05/10/12

Probability and Measure ①

1·1 Definitions Let E be a set. A *σ-algebra* $\mathcal{E}$ on E is a set of subsets of E such that $\emptyset \in \mathcal{E}$ and for all $A \in \mathcal{E}$ and all sequences $(A_n : n \in \mathbb{N})$ in $\mathcal{E}$

$A^c \in \mathcal{E}$, $\bigcup_n A_n \in \mathcal{E}$

The pair $(E, \mathcal{E})$ is called a *measurable space*. The elements of $\mathcal{E}$ are called $(\mathcal{E}-)$ *measurable sets*.

A measure μ on $(E, \mathcal{E})$ is a function $\mu : \mathcal{E} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ and for all sequences $(A_n : n \in \mathbb{N})$ of *disjoint* sets in $\mathcal{E}$, $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$. This last property is called *countable additivity*.

The triple $(E, \mathcal{E}, \mu)$ is called a *measure space*.

1·2 Discrete Measure Spaces

Let E be a countable set, and take $\mathcal{E}$ to be the set of all subsets of E. A *mass function* m on E is any function $m : E \to [0, \infty]$. Given a measure μ on E, by countable additivity, $\mu(A) = \sum_{x \in A} \mu(\{x\})$. So there is a one-to-one correspondence between measures and mass functions.

$\mu(A) = \sum_{x \in A} m(x)$, $m(x) = \mu(\{x\})$

Extension ⟷ Carathéodory, Uniqueness ⟷ Dynkin

1.3 Generated σ-algebras

Let $\mathcal{A}$ be a set of subsets of E. Define $\sigma(\mathcal{A}) :=$
$\{A \subseteq E : A \in \mathcal{E}$ for all σ-algebras $\mathcal{E}$ containing $\mathcal{A}\}$
Then (exercise) $\sigma(\mathcal{A})$ is a σ-algebra. We call $\sigma(\mathcal{A})$ the σ-algebra generated by $\mathcal{A}$ and note that $\sigma(\mathcal{A})$ is the smallest σ-algebra containing $\mathcal{A}$.

1.4 π-systems and d-systems



Let $\mathcal{A}$ be a set of subsets of E. We say that $\mathcal{A}$ is a π-system if $\emptyset \in \mathcal{A}$ and for all $A, B \in \mathcal{A}$, we have $A \cap B \in \mathcal{A}$.
We say that $\mathcal{A}$ is a d-system if $E \in \mathcal{A}$, and for all $A, B \in \mathcal{A}$ with $A \subseteq B$, and all increasing sequences $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$, we have $B \setminus A \in \mathcal{A}$, and $\bigcup_n A_n \in \mathcal{A}$
(Here $(A_n : n \in \mathbb{N})$ increasing means $A_n \subseteq A_{n+1}$ for all n).
Exercise: If $\mathcal{A}$ is both a π-system and d-system, then $\mathcal{A}$ is a σ-algebra.

Lemma 1.4.1 (Dynkin's Lemma)
Let $\mathcal{A}$ be a π-system. Then, any d-system containing $\mathcal{A}$ also contains the σ-algebra containing $\mathcal{A}$.

Proof
Consider $\mathcal{D} = \{A \subseteq E : A \in \tilde{\mathcal{D}}$ for all d-systems $\tilde{\mathcal{D}}$ containing $\mathcal{A}\}$
Then $\mathcal{D}$ is a d-system (exercise). We shall see that $\mathcal{D}$ is also a π-system, ~~which proves the lemma~~, hence is a σ-algebra, proving the lemma.

05/10/12 Probability and Measure ①

We shall see that $\mathcal{D}'$ is a d-system.

Consider $\mathcal{D}' = \{B \in \mathcal{D} : B \cap A \in \mathcal{D} \text{ for all } A \in \mathcal{A}\}$

Then $\mathcal{A} \in \mathcal{D}'$ because $\mathcal{A}$ is a π-system. Also $E \in \mathcal{D}'$.

Suppose $B_1, B_2 \in \mathcal{D}'$ with $B_1 \subseteq B_2$. Then, for all $A \in \mathcal{A}$, $B_1 \cap A$, $B_2 \cap A \in \mathcal{D}$ so $(B_2 \cap A) \setminus (B_1 \cap A) = \underbrace{(B_2 \setminus B_1) \cap A}_{\in \mathcal{D}}$

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

Suppose $(B_n : n \in \mathbb{N})$ is an increasing sequence in $\mathcal{D}'$. Then, for all $A \in \mathcal{A}$, $A \cap B_n \in \mathcal{D}$, so $(\bigcup_n B_n) \cap A = \underbrace{\bigcup_n (B_n \cap A)}_{\in \mathcal{D}}$

So $\bigcup_n B_n \in \mathcal{D}'$.

Hence $\mathcal{D}$ is a d-system as claimed, so $\mathcal{D}' = \mathcal{D}$

Now consider $\mathcal{D}'' = \{B \in \mathcal{D} : B \cap A \in \mathcal{D} \text{ for all } A \in \mathcal{D}\}$.

Then $\mathcal{A} \subseteq \mathcal{D}''$ because $\mathcal{D} = \mathcal{D}'$. We can check that $\mathcal{D}''$ is a d-system, just as for $\mathcal{D}'$. Hence $\mathcal{D} = \mathcal{D}''$ so $\mathcal{D}$ is a π-system, as promised. $\square$

08/10/12

Probability and Measure (2)

1·5 Set functions and Properties

Let $\mathcal{A}$ be a set of subsets of E containing $\emptyset$. A <u>set function</u> on $\mathcal{A}$ is any function $\mu : \mathcal{A} \to [0,\infty]$ with $\mu(\emptyset) = 0$.

μ is <u>increasing</u> if $A, B \in \mathcal{A}$, $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$.

μ is <u>additive</u> if $A, B \in \mathcal{A}$ disjoint, $A \cup B \in \mathcal{A}$, $\mu(A \cup B) = \mu(A) + \mu(B)$

μ is <u>countably additive</u> if $A_n \in \mathcal{A}$ $\forall n$, disjoint, with $\bigcup_n A_n \in \mathcal{A} \Rightarrow \mu(\bigcup_n A_n) = \sum_n \mu(A_n)$.

μ is <u>countably sub-additive</u> if $A_n \in \mathcal{A}$ $\forall n$, $\bigcup_n A_n \in \mathcal{A}$ $\Rightarrow \mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$.

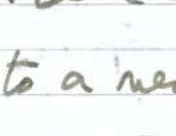
1·6 Construction of measures

Let $\mathcal{A}$ be a set of subsets of E, containing $\emptyset$. We say that

ring ~~[ ] [ ]~~ $\mathcal{A}$ is a <u>ring</u> if $\forall A, B \in \mathcal{A}$, $B \setminus A \in \mathcal{A}$, $A \cup B \in \mathcal{A}$. We say that $\mathcal{A}$ is an <u>algebra</u>, if for all $A, B \in \mathcal{A}$, $A^c \in \mathcal{A}$, $A \cup B \in \mathcal{A}$.

<u>Theorem 1·6·1</u> (Carathéodory's Extension Theorem)

Let $\mathcal{A}$ be a ring of subsets of E and let μ be a countably additive set function on E. Then μ extends to a measure on the σ-algebra generated by $\mathcal{A}$.



cover B, with sets for which μ^* is defined

<u>Proof</u>

Define for any subset $B \subseteq E$ the <u>outer measure</u> $\mu^*(B)$

$\mu^*(B) = \inf \sum_n \mu(A_n)$ 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.

Say that $A \subseteq E$ is <u>μ^*-measurable</u> if $\forall B \subseteq E$,
$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$
Write $\mathcal{M}$ for the set of all μ^*-measurable sets.
We'll show that $\mu^* = \mu$ on $\mathcal{A}$, that $\mathcal{A} \subseteq \mathcal{M}$, and that $\mathcal{M}$ is a σ-algebra, $\mu^*|_{\mathcal{M}}$ is a measure, proving the theorem.
Note that $\mu^*(\emptyset) = 0$ and μ^* is increasing.

<u>Step I</u>

We show that μ^* is countably sub-additive. We have to show that for $B \subseteq \bigcup_n B_n$, we have $\mu^*(B) \leq \sum_n \mu^*(B_n)$.
It will suffice to consider the case $\mu^*(B_n) < \infty$ $\forall n$.
Then Given $\varepsilon > 0$, $\exists$ sequences $(A_{nm} : m \in \mathbb{N})$ such that
$B_n \subseteq \bigcup_m A_{nm}$, $\mu^*(B_n) + \varepsilon/2^n \geq \sum_m \mu(A_{mn})$
Then $B \subseteq \bigcup_{m,n} A_{mn}$ so
$\mu^*(B) \leq \sum_{n,m} \mu(A_{mn}) \leq \varepsilon + \sum_n \mu^*(B_n)$
Since $\varepsilon > 0$ was arbitrary, we are done.

<u>Step II</u>

We show that $\mu^* = \mu$ on $\mathcal{A}$.
Since $\mathcal{A}$ is a ring, and μ is countably additive, μ is also increasing and countably sub-additive (example sheet).
So, for any $A \in \mathcal{A}$, and for any sequence $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$ such that $A \subseteq \bigcup_n A_n$, $\mu(A) \leq \sum_n \mu(A_n \cap A)$
$\leq \sum_n \mu(A_n)$.

38/10/12

Probability and Measure ②

On taking the infemum over all such sequences we see that $\mu^*(A) \geq \mu(A)$. The reverse inequality is obvious so $\mu^* = \mu$ on $\mathcal{A}$.

Step III

We show that $\mathcal{A} \subseteq \mathcal{M}$. Take $A \in \mathcal{A}$ and $B \subseteq E$.

~~We want to~~ By sub-additivity, $\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c)$ so it suffices to show that $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$, and it suffices to do this when $\mu^*(B)$ is finite. Then, given $\varepsilon > 0$ there exists a sequence $(A_n : n \in \mathbb{N})$ in $\mathcal{A}$ with $B \subseteq \bigcup_n A_n$, and $\mu^*(B) + \varepsilon \geq \sum_n \mu(A_n)$.

Then $B \cap A \subseteq \bigcup_n (A_n \cap A)$, $B \cap A^c \subseteq \bigcup_n (A_n \cap A^c)$

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

$$= \sum_n \mu(A_n) \leq \mu^*(B) + \varepsilon$$

Since $\varepsilon > 0$ was arbitrary, we are done.

Step IV

We show that $\mathcal{M}$ is an algebra. It is clear that $\emptyset \in \mathcal{M}$, and $A^c \in \mathcal{M}$ whenever $A \in \mathcal{M}$. Take $A_1, A_2 \in \mathcal{M}$ and $B \subseteq E$.

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

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

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

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

So $A_1 \cap A_2 \in \mathcal{M}$. Hence $\mathcal{M}$ is an algebra.

<u>Step V</u>

We show that $\mathcal{M}$ is a σ-algebra, and that $\mu^*|_{\mathcal{M}}$ is a measure. Since we know that $\mathcal{M}$ is an algebra, it will suffice to show that for any disjoint sequence $(A_n : n \in \mathbb{N})$ in $\mathcal{M}$ we have $A = \bigcup_n A_n \in \mathcal{M}$ and $\mu^*(A) = \sum_n \mu^*(A_n)$.

So take $B \subseteq E$. We have $\mu^*(B) = \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c$

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

$$= \sum_{i=1}^{n} \mu^*(B \cap A_i) + \mu^*(\underbrace{B \cap A_1^c \cap \ldots \cap A_n^c}_{\supseteq B \cap A^c})$$

So letting $n \to \infty$

$$\mu^*(B) \geq \sum_n \mu^*(B \cap A_n) + \mu^*(B \cap A^c)$$

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

countable sub-additivity of μ^*.

Now, the reverse inequality holds also by sub-additivity:

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

So in fact equality holds throughout, so $A \in \mathcal{M}$, and by taking $B = A$, we see that $\mu^*(A) = \sum_n \mu^*(A_n)$ as required. $\square$

10/10/12

Probability and Measure ③

<u>Recap</u> E, a set

- $\mathcal{E}$, a σ-algebra, set of subsets of E.
- Closed under countable operations
- μ, measure $\mu : \mathcal{E} \Rightarrow [0, \infty]$, $\mu(\biguplus_n A_n) = \sum_n \mu(A_n)$ disjoint

<u>Dynkin</u>

$\mathcal{A} \subseteq \mathcal{D} \Rightarrow \sigma(\mathcal{A}) \subseteq \mathcal{D}$

π-system, d-system closed under relative complements, increasing unions

Suppose $A_n \uparrow A$ $(A_n \subseteq A_{n+1},\ \bigcup_n A_n = A)$

$B_1 = A_1$, $B_{n+1} = A_{n+1} \setminus A_n$, $A_n = B_1 \cup \ldots \cup B_n$, $A = \bigcup_n B_n$

$\mu(A_n) = \sum_{i=1}^{n} \mu(B_i) \to \sum_{i=1}^{\infty} \mu(B_i) = \mu(A)$.

<u>1.7 Uniqueness of measures</u>

<u>Theorem 1.7.1</u>

Let μ_1, μ_2 be measures on $(E, \mathcal{E})$ with $\mu_1(E) = \mu_2(E) < \infty$ finite. Suppose that $\mu_1(A) = \mu_2(A)$ for all sets $A \in \mathcal{A}$, where $\mathcal{A}$ is a π-system generating $\mathcal{E}$. Then $\mu_1 = \mu_2$.

<u>Proof</u>

Define $\mathcal{D} = \{A \in \mathcal{E} : \mu_1(A) = \mu_2(A)\}$. Then $\mathcal{A} \subseteq \mathcal{D}$ and $E \in \mathcal{D}$. For $A, B \in \mathcal{D}$ with $A \subseteq B$ we have

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

Since $\mu_1(A) = \mu_2(A) < \infty$, we get $\mu_1(B \setminus A) = \mu_2(B \setminus A)$ so $B \setminus A \in \mathcal{D}$.

For $A_n \in \mathcal{D}$ with $A_n \uparrow A$, we have $\mu_1(A) = \lim_n \mu_1(A_n)$

$\mu(A) = \lim_n \mu_1(A_n) = \lim_n \mu_2(A_n) = \mu_2(A)$, so $A \in \mathcal{D}$.

Hence $\mathcal{D}$ is a d-system, so $\mathcal{E} = \sigma(\mathcal{A}) \subseteq \mathcal{D}$. $\square$

1·8 Borel Sets and Borel Measures

Let E be a topological space. The σ-algebra generated by the set of open sets is called the <u>Borel σ-algebra of E</u> and is written $\mathcal{B}(E)$. A measure μ on $(E, \mathcal{B}(E))$ is called a <u>Borel Measure</u>. If $\mu(K) < \infty$ for all compact sets K, we say μ is a <u>Radon Measure</u>.

1·9 Probability, finite and σ-finite measures

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

If $\mu(E) = 1$, we say that μ is a <u>probability measure</u> and call $(E, \mathcal{E}, \mu)$ is a <u>probability space</u>. We usually use notation $(\Omega, \mathcal{F}, \mathbb{P})$ for this.

If $\mu(E) < \infty$, we say that μ is a <u>finite measure</u>.

If there exists sets $E_n \in \mathcal{E}$ with $\mu(E_n) < \infty$ for all n and $\bigcup_n E_n = E$, we say that μ is a <u>σ-finite measure</u>.

1·10 Lebesgue Measures

<u>Theorem 1·10·1</u>

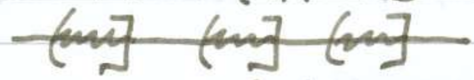
There exists a unique Borel measure μ on $\mathbb{R}$ (usual topology) such that $\mu((a,b]) = b - a$ for all $a, b \in \mathbb{R}$ with $a < b$.

10/10/12

Probability and Measure ③

Proof:

(Existence) Consider the ring $\mathcal{A}$ of disjoint unions of intervals of the form $A = (a_1, b_1] \cup (a_2, b_2] \cup \ldots \cup (a_n, b_n]$



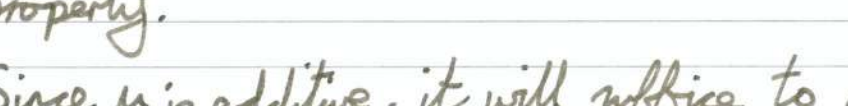
for all $n \in \mathbb{N}$. Define $\mu : \mathcal{A} \to [0, \infty]$ by

$$\mu(A) = \sum_{i=1}^{n} (b_i - a_i)$$

The presentation of A is not unique as $(a, b] \cup (b, c] = (a, c]$ for $a < b < c$. However, it is easy to see that μ is well defined and moreover additive. We shall show that μ is countably additive, so then by Carathéodory's Extension Theorem μ extends to a measure on $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$ with the claimed property.

Since μ is additive, it will suffice to show that for $A_n \in \mathcal{A}$ with $A_n \uparrow A \in \mathcal{A}$ that $\mu(A_n) \to \mu(A)$. Consider $B_n = A \setminus A_n$ ~~then $A_n B_n = \emptyset$~~ It will suffice to show that for $B_n \in \mathcal{A}$ decreasing with $\bigcap_n B_n = \emptyset$, that $\mu(B_n) \to 0$.

Suppose, for contradiction, that $\exists \varepsilon > 0$ such that $\mu(B_n) \geq 2\varepsilon \ \forall n$. We can find $C_n \in \mathcal{A}$ such that $\overline{C_n} \subseteq B_n$ and $\mu(B_n \setminus C_n) \leq \varepsilon / 2^n$



Note that $\mu(B_n \setminus (C_1 \cap \ldots \cap C_n)) = \mu((B_1 \setminus C_1) \cup \ldots \cup (B_n \setminus C_n))$

$$\leq \varepsilon \sum_n 2^{-n} = \varepsilon$$

Hence $\mu(C_1 \cap \dots \cap C_n) \geq \varepsilon$ so $C_1 \cap \dots \cap C_n \neq \emptyset$

Now $K_n = \overline{C_1 \cap \dots \cap C_n}$ is closed, non-empty, $K_n \supseteq K_{n+1}$,

so since $\mathbb{R}$ is complete, $\emptyset = \bigcap_n B_n \supseteq \bigcap_n K_n \neq \emptyset$ $\#$

So $\mu(B_n) \to 0$ as required.

12/10/12

Probability and Measure ④

Correction: $\mu(B \setminus (C_1 \cap \ldots \cap C_n)) = [\mu((B_n \setminus C_1) \cup \ldots \cup (B_n \setminus C_n)) \leq \mu((B_1 \setminus C_1) \cup \ldots \cup (B_n \setminus C_n))]$

Uniqueness

Suppose μ, ν are Borel measures on $\mathbb{R}$ with $\mu((a,b]) = \nu((a,b]) = b-a$ whenever $a<b$.

For $n \in \mathbb{Z}$ and $B \in \mathcal{B}$, define

$\mu_n(B) = \mu(B \cap (n, n+1])$, $\nu_n(B) = \nu(B \cap (n,n+1])$

Then, μ_n, ν_n are Borel probability measures on $\mathbb{R}$, and $\mu_n = \nu_n$ on the π-system of intervals $(a,b]$, which generates $\mathcal{B}$. So, by uniqueness of extension $\mu_n = \nu_n$ on $\mathcal{B}$. Then, by countable additivity, for $B \in \mathcal{B}$, $\mu(B) = \sum_n \mu(B \cap (n, n+1])$

$= \sum_n \nu(B \cap (n, n+1]) = \nu(B)$ $\square$

For $x \in \mathbb{R}$, and $B \in \mathcal{B}$, define $\mu_x(B) = \mu(B+x)$ "Lebesgue measure"

where $B+x := \{b+x \mid b \in B\}$. Then μ_x is a Borel measure. $\mu_x((a,b]) = (b+x) - (a+x) = b-a$.

So by uniqueness, $\mu_x = \mu$. Thus, Lebesgue measure is <u>translation invariant</u>. We can consider μ_0 as a probability measure on $(0,1]$. $\mu_0(B) = \mu(B \cap (0,1])$

Note that μ_0 is also translation invariant, if we define for $B \in \mathcal{B}[0,1]$, $B+x = \{b+x \bmod 1 \mid b \in B\} \subseteq (0,1]$

We used Carathéodory's Extension Theorem, so in fact we have constructed a σ-algebra $\mathcal{M} \supseteq \mathcal{B}$ and a measure

μ^* on $\mathcal{M}$ extending Lebesgue measure μ.

Recall that $\mathcal{M}$ is the set of "μ^*-measurable sets"

We call $\mathcal{M}$ the set of <u>Lebesgue - measurable sets</u>

In Ex 1·9, we see that $\mathcal{M}$ is concretely given by

Δ also works

$$\mathcal{M} = \{ B \cup N : B \in \mathcal{B}, N \in \mathcal{N} \}$$

$\in$ null sets

null sets $\mathcal{N} = \{ N \subseteq \mathbb{R} : N \subseteq A$ for some $A \in \mathcal{B}$ with $\mu(A) = 0 \}$

For $B \cup N \in \mathcal{M}$, we have $\mu(B \cup N) = \mu(B)$

1·11 A non-Lebesgue-measurable subset of $(0,1]$

For $x, y \in (0,1]$, write $x \sim y$ if $x - y \in \mathbb{Q}$.

