{\tilde{X}} = \mu_X$

$\tilde{X}_n$ has the same distribution as X_n $\forall n$ $\mu_{\tilde{X}_n} = \mu_{X_n}$

Let Ω_0 be the subset of $(0,1)$ where $\tilde{X}$ is continuous

$\tilde{X}$ is non-decreasing $\Rightarrow (0,1) \setminus \Omega_0$ is countable

$\Rightarrow \mathbb{P}(\Omega_0) = 1$

F_X is non-decreasing $\Rightarrow \mathbb{R} \setminus S$ is countable

$\Rightarrow S$ is dense

Probability and Measure

Given $\omega \in \Omega_0$ and $\varepsilon > 0$,

$\exists\ x^-, x^+ \in S$ with $x^- < \tilde{X}(\omega) < x^+$
and $x^+ - x^- < \varepsilon$

and, by right-continuity, $\exists\ \omega^+ \in (\omega, 1)$ such that $\tilde{X}(\omega^+) \leq x^+$

Then $F_X(x^-) < \omega < \omega^+ \leq F_X(x^+)$

So $\exists\ N \in \mathbb{N}$ such that $n \geq N \Rightarrow F_{X_n}(x^-) < \omega \leq F_{X_n}(x^+)$

$\Rightarrow x^- < \tilde{X}_n(\omega) \leq x^+$

$\Rightarrow |\tilde{X}_n(\omega) - \tilde{X}(\omega)| < \varepsilon$ $\square$

Lecture 13

2.6 Tail events

Let $(X_n : n \in \mathbb{N})$ be a sequence of random variables

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

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

Then $\mathcal{T}$ is a σ-algebra, the tail σ-algebra of $(X_n : n \in \mathbb{N})$

$\mathcal{T}$ contains the events which depend only on the limiting behaviour of the sequence.

Theorem 2.6.1 Kolmogorov's zero-one law

Suppose $(X_n : n \in \mathbb{N})$ is a sequence of random variables in $\mathbb{R}$

Then the tail σ-algebra $\mathcal{T}$ of $(X_n : n \in \mathbb{N})$ is trivial

ie $\mathbb{P}(A) \in \{0, 1\}$ $\forall A \in \mathcal{T}$

Moreover, if X is a $\mathcal{T}$-measurable random variable

then $X = c$ almost surely for some constant $c \in \mathbb{R}$

Proof Set $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$

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

Consider $A = \{X_1 \leq x_1, \dots, X_n \leq x_n\}$ where $x_1, \dots, x_n \in \mathbb{R}$

$B = \{X_{n+1} \leq x_{n+1}, \dots, X_{n+m} \leq x_{n+m}\}$ $x_{n+1}, \dots, x_{n+m} \in \mathbb{R}$, $m \in \mathbb{N}$

Then $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ $\forall$ such A and B, by independence.

Now the set of such A is a π-system generating $\mathcal{F}_n$

and the set of such B is a π-system generating $\mathcal{T}_n$

$\Rightarrow \mathcal{F}_n$ and $\mathcal{T}_n$ are independent (by Theorem 1.12.1)

But $\bigcup_n \mathcal{F}_n$ is a π-system generating $\mathcal{F}_\infty$

$\Rightarrow \mathcal{F}_\infty$ and $\mathcal{T}$ are independent (by Theorem 1.12.1 again)

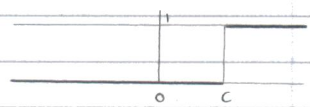
So for $A \in \mathcal{T} \subseteq \mathcal{F}_\infty$, $\mathbb{P}(A) = \mathbb{P}(A \cap A)$

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

$\Rightarrow \mathbb{P}(A) \in \{0, 1\}$

If X is a $\mathcal{T}$-measurable random variable

then $F_X(x) = \mathbb{P}(X \leq x) \in \{0, 1\}$



$\Rightarrow \mathbb{P}(X = c) = 1$

$X = c$ almost surely

where $c = \inf\{x \in \mathbb{R} : F_X(x) = 1\}$ $\square$

2.7 Large values in sequences of independent identically distributed random variables

Let $(X_n : n \in \mathbb{N})$ be a sequence of (real) independent identically distributed random variables each with distribution function F_X.

Suppose $F_X(x) < 1 \quad \forall x \in \mathbb{R}$

Then, almost surely, the sequence $(X_n : n \in \mathbb{N})$ is unbounded above

so $\limsup_n X_n = \infty$

A way to describe the occurence of large values in a sequence is to find a function $g : \mathbb{N} \to (0, \infty)$ such that, almost surely, $\limsup_n \left(\frac{X_n}{g(n)} \right) = 1$

example

Suppose $X_n \sim \exp(1)$

ie $\mathbb{P}(X_n \geq x) = e^{-x}$

Fix $\alpha > 0$ and consider the independent events $A_n = \{ X_n \geq \alpha \log n \}$

Then $\mathbb{P}(A_n) = \mathbb{P}(X_n \geq \alpha \log n)$

$= e^{-\alpha \log n}$

$= n^{-\alpha}$

so the series $\sum_n \mathbb{P}(A_n)$ converges $\Leftrightarrow \alpha > 1$

By Borel-Cantelli, $\mathbb{P}(A_n \text{ happens infinitely often}) = \begin{cases} 1 & \text{if } \alpha \leq 1 \\ 0 & \text{if } \alpha > 1 \end{cases}$

$\mathbb{P}(A_n \text{ happens infinitely often}) = 1$ if $\alpha \leq 1$

$\Rightarrow \limsup_n \left(\frac{X_n}{\log n} \right) \geq 1$ almost surely

$\mathbb{P}(A_n \text{ happens infinitely often}) = 0$ if $\alpha > 1$

$\Rightarrow \mathbb{P}\left(\frac{X_n}{\log n} \geq 1 + \varepsilon \text{ infinitely often} \right) = 0$

$\Rightarrow \forall m, \frac{X_n}{\log n} \leq 1 + \frac{1}{m}$ eventually almost surely

$\Rightarrow \limsup_n \left(\frac{X_n}{\log n} \right) \leq 1$ almost surely

Hence $\limsup_n \left(\frac{X_n}{\log n} \right) = 1$ almost surely

Probability and Measure

§3 Integration

3.1 Definition of the integral and basic properties

Let $(E, \mathcal{E}, \mu)$ be a measure space.
We shall define for non-negative measurable functions on E,
and (under a natural condition) for (real-valued) measurable functions f on E,
the integral of f.

Notation — The integral of f is $\mu(f) = \int_E f \, d\mu$
$= \int_E f(x) \, \mu(dx)$

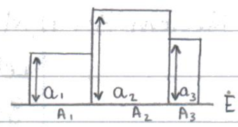
When $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ and μ is the Lebesgue measure,
$\mu(f) = \int_{\mathbb{R}} f(x) \, dx$

For $I \in \{(a,b), (a,b], [a,b), [a,b]\}$,
$\int_I f(x)\,dx = \int_{(a,b)} f(x)\,dx = \int_{\mathbb{R}} f \mathbb{1}_I(x)\,dx$
$= \int_a^b f(x)\,dx$

Note that the sets $\{a\}$, $\{b\}$ have measure 0

For a random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$,
the integral is the expectation of X, written $\mathbb{E}(X)$

Definition — Let $(E, \mathcal{E}, \mu)$ be a measure space
A simple function f is a function of the form $f = \sum_{k=1}^{m} a_k \mathbb{1}_{A_k}$
where $m \in \mathbb{N}$, $a_k \in [0, \infty)$, $A_k \in \mathcal{E}$ $\forall k$



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

Convention — $0 \cdot \infty = 0$

The representation of f is not unique, but $\mu(f)$ is well-defined

For simple functions f, g, and constants $\alpha, \beta \in [0, \infty)$,
i. $\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)$
ii. $f \le g \Rightarrow \mu(f) \le \mu(g)$
iii. $\mu(f) = 0 \Leftrightarrow f = 0$ almost everywhere

Definition — For a non-negative measurable function $f : E \to [0, \infty]$,
$\mu(f) = \sup\{\mu(g) : g \text{ is simple}, g \le f\}$

By ii, this is consistent with the definition for simple functions

Note — Property ii holds for non-negative measurable functions

Notation — $a \vee b = \max\{a, b\}$
$a \wedge b = \min\{a, b\}$

Lecture 10

Definition An integrable function is a measurable function $f : E \to \mathbb{R}$ such that $\mu(|f|) < \infty$

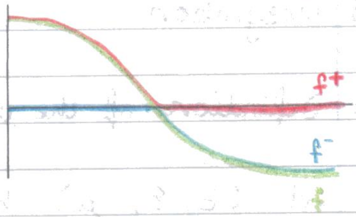
Let $f^+(x) = \max\{f(x), 0\}$

$f^-(x) = \max\{-f(x), 0\}$

Then $f = f^+ - f^-$

$|f| = f^+ + f^-$

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



Note f is measurable $\Rightarrow$ $f^+, f^-, |f|$ are also measurable

Also $f^{\pm} \leq |f|$ $\Rightarrow$ $\mu(f^{\pm}) \leq \mu(|f|) < \infty$

so f is integrable $\Rightarrow$ f^+, f^- are also integrable

$|\mu(f)| \leq \mu(f^+) + \mu(f^-)$

$= \mu(|f|)$

Convention When f is not integrable, but one of $\mu(f^+)$ or $\mu(f^-)$ is finite, we sometimes still define $\mu(f) = \mu(f^+) - \mu(f^-)$

In such cases the integral takes the value ∞ or $-\infty$

Lecture 9

Notation Let $x \in [0, \infty]$, $(x_n : n \in \mathbb{N})$ be a sequence in $[0, \infty]$

Then $x_n \uparrow x$ means $x_n \leq x_{n+1}$ $\forall n$, and $x_n \to x$ as $n \to \infty$

Let f, $(f_n : n \in \mathbb{N})$ be non-negative measurable functions on E

Then $f_n \uparrow f$ means $f_n(x) \uparrow f(x)$ $\forall x \in E$

Theorem 3.1.1 Monotone convergence

Let f, $(f_n : n \in \mathbb{N})$ be non-negative measurable functions

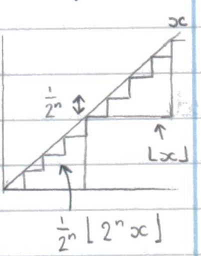
Suppose $f_n \uparrow f$

Then $\mu(f_n) \uparrow \mu(f)$

Proof Set $M = \sup_n \mu(f_n)$

We know that $\mu(f_n) \leq \mu(f_{n+1}) \uparrow M \leq \mu(f)$

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



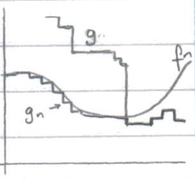
so it is sufficient to show that $\mu(g) \leq M$ $\forall$ simple functions $g \leq f$

Let $g = \sum_{k=1}^{m} a_k 1_{A_k}$ be a simple function, $g \leq f$

wlog $a_k > 0$ $\forall k$

and the sets $A_k \in \mathcal{E}$ are all disjoint

Set $g_n(x) = \min\{\frac{1}{2^n}\lfloor 2^n f_n(x) \rfloor, g(x)\}$



Then g_n is simple and $g_n \leq f_n$ $\forall n$

Fix $0 < \varepsilon < 1$ and consider the measurable sets $A_k(n) = \{1_{A_k} g_n \geq (1-\varepsilon) a_k\}$

Now $g_n \uparrow g$ $\Rightarrow$ $A_k(n) \uparrow A_k$

$\Rightarrow$ $\mu(A_k(n)) \uparrow \mu(A_k)$ by countable additivity

Note $1_{A_k} \cdot g_n \geq (1-\varepsilon) a_k 1_{A_k(n)}$

$\Rightarrow \mu(1_{A_k} \cdot g_n) \geq (1-\varepsilon) a_k \mu(A_k(n))$ $\quad$ g_n is simple

Then, since g_n is simple, $M \geq \mu(f_n)$

$\geq \mu(g_n)$

$= \sum_{k=1}^{m} \mu(1_{A_k} g_n)$ $\quad$ $g_n = \sum_{k=1}^{m} 1_{A_k} g_n$

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

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

$= (1-\varepsilon)\mu(g)$

ε was arbitrary $\Rightarrow$ $M \geq \mu(g)$ $\square$

Probability and Measure

Theorem 3·1·2 For non-negative measurable functions f, g, and constants $\alpha, \beta \in [0, \infty)$,
i. $\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)$
ii. $f \geq g \Rightarrow \mu(f) \leq \mu(g)$
iii. $\mu(f) = 0 \Leftrightarrow f = 0$ almost everywhere

Proof Set $f_n = \min\{\frac{1}{2^n}\lfloor 2^n f \rfloor, n\}$
$g_n = \min\{\frac{1}{2^n}\lfloor 2^n g \rfloor, n\}$
Then $f_n \uparrow f$, $g_n \uparrow g$, $\alpha f_n + \beta g_n \uparrow \alpha f + \beta g$
So by monotone convergence, $\mu(f_n) \uparrow \mu(f)$
$\mu(g_n) \uparrow \mu(g)$
$\mu(\alpha f_n + \beta g_n) \uparrow \mu(\alpha f + \beta g)$

i. $\mu(\alpha f_n + \beta g_n) = \alpha \mu(f_n) + \beta \mu(g_n) \uparrow \alpha \mu(f) + \beta \mu(g)$

ii. Clear from the definition of the integral

iii. $\mu(f) = 0 \Leftrightarrow \mu(f_n) = 0 \quad \forall n$
$f_n = 0$ almost everywhere $\forall n$
$f_n = 0 \quad \forall n$ almost everywhere
$f = 0$ almost everywhere $\square$

Lecture 10

Theorem 3·1·3 For integrable functions f, g, and constants $\alpha, \beta \in \mathbb{R}$,
i. $\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)$
ii. $f \leq g \Rightarrow \mu(f) \leq \mu(g)$
iii. $f = 0$ almost everywhere $\Rightarrow \mu(f) = 0$

Proof i. Note that $(-f)^+ = f^-$, $(-f)^- = f^+$
so $-f$ is integrable, with $\mu(-f) = -\mu(f)$
For $\alpha \geq 0$, $(\alpha f)^+ = \alpha f^+$, $(\alpha f)^- = \alpha f^-$
so αf is integrable, with $\mu(\alpha f) = \mu(\alpha f^+) - \mu(\alpha f^-)$
$= \alpha \mu(f^+) - \alpha \mu(f^-)$
$= \alpha \mu(f)$

Set $h = f + g$
Then h is measurable, and $|h| \leq |f| + |g| \Rightarrow \mu(|h|) \leq \mu(|f|) + \mu(|g|) < \infty$
$\Rightarrow h$ is integrable
Note that $h^+ - h^- = f^+ - f^- + g^+ - g^-$
$\Rightarrow h^+ + f^- + g^- = h^- + f^+ + g^+$
$\Rightarrow \mu(h^+) + \mu(f^-) + \mu(g^-) = \mu(h^-) + \mu(f^+) + \mu(g^+)$
$\Rightarrow \mu(h) = \mu(f) + \mu(g)$

ii. $f \leq g \Rightarrow g - f \geq 0$
$\mu(g) - \mu(f) = \mu(g - f)$ by i
≥ 0

iii. $f = 0$ almost everywhere $\Rightarrow f^+, f^- = 0$ almost everywhere
$\Rightarrow \mu(f^+), \mu(f^-) = 0$
$\Rightarrow \mu(f) = 0$ $\square$

Note — We lost the reverse implication in iii

Proposition 3.1.4 Let $\mathcal{A}$ be a π-system containing E and generating $\mathcal{E}$.
If f is an integrable function and $\mu(f 1_A) = 0 \quad \forall A \in \mathcal{A}$
then $f = 0$ almost everywhere

Some minor variations of the monotone convergence theorem

Proposition 3.1.5 Let $(f_n : n \in \mathbb{N})$ be a sequence of non-negative measurable functions
f a non-negative measurable function such that $f_n \uparrow f$ almost everywhere
Then $\mu(f_n) \uparrow \mu(f)$

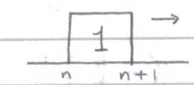
Proposition 3.1.6 Let $(g_n : n \in \mathbb{N})$ be a sequence of non-negative measurable functions
Then $\sum_{n=1}^{\infty} \mu(g_n) = \mu\left(\sum_{n=1}^{\infty} g_n\right)$ — This is just monotone convergence with $f_n = g_1 + \dots + g_n$

Monotone convergence is the counterpart for the integration of functions to the countable additivity property of the measure on sets

3.2 Integrals and limits

Lemma 3.2.1 Fatou's Lemma
Let $(f_n : n \in \mathbb{N})$ be a sequence of non-negative measurable functions
Then $\mu\left(\liminf_{n \to \infty} f_n\right) \leq \liminf_{n \to \infty} \mu(f_n)$

example
$f_n = 1_{[n, n+1]}$
Then $\mu\left(\liminf_{n \to \infty} f_n\right) = 0$
$\liminf_{n \to \infty} \mu(f_n) = 1$



Proof
Set $g_n = \inf_{m \geq n} f_m$
$g = \liminf_{n \to \infty} f_n$
Then $g_n \leq f_m \quad \forall m \geq n$
$\Rightarrow \mu(g_n) \leq \inf_{m \geq n} \mu(f_m)$
Also $g_n \uparrow g$, so by monotone convergence $\mu(g_n) \uparrow \mu(g)$
Hence $\mu\left(\liminf_{n \to \infty} f_n\right) \leq \liminf_{n \to \infty} \mu(f_n)$ $\square$

Theorem 3.2.2 Dominated Convergence
Let f be a measurable function (real-valued)
$(f_n : n \in \mathbb{N})$ be a sequence of measurable functions (real-valued)
Suppose $f_n(x) \to f(x)$ as $n \to \infty \quad \forall x \in E$
and $\exists$ an integrable function g such that $|f_n| \leq g \quad \forall n$
Then f, f_n are integrable
and $\mu(f_n) \to \mu(f)$ as $n \to \infty$

Probability and Measure

Proof f, f_n are measurable
and $|f|, |f_n| \leq g \Rightarrow \mu(|f|), \mu(|f_n|) \leq \mu(g) < \infty$
$\Rightarrow f, f_n$ are integrable

Note that $0 \leq g \pm f_n \to g \pm f$
so $\liminf (g \pm f_n) = g \pm f$

Now
$$\begin{aligned}
\mu(g) + \mu(f) &= \mu(g+f) \\
&= \mu(\liminf (g + f_n)) \\
&\leq \liminf \mu(g + f_n) \qquad \text{by Fatou's Lemma} \\
&= \liminf (\mu(g) + \mu(f_n)) \\
&= \mu(g) + \liminf \mu(f_n)
\end{aligned}$$
$\mu(g) < \infty \Rightarrow \mu(f) \leq \liminf \mu(f_n)$

