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

$A^c \in E, \bigcup_n A_n \in E$

The pair $(E, E)$ is called a measurable space. The elements of $E$ are called (E-) measurable sets.

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

The triple $(E, E, \mu)$ is called a measure space.

1.2 Discrete Measure Spaces

Let $E$ be a countable set, and take $E$ to be the set of all subsets of $E$. A mass function $m$ on $E$ is any function $m : E \rightarrow [0, \infty]$. Given a measure $\mu$ 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.

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

Extension $\leftrightarrow$ Carathéodory, Uniqueness $\leftrightarrow$ Dynkin
3 Generated $\sigma$-algebras

Let $\mathcal{A}$ be a set of subsets of $E$. Define $\sigma(\mathcal{A}) := \{A \subseteq E : \exists E, \mathcal{G} \subseteq \sigma \text{ containing } \mathcal{A} \}$

Then (exercise) $\sigma(\mathcal{A})$ is a $\sigma$-algebra. We call $\sigma(\mathcal{A})$ the $\sigma$-algebra generated by $\mathcal{A}$ and note that $\sigma(\mathcal{A})$ is the smallest $\sigma$-algebra containing $\mathcal{A}$.

1.4 $\pi$-systems and $\delta$-systems

Let $\mathcal{A}$ be a set of subsets of $E$. We say that $\mathcal{A}$ is a $\pi$-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 $\delta$-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 $\bigcup A_n \in \mathcal{A}$.

(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 $\pi$-system and $\delta$-system, then $\mathcal{A}$ is a $\sigma$-algebra.

Lemma 1.4.1 (Dynkin's Lemma)

Let $\mathcal{A}$ be a $\pi$-system. Then, any $\delta$-system containing $\mathcal{A}$ also contains the $\sigma$-algebra containing $\mathcal{A}$.

Proof:

Consider $\mathcal{D} = \{A \subseteq E : A \in \mathcal{D} \text{ for all } \delta$-systems $\mathcal{D} \text{ containing } \mathcal{A} \}$.

Then $\mathcal{D}$ is a $\delta$-system (exercise). We shall see that $\mathcal{D}$ is also a $\pi$-system, which proves the lemma, hence it is a $\sigma$-algebra, proving the lemma.
We shall see that $D'$ is a $\pi$-system.
Consider $D' = \{ B \in D : B \cap A \in D \text{ for all } A \in A \}$
Then $A \in D'$ because $A$ is a $\pi$-system. Also $E \in D'$.
Suppose $B_1, B_2 \in D'$ with $B_1 \subseteq B_2$. Then, for all $A \in A$,
$B_1 \cap A, B_2 \cap A \in D$ so $(B_2 \cap A) \setminus (B_1 \cap A) = (B_2 \setminus B_1) \cap A \in D$
So $B_2 \setminus B_1 \in D'$.

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

Hence $D$ is a $\pi$-system as claimed, so $D' = D$.

Now consider $D'' = \{ B \in D : B \cap A \in D \text{ for all } A \in D \}$.
Then $A \in D''$ because $D = D'$. We can check that
$D''$ is a $\pi$-system, just as for $D'$. Hence $D = D''$ so
$D$ is a $\pi$-system, as promised. $\square$
1.5 Set functions and properties

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

- $\mu$ is increasing if $A, B \in A$, $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$.
- $\mu$ is additive if $A, B \in A$ disjoint, $A \cup B \in A$, $\mu(A \cup B) = \mu(A) + \mu(B)$.
- $\mu$ is countably additive if $\{ A_n \} \in A$ disjoint, with $\bigcup \{ A_n \} \in A$ and $\forall n \in \mathbb{N}$, $\mu\left( \bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n)$.
- $\mu$ is countably sub-additive if $\{ A_n \} \in A$ disjoint, $\forall n \in \mathbb{N}$, $\bigcup_{n=1}^{\infty} A_n \in A$ and $\mu\left( \bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

1.6 Construction of measures

Let $A$ be a set of subsets of $E$, containing $\emptyset$. We say that $A$ is a ring if $\forall A, B \in A$, $B \subseteq A$, $A \cup B \in A$. We say that $A$ is an algebra if for all $A, B \in A$, $A \subseteq A$, $A \cup B \in A$.

Theorem 1.6.1 (Carathéodory's Extension Theorem)

Let $A$ be a ring of subsets of $E$ and let $\mu$ be a countably additive set function on $E$. Then $\mu$ extends to a measure on the $\sigma$-algebra generated by $A$.

Proof

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

$\mu^*(B) = \inf \sum \mu(A_n)$ where the infimum is taken over all sequences $(A_n : n \in \mathbb{N})$ in $A$ such that $B \subseteq \bigcup_{n=1}^{\infty} A_n$ and $\mu(A_n) < \infty$.
Say that \( A \subseteq E \) is \( \mu^* \)-measurable if \( \forall B \subseteq E \),
\[
\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)
\]
Write \( M \) for the set of all \( \mu^* \)-measurable sets.
We'll show that \( \mu^* = \mu \) on \( A \), that \( A \subseteq M \), and that
\( M \) is a \( \sigma \)-algebra, \( \mu^* | M \) is a measure, proving the theorem.
Note that \( \mu^*(\emptyset) = 0 \) and \( \mu^* \) is \( \sigma \) increasing.

**Step I**

We show that \( \mu^* \) 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 \( (A_{mn} : m, n \in \mathbb{N}) \) such that
\[
B_n \subseteq \bigcup_m A_{mn}, \quad \mu^*(B_n) + \frac{\varepsilon}{2^n} \geq \sum_m \mu (A_{mn})
\]
Then \( B \subseteq \bigcup_{n, m} A_{mn} \), 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.