This is an equivalence relation. Use the Axiom of Choice, to choose a set $S \subseteq (0,1]$ containing exactly one element of each equivalence class. Then, we can check that $(0,1]$ is equal to a disjoint $\bigcup_{q \in (0,1] \cap \mathbb{Q}} (S + q)$ addition modulo 1.

$$x = s + q = s' + q \Rightarrow s - s' \in \mathbb{Q}$$

Suppose that S is Lebesgue measurable. Then, by translation invariance, $\mu(S+q) = \mu(S)$ $\forall q \in \mathbb{Q}$. Then, by countable additivity $1 = \mu((0,1]) = \sum_{q \in (0,1] \cap \mathbb{Q}} \mu(S+q) = \infty \cdot \mu(S)$

This is impossible, so we conclude that S is not Lebesgue-measurable, $S \notin \mathcal{M}$.

Leb. measure on rationals?

1·12 Independence

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We use this to model an experiment whose outcome is subject to chance.

Ω models possible outcomes.

$\mathcal{F}$ is the set of observable sets of outcomes, called events.

12/10/12

Probability and Measure ④

$\mathbb{P}(A)$ is the probability of the event A.

Let I be a countable set. Say that a family of events $(A_i : i \in I)$ is independent if for all finite $J \subseteq I$,

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

Exercise: Show that this always extends to J countable infinite.

We say that a family $(\mathcal{A}_i : i \in I)$ of sub-σ-algebras of $\mathcal{F}$ is <u>independent</u> if $(A_i : i \in I)$ is independent whenever $A_i \in \mathcal{A}_i \ \forall i$.

<u>Theorem 1.12.1</u>

Let $\mathcal{A}_1, \mathcal{A}_2$ be π-systems in $\mathcal{F}$ and suppose $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2) \ \forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$.

The conclusion is that $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2)$ are independent.

<u>Proof</u>

Fix $A_1 \in \mathcal{A}_1$ and define for $B \in \mathcal{F}$

$\mu(B) = \mathbb{P}(A_1 \cap B)$, $\nu(B) = \mathbb{P}(A_1)\mathbb{P}(B)$.

Then $\mu(\Omega) = \nu(\Omega) = \mathbb{P}(A_1) < \infty$, and $\mu = \nu$ on $\mathcal{A}_2$.

So by uniqueness of extension, $\mu = \nu$ on $\sigma(\mathcal{A}_2)$.

Now fix $A_2 \in \sigma(\mathcal{A}_2)$ and define for $B \in \mathcal{F}$

$\tilde{\mu}(B) = \mathbb{P}(B \cap A_2)$, $\tilde{\nu}(B) = \mathbb{P}(B)\mathbb{P}(A_2)$

Then $\tilde{\mu}(\Omega) = \tilde{\nu}(\Omega) = \mathbb{P}(A_2) < \infty$ and $\tilde{\mu} = \tilde{\nu}$ on $\mathcal{A}_1$ by the first part. So by uniqueness of extension, $\tilde{\mu} = \tilde{\nu}$ on $\sigma(\mathcal{A}_1)$, proving the result. $\square$

15/10/12

Probability and Measure Theory 5

1.13 Borel - Cantelli Lemmas

Let $(A_n : n \in \mathbb{N})$ be a sequence of events. Define

$\{A_n \text{ infinitely often}\} := \limsup_n A_n := \bigcap_n \bigcup_{m \geq n} A_m \quad (\in \mathcal{F})$

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

Lemma 1.13.1 (First B-C lemma)

i.o.

Suppose $\sum_n \mathbb{P}(A_n) < \infty$. Then $\mathbb{P}(A_n \text{ infinitely often}) = 0$

Proof

We have $\mathbb{P}(A_n \text{ i.o}) \leq \mathbb{P}(\bigcup_{m \geq n} A_m) \leq \sum_{m \geq n} \mathbb{P}(A_m) \to 0$ □

Lemma 1.13.1 (Second B-C lemma)

Assume that the events $(A_n : n \in \mathbb{N})$ are independent.

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

Proof

We use the inequality $1 - x \leq e^{-x}$ for $x \geq 0$

The sequence of events $(A_n^c : n \in \mathbb{N})$ is also independent.

We have $\mathbb{P}(\bigcap_{m=n}^{N} A_m^c) = \prod_{m=n}^{N} (1 - \mathbb{P}(A_m)) \leq \exp(-\sum_{m=n}^{n} \mathbb{P}(A_m$

which $\to 0$ as $n \to \infty$, so $\mathbb{P}(\bigcap_{m=n} A_m^c) = 0 \ \forall n$.

So $\mathbb{P}(\bigcup_n \bigcap_{m \geq n} A_m^c) = 0$. But $\{A_n \text{ infinitely often}\}$

$= \bigcap_n \bigcup_{m \geq n} A_m = (\bigcup_n \bigcap_{m \geq n} A_m^c)^c$. So $\mathbb{P}(A_n \text{ i.o.}) = 1$

2.7 Large values in sequences of i.i.d. r.v. s

Let $(X_n : n \in \mathbb{N})$ be a sequence of i.i.d r.v. s. Suppose that $\mathbb{P}(X \geq x) > 0$ for all $x \geq 0$. ~~Suppose that~~

Take $x_n \uparrow \infty$ and set $A_n = \{X_n \geq x_n\}$ and

$a_n = \mathbb{P}(A_n)$. Then the events $(A_n : n \in \mathbb{N})$ are independent.

So $\mathbb{P}(X_n \geq x_n \text{ i.o}) = \mathbb{P}(A_n \text{ i.o}) = \begin{cases} 1 & \sum a_n = \infty \\ 0 & \sum a_n < \infty \end{cases}$

<u>Example</u>

$X \sim E(1)$, $\mathbb{P}(X \geq x) = e^{-x}$

Take $x_n = \alpha \log n$, then $a_n = e^{-x_n} = \frac{1}{n^\alpha}$

So if $\alpha = 1$, $\sum_n a_n = \infty$, whereas if $\alpha > 1$, $\sum_n a_n < \infty$. Hence $\mathbb{P}(X_n \geq \log n \text{ i.o}) = 1$, $\mathbb{P}(X_n \geq \alpha \log n \text{ i.o}) = 0 \ \forall \alpha > 1$.

So $\limsup_{n \to \infty} X_n / \log n = 1$ almost surely.

2. Measurable Functions and Random Variables

2.1 Measurable Functions

Let $(E, \mathcal{E})$ and $(G, \mathcal{G})$ be measurable spaces, and let $f : E \to G$. We say that f is <u>measurable</u> if $f^{-1}(A) \in \mathcal{E}$ for all $A \in \mathcal{G}$. Here, $f^{-1}(A)$ is the <u>inverse image</u> $f^{-1}(A) = \{x \in E : f(x) \in A\}$

Since $(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A))$, ~~the complete~~ composition of measurable functions is also measurable. Often $G = \mathbb{R}$ or $[-\infty, \infty]$. Then we <u>always</u> take $\mathcal{G} = \mathcal{B}$

For $A \subseteq E$, define the <u>indicator function</u> $1_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$

Then 1_A is measurable $\iff A \in \mathcal{E}$ (exercise)

Inverse images preserve set operations, take ANY $f : E \to G$

$f^{-1}(G \setminus A) = E \setminus f^{-1}(A)$, $f^{-1}(\bigcup_i A_i) = \bigcup_i f^{-1}(A_i)$

Hence $\{f^{-1}(A) : A \in \mathcal{G}\}$ is a σ-algebra on E, and $\{A \subseteq G : f^{-1}(A) \in \mathcal{E}\}$ is a σ-algebra on G.

Suppose we can show that $f^{-1}(A) \in \mathcal{E}$ for all $A \in \mathcal{A}$, for some set of subsets $\mathcal{A}$ generating $\mathcal{G}$. Then, $\{A \subseteq G : f^{-1}(A) \in \mathcal{E}\}$ contains $\mathcal{A}$, so also contains $\sigma(\mathcal{A}) = \mathcal{G}$.
So f is measurable. For example, $\mathcal{B}$ is generated by $\{(-\infty, a] : a \in \mathbb{R}\}$ so $f : E \to \mathbb{R}$ is measurable $\Leftrightarrow \{x \in E : f(x) \leq a\} \in \mathcal{E} \; \forall a \in \mathbb{R}$.

In the case where E, G are topological spaces and $f : E \to G$ is continuous, then $f^{-1}(U)$ is open in E, and hence is Borel-measurable, for all open sets U in G. Since such sets U generate $\mathcal{B}(G)$ we see that every continuous function is Borel-measurable.

<u>Proposition 2.1.1</u>

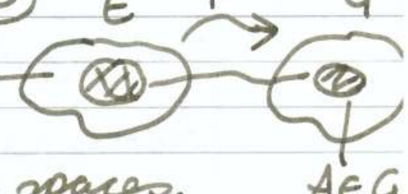
Let $(f_n : n \in \mathbb{N})$ be a sequence of measurable functions on $(E, \mathcal{E})$. Then the following functions are also measurable:
$f_1 + f_2$, $f_1 f_2$, $\inf_n f_n$, $\sup_n f_n$, $\liminf_n f_n$, $\limsup_n f_n$

<u>Proof</u> (Exercise)

$(f_1 + f_2)(x) > a \Leftrightarrow f_1(x) > q$, $f_2(x) > a - q$ for some $q \in \mathbb{Q}$.

17/10/12 Probability and Measure ⑥

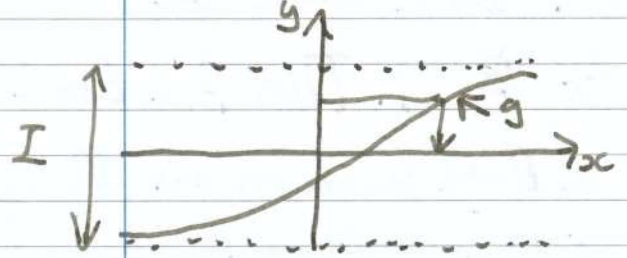
2.2 Image Measures

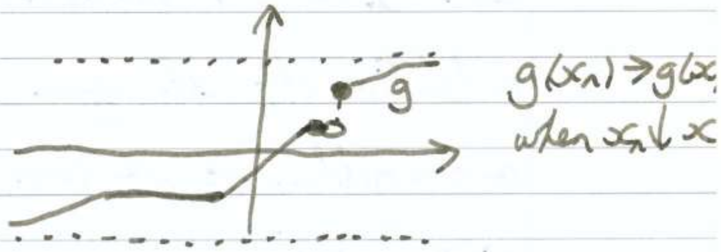


Let $(E, \mathcal{E})$ and $(G, \mathcal{G})$ be measurable spaces. Let μ be a measure on $\mathcal{E}$. Then, for any measurable function $f: E \to G$ we can define the <u>image measure</u> $\nu = \mu \circ f^{-1}$ by $\nu(A) = \mu(f^{-1}(A))$, $A \in \mathcal{G}$.

<u>Lemma 2.2.1</u>

Let $g: \mathbb{R} \to \mathbb{R}$ be non-constant, right continuous, and non-decreasing. Set $g(\pm\infty) = \lim_{x \to \pm\infty} g(x)$, and set $I = (g(-\infty, +\infty))$. Define $f: I \to \mathbb{R}$ by ~~$f(y) = \inf\{$~~
$f(y) = \inf\{x \in \mathbb{R} \mid y \leq g(x)\}$





Then f is non-decreasing, left continuous. Moreover, for $x \in \mathbb{R}$, $y \in I$, $f(y) \leq x \Leftrightarrow y \leq g(x)$.

<u>Proof</u>

For $y \in I$, define $J_y = \{x \in \mathbb{R} : y \leq g(x)\}$. Then $J_y \neq \emptyset$, $J_y \neq \mathbb{R}$. Since g is non-decreasing, if $x \in J_y$ and $x' \geq x$, then $x' \in J_y$. Since g is right continuous if $x_n \in J_y$ and $x_n \downarrow x$ then $x \in J_y$. Hence $J_y = [f(y), \infty)$.
So for $x \in \mathbb{R}$, $y \in I$, $f(y) \leq x \Leftrightarrow y \leq g(x)$.
If $y' \geq y$, then $J_{y'} \subseteq J_y$ so $f(y') \geq f(y)$. If $y_n \uparrow y$

then $J_y = \bigcap_n J_{y_n}$ so $f(y_n) \to f(y)$. Hence f is non decreasing and left continuous, as claimed. $\square$

Theorem 2.2.2

Let $g: \mathbb{R} \to \mathbb{R}$ be as in the lemma. Then, there exists a unique Radon measure ν on $\mathbb{R}$ such that

$\nu((a,b]) = g(b) - g(a)$, whenever $a, b \in \mathbb{R}$, $a < b$.

Moreover, all Radon measures on $\mathbb{R}$ are obtained in this way.

Often we write dg for ν and call dg the Lebesgue - Stieltjes Measure associated with g.

Proof:

Let I and f be as in the lemma. Let λ denote the restriction of Lebesgue measure to $\mathcal{B}(I)$. Then f is Borel measurable, so we can defined a Borel measure on $\mathbb{R}$ by $\nu = \lambda \circ f^{-1}$. We have $\nu((a,b]) = \lambda(\{x \in I : a < f(x) \leq b\})$

$= \lambda(\{x \in I : g(a) < x \leq g(b)\}) = g(b) - g(a)$

so ν is Radon and has the desired property.

The same argument used for uniqueness of Lebesgue measure shows that ν is the only such measure.

Suppose that ν is a Radon measure on $\mathbb{R}$. Define

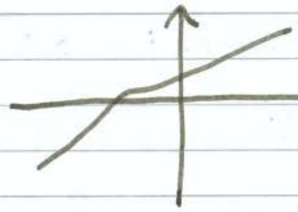
$g(x) = \begin{cases} \mu((0,x]) & x \geq 0 \\ -\mu((x,0]) & x \leq 0 \end{cases}$ Then g is non-decreasing and right continuous

Then $\mu((a,b]) = g(b) - g(a)$ for $a < b$, so

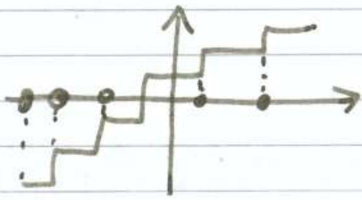
$\mu = dg$ $\square$

17/10/12

Probability and Measure 6



g continuous, dg has no atoms



dg is <u>all</u> atoms

2.3 Random Variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $(E, \mathcal{E})$ be a measurable space. A <u>random variable</u> on E is a measurable function $X : \Omega \to E$.

The image measure $\mu_X = \mathbb{P} \circ X^{-1}$ is a measure on $(E, \mathcal{E})$ is called the <u>law</u> or <u>distribution</u> of X. The law μ_X of a real valued rv is uniquely determined by the values on $\mu_X((-\infty, x]) = F_X(x)$

$= \mathbb{P}(\{x \in \Omega : X(\omega) \leq x\}) = \mathbb{P}(X \leq x)$, $x \in \mathbb{R}$

We call the function $F_X : \mathbb{R} \to [0,1]$ the <u>distribution function</u> of X. (We have $\mu_X = dF_X$). Note that $F = F_X$ satisfies:

- $F(x) \to 0$ as $x \to -\infty$, $F(x) \to 1$ as $x \to +\infty$
- F is non-decreasing and right continuous.

We call any function F on $\mathbb{R}$ with these properties a <u>distribution function</u>.

Let F be a distribution function. Take $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}(\Omega)$. Take $\mathbb{P}$ to be the restriction of Lebesgue measure on $\mathcal{F}$.

Define $X: \Omega \to \mathbb{R}$ by $X(\omega) = \inf\{x \in \mathbb{R} : \omega \leq F(x)\}$

Then X is a random variable (as left continuous) and

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\})$$

$$= \mathbb{P}(\{\omega \in \Omega : \omega \leq F(x)\})$$

Hence every distribution function is the distribution function of some random variable in $\mathbb{R}$.

18/10/12

Probability and Measure ⑦

$(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, $(E, \mathcal{E})$ a measurable space

Recap: An E-valued random variable X is a measurable function $X : \Omega \to E$. (We can ask: what is the probability that $X \in A$, for $A \in \mathcal{E}$. $\mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}(X^{-1}(A))$, well-defined.

Let $(X_i : i \in I)$ be a family of E-valued random variables.

Define $\sigma(X_i : i \in I) = \sigma(\{X_i^{-1}(A) : i \in I, A \in \mathcal{E}\})$ $(\subseteq \mathcal{F})$

This is called the σ-algebra generated by $(X_i : i \in I)$.

Note that $\sigma(X_i) = \{X_i^{-1}(A) : A \in \mathcal{E}\}$

We say that $(X_i : i \in I)$ are <u>independent</u> if the σ-algebras $(\sigma(X_i) : i \in I)$ are themselves independent. For a sequence of real r.v.s this is equivalent to, $\forall n$, all $x_1, \dots, x_n \in \mathbb{R}$

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

This comes from the fact that $\{(-\infty, x] : x \in \mathbb{R}\}$ is a π-system generating $\mathcal{B}$.

2.4 Rademacher Functions

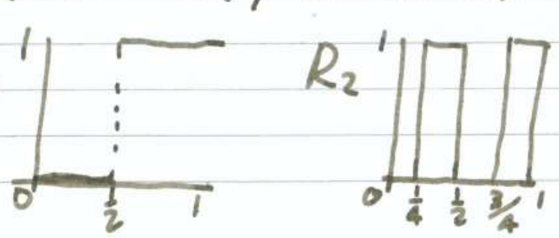
Take $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}(\Omega)$, $\mathbb{P}$ = Lebesgue

Each element $\omega \in \Omega$ has a unique binary expansion

$\omega = \sum_{n=1}^{\infty} \omega_n 2^{-n} = 0 \cdot \omega_1 \omega_2 \dots$, such that

$\omega_n = 1$ for infinitely many n (i.e. use $0111\dots$ instead of $1000\dots$)

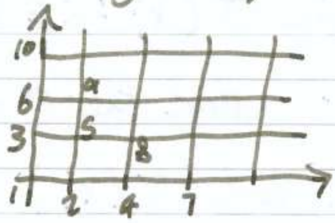
Define $R_n(\omega) = \omega_n$, the Rademacher functions.

Then R_1  etc

We have $\mathbb{P}(R_1 = a_1, \dots, R_n = a_n) = 2^{-n}$ for all n, all $a_k \in \{0,1\}$
Hence $(R_n : n \in \mathbb{N})$ is a sequence of independent Bernoulli $(\frac{1}{2})$ random variables $(\mathbb{P}(R=0) = \mathbb{P}(R=1) = \frac{1}{2})$

<u>Proposition 2.4.1</u>

Let $(F_n : n \in \mathbb{N})$ be a sequence of distribution functions. There exists a sequence $(X_n : n \in \mathbb{N})$ of independent r.v.s such that $F_{X_n} = F_n \;\forall n$.



<u>Proof</u>

Let $m : \mathbb{N}^2 \to \mathbb{N}$ be a bijection

Set $Y_{k,n} = R_{m(k,n)}$ and $Y_n = \sum_{k=1}^{\infty} 2^{-k} Y_{k,n}$. Then the sum converges everywhere to a well defined r.v. and ~~the~~ ~~σ-algebra generated by~~ $\sigma(Y_n) \subseteq \sigma(Y_{k,n} : k \in \mathbb{N})$ so ~~the~~ $(Y_n : n \in \mathbb{N})$ are independent.

For $y = 0 \cdot y_1 \dots y_k = i 2^{-k}$, $i \in \mathbb{Z}^+$

$\mathbb{P}(i 2^{-k} < Y_n \leq (i+1) 2^{-k}) = \mathbb{P}(Y_{1,n} = y_1, \dots, Y_{k,n} = y_k) = 2^{-k}$

$\mathbb{P}(Y_n \leq y) = y$ for all such y, and hence for all $y \in (0,1)$

Now define $G_n(y) = \inf\{x \in \mathbb{R} : y \leq F_n(x)\}$. Then by Lemma 2.2.1, G_n is Borel, and $G_n(y) \leq x \Leftrightarrow y \leq F_n(x)$

So we obtain a sequence of random variables $(X_n : n \in \mathbb{N})$ be setting $X_n = G_n(Y_n)$. Also $\sigma(X_n) \subseteq \sigma(Y_n)$ so $(X_n : n \in \mathbb{N})$ is independent.

18/10/12

Probability and Measure (7)

Finally $F_n(x) = \mathbb{P}(G_n(Y_n) \le x) = \mathbb{P}(Y_n \le F_n(x)) = F_n(x)$

2.5 Convergence of measurable functions and random variables

Let $(E, \mathcal{E}, \mu)$ be a measurable space. We often define a set $A \in \mathcal{E}$ by a property characterising its elements. If $\mu(A^c) = 0$ then we say that this property holds almost everywhere. (almost surely)

Thus we say that a sequence of measurable functions $(f_n : n \in \mathbb{N})$ converges almost everywhere to f if

$$\mu(\{x \in E : f_n(x) \not\to f(x)\}) = 0 \qquad (f_n \to f \text{ a.e.})$$

We say that $f_n \to f$ in measure (in probability) if $\forall \varepsilon > 0$,

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

Proposition 2.5.1 Let $(f_n : n \in \mathbb{N})$, f be measurable functions.