Similarly
$$\begin{aligned}
\mu(g) - \mu(f) &= \mu(g - f) \\
&= \mu(\liminf (g - f_n)) \\
&\leq \liminf \mu(g - f_n)) \qquad \text{by Fatou's Lemma} \\
&= \mu(g) - \limsup \mu(f_n) \qquad \inf(-a_n) = -\sup(a_n)
\end{aligned}$$
$\mu(g) < \infty \Rightarrow \mu(f) \geq \limsup \mu(f_n)$
$\Rightarrow \mu(f) \leq \liminf \mu(f_n) \leq \limsup \mu(f_n) \leq \mu(f)$

Hence $\mu(f_n) \to \mu(f)$ as $n \to \infty$ $\square$

3.3 Transformation of Integrals

$$\begin{tikzcd}
\text{measure space} \arrow[r] \arrow[d] & \text{new measure space} \arrow[d] \\
\text{integral} \arrow[r, dashed, no head] & \text{integral}
\end{tikzcd}$$

Proposition 3.3.1 Let $(E, \mathcal{E}, \mu)$ be a measure space
fix $A \in \mathcal{E}$
Set $\mathcal{E}_A = \{B \in \mathcal{E} : B \subseteq A\}$, the set of measurable subsets of A (this is a σ-algebra)
$\mu_A(B) = \mu(B)$ for $B \in A$, the restriction of μ to $\mathcal{E}_A$ (this is a measure)
Then $(A, \mathcal{E}_A, \mu_A)$ is a measure space
Moreover, given a non-negative measurable function f on E,
$f|_A$ is a $\mathcal{E}_A$-measure function on A
and we obtain all non-negative measurable functions in this way.
$\mu(f 1_A) = \mu_A(f|_A)$

Lecture 11

Proposition 3.3.2 Let $(E, \mathcal{E}, \mu)$ be a measure space
$(G, \mathcal{G})$ be a measurable space
$f : E \to G$ be a measurable function
Define an image measure on $(G, \mathcal{G})$ by $\nu = \mu \circ f^{-1}$
Then $\forall$ non-negative measurable functions g on G, $\nu(g) = \mu(g \circ f)$

In particular, for a G-valued random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$,
for any non-negative measurable function g on G,
$$\mathbb{E}(g(X)) = \mu_X(g)$$

Proposition 3·3·3 Let $(E, \mathcal{E}, \mu)$ be a measure space
let f be a non-negative measurable function on E
Define $\nu(A) = \mu(f 1_A)$, $\quad A \in \mathcal{E}$
Then ν is a measure on $(E, \mathcal{E})$
and, $\forall$ non-negative measurable functions g on E, $\quad \nu(g) = \mu(fg)$

Proof

$\nu(\emptyset) = 0$

For $A_n \in \mathcal{E}$ disjoint,
$$\begin{aligned} \nu\left(\bigcup_n A_n\right) &= \mu\left(f 1_{\bigcup_n A_n}\right) \\ &= \mu\left(\sum_n f 1_{A_n}\right) && A_n \text{ are disjoint} \\ &= \sum_n \mu(f 1_{A_n}) && \text{by monotone convergence} \\ &= \sum_n \nu(A_n) \end{aligned}$$

$\Rightarrow$ ν is a measure

$g = 1_A$: $\nu(g) = \mu(fg)$ by definition of ν

$g = \sum_{k=1}^{m} a_k 1_{A_k}$:
$$\begin{aligned} \nu(g) &= \nu\left(\sum_{k=1}^{m} a_k 1_{A_k}\right) \\ &= \sum_{k=1}^{m} a_k \nu(1_{A_k}) \\ &= \sum_{k=1}^{m} a_k \mu(f 1_{A_k}) \\ &= \mu\left(f \sum_{k=1}^{m} a_k 1_{A_k}\right) \\ &= \mu(fg) \end{aligned}$$

g non-negative measurable : set $g_n = \min\left\{ \frac{1}{2^n} \lfloor 2^n g \rfloor, n \right\}$

Then $g_n \uparrow g$, g_n is simple $\forall n$
$$\begin{aligned} \nu(g) &= \lim_{n\to\infty} \nu(g_n) && \text{by monotone convergence} \\ &= \lim_{n\to\infty} \mu(f g_n) \\ &= \mu(fg) && \text{by monotone convergence} \end{aligned}$$
$\square$

In particular, to each non-negative Borel function f on $\mathbb{R}$,
there corresponds a Borel measure μ on $\mathbb{R}$ given by $\mu(A) = \int_A f(x)\,dx$

Then, $\forall$ non-negative Borel functions g, $\quad \mu(g) = \int_{\mathbb{R}^n} g(x) f(x)\,dx$

μ has density f (with respect to Lebesgue measure)

If the law μ_X of a real-valued random variable X has a density f_X,
then f_X is a density function for X.

$\Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}^+$

$\mu_X = \mathbb{P} \circ X^{-1}$

Then
$$\begin{aligned} \mathbb{P}(X \in A) &= \mu_X(A) \\ &= \int_A f_X(x)\,dx \\ &= \text{Leb}(f 1_A) && \forall \text{ Borel sets } A \end{aligned}$$

and
$$\begin{aligned} E(g(X)) &= \mu_X(g) \\ &= \int_{\mathbb{R}} g(x) f_X(x)\,dx && \forall \text{ non-negative Borel functions } g \text{ on } \mathbb{R} \end{aligned}$$

Probability and Measure

3·4 Fundamental theorem of calculus

Theorem 3·4·1 Fundamental theorem of calculus

i. Let $f : [a,b] \to \mathbb{R}$ be a continuous function
Set $F_a(t) = \int_a^t f(x)\,dx$
Then F_a is differentiable on $[a,b]$, with $F_a' = f$

ii. Let $F : [a,b] \to \mathbb{R}$ be differentiable with continuous derivative f
Then $\int_a^b f(x)\,dx = F(b) - F(a)$

Proof

i. Fix $t \in [a,b)$ $0<\delta\leq b-t$
Given $\varepsilon > 0$, $\exists$ such that $t \leq x \leq t+\delta \Rightarrow |f(x) - f(t)| \leq \varepsilon$
So $0 < h \leq \delta \Rightarrow$
$$\left|\frac{F_a(t+h) - F_a(t)}{h} - f(t)\right| = \frac{1}{h}\left|\int_t^{t+h} f(x) - f(t)\,dx\right|$$
$$\leq \frac{1}{h}\int_t^{t+h} |f(x) - f(t)|\,dx$$
$$\leq \frac{1}{h}\cdot\varepsilon\int_t^{t+h} dx$$
$$= \varepsilon$$

$\Rightarrow$ F_a is differentiable on the right at t with derivative $f(t)$
Similarly $\forall\ t \in (a,b]$, F_a is differentiable on the left at t with derivative $f(t)$.
Hence F_a is differentiable on $[a,b]$, with $F_a' = f$

ii. $F - F_a$ is differentiable on (a,b), with $(F - F_a)' = 0$
So by the mean value theorem $F - F_a$ is constant
$$\Rightarrow F(b) - F(a) = F_a(b) - F_a(a)$$
$$= \int_a^b f(x)\,dx \qquad \square$$

example

$$\int_0^\infty e^{-x}\,dx = \lim_{n\to\infty}\int_0^\infty 1_{[0,n]}(x)\,e^{-x}\,dx \quad \text{by monotone convergence}$$
$$= \lim_{n\to\infty}\int_0^n e^{-x}\,dx$$
$$= \lim_{n\to\infty}(1 - e^{-n})$$
$$= 1$$

Proposition 3·4·2 Let $\phi : [a,b] \to \mathbb{R}$ be continuously differentiable and strictly increasing
Then, $\forall$ non-negative Borel functions g on $[\phi(a), \phi(b)]$,
$$\int_{\phi(a)}^{\phi(b)} g(y)\,dy = \int_a^b g(\phi(x))\,\phi'(x)\,dx$$

Sketch of proof

First, the case where g is the indicator function of an interval follows from the Fundamental Theorem of Calculus.
Next, the set of Borel sets B such that the conclusion holds for $g = 1_B$ is a d-system.
By Dynkin's lemma this set must be the whole Borel σ-algebra
By linearity, the identity extends to simple functions
By monotone convergence, taking $g_n = \min\{\frac{1}{2^n}\lfloor 2^n g\rfloor, n\}$,
the identity extends to non-negative measurable functions

3·5 Differentiation under the integral sign

Theorem 3·5·1 Differentiation under the integral sign

Let $I \subseteq \mathbb{R}$ be an open interval

Suppose $f : I \times E \to \mathbb{R}$ satisfies

i. $E \to \mathbb{R}$
$x \mapsto f(x, t)$ is integrable $\forall\, t \in I$

ii. $I \to \mathbb{R}$
$t \mapsto f(x, t)$ is differentiable $\forall\, x \in E$

and $\exists$ an integrable function g on E such that $\left|\frac{\partial f}{\partial t}(t, x)\right| \leq g(x)$ $\forall\, t \in I, x \in E$

Then the function $x \mapsto \frac{\partial f}{\partial t}(t, x)$ is integrable $\forall\, t \in I$

Define $F : I \to \mathbb{R}$
$$F(t) = \int_E f(t, x)\, \mu(dx)$$
Then F is differentiable, with $F'(t) = \int_E \frac{\partial f}{\partial t}(t, x)\, \mu(dx)$

Proof Fix $t \in I$ and a sequence $t_n \in I \setminus \{t\}$ with $t_n \to t$

Set $g_n(x) = \dfrac{f(t_n, x) - f(t, x)}{t_n - t} - \dfrac{\partial f}{\partial t}(t, x)$

By the Mean Value Theorem, $\exists\, c \in (t_n, t)$ such that
$$f(t_n, x) - f(t, x) = (t_n - t)\frac{\partial f}{\partial t}(c, x)$$
$$\Rightarrow |g_n(x)| \leq 2g(x) \quad \forall\, x$$
and $g_n(x) \to 0$ as $n \to \infty$

So $x \mapsto \frac{\partial f}{\partial t}(t, x)$ is the limit of measurable functions

$\Rightarrow$ it is measurable

$\Rightarrow$ it is integrable

Then $\dfrac{F(t_n) - F(t)}{t_n - t} - \displaystyle\int_E \frac{\partial f}{\partial t}(t, x)\, \mu(dx) = \int_E g_n(x)\, \mu(dx)$

$\to 0$ by dominated convergence $\square$

example
$$\hat{f}(u) = \int_{\mathbb{R}} e^{iux} f(x)\, dx$$
$$\hat{f}'(u) = \int_{\mathbb{R}} ix e^{iux} f(x)\, dx$$

Lecture 12

3·6 Product measure and Fubini's theorem

Let $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ be finite measure spaces.

The set $\mathcal{A} = \{A_1 \times A_2 : A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2\}$ is a π-system of subsets of $E = E_1 \times E_2$

Define the product σ-algebra $\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2$
$$= \sigma(\mathcal{A})$$

Lemma 3·6·1 Let $f : E \to G$ be $\mathcal{E}$-measurable

Then, $\forall\, x_1 \in E_1$, the function $E_2 \to G$, $x_2 \mapsto f(x_1, x_2)$ is $\mathcal{E}_2$-measurable

Probability and Measure

Proof Let $\mathcal{V} = \{ g : E \to G \mid g$ is bounded, $\mathcal{E}$-measurable, and $\forall x_1 \in E_1$, $x_2 \mapsto g(x_1, x_2)$ is $\mathcal{E}_2$-measurable$\}$

Then $\mathcal{V}$ is a vector space

$1_A \in \mathcal{V} \ \forall A \in \mathcal{A}$

If $g_n \in \mathcal{V} \ \forall n$ and $0 \le g_n \uparrow g$ with g bounded, then $g \in \mathcal{V}$

So by the Monotone Class Theorem, $\mathcal{V}$ contains all bounded $\mathcal{E}$-measurable functions.

For an $\mathcal{E}$-measurable function h that is unbounded,

let $h_n = \min\{ n, \max\{ -n, h \}\}$

Then $h_n \in \mathcal{V} \ \forall n \Rightarrow \forall x_1 \in E_1$, $x_2 \mapsto h_n(x_1, x_2)$ is $\mathcal{E}_2$-measurable and $h_n \to h$

Given $x_1 \in E_1$, let $f(x_2) = h(x_1, x_2)$

Then $f^{-1}(h(x_1, x_2)) = \{ y \in E_2 : f(y) = h(x_1, x_2) \}$

$= \{ y \in E_2 : f(y) = h_m(x_1, x_2) \}$ for $m \in \mathbb{N}$ sufficiently large

$= f^{-1}(h_m(x_1, x_2))$

$= f_m^{-1}(h_m(x_1, x_2))$

where $f_m(x_2) = h_m(x_1, x_2)$ is $\mathcal{E}_2$-measurable

$\Rightarrow h \in \mathcal{V}$ $\square$

Lemma 3·6·2 Let f be an $\mathcal{E}$-measurable function on E

For $x_1 \in E_1$, let $f_1(x_1) = \int_{E_2} f(x_1, x_2)\, \mu_2(dx_2)$

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

If f is non-negative then $f_1 : E_1 \to [0, \infty]$ is an $\mathcal{E}_1$-measurable function.

Proof Let $\mathcal{U} = \{ g \mid g_1 : E_1 \to \mathbb{R}$, $g_1(x_1) = \int_{E_2} g(x_1, x_2)\, \mu(dx_2)$ is a bounded $\mathcal{E}_1$-measurable function$\}$

Then by the Monotone Class Theorem, $\mathcal{U}$ contains all bounded $\mathcal{E}$-measurable functions. (since $\mu_2(E_2) < \infty$)

Let $\mathcal{V} = \{ g \mid g_1 : E_1 \to \mathbb{R}$, $g_1(x_1) = \int_{E_2} g(x_1, x_2)\, \mu(dx_2)$ is an $\mathcal{E}_1$-measurable function$\}$

Then $\mathcal{V} \supseteq \mathcal{U} \Rightarrow \mathcal{V}$ contains all bounded $\mathcal{E}$-measurable functions.

For an $\mathcal{E}$-measurable function h that is unbounded but non-negative,

let $h_n = \min\{ h, n \}$

Then $h_n \in \mathcal{V} \ \forall n$

and $h_n \uparrow h$

$\Rightarrow h \in \mathcal{V}$ $\square$

Theorem 3·6·3 Product measure

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

such that $\mu(A_1 \times A_2) = \mu_1(A_1)\, \mu_2(A_2) \quad \forall A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2$

Proof Existence

Define $\mu(A) = \int_{E_1} \left(\int_{E_2} 1_A(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1), \quad A \in \mathcal{E}$

Then $\mu(\emptyset) = 0$

For $A_n \in \mathcal{E}$, disjoint, $\mu(\bigcup_n A_n) = \int_{E_1} \left(\int_{E_2} \sum_n 1_{A_n}(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1)$

$= \sum_n \int_{E_1} \left(\int_{E_2} 1_{A_n}(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1)$

by Monotone Convergence twice

$= \sum_n \mu(A_n)$

$\Rightarrow \mu$ is a measure, $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$

Uniqueness

$\mathcal{A}$ is a π-system generating $\mathcal{E}$

and $\mu(E) = \mu_1(E_1)\mu_2(E_2) < \infty$

so by uniqueness of extension from π-systems, μ is unique □

Proposition 3.6.4 Let $\hat{E} = E_2 \times E_1$

$\hat{\mathcal{E}} = \mathcal{E}_2 \otimes \mathcal{E}_1$

$\hat{\mu} = \mu_2 \otimes \mu_1$

Define $\wedge : E \to \hat{E}$

$\wedge(x_1, x_2) = (x_2, x_1)$

Then $\wedge$ is $\mathcal{E}$ & $\hat{\mathcal{E}}$ - measurable

and $\hat{\mu} = \mu \circ \wedge^{-1}$

For a function f on $E_1 \times E_2$

let $\hat{f}(x_2, x_1) = f(x_1, x_2)$ be a function on $E_2 \times E_1$

If f is a non-negative $\mathcal{E}$-measurable function

then $\hat{f}$ is a non-negative $\hat{\mathcal{E}}$-measurable function

and $\hat{\mu}(\hat{f}) = \mu(f)$

Theorem 3.6.5 Fubini's Theorem

i. Let f be a non-negative $\mathcal{E}$-measurable function

Then $\mu(f) = \int_{E_1} \left(\int_{E_2} f(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1)$ ⊛

ii. Let f be a μ-integrable function on E

Define $A_1 = \{ x_1 \in E_1 : \int_{E_2} |f(x_1, x_2)| \mu_2(dx_2) < \infty \}$

and set $f_1(x_1) = \begin{cases} \int_{E_2} f(x_1, x_2) \mu_2(dx_2) & \text{if } x_1 \in A_1 \\ 0 & \text{if } x_1 \notin A_1 \end{cases}$

Then $\mu_1(E_1 \setminus A_1) = 0$

and f_1 is μ_1-measurable, with $\mu_1(f_1) = \mu(f)$

Notes 1. By Lemmas 3.6.1 and 3.6.2, the iterated integral in i. is well-defined $\forall$ bounded or non-negative measurable functions f

2. In combination with Proposition 3.6.4, Fubini's theorem allows us to interchange the order of integration in multiple integrals, whenever the integrand is non-negative or μ-integrable.

For $f : E_1 \times E_2 \to [0, \infty]$ which is $\mathcal{E}_1 \times \mathcal{E}_2$ measurable,

$\int_{E_1} \left(\int_{E_2} f(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1) = (\mu_1 \otimes \mu_2)(f)$

$= \int_{E_2} \left(\int_{E_1} f(x_1, x_2) \mu_1(dx_1) \right) \mu_2(dx_2)$

Proof i. if $f = 1_A$ for some $A \in \mathcal{E}$

then ⊛ holds by definition of the product measure μ

By linearity of the integrals, ⊛ extends to simple functions on E.