**Step II**

We show that \( \mu^* = \mu \) on \( A \).

Since \( A \) is a ring, and \( \mu \) is countably additive, \( \mu \) is also increasing and countably sub-additive (example sheet).
So, for any \( A \in A \), and for any sequence \( (A_n : n \in \mathbb{N}) \) in \( A \) such that \( A \subseteq \bigcup_n A_n \),
\[
\mu (A) \leq \sum_n \mu (A_n) \leq \sum_n \mu (A_n \cap A)
\]
On taking the infimum over all such sequences we see that 
\( \mu^*(A) \geq \mu(A) \). The reverse inequality is obvious so 
\( \mu^* = \mu \) on \( A \).

**Step III**

We show that \( A \in M \). Take \( A \in A \) and \( B \in E \).

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 \( A \) 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 \( M \) is an algebra. It is clear that \( \emptyset \in M \), and \( A^c \in M \) whenever \( A \in M \). Take \( A_1, A_2 \in M \) and \( B \in E \).

Then, since \( \mu^*(B) = \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \)
\[ = \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c) \]
\[ = \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2^c) \cap A_1) \]
\[ = \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2^c) \cap A_1^c) \]
So \( A_1 \cap A_2 \in M \). Hence \( M \) is an algebra.
Step V

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

So take $B \subseteq E$. We have $\mu^*(B) = \sum \mu^*(B \cap A_n) = \sum \mu^*(B \cap A_1) = \mu^*(B \cap A_1) + \mu^*(B \cap A_2) + \mu^*(B \cap A_1 \cap A_2) = \sum \mu^*(B \cap A_n) = \mu^*(B \cap \bigcap_{n=1}^{\infty} A_n^c)$

So letting $n \to \infty$

$n^* (B) \geq \sum \mu^*(B \cap A_n) + \mu^*(B \cap A_\infty^c)$

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

countable sub-additivity of $\mu^*$.

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

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

So in fact equality holds throughout, so $\sigma$ $A \subseteq M$, and by taking $B = A$, we see that $\mu^*(A) = \sum \mu^*(A_n)$ as required.
Recap $E$, a set
- $E$, a $\sigma$-algebra, set of subsets of $E$.
- Closed under countable operations
- $\mu$, measure $\mu : E \rightarrow [0, \infty]$, $\mu (\bigcup A_n) = \sum \mu (A_n)$
- Dynkin
- $A \subseteq D \Rightarrow \sigma (A) \subseteq D$
- $\pi$-system $\pi$-system closed under relative complements increasing unions

Suppose $A_n \uparrow A$ ($A_n \subseteq A_{n+1}$, $\bigcup A_n = A$)
$B_1 = A_1$, $B_{n+1} = A_{n+1} \setminus A_n$, $A_n = B_1 \cup \ldots \cup B_n$, $A = \bigcup B_n$
$\mu (A_n) = \sum_{i=1}^{\infty} \mu (B_i) \rightarrow \lim_{n \to \infty} \mu (B_i) = \mu (A)$.

1.7 Uniqueness of measures

Theorem 1.7.1
Let $\mu_1$, $\mu_2$ be measures on $(E, E)$ with $\mu_1 (E) = \mu_2 (E)$ finite. Suppose that $\mu_1 (A) = \mu_2 (A)$ for all sets $A \in A$, where $A$ is a $\pi$-system generating $E$. Then $\mu_1 = \mu_2$.

Proof
Define $D = \{ A \in E : \mu_1 (A) = \mu_2 (A) \}$. Then $A \subseteq D$ and $E \subseteq D$. For $A, B \in D$ 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 $B \setminus A \in D$.
For $A_n \in D$ with $A_n \uparrow A$, we have $\mu_1 (A) = \lim_{n \to \infty} \mu_1 (A_n)$.
\[ m(A) = \lim_{n \to \infty} m_n(A_n) = \lim_{n \to \infty} m_n(A) = m_2(A), \forall A \in D. \]

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

1.8 Borel Sets and Borel Measures

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

1.9 Probability, finite and \( \sigma \)-finite measures

Let \( (E, \mathcal{A}, m) \) be a measure space.

If \( m(E) = 1 \), we say that \( m \) is a probability measure and call \( (E, \mathcal{A}, m) \) a probability space. We usually use notation \( (\Omega, \mathcal{A}, P) \) for this.

If \( m(E) < \infty \), we say that \( m \) is a finite measure.

If there exist sets \( E_n \in \mathcal{A} \) with \( m(E_n) < \infty \) for all \( n \) and \( \bigcup_n E_n = E \), we say that \( m \) is a \( \sigma \)-finite measure.

1.10 Lebesgue Measures

Theorem 1.10.1

There exists a unique Borel measure \( \mu \) on \( \mathbb{R} \) (usual topology) such that \( \mu([a, b]) = b - a \) for all \( a, b \in \mathbb{R} \) with \( a < b \).
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 \( \mu : \mathcal{A} \rightarrow [0, \infty] \) by

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

The presentation of \( \mathcal{A} \) is not unique as \( (a, b] \cup (b, c] = (a, c] \)

for \( a < b < c \). However, it is easy to see that \( \mu \) is well defined and moreover additive. We shall show that \( \mu \) is countably additive, so then by Carathéodory's Extension Theorem \( \mu \) extends to a measure on \( \mathcal{B}(\mathbb{R}) \) with the desired property.