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

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

Proof

By considering $f_n - f$ if necessary we reduce to the case $f = 0$. ~~As~~

a) Assume $f_n \to 0$ almost everywhere.
Given $\varepsilon > 0$, $\mu(|f_n| \le \varepsilon) \ge \mu(\bigcap_{m \ge n} \{|f_m| \le \varepsilon\})$

$$\uparrow \mu(|f_n| \le \varepsilon \text{ eventually}) \ge \mu(f_n \to 0) \ge \mu(E)$$

Hence if $\mu(E) < \infty$, this forces $\mu(|f_n| > \varepsilon) \to 0$ so $f_n \to 0$ in measure.

3

b) Suppose that $f_n \to f$ in measure. Then there exist (n_k) such that $\mu(|f_{n_k}| > \frac{1}{k}) \leq 2^{-k}$

Then $\mu(|f_{n_k}| > \frac{1}{k}$ infinitely often$) = 0$ by the Borel-Cantelli Lemma I $(\sum 2^{-k} < \infty)$

So $|f_{n_k}| \leq \frac{1}{k}$ eventually almost everywhere, so $f_{n_k} \to 0$ almost everywhere $\square$

22/10/12

Probability and Measure (8)

Convergence of Random Variables

Let $X, (X_n : n \in \mathbb{N})$ be real random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $X_n \to X$ <u>almost surely</u> if

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

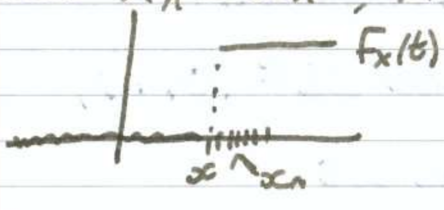
We say that $X_n \to X$ <u>in probability</u> if $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$ $\forall \varepsilon > 0$.

For real random variables $X, (X_n : n \in \mathbb{N})$ (defined possibly in different probability spaces) we say that $X_n \to X$ <u>in distribution</u> if $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$ for all $x \in \mathbb{R}$ such that F_X is continuous at x.

<u>Example</u>

Take $x, (x_n : n \in \mathbb{N})$ in $\mathbb{R}$ with $x_n \to x$ as $n \to \infty$.

Set $X_n = x_n$, $X = x$. Suppose $x_n > x$



$F_X(x) = 1$, $F_{X_n}(x) = 0$ $\forall n$.

<u>Example</u>

Take $X = \pm 1$ with probability $\frac{1}{2}$. Set $X_n = (-1)^n X$.

Here it is not true that $X_n \to X$ almost surely, or in probability, but $F_{X_n} = F_X$ $\forall n$ so $X_n \to X$ in distribution.

<u>Theorem 2.5.2</u>

Let $X, (X_n : n \in \mathbb{N})$ be real random variables.

a) Suppose $X, (X_n : n \in \mathbb{N})$ are all defined on the same space an $X_n \to X$. Then $X_n \to X$ in distribution.

b) If $X_n \to X$ in distribution, then there exist random variables $\tilde{X}, \tilde{X}_n$ having the same distributions as X, X_n respectively, and defined on the same space such that $\tilde{X}_n \to \tilde{X}$ almost surely.

Proof

Write S for the subset of $\mathbb{R}$ where F_X is continuous.

a) Suppose $X_n \to X$ in probability. Take $x \in S$ and $\varepsilon > 0$. There exists $\delta > 0$ such that $|F_X(y) - F_X(x)| \leq \varepsilon/2$ if $|y - x| \leq \delta$. Then, there exists N such that $\forall n \geq N$, $\mathbb{P}(|X_n - X| \geq \delta) \leq \varepsilon/2$ ~~$|F_{X_n}(x) - F_X(x)| \leq \varepsilon/2$~~. Then for $n \geq N$,

$$F_{X_n}(x) = \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x + \delta) + \mathbb{P}(|X_n - X| \geq \delta) \leq F_X(x) + \varepsilon$$

$$1 - F_{X_n}(x) = \mathbb{P}(X_n > x) \leq \mathbb{P}(X > x - \delta) + \mathbb{P}(|X_n - X| \geq \delta)$$

$$\leq 1 - F_X(x) + \varepsilon.$$

So $|F_{X_n}(x) - F_X(x)| \leq \varepsilon$, as required.

b) Suppose $X_n \to X$ in distribution. Take $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}(\Omega)$, $\mathbb{P} =$ Lebesgue, and set

$$\hat{X}_n(\omega) = \inf\{x \in \mathbb{R} : \omega \leq F_{X_n}(x)\},$$

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

Then $\tilde{X}, \hat{X}_n$ are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and $F_{\tilde{X}} = F_X$, $F_{\tilde{X}_n} = F_{\tilde{X}}$, $\forall n$.

Since $\tilde{X}$ is non-decreasing, it has at most countably many discontinuities. So the set $\Omega_0 \subseteq (0,1)$ where $\hat{X}$ is continuous is an event with $\mathbb{P}(\Omega_0 = 1)$.

22/10/12

Probability and Measure (8)

Since F_X is non-decreasing, $\mathbb{R} \setminus S$ is also countable, so S is dense in $\mathbb{R}$. Take $\omega \in \Omega_0$ and $\varepsilon > 0$.

There exist $x^{\pm} \in S$ such that $x^- < \hat{X}(\omega) < x^+$ with $x^+ - x^- < \varepsilon$. Then there exists $\omega^+ \in (\omega, 1)$ such that $\hat{X}(\omega^+) \leq x^+$ (Remember that $\tilde{X}(\omega) \leq x \Leftrightarrow \omega \leq F_X(x)$)

Now $x^- < \tilde{X}(\omega)$ so $F_X(x^-) < \omega < \omega^+ \leq F_X(x^+)$

$\hat{X}(\omega^+) \leq x^+$.

There exists N such that $\forall n \geq N$, $F_{X_n}(x^-) < \omega < F_{X_n}(x)$, that is to say that $x^- \leq \tilde{X}_n(\omega) \leq x^+$. Hence

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

2.6 Tail Events

Let $(X_n : n \in \mathbb{N})$ be a sequence of r.v.s on $(\Omega, \mathcal{F}, \mathbb{P})$.

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

We call $\mathcal{T}$ the <u>tail σ-algebra</u> of $(X_n : n \in \mathbb{N})$. We call each $A \in \mathcal{T}$ a <u>tail event</u>

<u>Example</u>

If $(X_n \to X)$ as $n \to \infty$ (everywhere) then X is $\mathcal{T}$ measurable. Generally, $\limsup X_n$ is $\mathcal{T}$-measurable.

<u>Theorem 2.6</u>

Let $(X_n : n \in \mathbb{N})$ be a sequence of independent r.v.s. The tail σ-algebra $\mathcal{T}$ of $(X_n : n \in \mathbb{N})$ is <u>trivial</u>, that is $\mathbb{P}(A) \in \{0, 1\}$ for all $A \in \mathcal{T}$. Moreover, for any

$\mathcal{T}$-measurable rv Y there exists $c \in \mathbb{R}$ such that $\mathbb{P}(Y = c) = 1$.

<u>Proof.</u>

Set $\mathcal{T}_n = \sigma(X_1, \dots, X_n)$. Then $\mathcal{T}_n$ is generated by the π-system of events $B = \{X_{n+1} \le x_{n+1}, \dots, X_{n+m} \le x_{n+m}\}$ $\forall m \in \mathbb{N}$. Now $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all such A, B, by independence of $(X_n : n \in \mathbb{N})$. By Theorem 1·12·1, the σ-algebras $\mathcal{T}_n$, $\mathcal{T}_m$ are independent. Hence $\mathcal{T}_n$ is independent of $\mathcal{T}$ for all n. Now $\bigcup_n \mathcal{T}_n$ is a π-system generating $\mathcal{T}_\infty = \sigma(X_n : n \in \mathbb{N})$. So by Theorem 1·12·1 again, $\mathcal{T}$ and $\mathcal{T}_\infty$ are independent. But $\mathcal{T} \subseteq \mathcal{T}_\infty$. So for $A \in \mathcal{T}$. $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A^2)$. So $\mathbb{P}(A) \in \{0, 1\}$.

If Y is a $\mathcal{T}$-measurable rv then $\mathbb{P}(Y \le y) \in \{0, 1\}$, for all y. Set $c = \inf\{y \in \mathbb{R} : F_Y(y) = 1\}$. Then $\mathbb{P}(Y = c) = 1$ $\square$

24/10/12 Probability and Measure (9)

3 Integration

3.1 Definitions and Basic Properties

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

$A_n \uparrow A$, $(A_n \subseteq A_{n+1}, \bigcup_n A_n = A) \Rightarrow \mu(A_n) \uparrow \mu(A)$

We shall define for measurable functions $f : E \to [-\infty, \infty]$ the integral, written $\mu(f) = \int_E f \, d\mu = \int_E f(x) \mu(dx)$

Integrals with respect to Lebesgue measure on $(\mathbb{R}, \mathcal{B})$ are usually written $\int_{\mathbb{R}} f(x) \, dx$.

For a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ call the integral the expectation and write $\mathbb{E}(X) = \int_\Omega X \, d\mathbb{P}$

A <u>simple function</u> f on $(E, \mathcal{E})$ is any function of the form

$f = \sum_{k=1}^{m} a_k 1_{A_k}$, $m \in \mathbb{N}$, $a_k \geq 0$, $A_k \in \mathcal{E}$.

These are precisely the measurable functions $f : E \to [0, \infty)$ taking finitely many values. Define for such functions f

$\mu(f) = \sum_{k=1}^{m} a_k \mu(A_k)$ (agree that $0 \cdot \infty = 0$)

This is well defined (see exercise 3.1) and satisfies for simple functions f, g: $(\alpha, \beta \in \mathbb{R}_{\geq 0})$

a) $\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)$

b) $f \leq g$ implies that $\mu(f) \leq \mu(g)$

(To see this note that $g - f$ is simple, $\mu(f) + \mu(g-f) = \mu(g)$ by a)

c) $\mu(f) = 0 \iff f = 0$ almost everywhere

For $f \geq 0$ measurable, we define

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

It is clear that $f \leq h \Rightarrow \mu(f) \leq \mu(h)$. The definition is consistent for simple functions by property b). Say that a measurable function f is <u>integrable</u> if $\mu(|f|) < \infty$. For integrable functions f we define $\mu(f) = \mu(f^+) - \mu(f^-)$ (*)

Here $f^\pm \overset{= \pm f \vee 0}{=} \max\{\pm f, 0\}$. Note that $f^\pm \leq |f|$, so $\mu(f^\pm) \leq \mu(|f|)$ so $|\mu(f)| \leq \mu(|f|)$. We can define $\mu(f)$ by (*) provided not both $\mu(f^+)$, $\mu(f^-)$ are infinity.

~~Write $x_n \uparrow x$~~ For $x, x_n \in [0, \infty]$, write $x_n \uparrow x$ if $x_n \leq x_{n+1}$ and $x_n \to x$ as $n \to \infty$. ($x = \infty$ allowed).

For functions $f, f_n : E \to [0, \infty]$, write $f_n \uparrow f$ if $f_n(x) \uparrow f(x)$ for all $x \in E$.

<u>Theorem</u> (Monotone Convergence)

Let f be a non-negative measurable function and let $(f_n : n \in \mathbb{N})$ be a sequence of such functions. Suppose that $f_n \uparrow f$. Then $\mu(f_n) \uparrow \mu(f)$.

<u>Proof</u>

Note that $\mu(f_n) \uparrow a \leq \mu(f)$ for some $a \in [0, \infty]$.

So our task is to show that $a = \mu(f)$.

<u>Case 1</u> ($f_n = 1_{A_n}$, $f = 1_A$)

$\mu(f_n) = \mu(A_n) \uparrow \mu(A) = \mu(f)$ by countable additivity.

Case 2 (f_n simple, $f = 1_A$)

Given $\varepsilon > 0$, consider $A_n = \{x \in E : f_n(x) > 1 - \varepsilon\}$.

Then $A_n \in \mathcal{E}$ and $A_n \uparrow A$ as $n \to \infty$. ~~So~~ Note that $f_n \geq (1-\varepsilon)1$

$f_n \geq (1-\varepsilon) 1_{A_n}$. So $\mu(f_n) \geq (1-\varepsilon)\mu(A_n) \uparrow (1-\varepsilon)\mu(A)$

but $(1-\varepsilon)\mu(A) = (1-\varepsilon)\mu(f)$. So since $\varepsilon > 0$ was arbitrary, we are done.

Case 3 (f_n simple, f simple)

Write $f = \sum_{k=1}^{m} a_k 1_{A_k}$ with $a_k > 0$, and $A_k \in \mathcal{E}$ disjoint

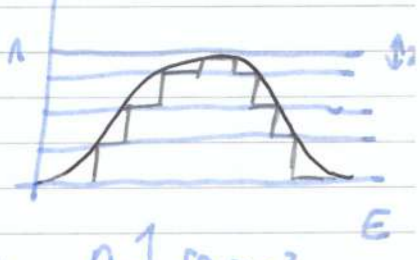
Then $a_k^{-1} 1_{A_k} f_n \uparrow 1_{A_k}$ as $n \to \infty$ for all k.

So $\mu(1_{A_k} f_n) \uparrow \mu(A_k)$ (case 2), note $f_n = \sum_{k=1}^{m} 1_{A_k} f_n$

so $\mu(f_n) = \sum_{k=1}^{m} \mu(1_{A_k} f_n) \uparrow \sum_{k=1}^{m} a_k \mu(A_k) = \mu(f)$

Case 4 (f_n simple, $f \geq 0$ measurable)

Take $g \leq f$ simple. Then $f_n \wedge g = \min\{f_n, g\} \uparrow g$ so $\mu(f_n) \geq \mu(f_n \wedge g) \uparrow \mu(g)$. Take the supremum over all such g to see that $\mu(f_n) \uparrow \mu(f)$.



Case 5 ($f_n, f \geq 0$ measurable)

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

$$= \sum_{i=0}^{n2^n - 1} i\, 2^{-n} 1_{\{x \in E : i2^{-n} \leq f_n(x) < (i+1)2^{-n}\}} + n 1_{\{f_n \geq n\}}$$

Then g_n is simple, ~~$g_n \geq f_n$~~ and $f_n \uparrow f$ forces $g_n \uparrow f$. Also, $g_n \leq f_n$, so $\mu(f_n) \geq \mu(g_n) \uparrow \mu(f)$ $\square$

Proposition 3.1.2

Let f, g be non-negative measurable functions. Then $(\alpha, \beta \in \mathbb{R}_{\geq 0})$

a) $\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)$

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

c) $\mu(f) = 0$ iff $f = 0$ almost everywhere.

Proofs

a) Consider $f_n = (2^{-n} \lfloor 2^n f \rfloor) \wedge n$, $g_n = (2^{-n} \lfloor 2^n g \rfloor) \wedge n$

then f_n, g_n simple, $f_n \uparrow f$, $g_n \uparrow g$, so $\alpha f_n + \beta g_n \nearrow \alpha f + \beta g$

So by monotone convergence $\mu(f_n) \uparrow \mu(f)$, $\mu(g_n) \uparrow \mu(g)$

$\mu(\alpha f_n + \beta g_n) \uparrow \mu(\alpha f + \beta g)$, letting $n \to \infty$.

b) We obtain b) as seen already.

c) $\mu(f) = 0$ iff $\mu(f_n) = 0$ for all n

iff $f_n = 0$ almost ~~every~~ everywhere for all n iff $f = 0$ almost everywhere $\square$

26/10/12

Probability and Measure (10)

Recap

Let $(E, \mathcal{E}, \mu)$ be a measure space. Write $\mathcal{M}^+$ for the set of non-negative measurable functions. Then there exists a unique map $\mu : \mathcal{M}^+ \to [0, \infty]$ such that $\mu(1_A) = \mu(A)$ for all $A \in \mathcal{E}$ and for every sequence $(f_n : n \in \mathbb{N})$ in $\mathcal{M}^+$ we have $\mu(\sum_n f_n) = \sum_n \mu(f_n)$.

Moreover, $\mu(\alpha f) = \alpha \mu(f)$ $\forall \alpha \in [0, \infty)$, $f \in \mathcal{M}^+$, and $\mu(f) = 0$ iff $f = 0$ almost everywhere.

Proposition 3.12

(We define for integrable functions (i.e. f measurable, $\mu(|f|) < \infty$)
$\mu(f) = \mu(f^+) - \mu(f^-)$. $f^\pm = (\pm f) \vee 0$)

Let f, g be integrable functions. Then

a) $\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)$ $\forall \alpha, \beta \in \mathbb{R}$.

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

c) $f = 0$ almost everywhere $\Rightarrow \mu(f) = 0$.

Proof

Since $(-f)^\pm = f^\mp$ we have $\mu(-f) = \mu(f^-) - \mu(f^+) = -\mu(f)$.
For $\alpha \geq 0$, $(\alpha f)^\pm = \alpha f^\pm$ so $\mu(\alpha f) = \mu(\alpha f^+) - \mu(\alpha f^-) = \alpha \mu(f)$

Suppose $f + g = h$. Then $f^+ + g^+ + h^- = f^- + g^- + h^+$.
So $\mu(f^+) + \mu(g^+) + \mu(h^-) = \mu(f^-) + \mu(g^-) + \mu(h^+)$
and so $\mu(f) + \mu(g) = \mu(h)$. This proves a).

For b), note for $f \leq g$, $g - f \geq 0$, so using a),
$\mu(g) - \mu(f) = \mu(g - f) \geq 0$.

For c), if $f = 0$ almost everywhere, then $f^\pm = 0$ almost everywhere, so $\mu(f) = \mu(f^+) - \mu(f^-) = 0$ $\square$

Proposition 3.1.3
Let f be a measurable function and let $(f_n : n \in \mathbb{N})$ be a sequence of such functions. Suppose $f_n(x) \geq 0$ for all n, and suppose $f_n(x) \uparrow f(x)$ as $n \to \infty$, for almost all x. Then $\mu(f_n) \uparrow \mu(f)$ (exercise)

3.2 Limits and Integrals

$\mu(\liminf f_n) = 0$
$\liminf \mu(f_n) = 1$

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

Proof
Set $g_n = \inf_{m \geq n} f_m$ and $g = \lim_{n \to \infty} g_n$ $(= \sup_n g_n)$
Then $g_n \leq f_m$ for all $m \geq n$, so $\mu(g_n) \leq \inf_{m \geq n} \mu(f_m)$ (*)
But $g_n \uparrow g$ so by monotone convergence,
$\mu(g_n) \uparrow \mu(g)$. Now let $n \to \infty$ in (*) to obtain
$\mu(\liminf f_n) = \mu(g) = \lim_n \mu(g_n) \leq \liminf_n \mu(f_n)$. $\square$

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

We call g a dominating function.

Proof

Since $|f_n| \leq g$, we have $\mu(|f_n|) \leq \mu(g) < \infty$.
Also, $|f| \leq g$ so $\mu(|f|) < \infty$. Hence f_n, f integrable.
$0 \leq g \pm f_n$, and $g \pm f = \liminf_n (g \pm f_n)$.

26/10/12

Probability and Measure (10)

By Fatou's Lemma, $\mu(g)+\mu(f)=\mu(g+f)=\mu(\liminf_n (g+f_n))$
$\leq \liminf_n \mu(g+f_n) = \mu(g)+\liminf_n \mu(f_n)$.

$\mu(g)-\mu(f)=\mu(g-f)=\mu(\liminf_n (g-f_n))$
$\leq \liminf_n \mu(g-f_n) = \mu(g) - \limsup_n \mu(f_n)$

Since $\mu(g)<\infty$, we obtain $\mu(f) \leq \liminf_n \mu(f_n)$
$\leq \limsup_n \mu(f_n) \leq \mu(f)$. Hence $\mu(f_n)\to\mu(f)$ as required.

3.3 Transformations of Integrals

Proposition 3.3.1 $(E,\mathcal{E},\mu)$ a measure space
~~Let (E~~ Let $A\in\mathcal{E}$. Write $\mathcal{E}_A = \{B\in\mathcal{E} : B\subseteq A\}$ and set $\mu_A=\mu|_{\mathcal{E}_A}$. Then $\mathcal{E}_A$ is a σ-algebra, μ_A is a measure on $\mathcal{E}_A$ and for measurable functions f on E, $f|_A$ is $\mathcal{E}_A$-measurable and $\mu_A(f|_A)=\mu(f1_A)$.

Examples

$E=\mathbb{R}$, $\mathcal{E}=\mathcal{B}$, $A=I$ an interval. We write
$\int_a^b f(x)\,dx = \int_I f(x)\,dx = \mu_I(f)$
$a=\inf I$, $b=\sup I$. The notation is ok since
$\mu(\{a\})=\mu(\{b\})=0$.