For f non-negative measurable,

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

Probability and Measure

Then f_n are simple $\Rightarrow$ $\circledast$ holds for f_n
and $f_n \uparrow f$
So by monotone convergence $\mu(f_n) \uparrow \mu(f)$
and $\forall x_1 \in E_1$, $\int_{E_2} f_n(x_1, x_2)\, \mu_2(dx_2) \uparrow \int_{E_2} f(x_1, x_2)\, \mu_2(dx_2)$

$$\Rightarrow \mu(f) = \lim_{n\to\infty} \mu(f_n)$$
$$= \lim_{n\to\infty} \int_{E_1} \left(\int_{E_2} f_n(x_1, x_2)\, \mu_2(dx_2)\right) \mu_1(dx_1)$$
$$= \int_{E_1} \left(\int_{E_2} f(x_1, x_2)\, \mu_2(dx_2)\right) \mu_1(dx_1)$$

ii. Let f be a μ-integrable function on E
Consider $F_1 : E_1 \to [0, \infty]$
$$F_1(x_1) = \int_{E_2} |f(x_1, x_2)|\, \mu_2(dx_2)$$
By Lemma 3·6·2, F_1 is $\mathcal{E}_1$-measurable
Now $A_1 = \{ x_1 \in E_1 : F_1(x_1) < \infty \}$
so A_1 is an $\mathcal{E}_1$-measurable set, $A_1 \in \mathcal{E}_1$
Also $\mu_1(F_1) = \mu(|f|)$
$< \infty$
$\Rightarrow \mu_1(A_1^c) = 0$
f_1 is well-defined
Consider $f_1^{(\pm)}(x_1) = \int_{E_2} f^{\pm}(x_1, x_2)\, \mu_2(dx_2)$
where $f^{\pm} = \max\{\pm f, 0\}$
(note that in general $f_1^{(\pm)} \neq f_1^{\pm}$)
Then $f_1^{(\pm)}$ are $\mathcal{E}_1$-measurable, and $\mu_1(f_1^{(\pm)}) = \mu(f^{\pm})$ by i.
Now $f_1 = (f_1^{(+)} - f_1^{(-)}) 1_{A_1}$ $\Rightarrow$ f_1 is μ_1-integrable
$$\mu_1(f_1) = \mu_1(f_1^{(+)}) - \mu_1(f^{(-)})$$
$$= \mu(f^+) - \mu(f^-)$$
$$= \mu(f) \qquad \square$$

The existence of product measure and Fubini's theorem (Theorems 3·6·3, 4, 5) extend to σ-finite measures μ_1, μ_2.
Take $E_k = \text{disjoint} \bigcup_n E_k^n$, where $\mu_k(E_k^n) < \infty$
Set $\mu_k^n(A) = \mu_k(A \cap E_k^n)$
then μ_k^n are finite measures
$$\mu_k = \sum_n \mu_k^n$$
$$\mu_1 \otimes \mu_2 = \sum_{n,m} \mu_1^n \otimes \mu_2^m$$

By a π-system uniqueness argument,
the operation of taking the product of two measure spaces is associative
ie $(\mathcal{E}_1 \otimes \mathcal{E}_2) \otimes \mathcal{E}_3 = \mathcal{E}_1 \otimes (\mathcal{E}_2 \otimes \mathcal{E}_3)$
$(\mu_1 \otimes \mu_2) \otimes \mu_3 = \mu_1 \otimes (\mu_2 \otimes \mu_3)$
So by induction we can define $\mu_1 \otimes \dots \otimes \mu_n$ without specifying order
If ~~μ_k~~ $\mu_k = (\mathbb{R}, \mathcal{B}, \text{Leb})$
we get $(\mathbb{R}^n, \mathcal{B}^{\otimes n}, \text{Leb}^{\otimes n})$, the Lebesgue measure on $\mathbb{R}^n$
The corresponding integral is written
$$\int_{\mathbb{R}^n} f(x)\, dx$$

example using Fubini's Theorem

Let $X \geq 0$ be a random variable

Then $\mathbb{E}(X) = \int_0^\infty x \, \mu_X(dx)$

$= \int_0^\infty \int_0^\infty 1_{y \leq x} \, dy \, \mu_X(dx)$

$= \int_0^\infty \int_0^\infty 1_{y \leq x} \, \mu_X(dx) \, dy$

$= \int_0^\infty \mathbb{P}(X \geq y) \, dy$

eg $X \sim E(1)$

$\mathbb{P}(X \geq y) = e^{-y}$

$\mathbb{E}(X) = \int_0^\infty e^{-y} \, dy$

$= 1$

3.7 Laws of independent random variables

Recall

A family $X_1, \ldots, X_n$ of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ is independent if the family of σ-algebras $\sigma(X_1), \ldots, \sigma(X_n)$ is independent

Proposition 3.7.1

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

Let $X_1, \ldots, X_n$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $(E_1, \mathcal{E}_1), \ldots, (E_n, \mathcal{E}_n)$, where $(E_k, \mathcal{E}_k)$ are measure spaces

Set $E = E_1 \times \ldots \times E_n$

$\mathcal{E} = \mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_n$

Define $X : \Omega \to E$

$X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$

Then X is $\mathcal{E}$-measurable (so it's a random variable)

Moreover, the following are equivalent

i. $X_1, \ldots, X_n$ are independent

ii. $\mu_X = \mu_{X_1} \otimes \ldots \otimes \mu_{X_n}$

iii. $\forall$ bounded measurable functions f_k on E_k, $1 \leq k \leq n$ we have $\mathbb{E}\left(\prod_{k=1}^n f_k(X_k)\right) = \prod_{k=1}^n \mathbb{E}(f_k(X_k))$

Proof

Consider the π-system $\mathcal{A} = \{A_1 \times \ldots \times A_n : A_k \in \mathcal{E}_k, 1 \leq k \leq n\}$ on E

Then $X^{-1}(A) = \bigcap_{k=1}^n \{\omega \in \Omega : X_k(\omega) \in A_k\} \in \mathcal{F}$

But $\mathcal{A}$ generates $\mathcal{E}$, so this shows X is $\mathcal{E}$-measurable

i ⇒ ii

Set $\nu = \mu_{X_1} \otimes \ldots \otimes \mu_{X_n}$

For $A \in \mathcal{A}$, $\nu(A) = \prod_{k=1}^n \mu_{X_k}(A_k)$

$= \prod_{k=1}^n \mathbb{P}(X_k \in A_k)$

$= \mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n)$ for $X_1, \ldots, X_n$ independent

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

$= \mu_X(A)$

The probability measures ν, μ_X agree on $\mathcal{A}$ so by uniqueness of extension, $\nu = \mu_X$

Probability and Measure

ii ⇒ iii $\mathbb{E}\left(\prod_{k=1}^{n} f_k(X_k)\right) = \mu_X\left(\prod_{k=1}^{n} f_k(X_k)\right)$

$$= \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}\left(\prod_{k=1}^{n} f_k(X_k)\right) \quad \text{for } \mu_X = \mu_{X_1} \otimes \ldots \otimes \mu_{X_n}$$

$$= \int_{E_1} \cdots \int_{E_n} \prod_{k=1}^{n} f(x_k)\, \mu_n(dx_n) \cdots \mu_1(dx_1)$$

$$= \prod_{k=1}^{n} \int_{E_k} f_k(x_k)\, \mu_k(dx_k) \quad \text{by Fubini}$$

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

iii ⇒ i Take $f_k = 1_{A_k}$ with $A_k \in \mathcal{E}_k$

Then $\mathbb{P}(\cancel{A_1 \in}\ X_1 \in A_1, \ldots, X_n \in A_n) = \mathbb{E}(1_{A_1}(X_1), \ldots, 1_{A_n}(X_n))$

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

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

$$= \prod_{k=1}^{n} \mathbb{P}(X_k \in A_k) \qquad \square$$

§4 Norms and inequalities

4.1 L^p - norms

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

Definitions For $1 \leq p < \infty$, the L^p - norm of f is $\|f\|_p = (\int_E |f|^p \, d\mu)^{\frac{1}{p}}$
$= \mu(|f|^p)^{\frac{1}{p}}$

$L^p = L^p(E, \mathcal{E}, \mu) = L^p(\mu)$ is the set of measurable functions having finite L^p-norm

The L^∞ - norm of f is $\|f\|_\infty = \inf\{\lambda \geq 0 : |f| \leq \lambda$ almost everywhere$\}$
$L^\infty = L^\infty(E, \mathcal{E}, \mu)$ is the set of measurable functions having finite L^∞-norm

Note $\|f\|_p \leq \mu(E)^{\frac{1}{p}} \|f\|_\infty \quad \forall\ 1 \leq p < \infty$

Definition For $1 \leq p \leq \infty$ and $f_n, f \in L^p$
f_n converges to f in L^p ($f_n \to f$ in L^p) if $\|f_n - f\|_p \to 0$

4.2 Chebyshev's inequality

Let f be a non-negative measurable function
$\lambda \geq 0$

Notation $\{f \geq \lambda\}$ is the set $\{x \in E : f(x) \geq \lambda\}$

Observe that $\lambda 1_{\{f \geq \lambda\}} \leq f$
so on integrating, we get Chebyshev's inequality
$$\lambda \mu(f \geq \lambda) \leq \mu(f)$$

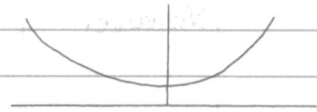
So for any measurable function g,
we can deduce inequalities for g by choosing some non-negative measurable function ϕ and apply Chebyshev's inequality to $f = \phi \circ g$

example If $g \in L^p$, $p < \infty$, and $\lambda > 0$
then
$$\mu(|g| \geq \lambda) = \mu(|g|^p \geq \lambda^p)$$
$$\leq \lambda^{-p} \mu(|g|^p)$$
$$< \infty$$

This gives the tail estimate $\mu(|g| \geq \lambda) = O(\lambda^{-p})$, as $\lambda \to \infty$

$\{g \geq \lambda\} = \{e^{\alpha g} \geq e^{\alpha \lambda}\}, \quad \alpha > 0$

4.3 Jenson's inequality



Let $I \subseteq \mathbb{R}$ be an interval

Definition Then a function $f : I \to \mathbb{R}$ is convex if $f(tx + (1-t)y) \leq t f(x) + (1-t) f(y) \quad \forall\, x, y \in I,\ t \in (0,1)$

For $f \in C^2$, f is convex $\iff f'' \geqslant 0$

Lemma 4·3·1 Let $f : I \to \mathbb{R}$ be convex
let m be an interior point of I
Then $\exists\, a, b \in \mathbb{R}$ such that $ax + b \leq f(x) \quad \forall\, x \in I$
and $am + b = f(m)$

Proof Given $x, y \in I$ with $x < m < y$, let $t = \frac{y-m}{y-x}$
By convexity of f, $f(tx + (1-t)y) \leq t f(x) + (1-t) f(y)$

$$f(m) \leq \frac{y-m}{y-x} f(x) + \frac{m-x}{y-x} f(y)$$

$$\Rightarrow \frac{f(m) - f(x)}{m - x} \leq \frac{f(y) - f(m)}{y - m}$$

Set $a = \sup\limits_{\substack{x \in I \\ x < m}} \left(\frac{f(m) - f(x)}{m - x} \right)$

$b = f(m) - am$

Then $f(m) - f(x) \leq a(m - x)$

$f(x) \geqslant ax + b \qquad \forall\ x \leq m$

Similarly $f(y) \geqslant ay + b \quad \forall\ y \geqslant m$ $\square$

Theorem 4·3·2 Jenson's inequality
Let X be an integrable random variables with values in an interval $I \subseteq \mathbb{R}$
let $f : I \to \mathbb{R}$ be a convex function
Then $\mathbb{E}(f(X))$ is well-defined
and $f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$

Proof Set $m = \mathbb{E}(X)$
Then $m \in I$
if $m = \inf I$ then $X - m \geqslant 0 \quad \forall\ X$ and $\mathbb{E}(X - m) = 0$

$\Rightarrow X = m$ almost surely

$\Rightarrow \mathbb{E}(f(X)) = f(m)$

$= f(\mathbb{E}(X))$

Similarly if $m = \sup I$ then $X = m$ almost surely
So assume m is an interior point of I
Choose $a, b \in \mathbb{R}$ as in Lemma 4·3·1
Then $aX + b \leq f(X)$

In particular $-f(X) \leq -aX - b$

$\Rightarrow f^-(X) \leq |aX - b|$

$\mathbb{E}(f^-(X)) \leq |a|\,\mathbb{E}(|X|) + |b|$

$< \infty$

$\Rightarrow \mathbb{E}(f(X))$ is well-defined

Moreover $f(\mathbb{E}(X)) = f(m)$
$= am + b$
$= \mathbb{E}(aX + b)$
$\leq \mathbb{E}(f(X))$ $\square$

Monotonicity of L^p-norms with respect to a probability measure $\|X\|_{L^p(\mathbb{P})}$
Let $1 \leq p < q < \infty$
$f(x) = |x|^{\frac{q}{p}}$, a convex function on $\mathbb{R}$
Then, for any $X \in L^p(\mathbb{P})$

$$\|X\|_p = (\mathbb{E}|X|^p)^{\frac{1}{p}}$$
$$= (f(\mathbb{E}|X|^p))^{\frac{1}{q}}$$
$$\leq (\mathbb{E}(f(|X|^p)))^{\frac{1}{q}} \quad \text{by Jensen's inequality}$$
$$= (\mathbb{E}|X|^q)^{\frac{1}{q}}$$
$$= \|X\|_q$$

$\|X\|_{L^q(\mathbb{P})}$ is non-decreasing in $q \in [1, \infty)$

In particular, $L^p(\mathbb{P}) \supseteq L^q(\mathbb{P})$

4.4 Hölder's inequality and Minkowski's inequality

Definition $p, q \in [1, \infty]$ are conjugate indices if $\frac{1}{p} + \frac{1}{q} = 1$

Theorem 4.4.1 Hölder's inequality
Let $p, q \in (1, \infty)$ be conjugate indices
let f, g be measurable functions
Then $\|fg\|_1 \leq \|f\|_p \|g\|_q$

$\mu(|fg|) \leq \|f\|_p \|g\|_q$

Proof If $\|f\|_p = 0$ then $f = 0$ almost everywhere
$\Rightarrow \|fg\|_1 = 0$
If $\|f\|_p = \infty$ and $\|g\|_q > 0$ then $\|f\|_p \|g\|_q = \infty$
So assume $\|f\|_p \in (0, \infty)$
Let $\hat{f} = \dfrac{f}{\|f\|_p}$
Define a probability measure $\mathbb{P}$ on $\mathcal{E}$ by $\mathbb{P}(A) = \mu(|\hat{f}|^p 1_A) = \int_A |\hat{f}| \, d\mu$
Note that for a random variable $X \geq 0$, $\mathbb{E}(X) = \mu(|\hat{f}|^p X)$
$\mathbb{E}(|X|) \leq (\mathbb{E}(|X|^q))^{\frac{1}{q}}$ by Jenson

Hence
$$\|\hat{f}g\|_1 = \mu(|\hat{f}g|)$$
$$= \mu\left(|\hat{f}|^p \frac{|g|}{|\hat{f}|^{p-1}} 1_{\{|\hat{f}|>0\}}\right)$$
$$= \mathbb{E}\left(\frac{|g|}{|\hat{f}|^{p-1}} 1_{\{|\hat{f}|>0\}}\right)$$
$$\leq \left(\mathbb{E}\left(\frac{|g|^q}{|\hat{f}|^{(p-1)q}} 1_{\{\hat{f}>0\}}\right)\right)^{\frac{1}{q}}$$
$$= (\mu(|g|^q))^{\frac{1}{q}} \quad (p-1)q = p$$
$$= \|g\|_q \|\hat{f}\|_p \quad \|\hat{f}\|_p = 1$$
$$\Rightarrow \|fg\|_1 = \|f\|_p \|g\|_q \quad \square$$

Probability and Measure

Theorem 4·4·2 Minkowski's inequality

Let $p \in [1, \infty)$

let f, g be measurable functions

Then $\| f + g \|_p \leq \| f \|_p + \| g \|_p$

Proof

The cases $\| f + g \|_p = 0$, $\| f \|_p = \infty$, $\| g \|_p = \infty$ are immediate

So suppose $\| f + g \|_p > 0$, $\| f \|_p < \infty$, $\| g \|_p < \infty$

Then $| f + g |^p \leq 2^p (| f |^p + | g |^p)$ — proof by induction, using $|f+g| \leq |f| + |g|$ by Cauchy-Schwarz

$\Rightarrow \mu (| f + g |^p) \leq 2^p (\mu (| f |^p) + \mu (| g |^p))$

$< \infty$

Let q be the conjugate pair of p, $\frac{1}{p} + \frac{1}{q} = 1$

Then
$$
\begin{aligned}
\mu (| f + g |^p) &= \mu (| f + g | \, | f + g |^{p-1}) \\
&\leq \mu (| f | \, | f + g |^{p-1}) + \mu (| g | \, | f + g |^{p-1}) \qquad |f+g| \leq |f| + |g| \\
&\leq (\| f \|_p + \| g \|_p) \, \| \, | f + g |^{p-1} \|_q \qquad \text{by Hölder} \\
&= (\| f \|_p + \| g \|_p) \left(\mu (| f + g |^{(p-1)q}) \right)^{\frac{1}{q}} \\
&= (\| f \|_p + \| g \|_p) \left(\mu (| f + g |^p) \right)^{1 - \frac{1}{p}}
\end{aligned}
$$

$\Rightarrow (\mu (| f + g |^p))^{\frac{1}{p}} \leq (\| f \|_p + \| g \|_p)$

$\| f + g \|_p \leq \| f \|_p + \| g \|_p$ $\square$

4·5 Approximation in L^p

Theorem 4·5·1 Let $\mathcal{A}$ be a π-system on E generating $\mathcal{E}$

such that $\mu(A) < \infty \quad \forall A \in \mathcal{A}$

and $E_n \uparrow E$ for some sequence $(E_n : n \in \mathbb{N})$ in $\mathcal{A}$

Consider $V_0 = \{ \sum_{k=1}^{n} a_k 1_{A_k} : a_k \in \mathbb{R}, \ A_k \in \mathcal{A}, \ n \in \mathbb{N} \}$

Let $p \in [1, \infty)$

Then V_0 is a dense subset of L^p

ie $V_0 \subseteq L^p$

and given $f \in L^p$, $\varepsilon > 0$, $\exists \ v \in V_0$ such that $\| f - v \|_p < \varepsilon$

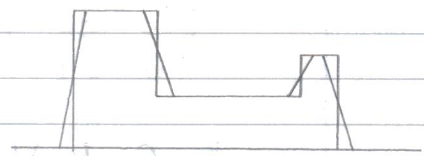
example

$(\mathbb{R}^n, dx)$

$\mathcal{A} = \{ \prod_{i=1}^{n} (a_i, b_i], \ a_i < b_i \}$

So step functions are dense in L^p

so $C_c(\mathbb{R}^n)$ are dense in L^p



Proof

$\forall A \in \mathcal{A}$, we have $\| 1_A \|_p = \mu(A)^{\frac{1}{p}} < \infty$