Since \( \mu \) is additive, it will suffice to show that for \( A_n \in \mathcal{A} \) with \( A_n \uparrow A \in \mathcal{A} \) that \( \mu(A_n) \rightarrow \mu(A) \). Consider \( B_n = A \backslash A_n \) then \( A_n \cap B_n = \emptyset \). It will suffice to show that for \( B_n \in \mathcal{A} \) decreasing with \( \bigcap_{n=1}^{\infty} B_n = \emptyset \), that \( \mu(B_n) \rightarrow 0 \).

Suppose, for contradiction, that \( \exists \varepsilon > 0 \) such that

\( \mu(B_n) \geq \varepsilon \) \( \forall n \). We can find \( C_n \in \mathcal{A} \) such that

\( C_n \subseteq B_n \) and \( \mu(B_n \setminus C_n) \leq \varepsilon / 2^n \)

Note that \( \mu(B_n \setminus (C_1 \cup \ldots \cup C_n)) = \mu(B_n \setminus C_1) \cup \ldots \cup \mu(B_n \setminus C_n) \leq \varepsilon \sum_{k=1}^{n} 2^{-k} = \varepsilon \).
Hence \( \mu\left(\bigcap_{n=1}^{\infty} C_n\right) \geq \varepsilon \Rightarrow \bigcap_{n=1}^{\infty} C_n \neq \emptyset \).

Now \( K_n = \bigcap_{n=1}^{\infty} C_n \) is closed, non-empty, \( K_n \supseteq K_{n+1} \),

so since \( \mathbb{R} \) is complete, \( \emptyset = \bigcap_{n=0}^{\infty} B_n \supseteq \bigcap_{n=1}^{\infty} K_n \neq \emptyset \). \( \ast \)

So \( \mu(B_0) > 0 \) as required.
Uniqueness

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

For \( n \in \mathbb{Z} \) and \( B \in \mathcal{B} \), define
\[
\mu_n(B) = \mu(B \cap [n,n+1]), \quad \nu_n(B) = \nu(B \cap [n,n+1])
\]

Then, \( \mu_n, \nu_n \) are Borel probability measures on \( \mathbb{R} \), and \( \mu_n = \nu_n \) on the \( \pi \)-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 \( B \in \mathcal{B} \),
\[
\mu(B) = \sum \mu(B \cap [n,n+1]) = \sum \nu(B \cap [n,n+1]) = \nu(B)
\]

For \( x \in \mathbb{R} \), and \( B \in \mathcal{B} \), define \( \mu_x(B) = \mu(B+x) \) where \( B+x := \{b+x \mid b \in B\} \). Then \( \mu_x \) is a Borel measure.
\[
\mu_x((a,b]) = (b+x) - (a+x) = b-a.
\]

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

Note that \( \mu_0 \) is also translation invariant, if we define for \( B \in \mathcal{B} \)
\[
B+x = \{b+x \mod 1 \mid b \in B\} \subseteq (0,1]
\]

We used Carathéodory's Extension Theorem, so in fact we have constructed a \( \sigma \)-algebra \( \mathcal{M} \) and a measure.
\( m^* \) on \( M \) extending Lebesgue measure \( m \).
Recall that \( M \) is the set of "\( m^* \)-measurable sets".
We call \( M \) the set of Lebesgue-measurable sets.
In Ex 1.9, we see that \( M \) is concretely given by
\[
\mathcal{M} = \{ B \in \mathcal{B}, N \in \mathcal{N} \}.
\]
We set
\[
\mathcal{N} = \{ N \subseteq [0,1] : N \subseteq A \text{ for some } A \in \mathcal{B} \text{ with } m(A) = 0 \}.
\]
For \( B \in \mathcal{M} \), we have \( m(B \cup N) = m(B) \).

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
\( \bigcup_{q \in \mathbb{Q}} (S + q) \) disjoint and addition modulo 1.

\[
x = S + q = S' + q' \Rightarrow S - S' \in \mathbb{Q}
\]
Suppose that \( S \) is Lebesgue measurable. Then, by translation
invariance, \( m(S + q) = m(S) \forall q \in \mathbb{Q} \).
Then, by
\[
\sum_{q \in \mathbb{Q}} m(S + q) = \infty.
\]
This is impossible, so we conclude that \( S \) is not
Lebesgue-measurable, \( S \notin \mathcal{M} \).

1.12 Independence
Let \( (\Omega, \mathcal{F}, P) \) be a probability space. We use this to
model an experiment whose outcome is subject to chance.
\( \Omega \) models possible outcomes.
\( \mathcal{F} \) is the set of observable sets of outcomes, called events.
$P(A)$ 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 $S \subseteq I$,

$$P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i).$$

Exercise: Show that this always extends to infinite.

We say that a family $\{A_i : i \in I\}$ of sub-$\sigma$-algebras of $\mathcal{F}$ is independent if $\{A_i : i \in I\}$ is independent whenever $A_i \in \mathcal{A} \forall i$.

**Theorem 1.12.1**

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

The conclusion is that $\mathcal{F}(A_1), \mathcal{F}(A_2)$ are independent.

**Proof**

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

$$\mu(B) = P(A_1 \cap B), \quad \nu(B) = P(A_1)P(B).$$

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

So by uniqueness of extension, $\mu = \nu$ on $\mathcal{F}(A_2)$.