Proposition 3.3.3
Let f be a non-negative measurable function. Define for $A\in\mathcal{E}$, $\nu(A)=\mu(f1_A)$. Then ν is a measure on $\mathcal{E}$, and for all non-negative measurable functions g, $\nu(g)=\mu(fg)$.

Proof

Take a sequence of disjoint sets $A_n\in\mathcal{E}$.
Then $\nu(\bigcup_n A_n) = \mu(f\sum 1_{A_n}) = \mu(\sum_n f1_{A_n})$
(monotone convergence) $=\sum\mu(f1_{A_n}) = \sum\nu(A_n)$

Hence, ν is a measure.

For $g \in 1_B$, $B \in \mathcal{E}$, we have $\nu(g) = \nu(B) = \mu(f 1_B)$
$= \mu(fg)$.

The identity $\nu(g) = \mu(fg)$ extends to simple functions by linearity.

For $g \geq 0$ measurable, recall $g_n = (2^{-n} \lfloor 2^n g \rfloor) \wedge n \uparrow g$ and g_n is simple. Then by monotone convergence,

$$\nu(g) = \lim_n \nu(g_n) = \lim_n \mu(f g_n) = \mu(fg) \quad \square$$

Proposition 3·3·4

Let $(E, \mathcal{E}, \mu)$ be a measurable space $(G, \mathcal{G})$ a measurable space, and $f: E \to G$ a measurable function.

Let $\nu = \mu \circ f^{-1}$ be the image measure. Then for all non-negative measurable functions g on G, we have $\nu(g) = \mu(g \circ f)$

Example

If X is an $(E, \mathcal{E})$-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, and if $g: E \to [0, \infty)$ is measurable, then

$\mu_X(g) = \int g \, d\mu_X = \mathbb{E}(g(X))$

We often use Proposition 3·3·3 to specify a measure on $\mathbb{R}$ by a density function f with respect to Lebesgue measure μ.

$\nu(A) = \int_A f(x) \mu(dx) = \int_A f(x) \, dx$

If we have f with $\int_{\mathbb{R}} f(x) \, dx = 1$, we call f a probability density function. We can specify a probability distribution ν on $\mathbb{R}$ in this way. $\frac{d\nu}{d\mu} = f$, $\nu(dx) = f(x) \, dx$.

If a random variable X has $\mu_X = \nu$, then for $g: \mathbb{R} \to [0, \infty)$ measurable, $\mathbb{E}(g(X)) = \int_{\mathbb{R}} g \, d\mu_X = \int_{\mathbb{R}} g(x) f(x) \, dx$

3·4 Fundamental Theorem of Calculus

Theorem 3·4·1

Let $f: [a, b] \to \mathbb{R}$ be a continuous function.

a) Define $F_a(t) = \int_a^t f(x) \, dx$, $t \in [a, b]$. Then F_a is differentiable on $[a,b]$ with $F' = f$.

b) Suppose $F : [a,b] \to \mathbb{R}$ is differentiable ~~on~~ with $F' = f$.

Then $\int_a^b f(x)\,dx = F(b) - F(a)$

Proof

Fix $t \in [a,b)$ and $\varepsilon > 0$. There exists $\delta > 0$ such that $[t, t+\delta] \subseteq [a,b]$ and for all $x \in [t, t+\delta]$, $|f(x) - f(t)| < \varepsilon$

Then for $0 < h \leq \delta$, $\left| \frac{F_a(t+h) - F_a(t)}{h} - f(t) \right|$

$$= \frac{1}{h} \left| \int_t^{t+h} (f(x) - f(t))\,dx \right| \leq \frac{1}{h} \int_t^{t+h} |f(x) - f(t)|\,dx < \varepsilon$$

This shows (since $\varepsilon > 0$ was arbitrary) that F_a is differentiable on the right at t with derivative f. A similar argument shows that F_a is differentiable on the left at every $t \in (a,b]$, with derivative f. Hence a) holds.

Consider $F - F_a$. Since $F - F_a$ is differentiable on $[a,b]$ with $(F - F_a)' = 0$, we have, by the Mean Value Theorem, that $F - F_a$ is constant. So $F(b) - F(a) = F_a(b) - F_a(a) = \int_a^b f(x)\,dx$

□

3.5 Differentiation under the Integral Sign

Theorem 3.5.1

Let $(E, \mathcal{E}, \mu)$ be a measure space and U be an open interval in $\mathbb{R}$. Suppose that $f : U \times E \to \mathbb{R}$ satisfies

I) $x \mapsto f(t,x) : E \to \mathbb{R}$ is ~~differentiable~~ integrable for all $t \in U$.

II) $t \mapsto f(t,x) : U \to \mathbb{R}$ is differentiable for all $x \in E$.

(III) There exists an integrable function g on E such that $\left| \frac{\partial f}{\partial t}(t,x) \right| \leq g(x)$ for all $t \in U$, $x \in E$.

(Note: we may choose U to make finding such a g easier)

Then $x \mapsto \frac{\partial f}{\partial t}(t,x) : E \to \mathbb{R}$ is integrable for all $t \in U$.

Moreover, if we define $F: U \to \mathbb{R}$ by $F(t) = \int_E f(t,x)\,\mu(dx)$ then F is differentiable on U with $F'(t) = \int_E \frac{\partial f}{\partial t}(t,x)\,\mu(dx)$

Proof

Fix $t \in U$ and a sequence (h_n) in $\mathbb{R} \setminus \{0\}$ such that $t + h_n \in U$ for all n and $h_n \to 0$ as $n \to \infty$. Define $g_n(x) = \frac{1}{h_n}(f(t+h_n,x) - f(t,x)) - \frac{\partial f}{\partial t}(t,x)$

Then $g_n(x) \to 0$ as $n \to \infty$ for all x by (II). By the Mean Value Theorem $|g_n(x)| = |\frac{\partial f}{\partial t}(\xi_n, x) - \frac{\partial f}{\partial t}(t,x)|$ for some $\xi_n \in (t, t+h_n)$. So $|g_n(x)| \leq 2g(x)$

Hence $x \mapsto \frac{\partial f}{\partial t}(t,x)$ is measurable (as the limit of measurable functions) and so also integrable (by III) and by dominated convergence $\frac{1}{h_n}(F(t+h_n) - F(t)) - \int_E \frac{\partial f}{\partial t}(t,x)\mu(dx) = \int_E g_n(x)\mu(dx) \to 0$ □

Proposition 3·4·2 (Integration by Substitution)

Let $\phi : [a,b] \to \mathbb{R}$ be strictly increasing, and differentiable with continuous derivative. Then, for all non-negative measurable functions g on $[\phi(a), \phi(b)]$, we have

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

Proof

Consider the measure on $[\phi(a), \phi(b)]$ given by $\nu = \mu \circ \phi^{-1}$
where $\frac{d\mu}{dx} = \phi'(x)$ $\left(\mu(A) = \int_A \phi'(x)\,dx \right)$
Then $\nu(g) = \mu(g \circ \phi) = \int_a^b g(\phi(x))\,\phi'(x)\,dx$
Now for $[c,d] \subseteq [a,b]$, $\nu([\phi(c), \phi(d)]) = \int_c^d \phi'(x)\,dx$
$= \phi(d) - \phi(c)$ by the Fundamental Theorem of Calculus.

Hence ν is Lebesgue by uniqueness of extension $\square$

3.6 Product Measure and Fubini's Theorem

Theorem 2.1.2 (Monotone class theorem)

Let $(E, \mathcal{E})$ be a measurable space and let $\mathcal{A}$ be a π-system on E ~~containing~~ ~~$\mathcal{E}$~~ with $E \in \mathcal{A}$ and $\sigma(\mathcal{A}) = \mathcal{E}$. Suppose that V is a vector space of bounded $\mathcal{E}$-measurable functions such that

i) $1_A \in V$ for all $A \in \mathcal{A}$

ii) If $f_n \in V$ for all n and $0 \leq f_n \uparrow f$ then $f \in V$

then V contains all bounded $\mathcal{E}$-measurable functions.

Proof

Consider $\mathcal{D} = \{A \in \mathcal{E} : 1_A \in V\}$. Then $\mathcal{A} \subseteq \mathcal{D}$ so $E \in \mathcal{D}$. For $A \subseteq B$, $1_{B \setminus A} = 1_B - 1_A$, so if $A, B \in \mathcal{D}$ then $B \setminus A \in \mathcal{D}$.
For $A_n \uparrow A$, $1_{A_n} \uparrow 1_A$, so if $A_n \in \mathcal{D}$ for all n, then $A \in \mathcal{D}$. Hence $\mathcal{D}$ is a d-system, so $\mathcal{D} = \sigma(\mathcal{A}) = \mathcal{E}$ by Dynkin's Lemma.
Now for $f \geq 0$, measurable, bounded, consider $f_n = 2^{-n} \lfloor 2^n f \rfloor$.
Then f is simple, so $f_n \in V$, and $0 \leq f_n \uparrow f$, so $f \in V$.
Finally, for f bounded, measurable, $f^{\pm} \in V$ so $f = f^+ - f^- \in V$ $\square$

Let $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ be finite (for now) measure spaces. ~~Write~~ Set $\mathcal{A} = \{A_1 \times A_2 : A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2\}$
Write $E = E_1 \times E_2$, $\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2 := \sigma(\mathcal{A})$. Note that $\mathcal{A}$ is a π-system on E with $E \in \mathcal{A}$ and $\sigma(\mathcal{A}) = \mathcal{E}$.

Lemma 3·6·1

Let $f : E \to \mathbb{R}$ be an $\mathcal{E}$-measurable function.

Then for $x_1 \in E_1$ the function $x_2 \mapsto f(x_1, x_2) : E_2 \to \mathbb{R}$ is $\mathcal{E}_2$-measurable.

Proof

Apply the monotone class theorem: with ~~write~~ V ~~for~~ the set of all bounded, $\mathcal{E}$-measurable functions f for which the conclusion holds.

Lemma 3·6·2

Let $f : E \to \mathbb{R}$ be a bounded, $\mathcal{E}$-measurable function and define

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

Then f_1 is an $\mathcal{E}_1$-measurable function.

Proof

Use the monotone class theorem. $\square$

Theorem 3·6·3 (Product Measure)

There exists a unique measure μ ($=: \mu_1 \otimes \mu_2$) on $\mathcal{E}$ such that $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for all $A_1 \in \mathcal{E}_1$, $A_2 \in \mathcal{E}_2$.

Proof

For uniqueness, since A is a π-system containing E, and we must have $\mu(E) = \mu(E_1)\mu(E_2) < \infty$, this follows by uniqueness of extension. For existence, define for $A \in \mathcal{E}$

$$\mu(A) = \int_{E_1} \left(\int_{E_2} 1_A(x_1, x_2)\, \mu(dx_2) \right) \mu(dx_1)$$

31/10/12

This is well defined by the lemma, and $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for all $A_1 \in \mathcal{E}_1$, $A_2 \in \mathcal{E}_2$. For $(A_n : n \in \mathbb{N})$ disjoint in $\mathcal{E}$,

arrows

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

by the monotone convergence theorem. $\square$

Proposition 3.6.4

Define $\hat{E} = E_2 \times E_1$, $\hat{\mathcal{E}} = \mathcal{E}_2 \otimes \mathcal{E}_1$, $\hat{\mu} = \mu_2 \otimes \mu_1$. For f with $f: E \to \mathbb{R}$, define $\hat{f}: \hat{E} \to \mathbb{R}$ by $\hat{f}(x_2, x_1) = f(x_1, x_2)$

Then suppose that f is ~~non-negative~~ bounded and $\mathcal{E}$-measurable. $\hat{f}$ is then $\hat{\mathcal{E}}$ measurable and $\hat{\mu}(\hat{f}) = \mu(f)$.

Proof

(If $0 \le f_n \uparrow f$ bounded, then $\hat{\mu}(\hat{f}) = \lim_n \hat{\mu}(\hat{f}_n)$ and $\mu(f) = \lim \mu(f_n)$) by monotone convergence

Use the monotone class theorem.

Theorem 3.6.5 (Fubini's Theorem)

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

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

Prop 3.6.4

b) Let f be a μ-integrable ($\mathcal{E}$-measurable) function, then

i) $x_2 \mapsto f(x_1, x_2) : E_2 \to \mathbb{R}$ is μ_2-integrable for almost all (with μ_1) $x_1 \in E_1$.

ii) $x_1 \mapsto \int_{E_2} f(x_1, x_2)\,\mu_2(dx_2)$ is μ_1-integrable and the formula in a) holds.

Proof

a) The case $f = 1_A$ for $A \in \mathcal{E}$ is true by definition. Then (*) extends to simple functions by linearity, and to non-negative measurable functions by monotone convergence.

(For $f \geq 0$, $\mathcal{E}$-measurable, consider $f_n = (2^{-n} \lfloor 2^n f \rfloor) \wedge n$.

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

b) For f a μ-integrable function, we have

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

So $\int_{E_2} |f(x_1, x_2)| \mu_2(dx_2) < \infty$ for almost everywhere x_1 (with μ_1

and $x_1 \mapsto \left| \int_{E_2} f(x_1, x_2) \mu_2(dx_2) \right| \left(\leq \int_{E_2} |f(x_1, x_2)| \mu_2(dx_2) \right)$ is μ_1 integrable

22/11/12

Probability and Measure (13)

Recall $(E_1, \mathcal{E}_1, \mu_1)$, $(E_2, \mathcal{E}_2, \mu_2)$ finite measure spaces

$E = E_1 \times E_2$, $\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2 = \sigma(A_1 \times A_2,\ A_1 \in \mathcal{E}_1,\ A_2 \in \mathcal{E}_2)$

$\mu = \mu_1 \otimes \mu_2$, product measure on $\mathcal{E}$: $\mu(A_1 \times A_2) = \mu(A_1)\mu(A_2)$

<u>Theorem 3.6.5</u>

b) Let f be μ-integrable. Define $A_1 = \{x_1 \in E_1 : \underset{\text{is } \mu_2 \text{ integrable}}{x_2 \mapsto f(x_1, x_2)}\}$

Then $A_1 \in \mathcal{E}_1$ and $\mu_1(A_1^c) = 0$. Define

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

Then f_1 is μ_1-integrable and $\mu_1(f_1) = \mu(f)$.

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

<u>Proof</u>

By Lemma 3.6.2 and part (a), the map

$x_1 \mapsto \int_{E_2} |f(x_1, x_2)|\, \mu_2(dx_2)$ is $\mathcal{E}_1$-measurable,

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

Hence $A_1 \in \mathcal{E}_1$ and $\mu_1(A_1^c) = 0$. Define

$f_1^{(\pm)}(x_1) = \int_{E_2} f^{\pm}(x_1, x_2)\mu_2(dx_2)$

Then $f_1^{(\pm)} < \infty$ on A_1 and $f_1 = (f^{(+)} - f^{(-)}) 1_{A_1}$.

Now $f_1^{(\pm)}$ are μ-measurable and by part (a)

$\mu_1(f_1^{(\pm)}) = \mu(f^{\pm}) < \infty$

Hence f_1 is μ_1-integrable and

$\mu_1(f_1) = \mu_1(f^{(+)}) - \mu_1(f^{(-)}) = \mu(f^+) - \mu(f^-) = \mu(f)$ $\square$

1

Extensions and non-extensions

- ~~All results ex~~ Fubini's Theorem extends easily to σ-finite measures

$\mu_k = \sum_n \mu_k^{(n)}$ with $\mu_k^{(n)}$ finite for all n, $k = 1, 2$.

$\mu = \mu_1 \otimes \mu_2 = \sum_{n,m} \mu_1^{(n)} \otimes \mu_2^{(m)}$. $f \geq 0$, measurable.

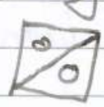
$\mu(f) = \sum_{n,m} \mu_1^{(n)} \otimes \mu_2^{(m)}(f) = \sum_{n,m}$ iterated intervals

$=$ iterated intervals (using monotone convergence)

- It does not extend to non-σ-finite measures

Take $E_1 = E_2 = [0,1]$, $\mu_1 =$ Lebesgue, $\mu_2 =$ counting

Consider ~~f:E~~ $f(x_1, x_2) = \begin{cases} 1 & x_1 = x_2 \\ 0 & \text{otherwise} \end{cases}$

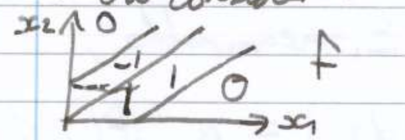


not σ-finite

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

$$\int_{E_2} \left(\int_{E_1} f(x_1, x_2)\, \mu_1(dx_1) \right) \mu_2(dx_2) = 0$$

Now consider:



$$\left.\begin{aligned} \int_0^\infty \left(\int_0^\infty f(x_1, x_2)\, dx_1 \right) dx_2 &= \tfrac{1}{2} \\ \int_0^\infty \left(\int_0^\infty f(x_1, x_2)\, dx_2 \right) dx_1 &= -\tfrac{1}{2} \end{aligned}\right) \neq$$

- Associativity: $\mu_1 \otimes (\mu_2 \otimes \mu_3) = (\mu_1 \otimes \mu_2) \otimes \mu_3$

This is true since both sides agree on $\{A_1 \times A_2 \times A_3, A_i \in \mathcal{E}_i\}$,

a π-system generating the σ-algebra. So:

$\mu_1 \otimes \ldots \otimes \mu_n$ is well defined without brackets.

- Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Exercise:

$$\mathcal{B}(\mathbb{R}^n) = \underbrace{\mathcal{B}(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}(\mathbb{R})}_{n \text{ times}}$$

Write λ for Lebesgue measure on $\mathbb{R}$. Then the n-fold

product measure $\lambda_n = \underbrace{\lambda \otimes \ldots \otimes \lambda}_{n \text{ times}}$ is a measure on $\mathbb{R}^n$,

22/11/12

and is the unique measure such that $\lambda_n\left(\prod_{k=1}^{n}[a_k, b_k]\right) = \prod_{k=1}^{n}(b_k - a_k)$

We will write $\lambda_n(f) = \int_{\mathbb{R}^n} f(x)\, dx$

3.7 Independent Random Variables

Recall that random variables $X_1, \ldots, X_n$ on $(\Omega, \mathcal{F}, \mathbb{P})$ are independent if the σ-algebras that they generate are independent

i.e. $\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{k=1}^{n} \mathbb{P}(X_k \in A_k)$

Here X_k is $(E_k, \mathcal{E}_k)$ valued and $A_k \in \mathcal{E}_k$.

Proposition 3.7.1

Let $X_1, \ldots, X_n$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, with X_k taking values in the measurable space $(E_k, \mathcal{E}_k)$. ~~Define~~ Set $E = E_1 \times \ldots \times E_k$ and $\mathcal{E} = \mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_k$. Define $X: \Omega \to E$ by $X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$. Then X is an $(E, \mathcal{E})$ valued random variable. Moreover, the following are equivalent.

a) $X_1, \ldots, X_n$ are independent

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

c) For all non-negative measurable functions f_k on E_k, 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 $\mathcal{A} = \{\prod_{k=1}^{n} A_k : A_k \in \mathcal{E}_k\}$. Then for $A = \prod_{k=1}^{n} A_k \in \mathcal{A}$

$X^{-1}(A) = \bigcap_{k=1}^{n} X_k^{-1}(A_k) \in \mathcal{F}$. But $\sigma(\mathcal{A}) = \mathcal{E}$ so X is $\mathcal{E}$-measurable.

Consider $\nu = \mu_{X_1} \otimes \dots \otimes \mu_{X_n}$

Suppose that a) holds. Take $A = \prod_{k=1}^{n} A_k \in \mathcal{A}$.

Then $\mu_X(A) = \mathbb{P}(X \in A) = \mathbb{P}\left(\bigcap_{k=1}^{n} \{X_k \in A_k\}\right)$

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

But $\mathcal{A}$ is a π-system generating $\mathcal{E}$. So $\mu_X = \nu$ on $\mathcal{E}$. Hence b) holds.

Suppose now that b) holds. Then

$$\mathbb{E}\left(\prod_{k=1}^{n} f_k(X_k)\right) = \int_E \prod_{k=1}^{n} f(x_k)\, \mu_X(dx) \overset{\text{Fubini}}{=} \prod_{k=1}^{n} \int_{E_k} f_k(x_k)\, \mu_{X_k}(dx_k)$$

$$= \prod_{k=1}^{n} \mathbb{E}(f_k(x_k))$$

Finally, if c) holds, then we ~~have~~ can deduce a) by taking $f_k = 1_{A_k}$ for $\mathcal{E}_k$-measurable sets A_k. $\square$

25/11/12 Probability and Measure 14

4. Norms and Inequalities

4.1 L^p-norms

Let $(E, \mathcal{E}, \mu)$ be a measure space. Define, for $1 \leq p \leq \infty$, $L^p = L^p(E, \mathcal{E}, \mu)$ as the set of all measurable functions f on E having finite L^p-norm. Here:

$\|f\|_p = \left(\int |f|^p d\mu\right)^{1/p}$ for $1 \leq p < \infty$

$\|f\|_\infty = \inf\{\lambda \geq 0,\ |f| \leq \lambda \ \mu\text{-almost everywhere}\}$

(e.g. on $(\mathbb{R}, \mathcal{B}, dx)$, $\|1_{\mathbb{Q}}\|_\infty = 0$)

Note $\|f\|_p \leq \mu(E)^{1/p}\|f\|_\infty$ so if $\mu(E) < \infty$, then $L^\infty \subseteq L^p$ for all p. Say $f_n \to f$ in L^p if $\|f_n - f\|_p \to 0$ as $n \to \infty$.