$\Rightarrow 1_A \in L^p$

So L^p is a vector space $\Rightarrow V_0 \subseteq L^p$

Let $V = \{ f \in L^p \mid \forall \varepsilon > 0, \exists v \in V_0 \text{ such that } \| f - v \|_p < \varepsilon \}$
$= Cl(V_0)$

Suppose $f, g \in V$, with $\| f - u \|_p < \varepsilon$
$\| g - v \|_p < \varepsilon$

Then $\| (f+g) - (u+v) \|_p \leq \| f - u \|_p + \| g - v \|_p$ by Minkowski's inequality
$\Rightarrow f + g \in V$
$\Rightarrow V$ is a vector space

Consider the case $E \in \mathcal{A}$
Define $\mathcal{D} = \{ A \in \mathcal{E} : 1_A \in V \}$
$\subseteq \mathcal{E}$

Then $\mathcal{A} \subseteq \mathcal{D} \Rightarrow E \in \mathcal{D}$
For $A, B \in \mathcal{D}$ with $A \subseteq B$,
$1_{B \setminus A} = 1_B - 1_A \in V_0 \subseteq V$
$\Rightarrow B \setminus A \in \mathcal{D}$
For $A_n \in \mathcal{D}$ with $A_n \uparrow A$
$\| 1_A - 1_{A_n} \|_p = \| 1_{A \setminus A_n} \|_p$
$= \mu(A \setminus A_n)^{\frac{1}{p}}$
$\to 0 \quad \Rightarrow 1_A \in V$

$\Rightarrow A \in \mathcal{D}$
$\Rightarrow \mathcal{D}$ is a d-system
So by Dynkin's π-system lemma, $\mathcal{E} \subseteq \mathcal{D}$
$\Rightarrow \mathcal{D} = \mathcal{E}$
Then $A \in \mathcal{E} \Rightarrow 1_A \in V$
So V is a vector space $\Rightarrow V$ contains all simple functions

Consider $f \in L^p$ with $f \geq 0$
Let $f_n = \min \{ \frac{1}{2^n} \lfloor 2^n f \rfloor, n \}$
$\uparrow f$
Then $|f|^p \geq |f - f_n|^p \to 0$ pointwise
so by dominated convergence $\| f - f_n \|_p \to 0$
Let g_n approximate f_n, $g_n \in V_0$
Then $\| f - g_n \| \leq \| f - f_n \| + \| f_n - g_n \|$ by Minkowski's inequality
$\to 0$

$\Rightarrow f \in V$
$\Rightarrow L^p \subseteq V$
$\Rightarrow V = L^p$

Now consider the general case, where it may be that $E \notin \mathcal{A}$
We now know that, $\forall f \in L^p, \ \forall n \in \mathbb{N}$,
$f 1_{E_n} \in V$
But $|f|^p \geq |f - f 1_{E_n}|^p \to 0$ pointwise,
so by dominated convergence $\| f - f 1_{E_n} \|_p \to 0$.
$\| f - g_n \| \leq \| f - f 1_{E_n} \| + \| f 1_{E_n} - g_n \|$
$\Rightarrow f \in V$
$\Rightarrow V = L^p$ $\square$

Probability and Measure

§5 Completeness of L^p and orthogonal projection

5.1 L^p as a Banach space

Lecture 16

Definition Let V be a vector space.
A map $V \to [0, \infty)$
$v \mapsto \|v\|$ is a norm on V if
i. $\|u+v\| \leq \|u\| + \|v\| \quad \forall\, u, v \in V$ (norm inequality)
ii. $\|\alpha v\| = |\alpha|\,\|v\| \quad \forall\, v \in V, \alpha \in \mathbb{R}$
iii. $\|v\| = 0 \Rightarrow v = 0$

Note By i, if $\|v_n - v\| \to 0$ then $\|v_n\| \to \|v\|$
$\|v_n\| \leq \|v\| + \|v_n - v\|$
$\|v\| \leq \|v_n\| + \|v - v_n\|$

Definition A symmetric bilinear map $V \times V \to \mathbb{R}$
$(u, v) \mapsto \langle u, v \rangle$
is an inner product on V if
$\langle v, v \rangle \geq 0$, with equality $\Leftrightarrow v = 0$

Given an inner product, there is an associated norm given by
$\|v\| = \langle v, v \rangle^{\frac{1}{2}}$
The norm inequality here follows from Cauchy-Schwarz.

Consider $V = L^p(E, \mathcal{E}, \mu)$
By Minkowski's inequality, V is a vector space
$\|f\|_p = (\int_E |f|^p \, d\mu)^{\frac{1}{p}}$ satisfies condition i
Also condition ii holds.
But condition iii may fail, since $\|f\|_p = 0 \Rightarrow f = 0$ almost everywhere (but not necessarily $f = 0$)
Define an equivalence relation $\sim$ on L^p
by $f \sim g$ if $f = g$ almost everywhere
$[f] = \{g \in L^p : g \sim f\}$, the equivalence class of f
Set $\mathcal{L}^p = \{[f] : f \in L^p\}$
Define $\|[f]\|_p = \|f\|_p$ (this is well-defined)
Then $\|\cdot\|_p$ is a norm on $\mathcal{L}^p$

The L^2-norm comes from an (almost) inner product $\|f\|_2^2 = \langle f, f \rangle$
where $\langle f, g \rangle = \int fg \, d\mu$
for $f, g \in L^2$
Then $\langle [f], [g] \rangle = \langle f, g \rangle$ defines an inner product on $\mathcal{L}^2$
$\Rightarrow \mathcal{L}^2$ is an inner product space
The notion of convergence in L^p defined in §4.1
is the usual notion of convergence in a normed space

Definition A normed vector space V is complete if every Cauchy sequence in V converges

ie given any sequence $(v_n : n \in V)$ in V such that $\| v_n - v_m \| \to 0$ as $n, m \to \infty$

$\exists\ v \in V$ such that $\| v_n - v \| \to 0$ as $n \to \infty$

A Banach space is a complete normed space

A Hilbert space is a complete inner product space

Lecture 15

Theorem 5·1·1 Completeness of L^p

Let $p \in [1, \infty]$

Let $(f_n : n \in \mathbb{N})$ be a sequence in L^p

such that $\| f_n - f_m \|_p \to 0$ as $n, m \to \infty$

Then $\exists\ f \in L^p$ such that $\| f_n - f \|_p \to 0$ as $n \to \infty$

Proof Consider the case $p < \infty$ (some adaptation is required for $p = \infty$)

$\exists$ a subsequence (n_k) such that $S = \sum_{k=1}^{\infty} \| f_{n_{k+1}} - f_{n_k} \|_p < 1$

By Minkowski's inequality, $\| \sum_{k=1}^{M} | f_{n_{k+1}} - f_{n_k} | \|_p \leq S \leq 1$

$\mu((\sum_{k=1}^{M} | f_{n_{k+1}} - f_{n_k} |)^p) \leq 1$ $\quad \forall\ M \in \mathbb{N}$

By monotone convergence $\quad \| \sum_{k=1}^{\infty} | f_{n_{k+1}} - f_{n_k} | \|_p \leq 1$

$\Rightarrow \sum_{k=1}^{\infty} | f_{n_{k+1}} - f_{n_k} | \leq 1$ almost everywhere

$\Rightarrow \sum_{k=1}^{\infty} | (f_{n_{k+1}} - f_{n_k})(x) \| \leq 1$

for almost all x

So, by completeness of $\mathbb{R}$, $(f_{n_k}(x))_{k \in \mathbb{N}}$ converges for almost all x

ie $(f_{n_k})_{k \in \mathbb{N}}$ converges almost everywhere

Let $f(x) = \begin{cases} \lim_{k \to \infty} f_{n_k}(x) & \text{if the limit exists} \\ 0 & \text{otherwise} \end{cases}$

Then f is measurable.

Given $\varepsilon > 0$, $\exists\ N \in \mathbb{N}$ such that

$m, n \geq N \Rightarrow \mu(| f_n - f_m |^p) < \varepsilon$

In particular, $\exists\ K \in \mathbb{N}$ such that

$k \geq K \Rightarrow \mu(| f_n - f_{n_k} |^p) < \varepsilon$

Now $| f_n - f_{n_k} | \to | f_n - f |$ almost everywhere, as $k \to \infty$

So $\mu(| f - f_n |^p) = \mu(\liminf_{k \to \infty} | f_n - f_{n_k} |^p)$

$\leq \liminf_{k \to \infty} \mu(| f_n - f_{n_k} |^p)$ by Fatou's Lemma

$\leq \varepsilon$

$\Rightarrow \| f - f_n \|_p \leq \varepsilon^{\frac{1}{p}}$

$\Rightarrow f \in L^p$

Since ε was arbitrary, $\| f - f_n \|_p \to 0$

$\Rightarrow f_n \to f$ in L^p $\square$

Lecture 16

Corollary 5·1·2 i. $\mathcal{L}^p$ is a Banach space, $\forall\ 1 \leq p \leq \infty$

ii. $\mathcal{L}^2$ is a Hilbert space

Probability and Measure

5.2 $\mathcal{L}^2$ as a Hilbert space

Applying some general Hilbert space arguments to L^2

Note

Pythagoras' rule

$$\|f+g\|_2^2 = \|f\|_2^2 + 2\langle f, g\rangle + \|g\|_2^2$$

Parallelogram law

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

f and g are orthogonal if $\langle f, g\rangle = 0$

Let V be a subset of L^2, $V \subseteq L^2$

V is closed if, for every sequence $(f_n : n \in \mathbb{N})$ in V with $f_n \to f$ in L^2

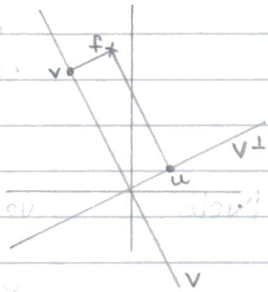
ie $\|f_n - f\| \to 0$ for some $f \in L^2$

$\exists\, v \in V$ with $v = f$ almost everywhere

(Equivalently, $\{[v] : v \in V\}$ is closed in $\mathcal{L}^2$)

Define $V^\perp = \{u \in L^2 : \langle u, v\rangle = 0 \quad \forall\, v \in V\}$

Theorem 5.2.1 Orthogonal projection



Let V be a closed subspace of L^2

Then $\forall\, f \in L^2 \ \exists\, v \in V,\ u \in V^\perp$ such that $f = v + u$

Moreover $\|f - v\|_2 \le \|f - g\|_2 \quad \forall\, g \in V$,

with equality $\Leftrightarrow$ $g = v$ almost everywhere.

The function v is called (a version of) the orthogonal projection of f on V.

Proof

$$d(f, V) = \inf_{k \in V} \|f - k\|_2$$

By definition of the infimum, $\exists$ a sequence $k_n \in V$ such that

$$\|f - k_n\|_2 \to d(f, V)$$

Then
$$4\,d(f,V)^2 \le 4\left\|f - \frac{k_n + k_m}{2}\right\|_2^2 + \|k_n - k_m\|_2^2$$
$$= \|k_n + k_m - 2f\|_2^2 + \|k_n - k_m\|_2^2$$
$$= 2(\|f - k_n\|_2^2 + \|f - k_m\|_2^2)$$

by the Parallelogram law

$$\to 4\,d(f,V)^2$$

$\Rightarrow \|k_n - k_m\|_2^2 \to 0$ as $n, m \to \infty$

L^2 is complete $\Rightarrow$ $\exists\, k \in L^2$ such that $\|k_n - k\|_2 \to 0$

V is closed $\Rightarrow$ $\exists\, v \in V$ such that $k = v$ almost everywhere

Hence
$$\|f - v\|_2 = \lim_{n\to\infty} \|f - k_n\|_2 = \inf_{k \in V} \|f - k\|_2$$

Suppose $\|f - v\|_2 = \|f - g\|_2$, some $g \in V$

Let $k_{2n+1} = v$

$k_{2n} = g$

Then (k_n) converges $\Rightarrow$ $v = g$ almost everywhere.

Set $u = f - v$

For any $h \in V$, $t \in \mathbb{R}$,

$$\begin{aligned} d(f, V)^2 &\le \| f - (v + th) \|_2^2 \\ &= \| (f - v) - th \|_2^2 \\ &= \| u - th \|_2^2 \\ &= \| u \|_2^2 - 2t \langle u, h \rangle + t^2 \| h \|_2^2 \end{aligned}$$

But $\| u \|_2^2 = \| f - v \|_2^2 = d(f, V)^2$

Taking t sufficiently small, so that $2 \langle u, h \rangle > t \| h \|_2^2$,

we see $\langle u, h \rangle = 0 \quad \forall h \in V$

$\Rightarrow u \in V^{\perp}$ $\square$

5.3 Variance, covariance and conditional expectation

In this section we look at some L^2 notions relevant to probability

Definitions Consider $X, Y \in L^2(\mathbb{P})$ with means $m_X = \mathbb{E}(X)$, $m_Y = \mathbb{E}(Y)$

$\text{var}(X) = \mathbb{E}[(X - m_X)^2]$ variance

$\text{cov}(X, Y) = \mathbb{E}[(X - m_X)(Y - m_Y)]$ covariance

$\text{corr}(X, Y) = \dfrac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}}$ correlation

Note $\text{var}(X) = 0 \Leftrightarrow X = m_X$ almost surely

X and Y are independent $\Rightarrow \text{cov}(X, Y) = 0$

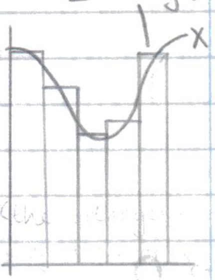
but $\text{cov}(X, Y) = 0 \not\Rightarrow X$ and Y are independent

Definition Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ be a random variable

Then $\text{var}(X) = (\text{cov}(X_i, X_j))_{i,j=1}^n$ is the covariance matrix

Proposition 5.3.1 Every covariance matrix is non-negative definite

$E(X|G)$ partial average of X



Suppose we are given a countable family of disjoint events $(G_i : i \in I)$ whose union is Ω

Set $\mathcal{G} = \sigma(G_i : i \in I)$, a sub-$\sigma$-algebra of $\mathcal{F}$

Let X be an integrable random variable, $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$

The conditional expectation of X given $\mathcal{G}$ is $\mathbb{E}(X | \mathcal{G}) = \sum_i \mathbb{E}(X | G_i) 1_{G_i}$

where $\mathbb{E}(X | G_i) = \dfrac{\mathbb{E}(X 1_{G_i})}{\mathbb{P}(G_i)}$ when $\mathbb{P}(G_i) > 0$

$\mathbb{E}(E(X|\mathcal{G})) = \mathbb{E}(X)$

$\mathbb{E}(X | G_i)$ is defined in some arbitrary way when $\mathbb{P}(G_i) = 0$

$\mathbb{E}(X | \mathcal{G}) \in L^2(\Omega, \mathcal{G}, \mathbb{P}) \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$

$L^2(\Omega, \mathcal{G}, \mathbb{P})$ is complete and hence closed

Proposition 5.3.2 If $X \in L^2$, then $\mathbb{E}(X | \mathcal{G})$ is a version of the orthogonal projection of X on $L^2(\Omega, \mathcal{G}, \mathbb{P})$

Probability and Measure

§6 Convergence in $L^1(\mathbb{P})$

6.1 Bounded convergence

Theorem 6.1.1 Bounded convergence

Let X be a random variable, $(X_n : n \in \mathbb{N})$ a sequence of random variables

Suppose $X_n \to X$ in probability

ie $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$

and $|X_n| \leq C$ almost surely $\forall n$, for some constant $C < \infty$

Then $X_n \to X$ in L^1

ie $\mathbb{E}(|X_n - X|) \to 0$

Proof By Theorem 2.5.1 ii, $\exists$ a subsequence (n_k) such that $X_{n_k} \to X$ almost surely

So $|X| = \lim_{k\to\infty} |X_{n_k}|$
$\leq C$ almost surely

Note that $|X_n - X| = \underbrace{|X_n - X|}_{\leq \frac{\varepsilon}{2}} 1_{\{|X_n - X| \leq \frac{\varepsilon}{2}\}} + \underbrace{|X_n - X|}_{\leq 2C} 1_{\{|X_n - X| > \frac{\varepsilon}{2}\}}$

$\leq \frac{\varepsilon}{2} + 2C\, 1_{\{|X_n - X| > \frac{\varepsilon}{2}\}}$

$\mathbb{E}|X_n - X| \leq \frac{\varepsilon}{2} + 2C\, \mathbb{P}(|X_n - X| > \frac{\varepsilon}{2})$

Given $\varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n \geq N$, $\mathbb{P}(|X_n - X| > \frac{\varepsilon}{2}) < \frac{\varepsilon}{4C}$

$\Rightarrow \mathbb{E}|X_n - X| < \varepsilon$ $\square$

6.2 Uniform integrability

Lemma 6.2.1 Let X be an integrable random variable

Given $\varepsilon > 0$, $\exists \delta > 0$ such that $A \in \mathcal{F}$, $\mathbb{P}(A) \leq \delta \Rightarrow \mathbb{E}(|X| 1_A) \leq \varepsilon$

Equivalently Set $I_X(\delta) = \sup\{\mathbb{E}(|X| 1_A) : A \in \mathcal{F}, \mathbb{P}(A) \leq \delta\}$

Then $I_X(\delta) \downarrow 0$ as $\delta \downarrow 0$

Proof Suppose not.