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

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

Then $\tilde{\mu}(\emptyset) = \tilde{\nu}(\emptyset) = P(A_2) < \infty$ and $\tilde{\mu} = \tilde{\nu}$ on $\mathcal{A}_1$, by the first part. So by uniqueness of extension, $\tilde{\mu} = \tilde{\nu}$ on $\mathcal{F}(A_1)$, proving the result. \(\square\)
1.13 Borel–Cantelli Lemmas

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

\[
\begin{align*}
\{A_n \text{ infinitely often}\} &:= \lim \sup A_n := \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_m \\
\{A_n \text{ eventually}\} &:= \lim \inf A_n := \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_m
\end{align*}
\]

Lemma 1.13.1 (First B.C. lemma)

Suppose \(\sum_{n} P(A_n) < \infty\). Then \(P(A_n \text{ infinitely often}) = 0\).

Proof

We have \(P(A_n \text{ i.o.}) \leq P\left(\bigcup_{m=n}^{\infty} A_m\right) \leq \sum_{m=n}^{\infty} P(A_m) \to 0\).

Lemma 1.13.2 (Second B.C. lemma)

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

Suppose \(\sum_{n} P(A_n) = \infty\). Then \(P(A_n \text{ i.o.}) = 1\).

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 \(P\left(\bigcap_{m=n}^{\infty} A_m^c\right) \leq \prod_{m=n}^{\infty} (1 - P(A_m)) \leq \exp(-\sum_{m=n}^{\infty} P(A_m)) \to 0\) as \(n \to \infty\), so \(P\left(\bigcap_{m=n}^{\infty} A_m^c\right) = 0\) for all \(n\).

So \(P\left(\bigcup_{m=n}^{\infty} A_m^c\right) = 0\). But \(\{A_n \text{ infinitely often}\} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = (\bigcup_{m=1}^{\infty} A_m)^c\), so \(P(A_n \text{ i.o.}) = 1\).

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 \(P(X \geq x) > 0\) for all \(x > 0\). Suppose that \(P(X \geq x) \to 0\) as \(x \to \infty\).

Take \(x_n \to \infty\) and set \(A_n = [X_n \geq x_n]\).
\( a_n = \text{Prob}(A_n). \) Then the events \((A_n : n \in \mathbb{N})\) are independent.

So \( \text{Prob}(X_n \geq x \text{ i.o}) = \text{Prob}(A_\infty \text{ i.o}) = \begin{cases} 1 & \sum a_n < \infty \\ 0 & \sum a_n = \infty \end{cases} \)

**Example**

\( X \sim \text{Exp}(1), \) \( \text{Prob}(X \geq x) = e^{-x} \)

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

\( \Rightarrow \) So if \( \alpha = 1, \) \( \sum a_n = \infty, \) whereas if \( \alpha > 1, \) \( \sum a_n < \infty. \) Hence \( \text{Prob}(X_n \geq \alpha \log n \text{ i.o}) = 1, \)

\( \text{Prob}(X_n \geq \alpha \log n \text{ i.o}) = 0 \) \( \forall \alpha > 1. \)

So \( \lim_{n \to \infty} \frac{X_n}{\log n} = 1 \) 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 measurable if

\( f^{-1}(A) \in \mathcal{E} \) for all \( A \in \mathcal{G}. \) Here, \( f^{-1}(A) \) is the inverse image \( f^{-1}(A) = \{x \in E : f(x) \in A\}. \)

Since \((g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A)),\) the composition of measurable functions is also measurable. Often \(G = \mathbb{R}\) or \([-\infty, \infty].\) Then we always take \(G = \mathbb{R}\)

For \( A \in \mathcal{E}, \) define the indicator function \(1_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}\)

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

Inverse images preserve set operations, take any \(f : E \to G\)

\( f^{-1}(G \cup A) = E \cup f^{-1}(A),\)

\( f^{-1}(G \cap A) = E \cap f^{-1}(A).\)
Hence \( \{ f^{-1}(A) : A \in \mathcal{G} \} \) is a \( \sigma \)-algebra on \( E \), and
\[ \{ A \in \mathcal{G} : f^{-1}(A) \in \mathcal{E} \} \] is a \( \sigma \)-algebra on \( \mathcal{G} \).

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,
\[ \{ A \in \mathcal{G} : f^{-1}(A) \in \mathcal{E} \} \] contains \( \mathcal{A} \), so also contains \( \mathcal{G}(A) = \mathcal{G} \).
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
\[ \iff \{ x \in E : f(x) \leq a \} \in \mathcal{E} \quad \forall a \in \mathbb{R} \).

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

**Proposition 2.1.**

Let \( (f_n : n \in \mathbb{N}) \) be a sequence of measurable functions on
\( (E, \mathcal{E}) \). Then the following functions are also measurable:
\( f_1 + f_2, \quad f_1 f_2, \quad \max f_n, \quad \sup f_n, \quad \liminf f_n, \quad \limsup f_n \)

**Proof.** (Exercise)

\( (f_1 + f_2)(x) > a \iff f_1(x) > q, f_2(x) > a - q \)
for some \( q \in \mathbb{Q} \).
2.2 Image Measures

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

Let \(\mu\) be a measure on \(E\). Then, for any measurable function \(f: E \to G\) we can define the image measure \(v = \mu \circ f^{-1}\) by \(v(A) = \mu(f^{-1}(A))\), \(A \in \mathcal{G}\).

Lemma 2.2.1

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

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

Proof

For \(y \in I\), define \(S_y = \{x \in \mathbb{R} : y \leq g(x)\}\). Then \(S_y \neq \emptyset\), \(S_y \neq \mathbb{R}\). Since \(g\) is non-decreasing, if \(x \in S_y\) and \(x' \geq x\), then \(x' \in S_y\). Since \(g\) is right continuous, if \(x_n \in S_y\) and \(x_n \xrightarrow{n \to \infty} x\) then \(x \in S_y\). Hence \(S_y = [\inf S_y, \infty)\).