4.2 Chebyshev's Inequality

Let f be a non-negative measurable function. Write $\{f \geq \lambda\}$ for the set $\{x \in E : f(x) \geq \lambda\}$. Note for $\lambda \geq 0$

$\lambda 1_{\{f \geq \lambda\}} \leq f$, so integrating we obtain

$\lambda \mu(\{f \geq \lambda\}) \leq \mu(f)$ Chebyshev's Inequality

For any measurable function g on E and any non-negative Borel-function φ on $\mathbb{R}$ we can apply this to $f = \varphi \circ g$ to get an inequality for g. For example (for $p \in [1, \infty)$), $\lambda > 0$

$\mu(|g| \geq \lambda) = \mu(|g|^p \geq \lambda^p) \leq \lambda^{-p}\mu(|g|^p)$

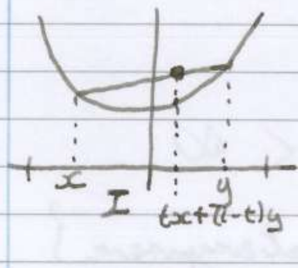
Note that for $g \in L^p$, we have $\mu(|g| \geq \lambda) = O(\lambda^{-p})$ as $\lambda \to \infty$.

1

4·3 Jensen's Inequality

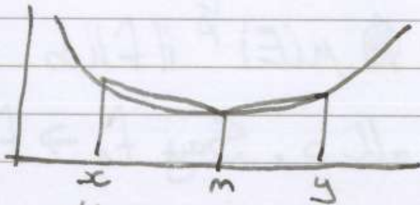
Let I be an interval in $\mathbb{R}$. Say that a function $f: I \to \mathbb{R}$ is <u>convex</u> if for all $x, y \in I$, and all $t \in [0,1]$,

$$f(tx + (1-t)y) \leq t f(x) + (1-t) f(y)$$



Lemma 4·3·1

Let $f: I \to \mathbb{R}$ be convex and let m be an interior point of I. Then there exist $a, b \in \mathbb{R}$ such that $ax + b \leq f(x)$ $\forall x \in I$ with equality at $x = m$.



<u>Proof</u>

There exist $x, y \in I$ with $x < m < y$. For all such x, y

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

. This is just a rearrangement of the basic convexity inequality, taking $m = tx + (1-t)y$.

So, $\exists\, a \in \mathbb{R}$ such that for all $x < m < y$,

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

Then $a(x - m) + f(m) \leq f(x)$ $\forall x \in I$.

So take $b = f(m) - am$ to obtain a, b with the required properties.

Theorem 4·3·2 (Jensen's Inequality)

Let X be an integrable random variable with values in I. Let $f: I \to \mathbb{R}$ be a convex function. Then $\mathbb{E}(f(X))$ is well defined and $f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$

25/11/12

Proof

The case where $X = c$ almost surely for some constant c is obvious. We exclude it. Then $m = \mathbb{E}(X)$ must be an interior point of I. ($X \geq \mathbb{E}(X)$ almost surely means $X - \mathbb{E}(X) \geq 0$, $\mathbb{E}(X - \mathbb{E}(X)) \neq 0$ ...)

We will see $f(\mathbb{E}(X)) = f(m) = am + b = \mathbb{E}(aX+b)$)

$ax + b \leq f(x)$ for all $x \in I$

so $aX + b \leq f(X)$ on Ω

Then $\mathbb{E}(f(X)^-) \leq |a| \mathbb{E}(X) + |b| < \infty$, so $\mathbb{E}(f(X))$ is well defined. Moreover

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

For $1 \leq p \leq q$, the function $x \mapsto |x|^{q/p}$ is convex on $\mathbb{R}$.

So if $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$, by Jensen's Inequality

$(\mathbb{E}(|X|^p))^{q/p} \leq \mathbb{E}(|X|^q)$ so

$\|X\|_p \leq \|X\|_q$

So the map $p \mapsto \|X\|_p$ is non-decreasing on $[1, \infty]$.

Hence $L^p \supseteq L^q$.

4.4 Hölder's and Minkowski's Inequality

We say that $p, q \in [1, \infty]$ are conjugate indices if

$\frac{1}{p} + \frac{1}{q} = 1$.

Theorem 4.4.1 (Hölder's Inequality)

Let f, g be measurable functions and let $p, q \in [1, \infty]$ by conjugate indices. Then $\mu(|fg|) \leq \|f\|_p \|g\|_q$

Proof

The cases $p \in \{1, \infty\}$ are clear; we exclude them.

The cases where $\|f\|_p \in \{0, \infty\}$ are clear. We exclude these too.

(N.B. $\alpha \|f\|_p = \|\alpha f\|_p$)

By scaling, we reduce to the case $\|f\|_p = 1$. Define

$\mathbb{P}(A) = \mu(1_A |f|^p)$, $A \in \mathcal{E}$.

Then $\mathbb{P}$ is a probability measure on $(E, \mathcal{E})$ and for all measurable functions X on E, $X \geq 0$, $\mathbb{E}(X) = \mu(X|f|^p)$, $\mathbb{E}(X) \leq \mathbb{E}(X^q)^{1/q}$

So $\mu(|fg|) = \mu\left(\frac{|g| 1_{|f|>0} |f|^p}{|f|^{p-1}}\right) = \mathbb{E}(X) \leq \mathbb{E}(X^q)^{1/q}$

(set $X = \frac{|g| 1_{|f|>0}}{|f|^{p-1}}$) $\quad = \mu\left(|g|^q 1_{|f|>0} \cancel{\frac{|f|^p}{|f|^{(p-1)q}}}\right)^{1/q}$

(since $\frac{1}{p} + \frac{1}{q} = 1$, $(p-1)q = p$)

$\leq \mu(|g|^q)^{1/q} = \|g\|_q = \|g\|_q \|f\|_p$ as required $\square$

07/11/12 Probability and Measure (15)

__Recap__

$\|f\|_p = \mu(|f|^p)^{1/p} \quad p \in [1, \infty)$

$\|f\|_\infty = \inf\{\lambda \geq 0 : |f| \leq \lambda \text{ almost everywhere}\}$

Hölder $\mu(|fg|) \leq \|f\|_p \|g_q\| \quad (\frac{1}{p} + \frac{1}{q} = 1)$

$\|\alpha f\|_p = |\alpha| \|f\|_p$

__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 for $p \in [1, \infty)$__

The cases $\|f+g\|_p = 0$ and $\|f\|_p = \infty$, and $\|g\|_p = \infty$ are obvious: we exclude them. Note

$|f+g| \leq |f| + |g| \leq 2\max\{|f|, |g|\}$

$|f+g|^p \leq 2^p \max\{|f|^p, |g|^p\} \leq 2^p(|f|^p + |g|^p)$

So $\mu(|f+g|^p) \leq 2^p(\|f\|_p^p + \|g\|_p^p) < \infty$

Now $|f+g|^p = |f+g||f+g|^{p-1} \leq (|f| + |g|)|f+g|^{p-1}$

so by Hölder's Inequality

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

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

$= \|f+g\|^{1-\frac{1}{p}}$ and the result follows. $\square$

5. Completeness of L^p and orthogonal projection

5.1 $\mathcal{L}^p$ as a Banach Space

Let V be a real vector space. A norm of V is a map $\| \cdot \| : V \to [0, \infty)$ such that:

i) $\| u + v \| \leq \| u \| + \| v \|$ for all $u, v \in V$

ii) $\| \alpha u \| = |\alpha| \| u \|$ for all $\alpha \in \mathbb{R}$, $u \in V$

iii) $\| u \| = 0 \Rightarrow u = 0$ for $u \in V$.

For each $p \in [1, \infty)$, the L^p-norm on L^p satisfies i) and ii). But iii) fails because $\|f\|_p = 0 \Rightarrow f = 0$ almost everywhere, but not $f \equiv 0$. Define an equivalence relation $\sim$ on L^p with $f \sim g$ if $f = g$ almost everywhere.

Then the set of equivalence classes $\mathcal{L}^p$ is a real-vector-space (exercise) and we can define consistently a norm on $\mathcal{L}^p$ by $\| [f] \|_p = \| f \|_p$. Actually, we prefer not to do this and to work directly with functions, not equivalence classes of functi

In any normed vector-space, we can define a metric $d(u, v) = \| u - v \|$. If this metric is complete, i.e. if every Cauchy sequence has a limit, then we say that $(V, \| \cdot \|)$ is a Banach space.

<u>Theorem 5.1.1</u> (Completeness of L^p)

Let $p \in [1, \infty)$, and $(f_n : n \in \mathbb{N})$ 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 $n \to \infty$.

(This says exactly that $(L^p, \|\cdot\|)$ is complete)

<u>Proof</u> (for $p \in [1, \infty)$)

There exists a subsequence (n_k) such that

$S = \sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p < \infty$ (fast Cauchy subsequence)

By Minkowski's inequality, for any $K \in \mathbb{N}$,

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

By monotone convergence, this inequality remains true for $K = \infty$

$S^p \geqslant \mu\left(|\sum_{k=1}^{K-1} |f_{n_{k+1}} - f_{n_k}|^p\right) \uparrow \mu\left(\sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}|^p\right)$

Consider $A = \{x \in E : \sum_{k=1}^{\infty} |f_{n_{k+1}}(x) - f_{n_k}(x)| < \infty\}$. Then $A \in \mathcal{E}$ and $\mu(E \setminus A) = 0$. By completeness of $\mathbb{R}$, $(f_{n_k}(x) : k \in \mathbb{N})$ has a limit for all $x \in A$. Define

$$f(x) = \begin{cases} \limsup_k f_{n_k}(x) & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Then f is measurable, and $f_{n_k}(x) \to f(x)$ for all $x \in A$.

Given $\varepsilon > 0$, choose N so that $\|f_n - f_m\| \leq \varepsilon$ for all $n, m \geqslant N$.

Then for $n \geqslant N$, we have $\|f_n - f_{n_k}\|_p \leq \varepsilon$ for all sufficiently large k.

Then $\mu(|f_n - f|^p) = \mu(\liminf_k |f_n - f_{n_k}|^p)$

$\mu(\liminf_k |f_n - f_{n_k}|^p) \leq \liminf_k \mu(|f_n - f_{n_k}|^p) \leq \varepsilon^p$

by Fatou's Lemma, for all $n \geq N$. Hence $f_n \to f$ in L^p $\square$.

Note that if $\|f_n - f\| \to 0$, $|\|f_n\| - \|f\|| \leq \|f_n - f\| \to 0$

5.2 L^2 as a Hilbert Space

Let V be a real vector space. An inner-product on V is a bilinear map $\langle \cdot , \cdot \rangle : V \times V \to \mathbb{R}$ such that

$\langle v, v \rangle \geq 0$ for all $v \in V$ with equality only if $v = 0$.

Given an inner product $\langle \cdot, \cdot \rangle$, we can define a norm by

$\|v\| = +\sqrt{\langle v, v \rangle}$ (Proof: exercise, Cauchy-Schwartz)

If the resulting normed space is a Banach space, then we call $(V, \langle \cdot, \cdot \rangle)$ a Hilbert space.

09/11/12 Probability and Measure (16)

Theorem 5·12 (Approximation in L^p)

Let $p \in [1, \infty)$, and let $\mathcal{A}$ be a π-system generating $\mathcal{E}$. Suppose $\mu(A) < \infty$ for all $A \in \mathcal{A}$. Consider

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

Then V_0 is a subspace of L^p, and for all $f \in L^p$, there exists a sequence $(f_n : n \in \mathbb{N})$ in V_0 with $\|f_n - f\|_p \to 0$.

Proof

Write V for the set of all $f \in L^p$ such that there exists $(f_n : n \in \mathbb{N})$ in V_0 with $f_n \to f$ in L^p. Let $\mathcal{D}$ denote the set of all $A \in \mathcal{E}$ such that $1_A \in V$. Then $\mathcal{A} \subseteq \mathcal{D}$ and....

(to be continued)

5·2 L^2 as a Hilbert space

Let V be a real vector space. A bilinear map $\langle \cdot , \cdot \rangle : V \times V$ is an inner product if $\langle v, v \rangle \geq 0$ for all $v \in V$ with equality only if $v = 0$. Given an inner product, we can define a norm on V by $\|v\| = \langle v, v \rangle^{\frac{1}{2}}$ (Cauchy-Schwartz)

Then $\|v+u\|^2 = \|v\|^2 + 2\langle v, u \rangle + \|u\|^2$ (Pythagoras)

and $\|v+u\|^2 + \|v-u\|^2 = 2\|v\|^2 + 2\|u\|^2$ (parallelogram law)

An inner-product space is a Hilbert Space if it is also complete.

For $f, g \in L^2$, define $\langle f, g \rangle = \mu(fg) = \int_E fg \, d\mu$

Then $\langle \cdot, \cdot \rangle : L^2 \times L^2 \to \mathbb{R}$ $\quad (\mu(|fg|) \leq \|f\|_2 \|g\|_2)$

is bilinear and $\langle f, f \rangle \geq 0$ but equality holds iff $f = 0$ almost everywhere. Note $\langle f, f \rangle^{\frac{1}{2}} = \|f\|_2$. Hence, since L^2 is complete, we see that $L^2 = \mathcal{L}^2/\sim$ is a Hilbert space.

For $f, g \in L^2$, we say f, g are <u>orthogonal</u> if $\langle f, g \rangle = 0$

Define for $V \subseteq L^2$, $V^\perp = \{ f \in L^2 : \langle f, g \rangle = 0 \text{ for all } g \in V \}$

We say that V is <u>closed</u> if for any sequence $(f_n : n \in \mathbb{N})$ in V such that $f_n \to f$ for some $f \in L^2$, there exists $g \in V$ such that $g = f$ almost everywhere.

<u>Theorem 5·2·1</u> (Orthogonal Projection)

Let V be a closed subspace of L^2. Define for $f \in L^2$

$d(f, V) = \inf \{ \|f - g\|_2 : g \in V \}$

For all $f \in L^2$, there exists $v \in V$ such that $\|f - v\|_2 = d(f, V)$

Moreover, for such v, $f - v \in V^\perp$.

<u>Proof</u>

There exists a sequence $(v_n : n \in \mathbb{N})$ in V such that $\|f - v_n\|_2 \to d(f, V)$. By the parallelogram law,

$\|2(f - \frac{1}{2}v_n - \frac{1}{2}v_m)\|_2^2 + \|v_n - v_m\|^2 = 2\|f - v_n\|^2 + 2\|f - v_m\|^2$

$4\|f - \frac{v_n + v_m}{2}\|_2^2$. Since $\frac{v_n + v_m}{2} \in V$, this is no less than $4\, d(f, V)^2$. Hence $\|v_n - v_m\| \to 0$ as $n, m \to \infty$.

By completeness of L^2, there exists $g \in L^2$ such that

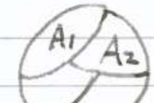
$v_n \to g$ in L^2. Since V is closed, there exists $v \in V$ such that $g = v$ almost everywhere. Then $\|f - v\|_2 = \lim_{n\to\infty} \|f - v_n\|_2 = d(f, V)$

Now for any $g \in V$ and $t \in \mathbb{R}$

$$0 \leq \|f - (v+tg)\|_2^2 - \|f - v\|_2^2$$
$$= 2\langle f-v, g\rangle t + \|g\|_2^2 t^2$$

So $\langle f-v, g\rangle = 0$. Hence $f - v \in V^\perp$ $\square$

5·3 Conditional Expectation $(\Omega, \mathcal{F}, \mathbb{P})$



Let $\mathcal{G}$ be a finite sub-σ-algebra of $\mathcal{F}$, generated by the partition $\mathcal{A}$ of Ω. Define for an integrable random variable X

$$\mathbb{E}(X|A) = \begin{cases} \mathbb{E}(X 1_A)/\mathbb{P}(A) & \text{if } \mathbb{P}(A) > 0 \\ 0 & \text{otherwise} \end{cases}$$

The (elementary) conditional expectation $\mathbb{E}(X|\mathcal{G})$ is defined by $\mathbb{E}(X|\mathcal{G}) = \sum_{A\in\mathcal{A}} \mathbb{E}(X|A) 1_A$

Theorem

orthogonal projection is only unique almost surely

The conditional expectation $\mathbb{E}(X, \mathcal{G})$ is (a version of) the orthogonal projection of X on $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

$L^2(\Omega, \mathcal{G}, \mathbb{P})$ is closed since it is finite dimensional, or as it is complete.

Proof

It is clear that $\overset{= Y}{\mathbb{E}(X, \mathcal{G})} \in L^2(\Omega, \mathcal{G}, \mathbb{P})$. Then it suffices to show that $\mathbb{E}((X-Y)Z) = 0$ for all $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Any such Z has the form $\sum_{A \in \mathcal{A}} a_A 1_A$, so we reduce to the case $Z = 1_A$, $A \in \mathcal{A}$. Then $\mathbb{E}(X 1_A) = \mathbb{E}(X | 1_A)\mathbb{P}(A) = \mathbb{E}(Y 1_A)$ as required. $\square$

Note that if $f \in L^2$ and $v \in V$ with $f - v \in V^\perp$ then $d(f, V) = \|f - v\|$ so v is a version of the orthogonal projection.

For, if $g \in V$, then

$$\|f - g\|^2 = \|f - v + (v - g)\|^2 + \|f - v\|^2 + \cancel{2\langle f - v, v - g\rangle} + \|v - g\|^2 \geq 0$$

12/11/12 Probability and Measure 17

See new section 4.5 in Online Notes.

6. Convergence in $L^1(\mathbb{P})$

6.1 Bounded Convergence

Theorem 6.1

Let X, $(X_n : n \in \mathbb{N})$ be random variables, with $X_n \to X$ in probability. Suppose there exists a constant $C < \infty$ such that $|X_n| \leq C$ almost surely for all n. Then $X_n \to X$ in L^1 ($\mathbb{E}(|X_n - X|) \to 0$).

Proof

There is a subsequence (n_k) such that $X_{n_k} \to X$ almost surely. So $|X| \leq C$ almost surely. Note that for any $\varepsilon > 0$ $|X_n - X| \leq \frac{\varepsilon}{2} + 2C 1_{|X_n - X| > \varepsilon/2}$

There exists N such that for all $n \geq N$, $\mathbb{P}(|X_n - X| > \frac{\varepsilon}{2}) \leq \varepsilon/4C$

So for $n \geq N$, $\mathbb{E}(|X_n - X|) \leq \varepsilon/2 + 2C\,\varepsilon/4C = \varepsilon$. □

6.2 Uniform Integrability

Lemma 6.2.1

Let X be an integrable random variable. Define

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

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

Proof

Suppose not. Then there exists $\varepsilon > 0$ and $A_n \in \mathcal{F}$ with $\mathbb{P}(A_n) \leq 2^{-n}$ and $\mathbb{E}(|X| 1_{A_n}) \geq \varepsilon$. By the First Borel-Cantelli Lemma, $\mathbb{P}(A_n \text{ infinitely often}) = 0$ $\left(\{A_n \text{ infinitely often}\} = \bigcap_n \bigcup_{m \geq n} A_m\right.$

Then by dominated convergence (using $|X|$)

$$\varepsilon \le \mathbb{E}(|X| 1_{\bigcup_{m \ge n} A_m}) \to \mathbb{E}(|X| 1_{\{A_n \text{ infinitely often}\}}) = 0$$

Let $\mathcal{X}$ be a family of random-variables. We say that $\mathcal{X}$ is <u>bounded in L^p</u> if $\sup_{X \in \mathcal{X}} \|X\|_p < \infty$. We say that $\mathcal{X}$ is <u>uniformly integrable</u> (UI) if $\mathcal{X}$ is bounded in L^1 and $I_{\mathcal{X}}(\delta) \to 0$ as $\delta \to 0$. Here, ~~$I_{\mathcal{X}}(\delta) = \sup\{\mathbb{E}(|X|$~~

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

Thus for $X \in L^1$, $\{X\}$ is UI.

Suppose that $\mathcal{X}$ is bounded in L^p for some $p > 1$. By Hölder's Inequality

$$\mathbb{E}(|X| 1_A) \le \|X\|_p \|1_A\|_q = \|X\|_p \mathbb{P}(A)^{\frac{1}{q}} \quad (p, q \text{ conjugate indices})$$

Hence, ~~bounded in L~~ $\mathcal{X}$ is uniformly integrable.

<u>Example</u>

On $[0,1]$ consider $X_n = n 1_{[0, \frac{1}{n}]}$. Then the sequence $\mathcal{X}$ $\{X_n : n \in \mathbb{N}\}$ is bounded in L^1 but $I_{\mathcal{X}}(\delta) = 1$ for all $\delta > 0$, so $\mathcal{X}$ is <u>not</u> uniformly integrable.