Then $\exists \varepsilon > 0$ and $A_n \in \mathcal{F}$

such that $\mathbb{P}(A_n) \leq \frac{1}{2^n}$, but $\mathbb{E}(|X| 1_{A_n}) \geq \varepsilon$, $\forall n$

By the first Borel-Cantelli lemma, $\mathbb{P}(A_n \text{ infinitely often}) = 0$

But then $|X| \geq |X| 1_{A_n} \to 0$

So by Dominated convergence $\varepsilon \leq \mathbb{E}(|X| 1_{A_n}) \to 0$ $\Rightarrow\Leftarrow$ $\square$

Let $\mathcal{X}$ be a family of random variables, $1 \leq p \leq \infty$

$\mathcal{X}$ is bounded in L^p if $\sup_{X \in \mathcal{X}} \|X\|_p < \infty$ $\quad \|X\|_p = (\mathbb{E}(|X|^p))^{\frac{1}{p}}$

$I_{\mathcal{X}}(\delta) = \sup\{\mathbb{E}(|X| 1_A) : X \in \mathcal{X}, A \in \mathcal{F}, \mathbb{P}(A) \leq \delta\}$

$\mathcal{X}$ is bounded in L^1 $\Leftrightarrow$ $I_{\mathcal{X}}(1) < \infty$

$\mathcal{X}$ is uniformly integrable if $\mathcal{X}$ is bounded in L^1 and $I_{\mathcal{X}}(\delta) \downarrow 0$ as $\delta \downarrow 0$

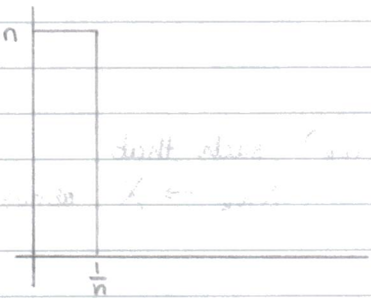
Jenson's inequality $\Rightarrow \|X\|_p \leq \|X\|_r$ whenever $p \leq r$

Holder's inequality $\Rightarrow \mathbb{E}(|X| 1_A) \leq \|X\|_p \, \mathbb{P}(A)^{\frac{1}{q}}$, where $\frac{1}{p} + \frac{1}{q} = 1$

So L^r - bounded $\Rightarrow L^p$ - bounded
$\Rightarrow$ uniformly integrable $\quad r \geq p > 1$
$\Rightarrow L^1$ - bounded

But L^1 - bounded $\not\Rightarrow$ uniformly integrable

example



$X_n = n 1_{(0, \frac{1}{n})}$

$\mathbb{E}|X_n| = 1 \quad \forall n$

but $\{X_n : n \in \mathbb{N}\}$ is not uniformly integrable

By Lemma 6·2·1, X is an integrable random variable (ie $X \in L^1$)
$\Rightarrow \{X\}$ is uniformly integrable

Similarly $X_1, \dots, X_n$ are integrable random variables
$\Rightarrow \{X_1, \dots, X_n\}$ is uniformly integrable

Suppose Y is an integrable random variable (ie $\mathbb{E}|Y| < \infty$)

Then $\mathcal{X} = \{X : X$ is a random variable, $|X| \leq Y\}$ is uniformly integrable

since $\mathbb{E}(|X| 1_A) \leq \mathbb{E}(Y 1_A) \quad \forall A$ (compare to dominated convergence)

Lemma 6·2·2 Let $\mathcal{X}$ be a family of random variables

Then $\mathcal{X}$ is uniformly integrable $\Leftrightarrow \sup_{X \in \mathcal{X}} \mathbb{E}(|X| 1_{|X| \geq K}) \to 0$ as $K \to \infty$

Proof $\Rightarrow$ Suppose $\mathcal{X}$ is uniformly integrable

By Chebyshev's inequality, $\mathbb{P}(|X| \geq K) \leq \frac{1}{K} \mathbb{E}|X|$
$\leq \frac{1}{K} I_{\mathcal{X}}(1) \qquad I_{\mathcal{X}}(1) = \sup_{X \in \mathcal{X}} \mathbb{E}|X|$

Given $\varepsilon > 0$, $\exists \, \delta > 0$ such that
$A \in \mathcal{F}, \mathbb{P}(A) \leq \delta \Rightarrow \mathbb{E}(|X| 1_A) \leq \varepsilon \quad \forall X \in \mathcal{X}$

Take $K = \frac{I_{\mathcal{X}}(1)}{\delta}$

Then $\mathbb{P}(|X| \geq K) \leq \delta \Rightarrow \mathbb{E}(|X| 1_{|X| \geq K}) \leq \varepsilon \quad \forall X \in \mathcal{X}$

$\Leftarrow$ Note that $|X| 1_A \leq K 1_A + |X| 1_{|X| \geq K}$
$\Rightarrow \mathbb{E}|X| \leq K + \mathbb{E}(|X| 1_{|X| \geq K})$
$\leq K + 1$ for K sufficiently large
$\Rightarrow \mathcal{X}$ is L^1 - bounded

Also $\sup_{A \in \mathcal{F}, \mathbb{P}(A) \leq \delta} \mathbb{E}(|X| 1_A) \leq K\delta + \sup_{X \in \mathcal{X}} \mathbb{E}(|X| 1_{|X| \geq K})$

Given $\varepsilon > 0$, choose K so that $\mathbb{E}(|X| 1_{|X| \geq K}) \leq \frac{\varepsilon}{2} \quad \forall X \in \mathcal{X}$

Then set $\delta = \frac{\varepsilon}{2K}$, to give $\mathbb{E}(|X| 1_A) \leq K \cdot \frac{\varepsilon}{2K} + \frac{\varepsilon}{2}$
$= \varepsilon$

$\Rightarrow I_{\mathcal{X}}(\delta) \downarrow 0$ as $\delta \downarrow 0 \qquad \square$

Probability and Measure

Theorem 6·2·3 Let X be a random variable, $(X_n : n \in \mathbb{N})$ a sequence of random variables. Then the following are equivalent

i. $X_n \in L^1 \ \forall n$, $X \in L^1$, and $X_n \to X$ in L^1
(ie $\mathbb{E}|X_n - X| \to 0$)

ii. $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable, and $X_n \to X$ in probability
(ie $\mathbb{P}(|X_n - X| \geq \varepsilon) \to 0$ as $\varepsilon \to 0$)

Proof i ⇒ ii By Chebyshev's inequality, for $\varepsilon > 0$ (ε fixed)

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon}\mathbb{E}|X_n - X|$$
$$\to 0$$

$\Rightarrow X_n \to X$ in probability

Given $\varepsilon > 0$, choose $N \in \mathbb{N}$ so that $\mathbb{E}|X_n - X| \leq \frac{\varepsilon}{2} \ \forall n > N$
and choose $\delta > 0$ so that

$$A \in \mathcal{F},\ \mathbb{P}(A) \leq \delta \Rightarrow \mathbb{E}(|X_n| 1_A) \leq \varepsilon \quad \forall 1 \leq n \leq N$$
$$\text{and } \mathbb{E}(|X| 1_A) \leq \frac{\varepsilon}{2}$$

Then, for $n \geq N$, $|X_n| 1_A \leq |X| 1_A + |X - X_n|$

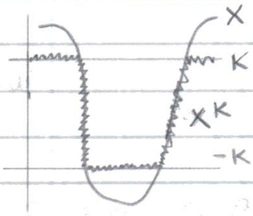
$$\Rightarrow \mathbb{E}(|X_n| 1_A) \leq \mathbb{E}(|X| 1_A) + \mathbb{E}|X - X_n|$$
$$\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2}$$
$$= \varepsilon$$

So $A \in \mathcal{F},\ \mathbb{P}(A) \leq \delta \Rightarrow \mathbb{E}(|X_n| 1_A) \leq \varepsilon \quad \forall n \in \mathbb{N}$

Also $\|X_n\|_1 \to \|X\|_1 \Rightarrow \{X_n : n \in \mathbb{N}\}$ is bounded in L^1

Hence $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable ii ⇒ i $\exists$ a subsequence (n_j) such that $X_{n_j} \to X$ almost surely

$$\Rightarrow \mathbb{E}|X| = \mathbb{E}(\liminf |X_{n_j}|)$$
$$\leq \liminf \mathbb{E}|X_{n_j}| \quad \text{by Fatou's Lemma}$$
$$< \infty$$



Consider the uniformly bounded sequence
$$X_n^K = \min\{\max\{-K, X_n\}, K\}$$
and set $X^K = \min\{\max\{-K, X\}, K\}$

Note that $|X^K| \leq K$

Now $|X - X_n| \leq \underbrace{|X - X^K|}_{\leq |X| 1_{|X| \geq K}} + \underbrace{|X^K - X_n^K|}_{\substack{\leq |X - X_n| \\ \to 0 \text{ in probability}}} + \underbrace{|X_n^K - X_n|}_{\leq |X_n| 1_{|X_n| \geq K}}$

$$\Rightarrow \mathbb{E}|X - X_n| \leq \mathbb{E}(|X| 1_{|X| \geq K}) + \mathbb{E}|X^K - X_n^K| + \mathbb{E}(|X_n| 1_{|X_n| \geq K})$$

Given $\varepsilon > 0$

$|X| \geq |X| 1_{|X| \geq K} \to 0$ almost surely

so by dominated convergence $\exists K_1$ such that $\mathbb{E}(|X| 1_{|X| \geq K_1}) \leq \frac{\varepsilon}{3}$

$\{X_n : n \in \mathbb{N}\}$ is uniformly integrable

$\Rightarrow \exists K_2$ such that $\mathbb{E}(|X| 1_{|X| \geq K_2}) \leq \frac{\varepsilon}{3}$

Let $K = \max\{K_1, K_2\}$

For fixed K, $X_n^K \to X^K$ in probability

so by dominated convergence, $\exists N \in \mathbb{N}$ such that
$$n > N \Rightarrow \mathbb{E}|X_n^K - X^K| < \frac{\varepsilon}{3}$$

$$\Rightarrow \mathbb{E}|X - X_n| \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}$$
$$= \varepsilon$$

$\Rightarrow X_n \to X$ in L^1 $\square$

Probability and Measure

§7 Fourier Transforms

Notation In this section only, for $p \in [1, \infty)$, write $L^p = L^p(\mathbb{R}^d)$ for the set of complex-valued Borel measurable functions on $\mathbb{R}^d$ such that

$$\|f\|_p = \left(\int_{\mathbb{R}^d} |f(x)|^p \, dx\right)^{\frac{1}{p}} < \infty$$

Definition For $f = u + iv \in L^1$

$$\int_{\mathbb{R}^d} f(x)\,dx = \int_{\mathbb{R}^d} u(x)\,dx + i\int_{\mathbb{R}^d} v(x)\,dx$$

Notes For $\alpha \in \mathbb{C}$, $\int_{\mathbb{R}^d} \alpha f(x)\,dx = \alpha \int_{\mathbb{R}^d} f(x)\,dx$

$$\begin{aligned}\left|\int_{\mathbb{R}^d} f(x)\,dx\right| &= e^{i\theta}\int_{\mathbb{R}^d} f(x)\,dx \qquad \text{for some } \theta \\ &= \int_{\mathbb{R}^d} \underbrace{e^{i\theta} f(x)\,dx}_{\in \mathbb{R}, \text{ by the first equality}} \\ &= \int_{\mathbb{R}^d} \mathrm{Re}(e^{i\theta} f(x))\,dx \\ &\leq \int_{\mathbb{R}^d} |f(x)|\,dx\end{aligned}$$

7.1 Definitions

Definition For a function $f \in L^1(\mathbb{R}^d)$ the Fourier transform $\hat{f}$ is $\hat{f}(u) = \int_{\mathbb{R}^d} f(x) e^{i\langle u, x\rangle}\,dx$, $u \in \mathbb{R}^d$.
where $\langle u, x\rangle = \sum_{i=1}^{n} u_i x_i$

Note $|\hat{f}(u)| \leq \|f\|_1 < \infty$

Suppose $u_n \to u$
Then $2 \geq |e^{i\langle u, x\rangle} - e^{i\langle u_n, x\rangle}| \to 0$ as $n \to \infty$, $\forall x$

$$\begin{aligned}\text{So } |\hat{f}(u_n) - \hat{f}(u)| &= \left|\int_{\mathbb{R}^d} f(x)(e^{i\langle u,x\rangle} - e^{i\langle u_n, x\rangle})\,dx\right| \\ &\leq \int_{\mathbb{R}^d} |f(x)||e^{i\langle u,x\rangle} - e^{i\langle u_n,x\rangle}|\,dx \\ &\to 0 \quad \text{by dominated convergence with } |f| \text{ as the dominating function}\end{aligned}$$

Hence $\hat{f} : \mathbb{R}^d \to \mathbb{C}$ is a continuous, bounded function

For $f \in L^1(\mathbb{R}^d)$ with $\hat{f} \in L^1(\mathbb{R}^d)$, we say that the Fourier inversion formula holds for f if

$$f(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \hat{f}(u) e^{-i\langle u, x\rangle}\,du \qquad \text{for almost all } x$$

For $f \in L^1 \cap L^2(\mathbb{R}^d)$, we say that the Plancherel identity holds for f if $\|\hat{f}\|_2 = (2\pi)^{\frac{d}{2}}\|f\|_2$

The main results of this section establish that, $\forall f \in L^1(\mathbb{R}^d)$, the inversion formula holds whenever $\hat{f} \in L^1(\mathbb{R}^d)$ and the Plancherel formula holds whenever $f \in L^2(\mathbb{R}^d)$

For a finite Borel measure μ on $\mathbb{R}^d$
the Fourier transform $\hat{\mu}$ is $\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} \mu(dx)$, $u \in \mathbb{R}^d$

$\hat{\mu}$ is a continuous function on $\mathbb{R}^d$ with $|\hat{\mu}(u)| \leq \mu(\mathbb{R}^d)$ $\forall u$
The definitions are consistent in that,
if μ has density f with respect to Lebesgue measure, then $\hat{\mu} = \hat{f}$

For a random variable X in $\mathbb{R}^d$
the characteristic function ϕ_X is $\phi_X(u) = \mathbb{E}(e^{i\langle u,X\rangle})$, $u \in \mathbb{R}^d$,
which is the Fourier transform of its law μ_X

So $\phi_X(u) = \mathbb{E}(e^{i\langle u,X\rangle})$
$= \hat{\mu}_X(u) = \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} \mu_X(dx)$

7.2 Convolutions

If f is Borel measurable on $\mathbb{R}^d$
then $(x,y) \mapsto f(x-y)$ is Borel measurable on $\mathbb{R}^d \times \mathbb{R}^d$
(by a monotone class argument)

Definition
Let $p \in [1,\infty)$, $f \in L^p(\mathbb{R}^d)$, ν be a probability measure on $\mathbb{R}^d$
Then convolution $f * \nu \in L^p(\mathbb{R}^d)$
is $f*\nu(x) = \int_{\mathbb{R}^d} f(x-y)\,\nu(dy)$ whenever the integral exists
ie if $f(x-.) \in L^1(\nu)$
$f*\nu(x) = 0$ otherwise

Note
$\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}|f(x-y)|\,\nu(dy)\right)^p dx \leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|f(x-y)|^p\,\nu(dy)\,dx$
by Jenson's inequality
$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|f(x-y)|^p\,dx\,\nu(dy)$
by Fubini's theorem
$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|f(x)|^p\,dx\,\nu(dy)$
$= \|f\|_p^{\,p}$
$< \infty$

so the integral defining the convolution exists for almost all x

$\|f*\nu\|_p^{\,p} = \int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d} f(x-y)\,\nu(dy)\right|^p dx$
$\leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|f(x-y)|^p\,\nu(dy)\,dx$ by Jenson's inequality
$= \|f\|_p^{\,p}$
$\Rightarrow f*\nu \in L^p(\mathbb{R}^d)$
with $\|f*\nu\|_p \leq \|f\|_p$

Notation
If ν has a density function g with respect to Lebesgue measure
then we write $f*g$ for $f*\nu$

Probability and Measure

Definition — Let μ, ν be probability measures on $\mathbb{R}^d$
X, Y be independent random variables with distributions μ, ν
Then the convolution $\mu * \nu$ is the distribution of $X + Y$
So $\mu * \nu (A) = \mathbb{P}(X + Y \in A)$
$= \int_{\mathbb{R}^d \times \mathbb{R}^d} 1_A (x+y) \mu(dx) \nu(dy), \quad A \in \mathcal{B}$

Note — Suppose μ has density function f with respect to Lebesgue measure.
Then $\mu * \nu (A) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1_A (x+y) f(x) dx \, \nu(dy)$
$= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1_A (x) f(x-y) dx \, \nu(dy)$ by Fubini's Theorem
$= \int_{\mathbb{R}^d} 1_A (x) f * \nu (x) dx$
so $\mu * \nu$ has density function $f * \nu$

exercise — By Fubini's Theorem, $\widehat{f * \nu}(u) = \hat{f}(u) \hat{\nu}(u)$
$\forall f \in L^1(\mathbb{R}^d)$
and $\forall$ probability measures ν on $\mathbb{R}^d$

Similarly, $\widehat{\mu * \nu}(u) = \mathbb{E}(e^{i \langle u, X+Y \rangle})$
$= \mathbb{E}(e^{i \langle u, X \rangle}) \mathbb{E}(e^{i \langle u, Y \rangle})$
$= \hat{\mu}(u) \hat{\nu}(u)$

7.3 Gaussians

Let $t \in (0, \infty)$
The centred Gaussian probability density function g_t on $\mathbb{R}^d$ of variance t
is $g_t(x) = \frac{1}{(2\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{2t}}, \quad x \in \mathbb{R}^d$

Consider d independent standard normal random variables $Z_1, \dots, Z_d$
$Z_k \sim N(0,1)$
Z_k has density $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ on $\mathbb{R}$
Set $Z = (Z_1, \dots, Z_d)$
Then Z has density g_1
and so $\sqrt{t} Z$ has density g_t

Computing $\hat{g}_t$

First consider the case $t = 1, d = 1$
Let X be a standard one-dimensional normal random variable, $X \sim N(0,1)$
$\hat{g}_1(u) = \mathbb{E}(e^{iuX})$
Since X is integrable, by Theorem 3.5.1, the characteristic function $\hat{g}_1$ is differentiable and we can differentiate under the integral sign to obtain
$\hat{g}_1'(u) = \mathbb{E}(iZ e^{iuZ})$
$= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iux} ix e^{-\frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} i(iue^{iux})(e^{-\frac{x^2}{2}}) dx$ integrating by parts
$= -u \hat{g}_1(u)$

So $\frac{d}{du}\left(e^{\frac{u^2}{2}}\hat{g}_1(u)\right) = 0$

$\Rightarrow \quad \hat{g}_1(u) = e^{-\frac{u^2}{2}}\hat{g}_1(0)$

$= e^{-\frac{u^2}{2}}$

Now take d independent standard normal random variables $X_1, \dots, X_d$ and set $X = (X_1, \dots, X_d)$