So for \(x \in \mathbb{R}\), \(y \in I\), \(f(y) \leq x \iff y \leq g(x)\).

If \(y' \geq y\), then \(S_{y'} \subseteq S_y\) so \(f(y') \geq f(y)\). If \(y_n \uparrow y\), then \(S_{y_n} \uparrow S_y\), so \(f(y_n) \uparrow f(y)\).
then $S_y = \bigcap_n S_{y_n}$ so $f(y_n) \to f(y)$. Hence $f$ is non-decreasing and left continuous, as claimed. \qed

**Theorem 2.2.2**

Let $g: \mathbb{R} \to \mathbb{R}$ be as in the lemma. Then, there exists a unique Radon measure $\nu$ 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 $\nu$ and call $dg$ the **Lebesgue–Stieltjes Measure** associated with $g$.

**Proof**

Let $I$ and $f$ be as in the lemma. Let $\lambda$ 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 $\nu$ is Radon and has the desired property.

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

Suppose that $\nu$ is a Radon measure on $\mathbb{R}$. Define $g(x) = \int_{(a, x)} \mu(dx)$. Then $g$ is non-decreasing and right continuous.

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

$\mu = dg$. \qed
Probability and Measure

2.3 Random Variables

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

The image measure $\mu_X = P \circ X^{-1}$ is a measure on $(\mathcal{E}, E)$ called the law or distribution of $X$. The law $\mu_X$ is uniquely determined by the values on $\mu_X((\infty, x]) = F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(X \leq x), x \in \mathbb{R}$

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

$-F(x) \geq 0$ as $x \rightarrow -\infty$, $F(x) \rightarrow 1$ as $x \rightarrow +\infty$

$F$ is non-decreasing and right continuous.

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

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

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

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

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

Hence every distribution function is the distribution function of some random variable in $\mathbb{R}$. 
Recap: An $E$-valued random variable $X$ is a measurable function $X: \Omega \rightarrow E$. (We can ask: what is the probability that $X \in A$, for $A \in E$. $\Pr(X \in A) = \Pr(\{\omega \in \Omega : x(\omega) \in A\}) = \Pr(x^{-1}(A))$.

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

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

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

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

We say that $(X_i : i \in I)$ are independent if the $\sigma$-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 $x_1, \ldots, x_n \in \mathbb{R}$, $\Pr(x_1 \leq x_1, \ldots, x_n \leq x_n) = \Pr(x_1 \leq x_n) \cdots \Pr(x_n \leq x_n)$.

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

2.4 Rademacher Functions

Take $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}(\Omega)$, $\Pr = \text{Lebesgue}$.

Each element $\omega \in \Omega$ has a unique binary expansion $\omega = \sum_{n=1}^{\infty} \omega_n 2^{-n} = 0, \omega_1, \omega_2, \ldots$, such that $\omega_n = 1$ for infinitely many $n$ (i.e. use $0111\ldots$ instead of $100\ldots$).

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

Then $R_1(\omega) = \omega_1$, $R_2(\omega) = \omega_2$, $R_3(\omega) = \omega_3$, etc.
We have \( P(R_i = a_1, \ldots, R_n = a_n) = 2^{-n} \) for all \( n \), all \( a_k \in \mathbb{R} \).

Hence \((R_n : n \in \mathbb{N})\) is a sequence of independent Bernoulli \((\frac{1}{2})\) random variables \((P(R = 0) = P(R = 1) = \frac{1}{2})\).

**Proposition 2.4.1**

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 \( F_{X_n} = F_n \forall n \).

**Proof**

Let \( m : \mathbb{N}^+ \to \mathbb{N} \) 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 \( \sigma \)-algebra generated by \( \sigma(Y_n) \subseteq \sigma (Y_{k,n} : k \in \mathbb{N}) \) so \((Y_n : n \in \mathbb{N})\) are independent.

For \( y = 0, y, \ldots, y_k = i \cdot 2^{-k}, i \in \mathbb{Z}^+ \)

\[
P(i2^{-k} < Y_n \leq (i+1)2^{-k}) = P(Y_{k,n} = y_k) = 2^{-k}
\]

\[
P(Y_n \leq y) = y \quad \text{for all } \text{ real } y \quad \text{and hence for all } y \in (0, 1]
\]

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 \( G_n(y) \leq x \iff y \leq F_n(x) \).

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.
Finally \( F_n(x) = \Pr \left( \frac{X_n}{n} \leq x \right) = \Pr \left( \frac{X_n}{n} \leq F_n(x) \right) = F(x) \).

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 the property holds almost everywhere (a.e.).

Thus we say that a sequence of measurable functions \((f_n : n \in \mathbb{N})\) converges almost everywhere to \(f\) if
\[
\mu \{ x \in E : f_n(x) \neq f(x) \} = 0 \quad (f_n \to f \text{ a.e.})
\]

We say that \(f_n \to f\) in measure (in probability) if \(\forall \varepsilon > 0, \lim_{n \to \infty} \mu(\{x \in E : |f_n(x) - f(x)| > \varepsilon\}) = 0\).

Proposition 2.5.1 Let \((f_n : n \in \mathbb{N})\) 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 \(f = 0\).