Suppose that $|X| \le Y$ for all $X \in \mathcal{X}$, for some (integrable) $Y \in L^1$. Then $I_{\mathcal{X}}(\delta) \le I_Y(\delta) \downarrow 0$ as $\delta \downarrow 0$, so such a family is uniformly integrable.

<u>Lemma 6.2.2</u>

Let $\mathcal{X}$ be a family of random variables. Then $\mathcal{X}$ is UI if and only if $\sup_{X \in \mathcal{X}} \mathbb{E}(|X| 1_{|X| \ge k}) \to 0$ as $k \to \infty$. (*)

12/11/12 Probability and Measure (17)

Proof:

Suppose that $\mathcal{X}$ is UI. Given $\varepsilon > 0$, $\exists \delta > 0$ such that $\mathbb{E}(|X| 1_A) \leq \varepsilon$ whenever $A \in \mathcal{F}$, $\mathbb{P}(A) \leq \delta$. Then for all $k \geq \delta^{-1} I_{\mathcal{X}}(1)$ we have by Chebyshev's Inequality, for all $X \in \mathcal{X}$ $\mathbb{P}(|X| \geq k) \leq k^{-1}\mathbb{E}(|X|) \leq \delta$, so $\mathbb{E}(|X| 1_{|X| \geq k}) \leq \varepsilon$.

Hence the condition holds.

Suppose on the other hand that (*) holds. There exists k_0 such that for all $X \in \mathcal{X}$, $\mathbb{E}(|X| 1_{|X| \geq k_0}) \leq 1$. Then for all $X \in \mathcal{X}$ we have $\mathbb{E}(|X|) \leq \mathbb{E}(|X| 1_{|X| \leq k_0}) + \mathbb{E}(|X| 1_{|X| > k_0}) \leq k_0 + 1$.

So $I_{\mathcal{X}}(1) \leq k_0 + 1$. Given $\varepsilon > 0$, note that for $A \in \mathcal{F}$ with $\mathbb{P}(A) \leq \delta$, $|X| 1_A \leq k 1_A + |X| 1_{|X| \geq k}$

Choose k so that $\mathbb{E}(|X| 1_{|X| > k}) \leq \varepsilon/2$ for all $X \in \mathcal{X}$. Then for $\delta \leq \varepsilon/2$, we have $\mathbb{E}(|X| 1_A) \leq k\mathbb{P}(A) + \mathbb{E}(|X| 1_{|X| \geq k}$

$\leq \frac{\varepsilon}{2} + \varepsilon/2 = \varepsilon$.

Hence $\mathcal{X}$ is UI $\square$

Theorem 6·2·3

Let X, $(X_n : n \in \mathbb{N})$ be random variable. The following are equivalent

a) $X \in L^1$, $X_n \in L^1$ for all n and $X_n \to X$ in L^1.

b) $X_n \to X$ in probability and $\{X_n : n \in \mathbb{N}\}$ is UI.

Proof:

Suppose that a) holds. By Chebyshev's Inequality, for all $\varepsilon > 0$

$\mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon^{-1} \mathbb{E}(|X_n - X|) \to 0$ as $n \to \infty$.

Hence $X_n \to X$ in probability. Given $\varepsilon > 0$, there exists N such that for $n > N$, $\mathbb{E}(|X_n - X|) \leq \varepsilon/2$. Also, there exists $\delta > 0$ such that for $n = 1, \dots, N$ and any $A \in \mathcal{F}$ with $\mathbb{P}(A) \leq \delta$, (by Lemma 6.2.1)

$\mathbb{E}(|X_n| 1_A) \leq \varepsilon$, and $\mathbb{E}(|X| 1_A) \leq \varepsilon/2$

But for $n > N$ we have for such A, $|X_n| \leq |X_n - X| + |X|$ so

$\mathbb{E}(|X_n| 1_A) \leq \mathbb{E}(|X_n - X|) + \mathbb{E}(|X| 1_A) \leq \varepsilon$.

Hence $\{X_n : n \in \mathbb{N}\}$ is UI, so b) holds.

Suppose on the other hand that b) holds. Note that for all $\varepsilon > 0$, $k <$

$$|X_n - X| \leq \varepsilon/4 + 2k 1_{|X_n - X| > \varepsilon/4} + |X_n| 1_{|X_n| > k} + |X| 1_{|X| > k}$$

This is because either $|X_n - X| \leq \varepsilon/4$, or $(|X_n - X| > \varepsilon/4, |X_n|, |X| \leq k)$ or $|X_n| > k$ or $|X| > k$.

There exists k such that $\mathbb{E}(|X_n| 1_{|X_n| \leq k}) \leq \varepsilon/4$ and $\mathbb{E}(|X| 1_{|X| \geq k}) \leq \varepsilon/4$. Then there exists N such that for all $n \geq N$, $\mathbb{P}(|X_n - X| > \varepsilon/4) \leq \varepsilon/8k$. Then for $n \geq N$ we have $\mathbb{E}(|X_n - X|) \leq \varepsilon/4 + 2k \frac{\varepsilon}{8k} + \varepsilon/4 + \varepsilon/4 = \varepsilon$

(We should have noted that there is a subsequence (n_k) such that $X_{n_k} \to X$ almost surely, so that

$\mathbb{E}(|X|) = \mathbb{E}(\liminf_k |X_{n_k}|) \leq \liminf \mathbb{E}(|X_{n_k}|) < \infty$ by UI)

$|X_n - X| \le \varepsilon + 2k \, 1_{|X_n - X| \ge \varepsilon} + |X| \, 1_{|X| \ge k} + |X_n| \, 1_{|X_n| \ge k}$

$x^k = (-k) \vee x \wedge k$

$|x^k - y^k| \le |x-y| \wedge (2k)$

$|X_n - X| \le |X_n - X_n^k| + |X_n^k - X^k| + |X^k - X|$

7 Fourier Transforms

7.1 Definitions

In this section, we write $L^p = L^p(\mathbb{R}^d)$ for the set of measurable functions $f : \mathbb{R}^d \to \mathbb{C}$ such that

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

For $f \in L^1$, $f = u + iv$ say, for λ = Lebesgue Measure on $\mathbb{R}^d$

$\lambda(f) := \lambda(u) + i\lambda(v)$

Note that $|\lambda(f)| = e^{i\alpha} \lambda(f)$, some α. $\tilde{f} = e^{i\alpha} f = \tilde{u} + i\tilde{v}$

$= \lambda(\tilde{f}) = \lambda(\tilde{u}) \le \lambda(|f|)$ $\qquad |\tilde{u}| \le |f|$

For $f \in L^1$ we define the <u>Fourier transform</u>

$\hat{f} : \mathbb{R}^d \to \mathbb{C}$ by $\hat{f}(u) = \int_{\mathbb{R}^d} e^{i\langle u, x\rangle} f(x) \, dx$, $u \in \mathbb{R}^d$

where $\langle u, x \rangle = \sum_{k=1}^{d} u_k x_k$, the usual scalar product.

Note that $|\hat{f}(u)| \le \|f\|_1$. Also, for $u_n \to u$, by the dominated convergence theorem, $\hat{f}(u_n) \to \hat{f}(u)$. So $\hat{f}$ is a continuous, bounded complex function on $\mathbb{R}^d$.

Let $f \in L^1$. If $\hat{f} \in L^1$, we say that the <u>Fourier inversion formula</u> holds for f if $f(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} \hat{f}(u) \, du$ for almost all x. We will show that this holds whenever $\hat{f} \in L^1$.

If $f \in L^2$, we say the Plancherelle identity holds for f if $\|\hat{f}\|_2 = (2\pi)^{d/2} \|f\|_2$. We will show that this holds whenever $f \in L^2$.

Let μ be a finite Borel measure on $\mathbb{R}^d$. Define $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ by $\hat{\mu}(u) = \int_{\mathbb{R}} e^{i\langle u, x\rangle} \mu(dx)$, $\mu \in \mathbb{R}^d$. If μ has density function f, then $\hat{\mu} = \hat{f}$.

Exercise: $|\hat{\mu}(u)| \leq \mu(\mathbb{R}^d)$, $\hat{\mu}$ is continuous on $\mathbb{R}^d$.

For X a random variable in $\mathbb{R}^d$, we define the <u>characteristic function</u> $\phi_X : \mathbb{R}^d \to \mathbb{C}$ by

$$\phi_X(u) = \hat{\mu}_X(u) = \mathbb{E}(e^{i\langle u, X\rangle})$$

<u>7.2 Convolutions</u>

Let $f \in L^p$ and let ν be a Borel probability measure on $\mathbb{R}^d$. We define the convolution $f * \nu$ in L^p by

$$f * \nu(x) = \begin{cases} \int_{\mathbb{R}^d} f(x-y)\,\nu(dy) & \text{if } f(x - \cdot) \in L^1(\nu) \\ 0 & \text{otherwise} \end{cases}$$

We will use the fact that $(x, y) \mapsto f(x-y)$, $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ is measurable.

By Jensen's Inequality and Fubini's Theorem,

$$\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-y)|\,\nu(dy)\right)^p dx \leq \underbrace{\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(x-y)|\,\nu(dy)\,dx}_{\text{Fubini: Can swap order}}$$

$$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(x)|^p dx\,\nu(dy) = \|f\|_p^p < \infty$$

So $f(x - \cdot) \in L^1(\nu)$ for almost all x and

$\|f * \nu\|_p = \left(\int_{\mathbb{R}^d} \left|\int_{\mathbb{R}^d} f(x-y)\,\nu(dy)\right|^p dx\right)^{1/p} \leq \|f\|_p$

If ν has a density g then we write $f * g$ for $f * \nu$

For probability measures μ, ν on $\mathbb{R}^d$, we define $\mu * \nu$ as the distribution of $X + Y$ for independent random variables having laws μ, ν. Thus $\mu * \nu(A) = \mathbb{P}(X+Y \in A) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1_A(x+y)\,\mu(dx)\,\nu(dy)$

Thus, if μ has a density f $(\in L^1(\mathbb{R}^d))$ then

$\mu * \nu(A) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1(x+y) f(x)\,dx\,\nu(dy)$

$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1_A(x) f(x-y)\,dx\,\nu(dy)$ Fubini: Swap order of integration, then bring $1_A(x)$ outside the middle integral

$= \int_{\mathbb{R}^d} 1_A(x)(f * \nu)(x)\,dx$

so $\mu * \nu$ has density $f * \nu$

7.3 Gaussians

Consider for $t \in (0, \infty)$ the centred Gaussian probability density function of variance t on $\mathbb{R}^d$ given by

$g_t(x) = (2\pi t)^{-\frac{d}{2}} e^{-|x|^2/2t}, \; x \in \mathbb{R}^d \quad (N(0, tI))$

We will compute $\hat{g}_t$. First consider the case $d=1$, $t=1$.

Take $Z \sim N(0,1)$ i.e. Z has density g_1 (in $\mathbb{R}$).

Then the characteristic function ϕ_Z is differentiable (because Z is integrable and we can differentiate under $\mathbb{E}$)

with $\phi_Z'(u) = \mathbb{E}(iZe^{iuZ}) = \int_{\mathbb{R}} \underbrace{ie^{iux}}_{v} \underbrace{\frac{1}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}}}_{u'} dx$ integrate by parts

$= -\int u e^{iux} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = -u\,\phi_Z(u)$

So $\frac{d}{du}\left(e^{\frac{u^2}{2}} \phi_Z(u)\right) = 0$, so $\phi_Z(u) = e^{-u^2/2}\phi_Z(0) = e^{-\frac{u^2}{2}}$

$\|f * \nu\|_p = \left(\int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(x-y)\,\nu(dy) \right|^p dx \right)^{1/p} \le \|f\|_p$

[If ν has a density g, then we write $f * g$ for $f * \nu$.]

For probability measures μ, ν on $\mathbb{R}^d$ we define $\mu * \nu$ as the distribution of $X + Y$ for independent random variables having laws μ, ν. Then $\mu * \nu(A) = \mathbb{P}(X+Y \in A) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1_A(x+y)\,\mu(dx)\,\nu(dy)$

Then, if μ has a density f ($f \in L^1(\mathbb{R}^d)$) 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, translation of integration, then bring $1_A(x)$ outside the inside integral]

$= \int_{\mathbb{R}^d} 1_A(x)\,(f * \nu)(x)\,dx$

so $\mu * \nu$ has density $f * \nu$.

7.3 Gaussians

Consider for $t \in (0, \infty)$ the centred Gaussian probability density function of variance t on $\mathbb{R}^d$ given by

$g_t(x) = (2\pi t)^{-d/2} e^{-|x|^2/2t}, \quad x \in \mathbb{R}^d \quad (N(0, tI))$.

We will compute $\hat{g}_t$. First consider the case $d = 1$, $t = 1$.

Take $Z \sim N(0,1)$ i.e. Z has density g_1 (on $\mathbb{R}$).

Then the characteristic function ϕ_Z is differentiable.

[Because Z is integrable and we can differentiate under the integral]

with $\phi_Z'(u) = \mathbb{E}(iZe^{iuZ}) = \int_{\mathbb{R}} \frac{ix\,e^{iux}}{\sqrt{2\pi}} e^{-x^2/2}\,dx$ integrating by parts

$= -\int \frac{u}{\sqrt{2\pi}} e^{iux} e^{-x^2/2}\,dx = -u\,\phi_Z(u)$

so $\frac{d}{du}\left(e^{u^2/2}\phi_Z(u)\right) = 0$ so $\phi_Z(u) = e^{-u^2/2}\phi_Z(0) = e^{-u^2/2}$

4.5 Approximation in L^p

Theorem 4.5.1

Let $\mathcal{A}$ be a π-system on E generating $\mathcal{E}$, with $\mu(A) < \infty$ for all $A \in \mathcal{A}$, and such that $E_n \uparrow E$ for some sequence $(E_n : n \in \mathbb{N})$ in $\mathcal{A}$. Define

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

Let $p \in [1, \infty)$. Then $V_0 \subseteq L^p$. Moreover, for all $f \in L^p$, for all $\varepsilon > 0$, there exists $v \in V_0$ such that $\|v - f\|_p \leq \varepsilon$ ("V_0 is dense in L^p")

Proof

For $A \in \mathcal{A}$, $\|1_A\|_p = \mu(A)^{1/p} < \infty$, so $1_A \in L^p$. So $V_0 \subseteq L^p$ because L^p is a vector-space.

Write V for the set of all $f \in L^p$ such that the conclusion holds

Note that V is a vector space, by Minkowski's inequality.

Assume for now that $E \in \mathcal{A}$. Define $\mathcal{D} = \{A \in \mathcal{E} : 1_A \in V\}$

Then $\mathcal{A} \subseteq \mathcal{D}$, $1 \in \mathcal{D}$; if $A, B \in \mathcal{D}$ with $A \subseteq B$, then $1_{B \setminus A} = 1_B - 1_A \in V$, so $B \setminus A \in \mathcal{D}$; for $A_n \in \mathcal{D}$, $A_n \uparrow A$ $\|1_A - 1_{A_n}\|_p = \mu(A \setminus A_n)^{1/p} \to 0$, so $1_A \in V$, so $A \in \mathcal{D}$.

Hence $\mathcal{D}$ is a d-system, so $d = \mathcal{E}$ by Dynkin's Lemma.

Since V is a vector space it must contain every simple function.

For $f \in L^p$ with $f \geq 0$, consider the simple functions

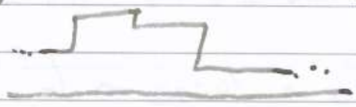
$f_n = (2^{-n}\lfloor 2^n f \rfloor) \wedge n$, then $f_n \uparrow f$, so

$|f|^p \geq |f - f_n|^p \to 0$, pointwise, so $\|f - f_n\|_p \to 0$

by dominated convergence, so $f \in V$.

Then $V \in L^p$ because V is a vector space.

Returning to the general case (where E may not be in A), the above argument shows that $f 1_{E_n}$ for any $f \in L^p$. Then $|f|^p \geq |f - f 1_{E_n}|^p \to 0$ pointwise. Hence $\|f - f 1_{E_n}\|_p \to 0$ by dominated convergence. So $f \in V$ □



Exercise

Show that step-functions are dense in L^p for all $p \in [1, \infty)$. Show also that continuous functions of compact support are dense in L^p, $p < \infty$

7.3 Gaussians (continued)

We showed that $\int_{\mathbb{R}} e^{iux} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = e^{-u^2/2}$, $u \in \mathbb{R}$

By a change of variable, $x = y/\sqrt{t}$, we obtain

$$\int_{\mathbb{R}} e^{iuy/\sqrt{t}} \frac{1}{\sqrt{2\pi t}} e^{-y^2/2t}\, dy = e^{-u^2 t/2}$$

Then by Fubini, for $g_t(x) = (2\pi t)^{-\frac{d}{2}} e^{-|x|^2/2t}$

$$\hat{g}_t(u) = \int_{\mathbb{R}^d} e^{i\langle u, x\rangle} g_t(x)\, dx = \prod_{k=1}^{d} \int_{\mathbb{R}} e^{iu_k x_k} \frac{1}{\sqrt{2\pi t}} e^{-x_k^2/2t}\, dx_k$$

$$= \prod_{k=1}^{d} e^{-u_k^2 t/2} = e^{-|u|^2 t/2}$$

So $\hat{g}_t = (2\pi)^{\frac{d}{2}} t^{-\frac{d}{2}} g_{1/t}$, so $\hat{\hat{g}}_t = (2\pi)^d g_t$.

So $g_t(x) = g_t(-x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle u, x\rangle} \hat{g}_t(u)\, dt$

Hence the Fourier Inversion Formula holds for g_t.

7·4 Gaussian Convolutions

By a <u>Gaussian Convolution</u>, we mean any function of the form $f * g_t$ where $f \in L^1$, $t \in (0, \infty)$

Note that $\widehat{f * g_t}(u) = \hat{f}(u) e^{-|u|^2 t/2}$

Recall $f * g_t(x) = \int_{\mathbb{R}^d} f(x-y) g(y)\, dy = \int_{\mathbb{R}^d} g_t(x-y) f(y)\, dy$

By dominated convergence, $f * g_t$ is continuous. Also

$\|f * g_t\|_1 \leq \|f\|_1$, $\|f * g_t\|_\infty \leq (2\pi t)^{-d/2} \|f\|$

$\|\widehat{f * g_t}\|_1 \leq \|\hat{f}\|_\infty \|\hat{g}_t\|_1 \leq \|f\|_1 (2\pi)^{\frac{d}{2}} t^{-\frac{d}{2}}$

$\|\widehat{f * g_t}\|_\infty \leq \|\hat{f}\|_\infty \|\hat{g}_t\|_\infty \leq \|f\|_1$

Lemma 7·4·1

The Fourier Inversion Formula holds for $f * g_t$ for all $f \in L^1$, $t \in (0, \infty)$.

Proof

We compute, using Fubini

$(2\pi)^d f * g_t(x) = (2\pi)^d \int_{\mathbb{R}^d} f(x-y) g_t(y)\, dy$

$= \int_{\mathbb{R}^d} f(x-y) \int_{\mathbb{R}^d} e^{-i\langle u, y\rangle} \hat{g}_t(u)\, du\, dy$

$= \int_{\mathbb{R}^d} \hat{g}_t(u) e^{-i\langle x, u\rangle} \int_{\mathbb{R}^d} f(x-y) e^{i\langle u, x-y\rangle}\, dy\, du$

(change of variable $x-y \mapsto y$) $= \int_{\mathbb{R}^d} \hat{g}_t(u) e^{-i\langle x, u\rangle} \int_{\mathbb{R}^d} f(y) e^{i\langle u, y\rangle}\, dy\, du$

$= \int_{\mathbb{R}^d} \hat{g}_t(u) e^{-i\langle x, u\rangle} \hat{f}(u)\, du = \int_{\mathbb{R}} \widehat{f * g_t}(u) e^{-i\langle x, u\rangle}\, du$

Lemma 7.4.2

Let $p \in (1, \infty)$ and let $f \in L^p$. Then $f * g_t \to f$ in L^p.

Proof

Given $\varepsilon > 0$, we can find a continuous function h of compact support such that $\|f - h\|_p \leq \varepsilon/3$. Then also

$\|f * g_t - h * g_t\|_p = \|(f-h) * g_t\|_p \leq \|f - h\|_p \leq \varepsilon/3$. Consider

$e(y) = \int_{\mathbb{R}^d} |h(x-y) - h(x)|^p dx$.

Then $e(y) \leq 2^p \|h\|_p^p$ and since $h(x-y) - h(x) \to 0$

as $y \to 0$, uniformly in x, we have $e(y) \to 0$ as $y \to 0$.

Now $\|h - h * g_t\|_p^p = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (h(x) - h(x-y)) g_t(y) dy \right|^p dx$

$\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |h(x) - h(x-y)|^p g_t(y) dy dx$

Jensen

$g_t(y) dy$ is a probability measure

$= \int_{\mathbb{R}^d} g_t(y) e(y) dy = \int_{\mathbb{R}^d} g_1(y) \underbrace{e(\sqrt{t} y)}_{\to 0} dy \to 0$ by bounded convergence

Hence $\|f - f * g_t\|_p \leq \|f - h\|_p + \|h - h * g_t\|_p + \|h * g_t - f * g_t\|$

Fubini

$\leq \varepsilon$ for sufficiently large t $\square$

7.5 Uniqueness and Inversion

Theorem 7.5.1

Let $f \in L^1$. Define for $t \in (0, \infty)$

$f_t(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} \hat{f}(u) e^{-|u|^2 t/2} du$

Then $f_t \to f$ in L^1. Moreover, if $\hat{f} \in L^1$, the Fourier Inversion Formula holds.