Then $\sqrt{t}\,X$ has density g_t

So $\hat{g}_t(u) = \mathbb{E}(e^{i\langle u, \sqrt{t}X\rangle})$

$= \mathbb{E}\left(\prod_{j=1}^{d} e^{iu_j\sqrt{t}X_j}\right)$

$= \prod_{j=1}^{d}\hat{g}_1(u_j\sqrt{t})$

$= e^{-|u|^2\frac{t}{2}}$ $\qquad |u|^2 = \sum_{k=1}^{d} u_k^2$

Hence $\hat{g}_t = (2\pi)^{\frac{d}{2}}\,t^{-\frac{d}{2}}\,g_{\frac{1}{t}}$

and $\hat{\hat{g}}_t = (2\pi)^d\,g_t$

Then $g_t(x) = g_t(-x)$

$= \frac{1}{(2\pi)^d}\hat{\hat{g}}_t(-x)$

$= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\hat{g}_t(u)e^{-i\langle u, x\rangle}\,du$

so the Fourier inversion formula holds for g_t

7.4 Gaussian convolutions

Definition

A Gaussian convolution is any function of the form $f * g_t$
where $f \in L^1(\mathbb{R}^d)$
g_t is a Gaussian, $t \in (0, \infty)$

Note

Suppose $x_n \to x$

Then $f * g_t(x_n) = \int_{\mathbb{R}^d} f(x_n - y)g_t(y)\,dy$

$= \int_{\mathbb{R}^d} g_t(x_n - y)\underbrace{f(y)}_{\in L^1}\,dy$

$\to \int_{\mathbb{R}^d} g_t(x - y)f(y)\,dy$ $\quad$ with $c|f|$ as the dominating function

by dominated convergence

$= f * g_t(x)$

$\Rightarrow \quad f * g_t$ is continuous

$\|f * g_t\|_1 \leq \|f\|_1$

$\|f * g_t\|_\infty \leq (2\pi t)^{-\frac{d}{2}}\|f\|_1$

Also $\widehat{f * g_t}(u) = \hat{f}(u)\,\hat{g}_t(u)$

$\Rightarrow \|\widehat{f * g_t}\|_1 \leq \|\hat{f}\|_\infty\|\hat{g}_t\|_1$

$\leq \|\hat{f}\|_\infty (2\pi)^{\frac{d}{2}}t^{-\frac{d}{2}}$

$\|\widehat{f * g_t}\|_\infty \leq \|\hat{f}\|_1$

Probability and Measure

exercise — By the parallelogram identity in $\mathbb{R}^d$, we have $g_s * g_s = g_{2s}$

So for any probability measure μ on $\mathbb{R}^d$ and any $s \in (0, \infty)$ we have $\mu * g_s \in L^1(\mathbb{R}^d)$

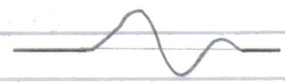
$$\Rightarrow \mu * g_{2s} = \mu * (g_s * g_s)$$
$$= (\mu * g_s) * g_s \quad \text{is a Gaussian convolution}$$

Lemma 7.4.1 — The Fourier inversion formula holds for all Gaussian convolutions

Proof — Let $f \in L^1(\mathbb{R}^d)$, $t > 0$

Then
$$(2\pi)^d f * g_t(x) = (2\pi)^d \int_{\mathbb{R}^d} f(x-y)\, g_t(y)\, dy$$
$$= \int_{\mathbb{R}^d} f(x-y) \int_{\mathbb{R}^d} \hat{g}_t(u)\, e^{-i\langle u, y\rangle}\, du\, dy$$
since the inversion formula holds for g_t
$$= \int_{\mathbb{R}^d} \hat{g}_t(u) \int_{\mathbb{R}^d} f(x-y)\, e^{i\langle u, x-y\rangle}\, dy\, e^{-i\langle u, x\rangle}\, du$$
by Fubini's theorem
$$= \int_{\mathbb{R}^d} \hat{g}_t(u)\, \hat{f}(u)\, e^{-i\langle u, x\rangle}\, du$$
$$= \int_{\mathbb{R}^d} \widehat{f * g_t}(u)\, e^{-i\langle u, x\rangle}\, du \qquad \square$$

Definition — A function of compact support is a function such that the closure of the set of points where the function is non-zero is compact



Lemma 7.4.2 — Let $f \in L^p(\mathbb{R}^d)$, where $p \in [1, \infty)$

Then $\| f * g_t - f \|_p \to 0$ as $t \to 0$

Proof — Given $\varepsilon > 0$ $\exists$ a continuous function h on $\mathbb{R}^d$ of compact support such that $\| f - h \|_p \leq \frac{\varepsilon}{3}$

Then
$$\| f * g_t - h * g_t \|_p = \| (f-h) * g_t \|_p$$
$$\leq \| f - h \|_p$$
$$\leq \frac{\varepsilon}{3}$$

Consider $e(y) = \int_{\mathbb{R}^d} | h(x-y) - h(x) |^p \, dx$

Then $e(y) \leq 2^p \| h \|_p^p < \infty \quad \forall y$

so by dominated convergence $e(y) \to 0$ as $y \to 0$ since h is continuous

Now
$$\| h - h * g_t \|_p^p = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (h(x) - h(x-y))\, g_t(y)\, dy \right|^p dx$$
$$\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} | h(x) - h(x-y) |^p\, g_t(y)\, dy\, dx$$
by Jenson's inequality
$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} | h(x) - h(x-y) |^p\, dx\, g_t(y)\, dy$$
by Fubini's theorem
$$= \int_{\mathbb{R}^d} e(y)\, g_t(y)\, dy$$
$$= \int_{\mathbb{R}^d} e(\sqrt{t}\, y)\, g_1(y)\, dy$$
$$\leq \left(\frac{\varepsilon}{3}\right)^p \quad \text{for } t \text{ sufficiently small}$$
(by dominated convergence)

$$\| f * g_t - f \|_p \leq \| f * g_t - h * g_t \|_p + \| h * g_t - h \|_p + \| h - f \|_p$$
$$\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \quad \text{for } t \text{ sufficiently small}$$
$$= \varepsilon \qquad \square$$

7·5 Uniqueness and inversion

Theorem 7·5·1 Let $f \in L^1(\mathbb{R}^d)$

For $t > 0$, $x \in \mathbb{R}^d$, set $f_t(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(u) e^{-|u|^2 \frac{t}{2}} e^{-i\langle u, x\rangle} du$

Then $\| f_t - f \|_1 \to 0$ as $t \to 0$

Moreover, the Fourier inversion formula holds for f whenever $f \in L^1(\mathbb{R}^d)$ and $\hat{f} \in L^1(\mathbb{R}^d)$

Proof Consider the Gaussian convolution $f * g_t$

Then $\widehat{f * g_t}(u) = \hat{f}(u)\, \hat{g_t}(u)$

$= \hat{f}(u)\, e^{-|u|^2 \frac{t}{2}}$

So $f_t = f * g_t$ by Lemma 7·4·1

$\Rightarrow \| f_t - f \|_1 \to 0$ as $t \to 0$ by Lemma 7·4·2

Moreover, if $\hat{f} \in L^1(\mathbb{R}^d)$, then by dominated convergence (dominated by $|\hat{f}|$)

$f_t(x) \to \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(u) e^{-i\langle u, x\rangle} dx$ as $t \to 0$

On the other hand, $\exists$ a sequence $t_n \to 0$ such that $f_{t_n}(x) \to f(x)$ almost everywhere

So the inversion formula holds for f. $\square$

7·6 Fourier transform in $L^2(\mathbb{R}^d)$

Recall Plancherel identity

$\| \hat{f} \|_2 = (2\pi)^{\frac{d}{2}} \| f \|_2$

Theorem 7·6·1 The Plancherel identity holds $\forall f \in L^1 \cap L^2(\mathbb{R}^d)$

Moreover there is a unique Hilbert space automorphism $F : \mathcal{L}^2 \to \mathcal{L}^2$ such that $F[f] = [(2\pi)^{-\frac{d}{2}} \hat{f}]$ $\forall f \in L^1 \cap L^2(\mathbb{R}^d)$

Note $F(\alpha u + \beta v) = \alpha F(u) + \beta F(v)$ $\forall \alpha, \beta \in \mathbb{R}$, $u, v \in \mathcal{L}^2$

$\| F(u) \|_2 = \| u \|_2$

Proof First suppose that $f \in L^1$ and $\hat{f} \in L^1$ $(\Rightarrow f, \hat{f} \in L^\infty)$

Then $(2\pi)^d \| f \|_2^2 = (2\pi)^d \int_{\mathbb{R}^d} f(x) \overline{f(x)}\, dx$

$= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \hat{f}(u) e^{-i\langle u, x\rangle} du \right) \overline{f(x)}\, dx$

by the inversion formula (since $f \in L^1$)

$= \int_{\mathbb{R}^d} \overline{\left(\int_{\mathbb{R}^d} f(x) e^{i\langle u, x\rangle} dx \right)} \hat{f}(u)\, du$

by Fubini's theorem (since $(x, u) \mapsto f(x)\hat{f}(u)$ is integrable on $\mathbb{R}^d \times \mathbb{R}^d$)

$= \int_{\mathbb{R}^d} \overline{\hat{f}(u)}\, \hat{f}(u)\, du$

$= \| \hat{f} \|_2^2$

Now suppose $f \in L^1 \cap L^2$

Consider for $t > 0$ the Gaussian convolution $f_t = f * g_t$

Then $f_t \to f$ in L^2 as $t \to 0$ by Lemma 7·4·2

$\Rightarrow \|f_t\|_2 \to \|f\|_2$

Also $\hat{f}_t(u) = \hat{f}(u) e^{-|u|^2 \frac{t}{2}}$

$$\Rightarrow \|\hat{f}_t\|_2^2 = \int_{\mathbb{R}^d} |\hat{f}(u)|^2 e^{-|u|^2 t} du$$
$$\uparrow \int_{\mathbb{R}^d} |\hat{f}(u)|^2 du \quad \text{as } t \downarrow 0 \quad \text{by Monotone convergence}$$
$$= \|\hat{f}\|_2^2$$

But $f_t, \hat{f}_t \in L^1 \Rightarrow$ the Plancherel identity holds for f_t and so, on letting $t \to 0$, we obtain the Plancherel identity for f

Define $F_0 : \mathcal{L}^1 \cap \mathcal{L}^2 \to \mathcal{L}^2$

by $F_0[f] = [(2\pi)^{-\frac{d}{2}} \hat{f}]$

Then by Plancherel's identity, F_0 is a linear isometry (of $\|\cdot\|_2$)

ie F_0 preserves the $\mathcal{L}^2$ norm

$\mathcal{L}^1 \cap \mathcal{L}^2$ is dense in $\mathcal{L}^2$ (by Gaussian convolutions)

$\Rightarrow F_0$ extends uniquely to a linear isometry F from $\mathcal{L}^2$ injectively into $\mathcal{L}^2$

Consider $V = \{[f] : f \in L^1, \hat{f} \in L^1\}$

By the inversion formula, $F_0(V) \subseteq V$

$F_0^4[f] = [f] \quad \forall [f] \in V$

But V contains all Gaussian convolutions

$\Rightarrow V$ is dense in $\mathcal{L}^2$

$\Rightarrow F$ is surjective onto $\mathcal{L}^2$ $\square$

7·7 Weak convergence and characteristic functions

Let μ be a Borel probability measure on $\mathbb{R}^d$

$(\mu_n : n \in \mathbb{N})$ be a sequence of Borel probability measures on $\mathbb{R}^d$

continuous bounded functions

Definition $\mu_n \to \mu$ weakly if $\mu_n(f) \to \mu(f) \ \forall f \in C_b(\mathbb{R}^d)$

differentiable with continuous derivative

exercise $\mu_n \to \mu$ weakly $\Leftrightarrow \mu_n(f) \to \mu(f) \ \forall f \in C_c^1(\mathbb{R}^d)$

compact support

exercise If $\mu_n \to \mu$ weakly and $\mu_n \to \nu$ weakly, then $\mu = \nu$

Let X be a random variable in $\mathbb{R}^d$

$(X_n : n \in \mathbb{N})$ be a sequence of random variables in $\mathbb{R}^d$

Definition $X_n \to X$ weakly if $\mu_{X_n} \to \mu_X$ weakly

exercise For $d = 1$, weak convergence is equivalent to convergence in distribution [as defined in section 2·5]

Note If X is a weak limit of the sequence of random variables $(X_n : n \in \mathbb{N})$ then so is any other random variable with the same distribution as X.

Theorem 7.7.1 Let X be a random variable in $\mathbb{R}^d$

Then the distribution μ_X of X is uniquely determined by its characteristic function

$$\phi_X(u) = \mathbb{E}(e^{i\langle u, x\rangle})$$

Moreover, if ϕ_X is integrable (ie $\phi_X \in L^1(\mathbb{R}^d)$)

then μ_X has a continuous bounded density function given by

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi_X(u) e^{-i\langle u, x\rangle} du$$

Moreover, if $(X_n : n \in \mathbb{N})$ is a sequence of random variables in $\mathbb{R}^d$ such that $\phi_{X_n}(u) \to \phi_X(u)$ as $n \to \infty$, $\forall u \in \mathbb{R}^d$,

then $X_n \to X$ weakly

Proof Let Z be a standard Gaussian random variable in $\mathbb{R}^d$, independent of X

So Z has density g_1

$\sqrt{t}\, Z$ has density g_t

$X + \sqrt{t}\, Z$ has density given by the Gaussian convolution $f_t = \mu_X * g_t$

We have $\hat{f}_t(u) = \phi_X(u) e^{-|u|^2 \frac{t}{2}}$, so, by the Fourier inversion formula,

$$f_t(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi_X(u) e^{-|u|^2 \frac{t}{2}} e^{-i\langle u, x\rangle} du$$

$\forall g \in C_b(\mathbb{R}^d)$ (ie $\forall$ continuous bounded functions g on $\mathbb{R}^d$)

$$\begin{aligned}\int_{\mathbb{R}^d} g(x) f_t(x)\, dx &= \mathbb{E}(g(X + \sqrt{t}\, Z)) \\ &\to \mathbb{E}(g(X)) \quad \text{by bounded convergence} \\ &= \int_{\mathbb{R}^d} g(x) \mu_X(dx)\end{aligned}$$

$\Rightarrow \phi_X$ determines μ_X uniquely

Suppose $\phi_X \in L^1$ (ie ϕ_X is integrable)

Then $|f_t(x)| \leq (2\pi)^d \|\phi_X\|_1 \quad \forall x$

$\Rightarrow f_t(x) \to f_X(x) \quad \forall x$ by dominated convergence (dominated by $|\phi_X|$)

So $f_t(x) \geq 0 \quad \forall t, x \Rightarrow f_X(x) \geq 0 \quad \forall x$

Consider $g \in C_c(\mathbb{R}^d)$ (ie g of compact support)

$$\begin{aligned}\int_{\mathbb{R}^d} g(x) f_X(x)\, dx &= \lim_{t \to 0} \int_{\mathbb{R}^d} g(x) f_t(x)\, dx \quad \text{by bounded convergence} \\ &= \int_{\mathbb{R}^d} g(x) \mu_X(dx)\end{aligned}$$

$\Rightarrow \mu_X$ has density function f_X

Suppose $(X_n : n \in \mathbb{N})$ is a sequence of random variables such that $\phi_{X_n}(u) \to \phi_X(u) \quad \forall u$

Consider $g \in C_c^1(\mathbb{R}^d)$ (so g is differentiable, with bounded derivative)

Given $\varepsilon > 0$, $\exists\, t > 0$ such that $\sqrt{t}\, \|\nabla g\|_\infty \mathbb{E}|Z| \leq \frac{\varepsilon}{3}$

Then $\mathbb{E}|g(X + \sqrt{t}\, Z) - g(X)| \leq \sqrt{t}\, \|\nabla g\|_\infty \mathbb{E}|Z|$ by the Mean Value Theorem

$\leq \frac{\varepsilon}{3}$ for t sufficiently small

Similarly $\mathbb{E}|g(X_n + \sqrt{t}\, Z) - g(X_n)| \leq \frac{\varepsilon}{3}$

Also $|\mathbb{E}(g(X_n + \sqrt{t}\, Z)) - \mathbb{E}(g(X + \sqrt{t}\, Z))|$

$$= \frac{1}{(2\pi)^d} \left| \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} \underbrace{(\phi_{X_n}(u) - \phi_X(u))}_{\substack{\to 0 \text{ by dominated convergence,} \\ \text{dominated by } 2}} \underbrace{e^{-|u|^2 \frac{t}{2}}}_{\in L^1(\mathbb{R}^d)} e^{-i\langle u, x\rangle} du \Big) \underbrace{g(x)}_{\in L^1(\mathbb{R}^d)} dx \right|$$

by the Fourier inversion formula

$\leq \frac{\varepsilon}{3}$ for n sufficiently large, by dominated convergence.