Assume \(f_n \to 0\) almost everywhere.

a) Given \(\varepsilon > 0\), \(\mu(\{f_n \leq \varepsilon\}) \geq \mu \left( \bigcap_{m \geq n} \{ |f_m| \leq \varepsilon \} \right) \uparrow \mu \left( \bigcap_{m \geq n} \{ |f_m| \leq \varepsilon \} \text{ eventually} \right) \geq \mu(f_n > 0) \uparrow \mu(E)

Hence if \(\mu(E) < \infty\), this forces \(\mu(\{f_n > \varepsilon\}) \to 0\) so \(f_n \to 0\) in measure.
b) Suppose that $f_n \to f$ in measure. Then there exist $(\lambda_k)$ such that $\mu \left( |f_n - f| > \frac{1}{k} \right) \leq 2^{-k}$

Then $\mathbb{N} \left( |f_n - f| > \frac{1}{k} \text{ infinitely often} \right) = 0$ by the Borel-Cantelli Lemma

$\sum (\mathbb{E} 2^{-k} < \infty)$

So $|f_n| \leq k$ eventually almost everywhere, so $f_n \to f$ almost everywhere \qed
Probability and Measure

Convergence of Random Variables

Let \( X_n : n \in \mathbb{N} \) be real random variables defined on \((\Omega, \mathcal{F}, \mathbb{P})\). We say that \( X_n \to X \) almost surely if
\[
\mathbb{P}(X_n \to X) = \mathbb{P}(\{ \omega \in \Omega : X_n(\omega) \to X(\omega) \}) = 1
\]
We say that \( X_n \to X \) in probability if \( \mathbb{P}(|X_n - x| > \varepsilon) \to 0 \) as \( n \to \infty \) for all \( \varepsilon > 0 \).

For real random variables \( X, X_n : n \in \mathbb{N} \) (defined possibly in different probability spaces) we say that \( X_n \to X \) in distribution 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 \( x \).

Example

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 \( x_n > x \)

\[
\begin{align*}
F_x(b) & \quad F_x(x) = 1, F_{X_n}(x) = 0 \quad \forall n.
\end{align*}
\]

Example

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 \( F_{X_n} = F_X \) for \( X_n \to X \) in distribution.

Theorem 2.5.2

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

a) Suppose \( X_n : n \in \mathbb{N} \) are all defined on the same space \( \Omega \). Then \( X_n \to X \) in distribution.
b) If \( X_n \xrightarrow{} X \) in distribution, then there exist random variables \( X, X_n \) having the same distribution as \( X, X_n \) respectively, and defined on the same space such that \( X_n \xrightarrow{} X \) almost surely.

**Proof**

Write \( S \) for the subset of \( \mathbb{R} \) where \( F_x \) is continuous.

a) Suppose \( X_n \xrightarrow{} X \) in probability. Take \( x \in S \) and \( \epsilon > 0 \).

There exists \( \delta > 0 \) such that \( |F_x(y) - F_x(x)| \leq \frac{\epsilon}{2} \) if \( |y - x| \leq \delta \). Then, there exists a \( N \) such that for all \( n \geq N \),

\[
|F_{X_n}(x) - F_x(x)| \leq \frac{\epsilon}{2}.
\]

Then for \( n \geq N \),

\[
F_{X_n}(x) = P(X_n \leq x) \leq P(X \leq x + \delta) + P(|X_n - x| \geq \delta) 
\leq F_x(x) + \epsilon.
\]

So \( |F_{X_n}(x) - F_x(x)| \leq \epsilon \), as required.

b) Suppose \( X_n \xrightarrow{} X \) in distribution. Take \( S = (0, 1) \),

\( T = B(S) \), \( \mu = \text{Lebesgue} \), and set

\[
X_n(w) = \inf \{ x \in \mathbb{R} : w \leq F_{X_n}(x) \},
\]

\[
X(w) = \inf \{ x \in \mathbb{R} : w \leq F_x(x) \}.
\]

Then \( X, X_n \) are random variables on \((S, \mathcal{S}, \mu)\) and

\[
F_{X_n} = F_X, \quad F_{X_n} = F_X, \quad \forall n.
\]

Since \( X \) is non-decreasing, it has a most countably many discontinuities. So the set \( S_0 = (0, 1) \), where \( X \) is continuous is an event with \( \mu(S_0) = 1 \).
Since $F_x$ is non-decreasing, $\mathbb{R} \setminus S$ is also countable, so $S$ is dense in $\mathbb{R}$. Take $\omega \in \mathbb{R}_0$ and $\varepsilon > 0$. There exist $x^+ \in S$ such that $x^- < x^+(\omega) < x^+$ with $x^+ - x^- < \varepsilon$. Then there exist $w^+ \in (\omega, 1)$ such that $x^+(\omega) < x^+$. Remember that $x^+(\omega) \leq x^+ < \varepsilon$. Now $x^- < x^+(\omega)$ so $F_x(x^-) < \omega < F_x(x^+)$. Hence

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

2.6 Tail Events

Let $(X_n : n \in \mathbb{N})$ be a sequence of r.v.s on $(\mathcal{S}, \mathcal{F}, \mathbb{P})$. Define $\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$, $\mathcal{T} = \bigwedge_n \mathcal{T}_n$. We call $\mathcal{T}$ the tail $\sigma$-algebra of $(X_n : n \in \mathbb{N})$. We call each $A \in \mathcal{T}$ a tail event.

Example

If $(X_n \Rightarrow x)$ as $n \to \infty$ (convergence) then $X$ is $\mathcal{T}$ measurable. Generally, lim sup $X_n$ is $\mathcal{T}$-measurable.

Theorem 2.6

Let $(X_n : n \in \mathbb{N})$ be a sequence of independent r.v.s. The tail $\sigma$-algebra $\mathcal{T}$ of $(X_n : n \in \mathbb{N})$ is trivial, that is $\mathbb{P}(A) \in \{0, 1\}$ for all $A \in \mathcal{T}$. Moreover, for any
There exist $c \in \mathbb{R}$ such that $P(Y = c) = 1$.