Since $\widehat{f * g_t}(u) = \hat{f}(u)\, e^{-|u|^2 t/2}$, by Lemma 7.4.1, we have

$f_t = f * g_t$. Hence $f_t \to f$ in L^1 by Lemma 7.4.2.

Then if $\hat{f} \in L^1$, we can apply dominated convergence with

$|\hat{f}|$ as dominating function to see that

$f_t(x) \to \int_{\mathbb{R}^d} e^{-i\langle u, x\rangle} \hat{f}(u)\, du$ as $t \to 0$.

But there is a sequence $t_n \to 0$ such that $f_{t_n} \to f$ a.e.

Hence the inversion formula holds for f $\square$

19/11/12

Probability and Measure (20)

7.6 Fourier Transform in $L^2(\mathbb{R}^d)$

Theorem 7.6.1

Let $f \in L^1 \cap L^2(\mathbb{R}^d)$. Then the Plancherel identity holds for f:

$\|\hat{f}\|_2 = (2\pi)^{d/2} \|f\|_2$

Moreover, there is a unique Hilbert space automorphism F on $\mathcal{L}^2$ such that for all $f \in L^1 \cap L^2(\mathbb{R}^d)$, $F[f] = [(2\pi)^{-\frac{d}{2}} \hat{f}]$

(Thus $F : \mathcal{L}^2 \to \mathcal{L}^2$ is a linear bijection with $\|Fh\| = \|h\| \ \forall h \in \mathcal{L}^2$)

Proof

Consider first the case where $\hat{f} \in L^1$. Then

$$(2\pi)^d \|f\|_2^2 = \int_{\mathbb{R}^d} (2\pi)^d f(x) \overline{f(x)}\, dx$$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{-i\langle x,u\rangle} \hat{f}(u)\, du\, \bar{f}(x)\, dx, \text{ by the inversion formula}$$

$$= \int_{\mathbb{R}^d} \hat{f}(u) \underline{\int_{\mathbb{R}^d} e^{-i\langle x,u\rangle} \bar{f}(x)\, dx}, \text{ by Fubini, since } (x,u) \mapsto |f(x)|\,|\hat{f}(u)| \text{ is integrable}$$

$$= \int_{\mathbb{R}^d} \hat{f}(u) \overline{\int_{\mathbb{R}^d} e^{i\langle x,u\rangle} f(x)\, dx}\, du$$

$$= \int_{\mathbb{R}^d} \hat{f}(u) \overline{\hat{f}(u)}\, du = \|\hat{f}\|_2^2$$

For general $f \in L^1 \cap L^2$, consider $f_t = f * g_t$ for $t \in (0,\infty)$

Then by Lemma 7.4.2, $f_t \to f$ in L^2, so $\|f_t\|_2 \to \|f\|_2$ as $t \to 0$.

Now $\underline{\hat{f}_t(u) = \hat{f}(u) e^{-|u|^2 t/2}}$, so $|\hat{f}_t|^2 \nearrow |\hat{f}|$ as $t \to 0$

So by monotone convergence, $\|\hat{f}_t\|_2^2 \to \|\hat{f}\|_2^2$ as $t \to 0$. We know $\hat{f}_t \in L^1$ and $\hat{f}_t(u) \leq \|f\|_1 e^{-|u|^2 t/2} \in L^1$. So $\|\hat{f}_t\|_2 = (2\pi)^{\frac{d}{2}} \|f_t\|$

and the desired identity for f follows on letting $t \to 0$.

Define F_0 on $\mathcal{L}^1 \cap \mathcal{L}^2$ by $F_0[f] = [(2\pi)^{-\frac{d}{2}} \hat{f}]$ for $f \in L^1 \cap L^2$.

Then F_0 is a well-defined linear map $\mathcal{L}^1 \cap \mathcal{L}^2$ with $\|F_0 h\|_2 = \|h\|_2$ for all $h \in \mathcal{L}^1 \cap \mathcal{L}^2$. Now $\mathcal{L}^1 \cap \mathcal{L}^2$ is dense in $\mathcal{L}^2$ (it contains all step functions), so F_0 extends uniquely to a linear isometry from $\mathcal{L}^2$ into $\mathcal{L}^2$.

It remains to show that F is surjective, $\mathcal{L}^2 \to \mathcal{L}^2$. Then $V \subseteq \mathcal{L}^1 \cap \mathcal{L}$

Consider $V = \{[f] : f \in L^1 \text{ with } \hat{f} \in L^1\}$. Then V is dense in $\mathcal{L}^2$ (it contains all Gaussian convolutions) and by the inversion formula $F^4 h = h$. So F is 'onto' V and hence F is onto $\mathcal{L}^2$ as required. $\square$

7.7 Weak Convergence and Characteristic Functions

Let μ, $(\mu_n : n \in \mathbb{N})$ be Borel probability measures on $\mathbb{R}^d$. We say that $\mu_n \to \mu$ weakly if $\mu_n(f) \to \mu(f)$ for all $f \in C_b(\mathbb{R}^d)$ (continuous bounded functions on $\mathbb{R}^d$). In fact, we could restrict the class of test functions to $C_c(\mathbb{R}^d)$ (compact support) or even to $C_c^1(\mathbb{R}^d)$ without affecting whether $\mu_n(f) \to \mu(f)$ for all f.

For random variables X, $(X_n : n \in \mathbb{N})$ in $\mathbb{R}^d$, say $X^n \to X$ weakly if $\mu_{X_n} \to \mu_X$ weakly (on $\mathbb{R}^d$). In the case $d = 1$ this is equivalent to convergence in distribution.

Theorem 7.7.1

Let X be a random variable in $\mathbb{R}^d$. Then the law μ_X of X is uniquely determined by its characteristic function ϕ_X $(= \hat{\mu}_X)$

Moreover, if ϕ_X is integrable, then X has a continuous bounded density (with respect to dx) given by

$f_X(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle x,u\rangle} \phi_X(u)\,du \quad , x \in \mathbb{R}^d.$

Moreover, if $(X_n : n \in \mathbb{N})$ is any sequence of random variables in $\mathbb{R}^d$ such that $\phi_{X_n}(u) \to \phi_X(u)$ for all $u \in \mathbb{R}^d$ as $n \to \infty$, then $X_n \to X$ weakly.

<u>Proof</u>

Let Z be a random variable in $\mathbb{R}^d$ independent of X with density g_1 (that is, $Z \sim N(0, I)$). Then $\sqrt{t}Z$ has density g_t and $X + \sqrt{t}Z$ has density function $f_t = \mu_X * g_t$. Now $f_t \in L^1$ and $\hat{f}_t = \phi_X \hat{g}_t$, $\hat{g}_t(u) = e^{-|u|^2 t/2}$, $\|\phi_X\|_\infty \leq 1$, so $\hat{f}_t \in L^1$.

So by inversion $f_t(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle u,x\rangle} e^{-t|u|^2/2} \phi_X(u)\,du$ (*)

So ϕ_X determines the distribution of $X + \sqrt{t}Z$ uniquely. By bounded convergence, for all $g \in C_b(\mathbb{R}^d)$

$\mu_X(g) = \mathbb{E}(g(X)) = \lim_{t \to 0} \mathbb{E}(g(X + \sqrt{t}Z))$

So ϕ_X determines $\mu_X(g)$ uniquely and hence μ_X.

Suppose now that $\phi_X \in L^1$. We use dominated convergence with dominating function $|\phi_X|$ to see that

$f_t(x) \to f_X(x)$ for all $x \in \mathbb{R}^d$, and $|f_t(x)| \leq (2\pi)^{-d} \|\phi_X\|_1$

Then for any $g \in C_c(\mathbb{R}^d)$, $\mathbb{E}(g(X)) = \lim_{t \to 0} \mathbb{E}(g(X + \sqrt{t}Z))$

$= \int g(y) f_t(y)\,dy \to \int g(y) f_X(x)\,dx$ by bounded convergence.

Note that $f_X(x) = \lim_{t \to 0} f_t(x) \geq 0 \ \forall x \in \mathbb{R}^d$. It follows that $\mu_X(dx) = f_X(x)\,dx$ (exercise)

Suppose now that $g \in C_c^1(\mathbb{R}^d)$ and $\phi_{X_n}(u) \to \phi_X(u) \ \forall u \in \mathbb{R}^d$

$|g(X) - g(X + \sqrt{t}Z)| \leq \|\nabla g\|_\infty \sqrt{t}\,|Z|$ (by the Mean Value Theorem)

so $|\mathbb{E}(g(X)) - \mathbb{E}(g(X + \sqrt{t}Z))| \leq \|\nabla g\|_\infty \sqrt{t}\,\mathbb{E}(|Z|) < \infty$

Similarly $|\mathbb{E}(g(X_n)) - \mathbb{E}(g(X_n + \sqrt{t}Z))| \leq \|\nabla g\|_\infty \sqrt{t}\,\mathbb{E}(|Z|)$

Now $(2\pi)^d \times \mathbb{E}(g(X_n + \sqrt{t}Z)) = \int_{\mathbb{R}^d} g(x) \int_{\mathbb{R}^d} e^{-i\langle x, u\rangle} e^{-|u|^2 t/2} \phi_{X_n}(u)\,du\,dx$

$\longrightarrow \int_{\mathbb{R}^d} g(x) \int_{\mathbb{R}^d} e^{-i\langle x, u\rangle} e^{-|u|^2 t/2} \phi_X(u)\,du\,dx = (2\pi)^d \mathbb{E}(g(X + \sqrt{t}Z))$

Here we used Fubini, and dominated convergence with dominating function $(x, u) \mapsto |g(x)|\,e^{-|u|^2 t/2}$ on $\mathbb{R}^d \times \mathbb{R}^d$.

Given $\varepsilon > 0$, choose t so that $\|\nabla g\|_\infty \sqrt{t}\,\mathbb{E}(|Z|) \leq \varepsilon/3$.

Then for sufficiently large n, we have

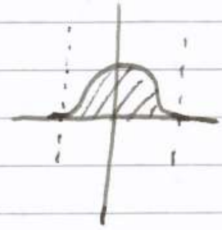
$|\mathbb{E}(g(X)) - \mathbb{E}(g(X_n))| \leq \varepsilon$ $\square$

21/11/12 Probability and Measure (21)

Lemma 7.7.2

Let μ be a Borel probability measure on $\mathbb{R}^d$ and let $(\mu_n : n \in \mathbb{N})$ be a sequence of such measures. Suppose $\mu_n(f) \to \mu(f)$ for all $C_c^\infty(\mathbb{R}^d)$. Then $\mu_n \to \mu$ weakly on $\mathbb{R}^d$ (i.e. $\mu_n(f) \to \mu(f)$ for all $f \in \underline{C_b(\mathbb{R}^d)}$ $\leftarrow$ note the different class of functions

Proof



Suppose that $f \in C_c(\mathbb{R}^d)$. Consider

$$\rho(x) = \begin{cases} c \exp\{-(1-|x|^2)^{-1}\} & \text{if } |x| < 1 \\ 0 & \text{if } |x| \geq 1 \end{cases}$$

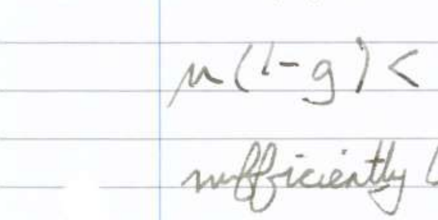
where c is chosen so that $\int_{\mathbb{R}^d} \rho(x)\,dx = 1$. Note that $\rho \in C_c^\infty(\mathbb{R}^d)$

Set $\rho_\tau(x) = \tau^{-d}\rho(x/\tau)$ for $\tau \in (0,\infty)$ ("$\tau = \varepsilon$")

and set $f_\tau = f * \rho_\tau$. Then $f_\tau \in C_c^\infty(\mathbb{R}^d)$ and $\|f_\tau - f\|_\infty \to 0$ as $\tau \to 0$.

Now $\mu_n(f_\tau) \to \mu(f_\tau)$ as $n \to \infty$ and $|\mu_n(f_\tau - f)| \leq \underset{\to 0}{\|f_\tau - f\|_\infty}$ as $\tau \to 0$. So $\mu_n(f) \to \mu(f)$ as $n \to \infty$.

Now suppose $f \in C_b(\mathbb{R}^d)$. Consider $g(x) = 0 \vee (R+1-|x|) \wedge 1$.



Then $g \in C_c(\mathbb{R}^d)$ so $\mu_n(g) \to \mu(g)$ as $n \to \infty$.

Since $\mu_n(1) = 1$ for all n, $\mu_n(1-g) \to \mu(1-g)$ as $n \to \infty$. Given $\varepsilon > 0$, there exists $R < \infty$ such that $\mu(1-g) < \frac{\varepsilon}{3}\|f\|_\infty$, so $\mu_n(1-g) \leq \frac{\varepsilon}{3}\|f\|_\infty$, for all sufficiently large n. Hence

$$|\mu_n(f) - \mu(f)| = |\mu_n(f(1-g))| + |\mu_n(fg) - \mu(fg)| + |\mu(f(1-g))|$$

$\leq \|f\|_\infty \mu_n(1-g) + 1 \cdot 1 + \|f\|_\infty \mu(1-g)$

$\leq \varepsilon/3 + \varepsilon/3 + \varepsilon/3$ for all sufficiently large n

(because $fg \in C_c(\mathbb{R}^d)$) $\square$

8: Gaussian Random Variables

8.1 Scalar Case

Let X be a random variable in $\mathbb{R}$. We say that X is Gaussian if for some $\mu \in \mathbb{R}$, either $X = \mu$ almost surely or for some $\sigma^2 \in (0, \infty)$, X has density $f_X(x) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Write $X \sim N(\mu, \sigma^2)$ ($\sigma^2 = 0$ when $X = \mu$ almost surely)

Proposition 8.1.1

Let $X \sim N(\mu, \sigma^2)$. Then $X \in L^2(\mathbb{P})$ and a) $\mathbb{E}(X) = \mu$,

b) $\mathrm{var}(X) = \sigma^2$, c) $\phi_X(u) = \mathbb{E}(e^{iuX}) = e^{iu\mu - \frac{1}{2}\sigma^2 u^2}$

d) For $a, b \in \mathbb{R}$, $aX + b \sim N(a\mu + b, a^2\sigma^2)$ (Exercise)

8.2 Gaussian random variables in $\mathbb{R}^d$

Let X be a random variable in $\mathbb{R}^d$. We say that X is Gaussian if for all $u \in \mathbb{R}^d$, $\langle u, X \rangle$ is Gaussian (in $\mathbb{R}$).

Example

Take $Y_1, \dots, Y_d$ independent $N(0,1)$. Set $Y = (Y_1, \dots, Y_d)$

Then for $u \in \mathbb{R}^d$, $\mathbb{E}(e^{i\langle u, Y\rangle}) = \prod_{k=1}^{d} \mathbb{E}(e^{iu_k y_k})$

$= \prod_{k=1}^{d} e^{-u_k^2/2} = e^{-|u|^2/2}$

So $\langle u, Y \rangle \sim N(0, |u|^2)$ by uniqueness of characteristic function

So Y is Gaussian.

21/11/12

Probability and Measure (21)

<u>Theorem 8.2.1</u>

Let X be a Gaussian random-variable in $\mathbb{R}^d$. Then $X \in L^2(\mathbb{P})$

Set $\mathbb{E}(X) = \mu \in \mathbb{R}^d$. $V = \mathrm{var}(X) = (\mathrm{cov}(X_j, X_k))_{jk} \in$ $d \times d$ matrix

Then a) $\phi_X(u) = e^{i\langle u,\mu\rangle - \frac{1}{2}\langle u, Vu\rangle}$, $u \in \mathbb{R}^d$.

b) The distribution of X is determined by μ and V (we write $X \sim N(\mu, V)$)

c) If A is an $m \times d$ matrix and $b \in \mathbb{R}^m$, then $AX + b$ is Gaussian in $\mathbb{R}^m$.

d) In the case that V is invertible, X has a density

$$f_X(x) = (2\pi)^{-\frac{d}{2}} (\det V)^{-\frac{1}{2}} e^{-\langle x-\mu, V^{-1}(x-\mu)\rangle/2}, \quad x \in \mathbb{R}^d$$

e) If $X = (X_1, X_2)$ with X_1 in $\mathbb{R}^{d_1}$, X_2 in $\mathbb{R}^{d_2}$, and if $\mathrm{cov}(X_1, X_2) = 0$, then X_1, X_2 are independent.

<u>Proof</u>

Each component $X_k \in L^2(\mathbb{P})$, so X is too. Note that

$$\mathbb{E}(\langle u, X\rangle) = \langle u, \mu\rangle, \quad \mathrm{var}(\langle u, X\rangle) = \langle u, Vu\rangle$$

Now $\langle u, X\rangle$ is Gaussian so $\mathbb{E}(e^{i\langle u,X\rangle}) = e^{i\langle u,\mu\rangle - \frac{1}{2}\langle u,Vu\rangle}$

Hence a) and b) hold by uniqueness of characteristic functions.

c) is left as an exercise.

Consider $Y = (Y_1, \dots, Y_d)$ as above, and since V is non-negative definite, there exists a symmetric $d \times d$ matrix Σ such that

$\Sigma^2 = V$ $\quad (V = U \operatorname{diag}(\lambda_k) U^T, \ \Sigma = U \operatorname{diag}(\sqrt{\lambda_k}) U^T)$

Consider $\hat{X} = \Sigma Y + \mu$. Then $\hat{X}$ is Gaussian and

$\mathbb{E}(\hat{X}) = \mu$, $\mathrm{var}(\hat{X}) = V$, so $\hat{X}$ has the same distribution as X.

3

Then, by a linear change of variables in $\mathbb{R}^d$, $\hat{X}$ has density

$$f_X(x) = f_{\hat{X}}(x) = (2\pi)^{-\frac{d}{2}} (\det V)^{-\frac{1}{2}} e^{-\langle x-\mu, V^{-1}(x-\mu)\rangle/2}$$

Finally, if $\mathrm{cov}(X_1, X_2) = 0$, then $V = \begin{pmatrix} V_{11} & 0 \\ 0 & V_{22} \end{pmatrix}$ so

$$\phi_X(u) = e^{i\langle u,\mu\rangle - \frac{1}{2}\langle u, Vu\rangle} = \phi_{X_1}\phi_{X_2}, \quad u = (u_1, u_2) \in \mathbb{R}^d \quad \square$$

10.3 Central Limit Theorem

Theorem

Let $(X_n : n \in \mathbb{N})$ be a sequence of square integrable, independent identically distributed random-variable (in $\mathbb{R}$). Set $\mu = \mathbb{E}(X)$, $\sigma^2 = \mathrm{Var}(X)$. Assume that $\sigma^2 > 0$. Set $S_n = X_1 + \dots + X_n$. $Z_n = (S_n - n\mu)/\sigma\sqrt{n}$. Then $Z_n \to Z$ weakly in $\mathbb{R}$ where $Z \sim N(0,1)$ (that is, $\mathbb{P}(Z_n \leq a) \to \int_{-\infty}^{a} (2\pi)^{-\frac{1}{2}} e^{-x^2/2}\,dx$ as $n \to \infty$ $\forall a \in \mathbb{R}$)

Proof

Set $\phi(u) = \mathbb{E}(e^{iuX_1})$ and $\phi_n(u) = \mathbb{E}(e^{iuZ_n})$. Since $X_1 \in L^2$, we can differentiate ϕ twice under $\mathbb{E}$ and obtain $\phi(0) = 1$, $\phi'(0)$, $\phi''(0) = 1$ in the case $\mu = 0$, $\sigma^2 = 1$. By Taylor's Theorem, $\phi(u) = 1 - \frac{u^2}{2} + o(u^2)$ as $u \to 0$. Then

$$\phi_n(u) = \left(\phi\left(\tfrac{u}{\sqrt{n}}\right)\right)^n = \left(1 - \tfrac{u^2}{2n} + o\left(\tfrac{u^2}{n}\right)\right)^n \text{ as } \tfrac{u^2}{n} \to 0$$

The complex logarithm satisfies $\log(1+z) = z + o(|z|)$ as $|z| \to 0$.

so $\log \phi_n(u) = n \log\left(1 - \frac{u^2}{2n} + o\left(\frac{u^2}{n}\right)\right) = -\frac{u^2}{2} + o(1)$ as $n \to \infty$ for all $u \in \mathbb{R}$.