Then $|\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))| \leq |\mathbb{E}(g(X_n)) - \mathbb{E}(g(X_n + \sqrt{t} Z))| + |\mathbb{E}(g(X_n + \sqrt{t} Z)) - \mathbb{E}(g(X + \sqrt{t} Z))|$
$+ |\mathbb{E}(g(X + \sqrt{t}\, Z)) - \mathbb{E}(g(X))|$

$\leq \varepsilon \Rightarrow \mu_{X_n}(g) \to \mu_X(g)$ as $n \to \infty$ $\blacksquare$

$$|\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))| \leq \mathbb{E}|g(X_n) - g(X_n + \sqrt{t}Z)| + |\mathbb{E}(g(X_n + \sqrt{t}Z)) - \mathbb{E}(g(X + \sqrt{t}Z))| + \mathbb{E}|g(X + \sqrt{t}Z) - g(X)|$$
$$\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}$$
$$= \varepsilon$$

$\Rightarrow \mathbb{E}(g(X_n)) \to \mathbb{E}(g(X))$

$\Rightarrow \mu_{X_n}(g) \to \mu_X(g)$

$\Rightarrow X_n \to X$ weakly as $n \to \infty$ $\square$

Theorem: Lévy's ~~Theorem~~ continuity theorem for characteristic functions

If $\phi_{X_n}(u)$ converges as $n \to \infty$, with limit $\phi(u)$, say, $\forall u \in \mathbb{R}$ and if ϕ is continuous in a neighbourhood of 0 then ϕ is the characteristic function of some random variable X and $X_n \to X$ in distribution

This is a stronger version of the final assertion of Theorem 7.7.1

Proof: omitted

Claim: Let $(\mu_n : n \in \mathbb{N})$ be a sequence of Borel probability measures on $\mathbb{R}^d$

Suppose $\mu_n(f) \to \mu(f) \quad \forall f \in C_c^1(\mathbb{R}^d)$

Then $\mu_n(f) \to \mu(f) \quad \forall f \in C_b(\mathbb{R}^d)$

ie $\mu_n \to \mu$ weakly

Proof: Fix $f \in C_b(\mathbb{R}^d)$ with $\|f\|_\infty \leq 1$

Given $\varepsilon > 0$, $\exists R < \infty$ such that $\mu(\{x \in \mathbb{R}^d : |x| \geq R\}) \leq \frac{\varepsilon}{4}$

Then $\exists g$ continuous with $1_{|x| \leq R} \leq g(x) \leq 1_{|x| \leq R+1}$

$\exists N \in \mathbb{N}$ such that, $\forall n \geq K$, $\mu_n(g) \geq \mu(g) - \frac{\varepsilon}{4}$

But $\mu(1-g) \leq \frac{\varepsilon}{4}$

$$\Rightarrow \mu_n(1-g) = 1 - \mu_n(g)$$
$$\leq 1 - \mu(g) + \frac{\varepsilon}{4}$$
$$= \mu(1-g) + \frac{\varepsilon}{4}$$
$$\leq \frac{\varepsilon}{2}$$

$\exists M \in \mathbb{N}$ such that, $\forall n \geq M$, $|\mu_n(fg) - \mu(fg)| \leq \frac{\varepsilon}{4}$

Let $N = \max\{K, M\}$

Then $\forall n \geq N$,

$$|\mu_n(f) - \mu(f)| \leq |\mu_n(f(1-g))| + |\mu_n(fg) - \mu(fg)| + |\mu(f(1-g))|$$
$$\leq \mu_n(1-g) + |\mu_n(fg) - \mu(fg)| + \mu(1-g)$$
$$\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4}$$
$$= \varepsilon \quad \square$$

§8 Gaussian random variables

8.1 Gaussian random variables in $\mathbb{R}$

Definition A real-valued random variable X is Gaussian if it has density function
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ for some $\mu \in \mathbb{R}$, $\sigma^2 \in (0,\infty)$

We also admit as Gaussian any random variable X such that $X = \mu$ almost surely, for some $\mu \in \mathbb{R}$.
This degenerate case corresponds to the case $\sigma^2 = 0$

Notation $X \sim N(\mu, \sigma^2)$

Proposition 8.1.1 Suppose $X \sim N(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$. Then
- i. $\mathbb{E}(X) = \mu$ $\quad (X \in L^2(\mathbb{P}))$
- ii. $Var(X) = \sigma^2$
- iii. $aX + b \sim N(a\mu + b, a^2\sigma^2)$
- iv. $\phi_X(u) = e^{iu\mu - \frac{1}{2}u^2\sigma^2}$

8.2 Gaussian random variables in $\mathbb{R}^n$

Definition A random variable X in $\mathbb{R}^n$ is Gaussian if $\langle u, X\rangle \ (= \sum_{i=1}^n u_i X_i)$ is Gaussian in $\mathbb{R}$, $\forall u \in \mathbb{R}^n$

example Let $X_1, \dots, X_n$ be independent $N(0,1)$ random variables
$X = (X_1, \dots, X_n)$
Then $\mathbb{E}(e^{i\langle u, X\rangle}) = \mathbb{E}\left(\prod_{k=1}^n e^{iu_k X_k}\right)$
$= e^{-\frac{1}{2}|u|^2}$

$\Rightarrow \langle u, X\rangle \sim N(0, |u|^2) \quad \forall u \in \mathbb{R}^n$

$\Rightarrow X$ is Gaussian

Theorem 8.2.1 Let X be a Gaussian random variable in $\mathbb{R}^n$
Let A be an $m \times n$ matrix, and let $b \in \mathbb{R}^m$. Then
- i. $AX + b$ is a Gaussian random variable in $\mathbb{R}^m$
- ii. $X \in L^2(\mathbb{P})$
  and the distribution μ_X of X is determined by its mean $\mu = \mathbb{E}(X)$
  and its variance $V = Var(X)$
  $= \mathbb{E}((X-\mu)(X-\mu)^T)$
- iii. $\phi_X(u) = e^{i\langle u, \mu\rangle - \frac{1}{2}\langle u, Vu\rangle}$
- iv. If V is invertible, then X has a density function on $\mathbb{R}^n$, given by
$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det V}} e^{-\frac{1}{2}\langle x-\mu, V^{-1}(x-\mu)\rangle}$$

Probability and Measure

v. Suppose $X = (X_1, X_2)$ with X_1 in $\mathbb{R}^{n_1}$, X_2 in $\mathbb{R}^{n_2}$
Then $\operatorname{cov}(X_1, X_2) = 0 \Rightarrow X_1, X_2$ are independent

Proof

i. Take $u \in \mathbb{R}^n$
Then $\langle u, AX + b \rangle = \langle A^T u, X \rangle + \langle u, b \rangle$
$= \langle v, X \rangle + \langle u, b \rangle$ where $v = A^T u$
which is Gaussian by Proposition 8·1·1
$\Rightarrow AX + b$ is Gaussian

iii. Let $e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i^{\text{th}}$ component

Then $\forall i$, $\langle e_i, X \rangle = X_i$ is Gaussian
$\Rightarrow X_i \in L^2(\mathbb{P})$
$\Rightarrow X \in L^2(\mathbb{P})$

For $u \in \mathbb{R}^n$, we have $\mathbb{E}(\langle u, X \rangle) = \langle u, \mu \rangle$
$\operatorname{var}(\langle u, X \rangle) = \operatorname{cov}(\langle u, X \rangle, \langle u, X \rangle)$
$= \langle u, Vu \rangle$

Since $\langle u, X \rangle$ is Gaussian, we must have
$\langle u, X \rangle \sim N(\langle u, \mu \rangle, \langle u, Vu \rangle)$ by Proposition 8·1·1
$\phi_X(u) = \mathbb{E}(e^{i\langle u, X \rangle})$
$= e^{i\langle u, \mu \rangle - \frac{1}{2}\langle u, Vu \rangle}$

ii. follows from iii. by uniqueness of characteristic functions

iv. Let $Y_1, \dots, Y_n$ be independent $N(0,1)$ random variables
Then $Y = (Y_1, \dots, Y_n)$ has density $f_Y(y) = \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}|y|^2}$

Set $\tilde{X} = V^{\frac{1}{2}} Y + \mu$

[writing $V = UDU^T$, where $D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$, $\lambda_i \geq 0$,
define $V^{\frac{1}{2}} = UD^{\frac{1}{2}}U^T$, where $D^{\frac{1}{2}} = \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix}$
then $(V^{\frac{1}{2}})^2 = V$]

Then $\tilde{X}$ is Gaussian, with $\mathbb{E}(\tilde{X}) = \mu$
$\operatorname{Var}(\tilde{X}) = V$

$\Rightarrow X$ and $\tilde{X}$ have the same distribution
If V is invertible, then, by a linear change of variables in $\mathbb{R}^n$, $\tilde{X}$, and hence X, have density $f_X(x)$

v. Suppose $X = (X_1, X_2)$ with $\mathrm{cov}(X_1, X_2) = 0$

We have $V = \left(\begin{array}{c|c} V_{11} & \\ \hline & V_{22} \end{array}\right)$ where $V_{11} = \mathrm{var}(X_1)$

$V_{22} = \mathrm{var}(X_2)$

Let $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$

Then $\langle u, Vu \rangle = (u_1, u_2) V \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$

$= u_1^T V_{11} u_1 + u_2^T V_{22} u_2$

$= \langle u_1, V_{11} u_1 \rangle + \langle u_2, V_{22} u_2 \rangle$

So $\phi_X(u) = e^{i\langle u, \mu \rangle - \frac{1}{2}\langle u, Vu \rangle}$

$= \phi_{X_1}(u_1)\, \phi_{X_2}(u_2)$

$\Rightarrow X_1$ and X_2 are independent. $\square$

Probability and Measure

§9 Ergodic Theory

9.1 Measure-preserving transformations

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

Definitions: A measurable function $\theta : E \to E$ is a measure-preserving transformation if $\mu \circ \theta^{-1} = \mu$
ie $\mu(\theta^{-1}(A)) = \mu(A) \quad \forall A \in \mathcal{E}$

A set $A \in \mathcal{E}$ is invariant if $\theta^{-1}(A) = A$

$\mathcal{E}_\theta = \{ A \in \mathcal{E} : A \text{ is invariant} \}$

exercise: $\mathcal{E}_\theta$ is a σ-algebra

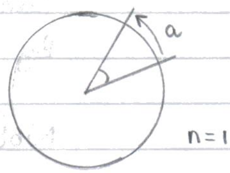
Definition: A measurable function f is invariant if $f = f \circ \theta$

exercise: Then f is invariant $\Leftrightarrow$ f is $\mathcal{E}_\theta$-measurable

Definition: θ is ergodic if $\forall A \in \mathcal{E}_\theta$, either $\mu(A) = 0$
or $\mu(A^c) = 0$

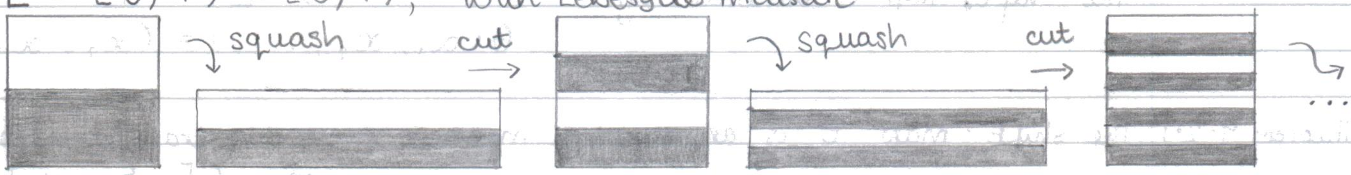
examples 1. $E = \{1, \dots, n\}$
$\mu(A) = |A|$
Then $\theta : E \to E$ is a measure-preserving transformation
$\Leftrightarrow$ θ is a bijection (ie a permutation)
and θ is ergodic $\Leftrightarrow$ θ is just one cycle

2. Rotations / Translation map on the torus
$E = [0, 1)^n$
with Lebesgue measure on its Borel σ-algebra



For each $a = (a_1, \dots, a_n) \in E$
set $\theta_a(x_1, \dots, x_n) = (x_1 + a_1 \pmod 1, \dots, x_n + a_n \pmod 1)$
Then θ_a is a measure-preserving transformation

3. Bakers' map
$E = [0,1) \times [0,1)$, with Lebesgue measure



$\theta(x) = 2x - \lfloor 2x \rfloor$
is a measure-preserving transformation.

Proposition 9.1.1 If f is integrable and θ is measure-preserving then $f \circ \theta$ is integrable and $\int_E f \, d\mu = \int_E f \circ \theta \, d\mu$

$f \in L^1(\mu)$
$f \circ \theta \in L^1(\mu)$
$\mu(f \circ \theta) = \mu(f)$

Proposition 9.1.2 If θ is an ergodic measure-preserving transformation and f is invariant then $\exists \; c \in \mathbb{R}$ such that $f = c$ almost surely

9.2 Bernoulli shifts

Let m be a Borel probability measure on $\mathbb{R}$
In section 2.4, we constructed a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which $\exists$ a sequence of independent random variables $(Y_n : n \in \mathbb{N})$, all having distribution m

Consider the infinite product space

$$E = \mathbb{R}^{\mathbb{N}} = \{ x = (x_n : n \in \mathbb{N}) : x_n \in \mathbb{R} \; \forall \, n \}$$

Define coordinate maps $X_n : E \to \mathbb{R}$ by $X_n(x) = x_n$
and set $\mathcal{E} = \sigma(X_n : n \in \mathbb{N})$, a σ-algebra on E
Then $\mathcal{E}$ is also generated by the π-system

$$\mathcal{A} = \prod_{n \in \mathbb{N}} = \{ x \in E : x_n \in A_n \; \forall \, n \}$$

where $A_n = \mathbb{R}$ $\forall$ but finitely many n

Define $Y : \Omega \to E$
by $Y(\omega) = (Y_n(\omega) : n \in \mathbb{N})$
Then Y is a measurable random variable
Let $\mu = \mathbb{P} \circ Y^{-1}$ be the image measure of Y
Then $\forall \, A \in \mathcal{A}$, $\mu(A) = \prod_{n \in \mathbb{N}} m(A_n)$
By uniqueness of extension, μ is the unique measure on $\mathcal{E}$ with this property

Note that, under the probability measure μ, the coordinate maps $(X_n : n \in \mathbb{N})$ are themselves a sequence of independent random variables with law m.

The probability space $(E, \mathcal{E}, \mu)$ is called the canonical model for such sequences.

The shift map is defined by $\theta : E \to E$

$$\theta(x_1, x_2, \ldots) = (x_2, x_3, \ldots)$$

Theorem 9.2.1 The shift-map θ is an ergodic measure-preserving transformation on $(E, \mathcal{E}, \mu)$

Recall

Tail σ-algebras

$\mathcal{T}_n = \sigma(X_m : m \geq n+1)$

$\mathcal{T} = \bigcap_{n \in \mathbb{N}} \mathcal{T}_n$

Proof that θ is ergodic

Consider $A = \prod_{n \in \mathbb{N}} A_n \in \mathcal{A}$

We have $\theta^{-n}(A) = \{X_{n+k} \in A_k \ \forall k\} \in \mathcal{T}_n$

Since $\mathcal{T}_n$ is a σ-algebra,

$\{A \in \mathcal{E} : \sigma^{-1}(A) \in \mathcal{T}_n\}$ is a σ-algebra containing $\mathcal{A}$

$\Rightarrow \theta^{-n}(A) \in \mathcal{T}_n \quad \forall A \in \mathcal{E}$

Hence if $A \in \mathcal{E}_\theta$

then $A = \theta^{-n}(A) \in \mathcal{T}_n$

$\Rightarrow A \in \bigcap_{n \in \mathbb{N}} \mathcal{T}_n \qquad \mathcal{E}_\theta \subseteq \mathcal{T}$

$\Rightarrow \mu(A) \in \{0, 1\}$ by Kolmogorov's zero-one law $\square$

9.3 Birkhoff's and von Neumann's ergodic theorems

Throughout this section, let $(E, \mathcal{E}, \mu)$ be a measure space

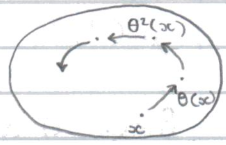
θ be a measure-preserving transformation

Given a measurable function f,

set $S_0 = 0$

and define $S_n = S_n(f)$

$= f + f \circ \theta + \dots + f \circ \theta^{n-1}$ for $n \geq 1$



Lemma 9.3.1 Maximal ergodic lemma

Let f be an integrable function on E

Set $S^* = S^*(f)$

$= \sup_{n \geq 0} S_n(f)$

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

Proof

Set $S_n^* = \max_{0 \leq m \leq n} S_m$

$A_n = \{S_n^* > 0\}$

Then, for $1 \leq m \leq n$, $\quad S_m = f + S_{m-1} \circ \theta$

$\leq f + S_n^* \circ \theta$

On A_n, $\quad S_n^* = \max_{1 \leq m \leq n} S_m$

$\Rightarrow S_n^* \leq f + S_n^* \circ \theta$

On A_n^c, $\quad S_n^* = 0$

$\leq S_n^* \circ \theta$

$\Rightarrow S_n^* \leq f 1_{A_n} + S_n^* \circ \theta$

$\int_E S_n^* \, d\mu \leq \int_{A_n} f \, d\mu + \int_E S_n^* \circ \theta \, d\mu$, $\quad$ S_n is integrable $\forall n$

$= \int_{A_n} f \, d\mu + \int_E S_n^* \, d\mu$ $\quad$ θ is measure-preserving

$< \infty$

$\Rightarrow \int_{A_n} f \, d\mu \geq 0 \quad \forall n$

Now $A_n \uparrow \{S^* > 0\}$ as $n \to \infty$, and f is integrable,
so by dominated convergence, dominated by $|f|$,

$$\int_{\{S^*>0\}} f \, d\mu = \lim_{n\to\infty} \int_{A_n} f \, d\mu \geq 0 \qquad \square$$

Theorem 9·3·2 Birkhoff's almost everywhere ergodic theorem

Suppose that $(E, \mathcal{E}, \mu)$ is σ-finite and that f is an integrable function on E

Then $\exists$ an invariant function $\bar{f}$ such that $\mu(|\bar{f}|) \leq \mu(|f|)$ and $\frac{S_n(f)}{n} \to \bar{f}$ almost everywhere as $n \to \infty$

Proof The functions $\liminf_n \left(\frac{S_n}{n}\right)$ and $\limsup_n \left(\frac{S_n}{n}\right)$ are invariant

$\Rightarrow$ for $a < b$,

$$D = D(a,b) = \left\{ \liminf_n \left(\frac{S_n}{n}\right) < a < b < \limsup_n \left(\frac{S_n}{n}\right) \right\} \text{ is invariant}$$

Claim $\mu(D) = 0$

Proof Since D is invariant, we can restrict μ, θ to D and reduce to the case $D = E$

Either $b > 0$ or $a < 0$; replacing f by $-f$ if necessary, assume $b > 0$

Let $B \in \mathcal{E}$ with $\mu(B) < \infty$

Then $g = f - b 1_B$ is integrable
and, for each $x \in D$, for some $n \in \mathbb{N}$,

$$S_n(g)(x) \geq S_n(f)(x) - nb > 0$$

$\Rightarrow$ $S^*(g) > 0$ everywhere on D

So by the Maximal ergodic lemma,

$$0 \leq \int_D g \, d\mu = \int_D f \, d\mu - b\mu(B)$$

Since μ is σ-finite, $\exists$ a sequence of sets $B_n \in \mathcal{E}$ with $\mu(B_n) < \infty$ $\forall n$ and $B_n \uparrow D$

$$\Rightarrow \quad b\mu(D) = \lim_{n\to\infty} b\mu(B_n) \leq \int_D f \, d\mu$$

$\Rightarrow$ $\mu(D) < \infty$

A similar argument, applied to $-f$ and $-a$, with $B = D$,
gives

$$(-a)\mu(D) \leq \int_D (-f) \, d\mu$$

Hence $b\mu(D) \le \int_D f \, d\mu \le a\mu(D)$
Since $a < b$ and $\int_D f \, d\mu < \infty$,
this $\Rightarrow \mu(D) = 0$

Set $\Delta = \{ \liminf_n \left(\frac{S_n}{n}\right) < \limsup_n \left(\frac{S_n}{n}\right) \}$

Then Δ is invariant
Also $\Delta = \bigcup_{\substack{a,b \in \mathbb{Q} \\ a<b}} D(a,b)$