**Proof.**

Set $T_n = \sigma(x_1, \ldots, x_n)$. Then $T_n$ is generated by the $\pi$-system of events $B = \{x_{m+1} \leq x_{m+1}, \ldots, x_{m+n} \leq x_{m+n+1} \}$ for all $m \in \mathbb{N}$. Now $P(A \cap B) = P(A)P(B)$ for all such $A, B$, by independence of $(X_n : n \in \mathbb{N})$. By Theorem 1.12.1, the $\sigma$-algebras $T_n$, $T_m$ are independent. Hence $T_n$ is independent of $Y$ for all $n$. Now $\cup_n T_n$ is a $\pi$-system generating $T_\infty = \sigma(x_n : n \in \mathbb{N})$. So by Theorem 1.12.1 again, $Y$ and $T_\infty$ are independent. But $T_\infty = Y \in T_\infty$. So for $A \in Y$, $P(A) = P(A \cap A) = P(A^2)$. So $P(A) \in \{0, 1\}$.

If $Y$ is $Y$-measurable, then $P(Y \leq y) \in \{0, 1\}$ for all $y$. Set $c = \inf \{y \in \mathbb{R} : F_Y(y) = 1\}$. Then $P(Y = c) = 1$. \qed
3 Integration

3.1 Definitions and Basic Properties

Let \((E, E, \mu)\) be a measure space.

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

We shall define for measurable functions \(f : E \to [0, \infty]\)

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 \(\int f(x) \, dx\).

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 simple function \(f\) on \((E, E)\) is any function of the form

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

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

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

This is well defined (see exercise 3.1.1) and satisfies for

simple functions \(f, g\): \(a, \beta \in \mathbb{R}_{\geq 0}\)

a) \(\mu(af + \beta g) = a \mu(f) + \beta \mu(g)\)

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) \(\mu(f) = 0 \quad \iff \quad f = 0\) almost everywhere.
For $f \geq 0$ measurable, we define

$$
\mu(f) = \inf \left\{ \mu(g) : g \text{ simple, } g \leq f \right\}
$$

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 integrable if $\mu(|f|) < \infty$. For integrable functions $f$, we define

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

Here $f^+ = \max\{f, 0\}$. Note that $f^+ \leq |f|$, so

$$
\mu(f^+) \leq \mu(|f|) \Rightarrow |\mu(f)| \leq \mu(|f|).
$$

We can define $\mu(f)$ by (8) provided not both $\mu(f^+), \mu(f^-)$ are infinite.

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 $x \in E$.

**Theorem (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 $\mu(f_n) \uparrow \mu(f)$.

**Proof:**

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

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

**Case 1** $(f_n = 1_{A_n}, f = 1_A)$

$\mu(f_n) = \mu(A_n) \vee \mu(A) = \mu(f)$ by countable additivity.
Case 2 (f_n simple, f = \mathbb{1}_A)

Given \epsilon > 0, consider \( A_n = \{ x \in E : f_n(x) > 1 - \epsilon \} \). Then \( A_n \in E \) and \( A_n \uparrow A \) as \( n \to \infty \). Note that \( f_n > (1 - \epsilon) \mathbb{1}_{A_n} \). So \( \mu(f_n) > (1 - \epsilon) \mu(A_n) \uparrow (1 - \epsilon) \mu(A) \) but \( (1 - \epsilon) \mu(A) = (1 - \epsilon) \mu(E) \). So since \( \epsilon > 0 \) was arbitrary, we are done.

Case 3 (f_n simple, f simple)

Write \( f = \sum_{k=1}^{m} a_k \mathbb{1}_{A_k} \) with \( a_k > 0 \) and \( A_k \in E \) disjoint. Then \( a_k \mathbb{1}_{A_k} f_n \uparrow \mathbb{1}_{A_k} \) as \( n \to \infty \) for all \( k \).

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

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

Case 4 (f_n simple, f > 0 measurable)

Take \( g \leq f \) simple. Then \( f_n \wedge g = \min \{ f_n, g \} \). \( g \geq 0 \)

\( \mu(f_n) > \mu(f_n \wedge g) \uparrow \mu(g) \). Take the supremum over all such \( g \) to see that \( \mu(f_n) \uparrow \mu(f) \).

Case 5 (f_n f > 0 measurable)

Define \( g_n = (2^{-n}/2^n f_n) \wedge 1 \)

\[ g_n = \sum_{i=0}^{n} 2^{-n} \mathbb{1}_{\{ x \in E : 2^{-n} f_n(x) \leq (2^{i+1})2^{-n} \}} + n \mathbb{1}_{\{ f_n \geq 2^n \}} \]

Then \( g_n \) is simple, \( g_n \leq f_n \) and \( g_n \uparrow f \) forces \( g_n \uparrow f \). Also, \( g_n \leq f_n \), so \( \mu(f_n) > \mu(g_n) \uparrow \mu(f) \).
Proposition 3.1.2

Let \( f, g \) be non-negative measurable functions. Then \( (x, y) \in \mathbb{R}^2 \):

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.