So $\phi_n(u) \to e^{-u^2/2}$ as $n \to \infty$ for all $u \in \mathbb{R}$. But $e^{-u^2/2} = \phi_Z(u)$

so $Z_n \to Z$ weakly as required. The general case $\mu \in \mathbb{R}$, $\sigma^2 > 0$ reduces to the case $\mu = 0$, $\sigma^2 = 1$ by a linear change of variables (exercise)

23/11/12

Probability and Measure (22)

9. Ergodic Theory : Definitions

Let $(E, \mathcal{E}, \mu)$ be a measure space. A measurable function $\Theta : E \to E$ is a <u>measure preserving transformation</u> if $\mu \circ \Theta^{-1} = \mu$ $\left(\mu(\Theta^{-1}(A)) = \mu(A) \text{ for all } A \in \mathcal{E} \right)$

A set $A \in \mathcal{E}$ is <u>invariant</u> if $\Theta^{-1}(A) = A$. The set $\mathcal{E}_\Theta$ of all invariant sets is a σ-algebra. A measurable function f on E is <u>invariant</u> if $f = f \circ \Theta$. Then (exercise) f is invariant if and only if f is $\mathcal{E}_\Theta$-measurable.

We say that Θ is <u>ergodic</u> if $\mathcal{E}_\Theta$ contains only sets of measure 0 and their complements (we say that $\mathcal{E}_\Theta$ is trivial)

<u>Examples</u>

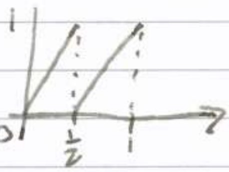
i) Shift Map on the torus. Fix $a \in E$.

$E = (0,1]^d$, $\mathcal{E} = \mathcal{B}(E)$, $\mu = \text{Lebesgue}^{\otimes d}$

Define $\Theta : E \to E$ by $\Theta(x_1, \dots, x_d) = (x_1 + a_1, \dots, x_1 + a_d)$ addition mod 1

It is easy to check that for $A = (b_1, c_1] \times \dots \times (b_d, c_d]$ we have $\mu(\Theta^{-1}(A)) = \mu(A)$. So μ and $\mu \circ \Theta^{-1}$ agree on the π-system of such sets, which generates $\mathcal{E}$. So Θ is a measure preserving transformation.

ii) Bakers' Map. $E = (0,1]$, $\mathcal{B}(E)$, Lebesgue



$\Theta(x) = 2x - \lfloor 2x \rfloor$

It is easy to see that $\mu(\Theta^{-1}(a,b]) = b - a$, so μ is a measure preserving transformation by the same π-system argument.

Proposition 9.1.1

Let f be an integrable function, θ a measure-preserving transformation. Then $f \circ \theta$ is integrable and $\mu(f\circ\theta) = \mu(f)$.

Proposition 9.1.2

Let f be invariant and suppose θ is ergodic. Then there exists a constant $c \in \mathbb{R}$ ~~with~~ such that $f = c$ almost everywhere.

Idea : $\{x \in E : f(x) \leq y\} \in \mathcal{E}_\theta$ so $\mu(f \leq y) = 0$ or $\mu(f > y) =$

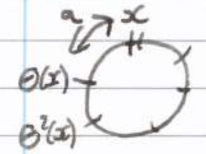
9.3 Birkhoff's and von Neumann's Ergodic Theorems

Let f be a measurable function $E \to \mathbb{R}$. Define $S_0 = 0$ and

$S_n = S_n^f = f + f\circ\theta + .. + f\circ\theta^{n-1}$

Thus for $x \in E$, $\frac{S_n(x)}{n} = \frac{1}{n}\left(f(x) + f(\theta(x)) + \dots + f(\theta^{n-1}(x))\right)$

which is the empirical average of f along the ~~point~~ orbit of θ from x.



Birkhoff's Theorem

σ-finite μ, θ a measure preserving transformation

$\Rightarrow S_n^f$ converges almost everywhere.

9.3.1 Lemma (Maximal Ergodic Lemma)

Let θ be a measure preserving transformation, and f be an integrable function. Define $S^*(x) = \sup_{n\geq 0} S_n^f(x)$

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

Proof

Consider $S_n^* = \max_{0\leq m\leq n} S_m$ and set $A_n = \{S_n^* > 0\}$

23/11/12

Probability and Measure (22)

For $m = 1, \dots, n$, we have $S_m = f + S_{m-1} \circ \theta \leq f + S_n^* \circ \theta$

On A_n we have $S_n^* = \max_{1 \leq m \leq n} S_m$ because $S_0 = 0 < S_n^*$

so $S_n^* \leq f + S_n^* \circ \theta$

On A_n^c we have $S_n^* = 0 \leq S_n^* \circ \theta$

Integrating and adding, we obtain

$$\int_E S_n^* \, d\mu \leq \int_{A_n} f \, d\mu + \int_E S_n^* \circ \theta \, d\mu$$

Now S_n^* is integrable, so this forces $\int_{A_n} f \, d\mu \geq 0$

Then $A_n \nearrow \{S^* > 0\}$ as $n \to \infty$ so by dominated convergence with dominating function $|f|$.

$$\int_{\{S^* > 0\}} f \, d\mu = \lim_{n \to \infty} \int_{A_n} f \, d\mu \geq 0 \qquad \square$$

10.1 Strong Law of Large Numbers with finite fourth moment.

Theorem 10.2.1

Let $(X_n : n \in \mathbb{N})$ be a sequence of independent random variables, and suppose that $\mathbb{E}(X_n) = 0$, $\mathbb{E}(X_n^4) \leq M$ for all n for some $M < \infty$. Set $S_n = X_1 + \dots + X_n$. Then $S_n/n \to 0$ almost surely.

Proof

~~Now~~ Observe that

$$\mathbb{E}(S_n^4) = \sum_{k=1}^{n} \mathbb{E}(X_k^4) + \binom{4}{2} \sum_{1 \leq j < k \leq n} \mathbb{E}(X_j^2 X_k^2)$$

because $\mathbb{E}(X_i X_j X_k X_l) = \mathbb{E}(X_i X_j^3) = \mathbb{E}(X_i X_j X_k^2)$ for all distinct indices i, j, k, l.

Now $\mathbb{E}(X_j^2 X_k^2) \leq \mathbb{E}(X_j^4)^{\frac{1}{2}} \mathbb{E}(X_k^4)^{\frac{1}{2}} \leq M$

So $\mathbb{E}(S_n^4) \leq nM + 6 \frac{1}{2} n(n-1) M \leq 3n^2 M$

So $\mathbb{E}\left(\sum_{n=1}^{\infty} (S_n/n)^4 \right) = \sum_{n=1}^{\infty} \mathbb{E}((S_n/n)^4) \leq \sum_{n=1}^{\infty} n^{-4} 3n^2 M$

$= 3M \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$

So $\sum_{n=1}^{\infty} (S_n/n)^4 < \infty$ almost surely so $S_n/n \to 0$ almost surely

$\square$

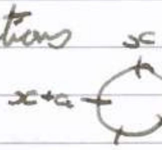
26/11/12

Probability and Measure (23)

Recap

$(E, \mathcal{E}, \mu)$ a measure space. We say that a measurable function $\theta : E \to E$ is a <u>measure-preserving transformation</u> if $\mu = \mu \circ \theta^{-1}$.

We say that a measurable function $f : E \to \mathbb{R}$ is <u>invariant</u> if $f = f \circ \theta$. We are interested in the averages of functions along orbits of θ.



$$S_n(x) = S_n^f(x) = f(x) + f(\theta(x)) + \dots + f(\theta^{n-1}(x))$$

<u>Maximal Ergodic Lemma</u>

Let θ be a measure-preserving transformation and let f be an integrable function on E. Define $S^* = \sup_n S_n$. Then $\int_{\{S^* > 0\}} f \, d\mu \geq 0$.

<u>Theorem 9.3.2</u> (Birkhoff's almost-everywhere ergodic theorem)

Assume that μ is σ-finite. Let θ be a measure-preserving transformation and let f be an integrable function on E. Then there exists an invariant function $\bar{f}$ on E such that $\mu(|\bar{f}|) \leq \mu(|f|)$ and $\frac{S_n}{n}(x) \to \bar{f}(x)$ almost everywhere.

<u>Proof</u>

Consider for $a, b \in \mathbb{R}$, with $a < b$. $D = D(a,b)$

$$D = \left\{ \liminf_n S_n/n < a, \ b < \limsup_n S_n/n \right\}$$

The functions $\liminf_n S_n/n$, $\limsup_n S_n/n$ are invariant.

$$S_{n+1} = S_n \circ \theta + f$$
$$\frac{S_{n+1}}{n+1} = \frac{n}{n+1}\left(\frac{S_n}{n}\right) \circ \theta + \frac{f}{n+1}$$
$$\limsup_n S_n/n = \left(\limsup_n S_n/n\right) \circ \theta$$

So D is invariant since $\liminf_n S_n/n$, $\limsup_n S_n/n$ are.

We shall show that $\mu(D) = 0$.

By restricting μ, θ, f to D if necessary, we reduce to the case where $D = E$. Either $a < 0$ or $b > 0$. We interchange cases by considering $-f, -b, -a$ in place of f, a, b.

Hence we reduce to the case $b > 0$.

Fix $B \in \mathcal{E}$ with $\mu(B) < \infty$ and consider $g = f - b1_B$.

Then $S_n^g(x) = S_n^f(x) - bS_n^{1_B}(x) \geq S_n^f(x) - nb > C$

(>0 for some $n \in \mathbb{N}$, for all x). That is, $S^{g*} > 0$ on D.

Note that g is integrable.

By the maximal ergodic lemma,

$$0 \leq \int_D g\, d\mu = \int_D f\, d\mu - b\mu(B).$$

There exists a sequence $B_n \in \mathcal{E}$ with $\mu(B_n) < \infty$ and $B_n \uparrow D$. Then $b\mu(D) = \lim_n b\mu(B_n) \leq \int_D f\, d\mu < \infty$

In particular $\mu(D) < \infty$.

Now we can repeat the argument with f replaced by $-f$ and b by $-a$ to obtain, taking $B = D$,

$$(-a)\mu(D) \leq \int_D (-f)\, d\mu$$

We now have $b\mu(D) \leq \int_D f\, d\mu \leq a\mu(D)$

This can only be true if $\mu(D) = 0$, since ~~$b \leq a$~~ $b > a$.

Consider $\Delta = \{\liminf_n S_n/n < \limsup_n S_n/n\}$. Then

$\Delta = \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} D(a,b)$, so Δ is invariant and $\mu(\Delta)=0$.

So we can define an invariant function $\tilde{f}: E \to [-\infty, \infty]$ by

$$\tilde{f}(x) = \begin{cases} \lim_n S_n/n\,(x) & x \in \Delta^c \\ 0 & x \in \Delta \end{cases}$$

Note that $\mu(|f| \circ \theta^n) = \mu(|f|)$ so

$$\mu(|S_n|) \leq \sum_{k=0}^{n-1} \mu(|f| \circ \theta^k) \leq n\mu(|f|)$$

By Fatou's Lemma, $\mu(|\tilde{f}|) = \mu(\liminf_n |S_n|/n) \leq \liminf_n \mu(|S_n|/n)$

$\leq \mu(|f|)$.

Hence $\tilde{f}(x) \in \mathbb{R}$ for μ-almost all x. Define

$$\bar{f}(x) = \begin{cases} \tilde{f}(x) & \text{if } \tilde{f}(x) \in \mathbb{R} \\ 0 & \text{otherwise} \end{cases}$$

Then $\bar{f}$ is invariant, $\mu(|\bar{f}|) = \mu(|\tilde{f}|) \leq \mu(|f|)$

and $S_n/n \to \bar{f}$ almost everywhere. $\square$

<u>Theorem 9.3.3</u> (von Neumann's L^p-ergodic theorem)

Assume $\mu(E) < \infty$. Let $p \in [1, \infty)$. Let θ be a measure-preserving transformation and let $f \in L^p$. Then $S_n/n \to \bar{f}$.

<u>Proof</u>

We have $\mu(|f|^p \circ \theta^n) = \mu(|f|^p)$. So by Minkowski's Inequality, $\|S_n\|_p \leq \sum_{k=0}^{n-1} \|f \circ \theta^k\|_p = n\|f\|_p$

Hence $\|S_n/n\|_p \leq \|f\|_p$. By Fatou's Lemma,

$\mu(|\bar{f}|^p) = \mu(\liminf_n |S_n/n|^p) \leq \liminf_n \mu(|S_n/n|^p)$

$\liminf_n \mu(|S_n^f/n|^p) \le \mu(|f|^p)$, so $\|\bar{f}\|_p \le \|f\|_p$

Fix $k \in (0, \infty)$, and consider $g = (-k) \wedge f \vee k$. By Dominated Convergence (using $|f|^p$ as dominating function) $\mu(|f-g|^p) \le \mu(|f|^p 1_{|f|>k}) \to 0$ as $k \to \infty$.

Given $\varepsilon > 0$, choose k so that $\|f-g\|_p \le \varepsilon/3$. Note that $f \in L^p(\mu) \subseteq L^1(\mu)$ as μ is a finite measure

Now $S_n^f - S_n^g = S_n^{f-g}$ and taking limits, $\bar{f} - \bar{g} = \overline{f-g}$ by Birkhoff's Theorem.

$$\|S_n^f/n - S_n^g/n\|_p = \|S_n^{f-g}/n\|_p \le \varepsilon/3,$$

$$\|\bar{f} - \bar{g}\|_p = \|\overline{f-g}\|_p \le \varepsilon/3$$

Now $S_n^g/n \to \bar{g}$ almost everywhere and $|S_n^g/n| \le k$, so by bounded convergence, $\mu(|S_n^g/n - g|^p) \to 0$ as $n \to \infty$.

So we can choose N so that $\forall n \ge N$,

$$\|S_n^g/n - \bar{g}\|_p \le \varepsilon/3$$

So by Minkowski's Inequality.

$$\|S_n^f/n - \bar{f}\|_p \le \|S_n^f/n - S_n^g/n\|_p + \|S_n^g/n - \bar{g}\|_p + \|\bar{g} - \bar{f}\|_p \le \varepsilon$$

□

28/11/12 Probability and Measure (24)

9.2 Bernoulli Shifts

Let m be a Borel probability measure on $\mathbb{R}$. There exists a sequence $(Y_n : n \in \mathbb{N})$ of independent random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ all having distribution m.

Consider $E = \mathbb{R}^{\mathbb{N}} = \{x = (x_n : n \in \mathbb{N}) : x_n \in \mathbb{R} \text{ for all } n\}$

Define coordinate maps $X_n : E \to \mathbb{R}$ by $X_n(x) = x_n$.

Set $\mathcal{E} = \sigma(X_n : n \in \mathbb{N})$. Then $\mathcal{E}$ is also generated by i.e. $\mathcal{E} = \sigma(\mathcal{A})$

$\mathcal{A} = \{A = \prod_{n \in \mathbb{N}} A_n : A_n \in \mathcal{B}, A_n = \mathbb{R} \text{ for all but finitely many } n\}$

Note that $\mathcal{A}$ is a π-system. We can define a random variable

$Y : \Omega \to E$ by $Y(\omega) = (Y_n(\omega) : n \in \mathbb{N})$

(Check $X_n \circ Y = Y_n : \Omega \to \mathbb{R}$ is measurable).

Define a probability measure μ on $(E, \mathcal{E})$ by $\mu = \mathbb{P} \circ Y^{-1}$

thus μ is the law of Y.

Call $(E, \mathcal{E}, \mu)$ the canonical model for i.i.d. sequences with distribution m. Consider the shift map $\Theta : E \to E$ by

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

Note that on $(E, \mathcal{E}, \mu)$, the coordinate maps $(X_n : n \in \mathbb{N})$ form a sequence of independent random variables all having distribution m.

For $A = \prod_{n \in \mathbb{N}} A_n \in \mathcal{A}$, $\Theta^{-1}(A) = \mathbb{R} \times A_1 \times A_2 \times \dots \in \mathcal{A}$.

and $\mu(A) = \prod_{n \in \mathbb{N}} m(A_n) = \mu(\Theta^{-1}(A))$

Since $\mathcal{A}$ generates $\mathcal{E}$, this shows that θ is measurable, and measure preserving on $(E, \mathcal{E}, \mu)$

Theorem 9·2·1

The shift map is ergodic. $(A = \theta^{-1}(A) \Rightarrow \mu(A) \in \{0, 1\})$

Proof

Define $\mathcal{Y}_n = \sigma(X_{n+k} : k \in \mathbb{N})$, $\mathcal{Y} = \bigcap_n \mathcal{Y}_n$

For $A = \prod_n A_n \in \mathcal{A}$, $\theta^{-n}(A) = \{X_{n+k} \in A_k \text{ for all } k \in \mathbb{N}\} \in \mathcal{Y}_n$

Since $\mathcal{Y}_n$ is a σ-algebra, and $\mathcal{A}$ generates $\mathcal{E}$, this implies that $\theta^{-n}(A) \in \mathcal{Y}_n$ for all $A \in \mathcal{E}$.

Suppose that $A \in \mathcal{E}_\theta$ (A is invariant), thus $A = \theta^{-1}(A) = \theta^{-n}(A) \in \mathcal{Y}_n$

Hence $A \in \mathcal{Y}$. But by Kolmogorov's zero-one law, for all $A \in \mathcal{Y}$, $\mathbb{P}(A) \in \{0, 1\}$ $\square$

10·2 Strong Law of Large numbers

Theorem 10·2·1 $\quad (S_n^f(x) = f(x) + f(\theta(x)) + \dots + f(\theta^{n-1}(x)))$

Assume that $\int_{\mathbb{R}} |x| \, m(dx) < \infty$, and $\int_{\mathbb{R}} x \, m(dx) = \nu$

Then $\mu(\{x \in E : \frac{x_1 + \dots + x_n}{n} \to \nu \text{ as } n \to \infty\}) = 1$.

Proof

We ~~apply~~ consider the function X_1 on $(E, \mathcal{E}, \mu)$. Then X_1 is integrable

$\mu(|X_1|) = \int_{\mathbb{R}} |x| m(dx) < \infty$ and $\mu(X_1) = \int_{\mathbb{R}} x \, m(dx) = \nu$

Consider the shift map θ a m.p.t.

Note that $S_n^{X_1}(x) = x_1 + \dots + x_n$. By Birkhoff's Theorem, there is an invariant function $\bar{X}_1$ such that $\frac{S_n^{X_1}}{n}(x) \to \bar{X}_1(x)$ for μ almost-all x.

28/11/12

Probability and Measure (24)

Since θ is ergodic, there exists a constant $c \in \mathbb{R}$ such that $\bar{X}_1 = c$ almost everywhere. But $S_n^{X_1}/n \to \bar{X}_1$ in L^1 (von Neumann ergodic theorem)

So $c = \mu(\bar{X}_1) = \lim_{n\to\infty} \mu\left(\frac{S_n^{X_1}}{n}\right) = \mu(X_1) = \nu$ $\square$

<u>Theorem 10.2.2</u> (Strong Law of Large Numbers)

Let $(Y_n : n \in \mathbb{N})$ be a sequence of integrable, i.i.d. r.v.s having mean ν. Set $S = Y_1 + \dots + Y_n$. Then $S_n/n \to \nu$ almost surely.

<u>Proof</u>

Observe $\{S_n/n \to \nu\} = \{Y \in E_0\}$ where $E_0 = \{x \in E : \frac{x_1 + \dots + x_n}{n} \to \nu\}$. So $\mathbb{P}(S_n/n \to \nu) = \mu(E_0) = 1$ $\square$

<u>Convergence of $X_n \to X$</u>

$$\begin{tikzcd}
\text{Almost Everywhere} \arrow[r, Rightarrow] \arrow[dr, dashed, "\text{Dominating Function}"'] & \text{In Probability} \arrow[r, Rightarrow] \arrow[l, dashed, bend right, "\exists\text{ a subsequence}"'] & \text{In Distribution} \arrow[d, Leftrightarrow] \\
\text{In } L^p\ (p>1) \arrow[r, Rightarrow] & \text{In } L^1 \arrow[u, Rightarrow] & \text{Weak Convergence} \arrow[r, Leftrightarrow] & \text{Convergence of Characteristic Functions}
\end{tikzcd}$$

We can return from $X_n \to X$ in distribution to $X_n \to X$ a.e. by changing the probability space.

$X,\ \underset{X_1}{\underset{\parallel}{X+\varepsilon_1}}, \dots, \underset{X_n}{\underset{\parallel}{X+\varepsilon_n}}$, $\mathbb{E}|\varepsilon_n| < \infty$ i.i.d. $\mathbb{E}\varepsilon_n = 0$

$S_n = \frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{a.s.} X$

X $X_1, X_2, \dots$ independent copies of X.

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} 1_{X_k \le x} \to F(x) \quad \text{a.e.}$$

$X_1, X_2, \dots$ i.i.d. r.vs $\mathbb{E}(X_1^2) < \infty$

Want to estimate $\mathbb{E}(X_1) = \mu$. Use $S_n/n \to \mu$ a.s.

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \le \varepsilon\right) = \mathbb{P}\left(\left|\frac{S_n - n\mu}{\sigma\sqrt{n}}\right| \le \frac{\varepsilon}{\sigma\sqrt{n}}\right) \xrightarrow{CLT} \int_{\{|x| \le \varepsilon/\sigma\}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, d$$