$\Rightarrow \mu(\Delta) = 0$

On Δ^c, $\frac{S_n}{n}$ converges in $[-\infty, \infty]$
so we can define an invariant function $\bar{f}$ by

$$\bar{f} = \begin{cases} \lim_{n\to\infty} \left(\frac{S_n}{n}\right) & \text{on } \Delta^c \\ 0 & \text{on } \Delta \end{cases}$$

Then $\frac{S_n}{n} \to \bar{f}$ almost everywhere

Finally, $\mu(|f \circ \theta^n|) = \mu(|f| \circ \theta^n)$
$= \mu(|f|)$
$\Rightarrow \mu(|S_n(f)|) = \mu(|S_n|)$
$\le n\mu(|f|) \quad \forall n$

Hence $\mu(|\bar{f}|) = \mu(\liminf_n |\frac{S_n}{n}|)$
$\le \liminf_n \mu(|\frac{S_n}{n}|)$ by Fatou's Lemma
$\le \mu(|f|) < \infty$ $\square$

Theorem 9·3·3 von Neumann's L^p ergodic theorem
Suppose $(E, \mathcal{E}, \mu)$ is a finite measure space ie $\mu(E) < \infty$
Let $p \in [1, \infty)$
Then, $\forall f \in L^p(\mu)$, $\frac{S_n(f)}{n} \to \bar{f}$ in L^p

Proof

Note that $\|f \circ \theta^n\|_p = \left(\int_E |f|^p \circ \theta^n \, d\mu\right)^{\frac{1}{p}}$
$= \|f\|_p$

So, by Minkowski's inequality, $\left\| \frac{S_n(f)}{n} \right\|_p \le \frac{1}{n} \sum_{k=0}^{n-1} \|f \circ \theta^k\|_p$
$= \|f\|_p$

Given $\varepsilon > 0$, choose $K < \infty$ such that $\|f - g\|_p < \frac{\varepsilon}{3}$,
where $g = \min\{\max\{-K, f\}, K\}$

By Birkhoff's theorem, $\frac{S_n(g)}{n} \to \bar{g}$ almost everywhere

we have $\left|\frac{S_n(g)}{n}\right| \le K \quad \forall n$

So, by bounded convergence, $\exists N \in \mathbb{N}$ such that $\forall n \ge N$,
$\left\| \frac{S_n(g)}{n} - \bar{g} \right\|_p < \frac{\varepsilon}{3}$

Now $\|\bar{f} - \bar{g}\|_p^p = \int_E \liminf_n \left| \frac{S_n(f-g)}{n} \right|^p d\mu$

$\leq \liminf_n \int_E \left| \frac{S_n(f-g)}{n} \right|^p d\mu$ by Fatou's Lemma

$\leq \|f-g\|_p^p$

So, $\forall n \geq N$,

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

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

$$< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}$$

$$= \varepsilon \qquad \square$$

Probability and Measure

§10 Sums of Independent Random Variables

10·0 Weak Law for Random Variables in $L^2(\mathbb{P})$

Let $(X_n : n \in \mathbb{N})$ be a sequence of random variables in $L^2(\mathbb{P})$
$\mathbb{E}(X_n^2) < \infty \quad \forall n$

Suppose
$$\mathbb{E}(X_n) = \mu_n$$
$$\mathrm{Var}(X_n) = \sigma_n^2$$
$$\mathrm{cov}(X_i, X_j) = 0 \qquad \forall\, i \neq j$$

Set $S_n = X_1 + \dots + X_n$

Then
$$\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 + \sum_{i \neq j} \mathrm{cov}(X_i, X_j)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2$$

Recall $\mathrm{Var}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2$

$$\Rightarrow \mathbb{E}\left(\left(\frac{S_n}{n} - \frac{1}{n}\sum_{i=1}^{n} \mu_i\right)^2\right) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2$$

So, if $\mu_n \equiv \mu$
and $\sigma_n^2 \equiv \sigma^2$

then
$$\mathbb{E}\left(\left(\frac{S_n}{n} - \mu\right)^2\right) = \frac{1}{n}\sigma^2$$
$$\frac{S_n}{n} \to \mu \quad \text{in } L^2(\mathbb{P})$$
and hence also in probability

10·1 Strong Law of Large Numbers for Random Variables of Finite Fourth Moment

Theorem
10·1·1 Let $(X_n : n \in \mathbb{N})$ be a sequence of independent random variables
with
$$\mathbb{E}(X_n) = \mu$$
$$\mathbb{E}(X_n^4) \leq M \qquad \forall n, \qquad \text{some } \mu \in \mathbb{R},\ M < \infty$$

Set $S_n = X_1 + \dots + X_n$

Then $\frac{S_n}{n} \to \mu$ almost surely as $n \to \infty$

Proof

Consider $Y_n = X_n - \mu$

Then $(Y_n : n \in \mathbb{N})$ is a sequence of independent random variables

Also $Y_n^4 \leq 2^4(X_n^4 + \mu^4)$
$$\Rightarrow \mathbb{E}(Y_n^4) \leq 16(M + \mu^4) \qquad \forall n$$
and $\mathbb{E}(Y_n) = 0$

It sufficies to show that $\frac{Y_1 + \dots + Y_n}{n} \to 0$ almost surely

So we are reduced to the case $\mu = 0$

Note that X_n, X_n^2, X_n^3 are all integrable, since X_n^4 is integrable.
Since $\mu = 0$, by independence,

$$\mathbb{E}(X_i X_j^3) = \mathbb{E}(X_i X_j X_k^2) = \mathbb{E}(X_i X_j X_k X_l) = 0$$

for distinct indices i, j, k, l

Also $\mathbb{E}(X_i^4) \leq M$

and
$$\mathbb{E}(X_i^2 X_j^2) = \mathbb{E}(X_i^2)\mathbb{E}(X_j^2) \quad \text{for } i \neq j \text{, by independence}$$
$$\leq (\mathbb{E}(X_i^4))^{\frac{1}{2}}(\mathbb{E}(X_j^4))^{\frac{1}{2}}$$
by the Cauchy-Schwarz inequality
$$\leq M$$

So
$$\mathbb{E}(S_n^4) = \mathbb{E}\Big(\sum_{1 \leq i \leq n} X_i^4 + \binom{4}{2} \sum_{1 \leq i < j \leq n} X_i^2 X_j^2\Big)$$
$$= \sum_{1 \leq i \leq n} \mathbb{E}(X_i^4) + 6 \sum_{1 \leq i < j \leq n} \mathbb{E}(X_i^2 X_j^2)$$
$$\leq nM + 6 \cdot \tfrac{1}{2} n(n-1)M$$
$$\leq 3n^2 M$$

$$\Rightarrow \mathbb{E}\Big(\sum_{n=1}^{\infty}\Big(\frac{S_n}{n}\Big)^4\Big) \leq \sum_{n=1}^{\infty} \frac{3n^2 M}{n^4} \quad \text{by Monotone convergence}$$
$$= 3M \sum_{n=1}^{\infty} \frac{1}{n^2}$$
$$< \infty$$

$$\Rightarrow \sum_{n=1}^{\infty}\Big(\frac{S_n}{n}\Big)^4 < \infty \quad \text{almost surely}$$

$$\Rightarrow \frac{S_n}{n} \to 0 \quad \text{almost surely}$$

10.2 Strong Law of Large Numbers

Recall: If $(E, \mathcal{E}, \mu)$ is the canonical model for sequences of independent random variables having law m

then
$$E = \mathbb{R}^{\mathbb{N}} = \{x = (x_n : n \in \mathbb{N}) : x_n \in \mathbb{R} \ \forall n\}$$
$$X_n(x) = x_n$$
$$\mathcal{E} = \sigma(X_n : n \in \mathbb{N})$$
$$\mu\Big(\prod_{n \in \mathbb{N}} A_n\Big) = \prod_{n \in \mathbb{N}} m(A_n)$$

Theorem 10.2.1 Let m be a Borel probability measure on $\mathbb{R}$ such that
$$\int_{\mathbb{R}} |x|\, m(dx) < \infty$$
$$\nu = \int_{\mathbb{R}} x\, m(dx)$$
Let $(E, \mathcal{E}, \mu)$ be the canonical model for a sequence of independent random variables with law m

Then
$$\mu\Big(\Big\{x \in E : \frac{x_1 + \dots + x_n}{n} \to \nu \text{ as } n \to \infty\Big\}\Big) = 1$$

Proof

Consider the shift map $\theta : E \to E$
given by $\theta(x_1, x_2, \dots) = (x_2, x_3, \dots)$

Then θ is measure-preserving and ergodic, by Theorem 9·2·1

The coordinate function $f = X_1$ is integrable
since $\mu(|f|) = \int_{\mathbb{R}} |x| \, m(dx) < \infty$

Now $S_n(f) = f + f \circ \theta + \dots + f \circ \theta^{n-1}$
$= X_1 + \dots + X_n$

So by Birkhoff's theorem (and von Neumann's theorem)
$\exists$ an invariant function $\bar{f}$ such that $\frac{S_n(f)}{n} \to \bar{f}$ almost everywhere and in L^1

Since θ is ergodic, $\bar{f} = c$ almost everywhere for some $c \in \mathbb{R}$

Then
$$\begin{aligned} c &= \mu(\bar{f}) \\ &= \lim_{n\to\infty} \mu\left(\frac{S_n(f)}{n}\right) \\ &= \mu(X_1) \\ &= \int_{\mathbb{R}} x \, m(dx) \\ &= \nu \end{aligned}$$
$\square$

Theorem 10·2·2 Strong Law of Large Numbers

Let $(Y_n : n \in \mathbb{N})$ be a sequence of independent, identically distributed random variables in $\mathbb{R}$
such that $\mathbb{E}|Y_1| < \infty$ ie the variables are integrable
$\mathbb{E}(Y_1) = \nu$

Set $S_n = Y_1 + \dots + Y_n$

Then $\frac{S_n}{n} \to \nu$ almost surely as $n \to \infty$

Proof

In the notation of Theorem 10·2·1,
take m to be the law of the random variables Y_n.

Define $Y : \Omega \to E$
by $Y(\omega) = (Y_n(\omega) : n \in \mathbb{N})$

Then $\mu = \mathbb{P} \circ Y^{-1}$ ie μ is the distribution of Y

Set $A = \left\{ x \in E : \frac{x_1 + \dots + x_n}{n} \to \nu \text{ as } n \to \infty \right\}$

Then $\left\{ \frac{S_n}{n} \to \nu \text{ as } n \to \infty \right\} = \{ Y \in A \}$

$$\Rightarrow \mathbb{P}\left(\frac{S_n}{n} \to \nu \text{ as } n \to \infty\right) = \mathbb{P}(Y \in A)$$
$$= \mu(A)$$
$$= 1$$
$\square$

10.3 Central Limit Theorem

Let $(X_n : n \in \mathbb{N})$ be a sequence of independent random variables $X_n \in L^2(\mathbb{P})$ having means μ_n and variances σ_n^2 (in $\mathbb{R}$)

Set $m_n = \mu_1 + \ldots + \mu_n$

$s_n^2 = \sigma_1^2 + \ldots + \sigma_n^2$

$S_n = X_1 + \ldots + X_n$

Let $Z_n = \dfrac{S_n - m_n}{s_n}$

Then $\mathbb{E}(Z_n) = 0$

$Var(Z_n) = 1$

Lindelberg condition

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}((X_k - \mu_k)^2 \mathbb{1}_{|X_k - \mu_k| > \varepsilon s_n}) \to 0 \text{ as } n \to 0, \quad \forall \varepsilon > 0 \qquad (L)$$

Fact $\quad (L) \Rightarrow Z_n \to Z$ weakly in distribution as $n \to \infty$, where $Z \sim N(0,1)$

Theorem 10.3.1 Central Limit Theorem

Let $(X_n : n \in \mathbb{N})$ be a sequence of independent, identically distributed random variables with mean 0 and variance 1

Set $S_n = X_1 + \ldots + X_n$

Then, $\forall x \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{S_n}{\sqrt{n}} \leq x\right) \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \, dy \quad \text{as } n \to \infty$$

Proof $\quad$ By Theorem 7.7.1 it is sufficient to show that

$$\phi_{\frac{S_n}{\sqrt{n}}}(u) \to e^{-\frac{u^2}{2}} \quad \forall u \in \mathbb{R}$$

Since $\mathbb{E}(X_1^2) < \infty$ $[X_1 \in L^2(\mathbb{P})]$

we can differentiate $\phi_{X_1}(u) = \mathbb{E}(e^{iuX_1})$ twice under the expectation, to obtain

$\phi_{X_1}(0) = 1$

$\phi_{X_1}'(0) = 0 \qquad\qquad \phi_{X_1}' = i\,\mathbb{E}(X_1 e^{iuX_1})$

$\phi_{X_1}''(0) = -1 \qquad\qquad \phi_{X_1}'' = -\mathbb{E}(X_1^2 e^{iuX_1})$

Hence, by Taylor's Theorem, $\phi_{X_1}(u) = 1 - \frac{1}{2}u^2 + o(u^2)$ as $u \to 0$

Now
$$\begin{aligned}
\phi_{\frac{S_n}{\sqrt{n}}}(u) &= \mathbb{E}(e^{iu \cdot \frac{1}{\sqrt{n}}(X_1 + \ldots + X_n)}) \\
&= (\mathbb{E}(e^{i\frac{u}{\sqrt{n}}X_1}))^n \\
&= (\phi_{X_1}(\tfrac{u}{\sqrt{n}}))^n \\
&= (1 - \tfrac{u^2}{2n} + o(\tfrac{u^2}{n}))^n
\end{aligned}$$

complex

The complex logarithm satisfies $\log(1+z) = z + o(|z|)$ as $z \to 0$

So, for each $u \in \mathbb{R}$,
$$\begin{aligned}
\log(\phi_{\frac{S_n}{\sqrt{n}}}(u)) &= n \log(1 - \tfrac{u^2}{2n} + o(\tfrac{u^2}{n})) \\
&= -\tfrac{u^2}{2} + o(1) \\
&\to -\tfrac{u^2}{2} \quad \text{as } n \to \infty \\
&= \log(e^{-\frac{u^2}{2}})
\end{aligned}$$

$\Rightarrow \phi_{\frac{S_n}{\sqrt{n}}}(u) \to e^{-\frac{u^2}{2}}$

$\Rightarrow \frac{S_n}{\sqrt{n}} \to N(0,1)$ in distribution $\square$

Probability and Measure

The Lebesgue σ-algebra $\mathcal{L}$ is the completion $\overline{\mathcal{B}}$ of the Borel σ-algebra with respect to the Lebesgue measure (on $\mathbb{R}$)

We know $\mathcal{B} \subseteq \mathcal{L}$

To show $\overline{\mathcal{B}} \subseteq \mathcal{L}$ it is sufficient to show that for $A \subseteq \mathbb{R}$, $N \subseteq B$, where $B \in \mathcal{B}$ is such that $\mu(B) = 0$

$$\mu^*(A) = \mu^*(A \cap \overset{N}{\cancel{B}}) + \mu^*(A \cap N^c)$$

By countable sub-additivity, $\mu^*(A) \leq \mu^*(A \cap N) + \mu^*(A \cap N^c)$

Also $\left.\begin{aligned}\mu^*(A \cap N) &\leq \mu^*(A \cap B) \\ &\leq \mu^*(B)\end{aligned}\right\}$ again by sub-additivity

$= 0$

and $\mu^*(A \cap N^c) \leq \mu^*(A)$

$\Rightarrow \mu^*(A) \geq \mu^*(A \cap N) + \mu^*(A \cap N^c)$

Lecture 4

Approximation of Borel sets

Suppose $B \in \mathcal{B}$ and $\mu(B) < \infty$

Then $\mu(B) = \inf \sum_n \mu(A_n)$

where the infimum is taken over all sequences $(A_n : n \in \mathbb{N})$ of finite unions of disjoint intervals such that $B \subseteq \bigcup_n A_n$

Given $\varepsilon > 0$, choose such a sequence $(A_n : n \in \mathbb{N})$ so that $\sum_n \mu(A_n) \leq \mu(B) + \frac{\varepsilon}{2}$

Then $\exists\ N \in \mathbb{N}$ such that $\sum_{n > N} \mu(A_n) \leq \frac{\varepsilon}{2}$

Let $A = \bigcup_{n=1}^{N} A_n$

Then $\mu(B \triangle A) \leq \varepsilon$, where $B \triangle A = (B \setminus A) \cup (A \setminus B)$

Note that A is also a finite union of disjoint intervals

Lebesgue σ-algebra

From CET we get not only a Borel measure but a further extension to the set $\mathcal{M}$ of outer measurable sets.

In this context $\mathcal{M}$ is called the Lebesgue σ-algebra

In fact $\mathcal{M} = \{A \cup N : A \in \mathcal{B},\ N \subseteq B,$ where $B \in \mathcal{B},\ \mu(B) = 0\}$

or $\triangle$ null

and $\mu(A \cup N) = \mu(A)$

Note $\mathcal{M} \neq 2^{\mathbb{R}}$

Let B be a ~~Borel~~ Lebesgue measurable set.
Suppose for now $B \subseteq I$ for some finite interval I,
and set $B' = I \setminus B$
Then $\mu(I) = \mu(B) + \mu(B')$
$\exists$ open sets A_k^n such that $B \subseteq \bigcup_k A_k^n$
and $\sum_k \mu(A_k^n) \leq \mu(B) + \frac{1}{n}$ (by definition of outer measure)

Set $G_n = \bigcup_k A_k^n$
Then G_n is open, $B \subseteq G_n$
$\mu(G_n) \leq \mu(B) + \frac{1}{n}$
Set $G = \bigcap_n G_n$
Then $B \subseteq G$, $\mu(G) = \mu(B)$
Construct G' similarly for B'
Then we must have $\mu(G \cap G') = 0$
and $G \setminus B \subseteq G \cap G'$

What is the regularity of indefinite integrals?
Take $f \geq 0$ integrable
Set $F_0(t) = \int_0^t f(x)\,dx$
Then for $t_n \uparrow t$, $F_0(t_n) = \int_0^\infty 1_{[0,t_n]}(x) f(x)$
$\rightarrow F(t)$ by monotone convergence

Similarly for $t_n \downarrow k$
So F is always continuous

$\exists$ a continuous, non-decreasing function F on $\mathbb{R}$
which is differentiable almost everywhere with derivative 0
and such that $F(0) = 0$, $F(1) = 1$
So $F(1) - F(0) \neq \int_0^1 F'(x)\,dx = 0$ F is not always differentiable