Proof:

a) Consider \( f_n = (2^{-n} L2^n f) \chi_{A^n}, g_n = (2^{-n} L2^n g) \chi_{A^n} \) then \( f_n, g_n \) simple, \( f_n \uparrow f, g_n \uparrow g \), so \( \alpha f_n + \beta g_n \)

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

\( \mu(\alpha f_n + \beta g_n) \uparrow \mu(\alpha f + \beta g) \), 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 everywhere for all \( n \) iff \( f = 0 \) almost everywhere.
Let \((E, \mathcal{E}, \mu)\) be a measure space. Write \(M^+\) for the set of non-negative measurable functions. Then there exists a unique map \(\mu : M^+ \to [0, \infty]\) such that \(\mu(1_A) = \mu(A)\) for all \(A \in \mathcal{E}\) and for every sequence \((f_n : n \in \mathbb{N}) \subseteq M^+\), we have \(\mu\left(\sum f_n\right) = \sum \mu(f_n)\).

Moreover, \(\mu(\alpha f) = \alpha \mu(f)\) for all \(\alpha \in [0, \infty]\) and \(f \in M^+\), and \(\mu(f) = 0\) if \(f = 0\) almost everywhere.

**Proposition 3.12**

We define for integrable functions (i.e., \(f\) measurable, \(\mu(\{f \leq 0\}) < \infty\))

\[
\mu(f) = \mu(f^+) - \mu(f^-), \quad f = (f^+)^* - \mu(f^-)^* \quad (0)
\]

Let \(f, g\) be integrable functions. Then

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

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

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

**Proof.**

Since \((-f)^+ = f^-\) we have \(\mu(-f) = \mu(f^-) - \mu(f^+) = -\mu(f)\).

For \(\alpha > 0\), \((\alpha f)^+ = \alpha f^+\) so \(\mu(\alpha f) = \mu\left((\alpha f)^+ - \mu((\alpha f)^-)\right) = \alpha \mu(f)\).

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 \(f^+ = 0\) almost everywhere, so \(\mu(f) = \mu(f^+) - \mu(f^-) = 0\).
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 \( f_n(x) \to f(x) \) as \( n \to \infty \) for almost all \( x \). Then \( \mu(f_n) \to \mu(f) \) (exercise)

3.2 Limits and Integrals
Lemma 3.2.1 (Fator's Lemma)
Let \( (f_n : n \in \mathbb{N}) \) be a sequence of non-negative measurable functions. Then \( \mu(\liminf f_n) \leq \liminf \mu(f_n) \)

Proof
Set \( g_n = \inf f_m \) and \( g = \liminf g_n \)
Then \( g_n \leq f_m \) for all \( m \geq n \), so \( \mu(g_n) \leq \inf \mu(f_n) \)
But \( g_n \leq g \) so by monotone convergence,
\( \mu(g_n) \to \mu(g) \). Now let \( n \to \infty \) in (*) to obtain
\( \mu(\liminf f_n) = \mu(g) = \lim \mu(g_n) \leq \liminf \mu(f_n) \).

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 exist an integrable function \( g \) such that \( |f_n(x)| \leq g(x) \) for all \( x \in E \). Then \( f_n \) 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.
\( 0 \leq g \leq f_n \), and \( g + f = \liminf (g + f_n) \).
Probability and Measure

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

\( \mu(\emptyset) - \mu(\Omega) = \mu(g - f) = \liminf (g - f_n) \)
\leq \liminf \mu (g - f_n) = \mu(\emptyset) - \limsup \mu (f_n).

Since \( \mu(\emptyset) < \infty \), we obtain \( \mu(f) \leq \liminf \mu (f_n) \)
\leq \limsup \mu (f_n) \leq \mu(f) \), Hence \( \mu(f_n) \to \mu(f) \) as required.

3.3 Transformation of Integrals

Proposition 3.3.1

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

Let \(A \subseteq E\). Write \(E_A = \{B \in \mathcal{E} : B \subseteq A\}\)
and set \(\mu_A = \mu\vert_{E_A}\). Then \(E_A\) is a \(\sigma\)-algebra, \(\mu_A\) is a
measure on \(E_A\) and for measurable functions \(f\) on \(E\),
\(f \vert_{E_A}\) in \(E_A\) - measurable and \(\mu_A(f \vert_{E_A}) = \mu (f \vert_{E_A})\).

Examples

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

Proposition 3.3.3

Let \(f\) be a non-negative measurable function. Defined for
\(A \subseteq E\), \(\nu(A) = \mu(f \vert_{E_A})\). Then \(\nu\) is a measure on \(E\),
and for all non-negative measurable functions \(g\),
\(\nu(g) = \mu(f \vert_{E_A})\).

Proof

Take a sequence of disjoint sets \(A_n \subseteq E\).
\[ \nu \left( \bigcup_n A_n \right) = \mu \left( \bigcup_n f \vert_{E_A} \right) = \mu \left( \sum_n f \vert_{E_A} \right) \]
(monotone convergence) = \(\sum_n \mu \left( f \vert_{E_A} \right) = \sum \nu \left( A_n \right) \)
Hence, $\nu$ is a measure.

For $g \in \mathcal{L}_1$, $B \in \mathcal{E}$, we have $\nu(g) = \nu(B) = \mu(f^{-1}g) = \mu(fg)$. The identity $\nu(g) = \mu(fg)$ extends to simple functions by linearity.

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

$$\nu(g) = \lim_n \nu(g_n) = \lim_n \mu(fg_n) = \mu(fg).$$
Proposition 3.3.4

Let \((E, E, \mu)\) be a measurable space \((G, G)\) a measurable space, and \(f: E \rightarrow G\) 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 \((E, E)\)-valued random variable on \((\Omega, \mathcal{F}, \mathbb{P})\),

and if \(g: E \rightarrow [0, \infty)\) is measurable, then

\[\mu_x(g) = \int g \, d\mu_x = E(g(x))\]

We often use Proposition 3.3.3 to specify a measure on \(\mathbb{R}\) by a density function \(f\) with respect to Lebesgue measure \(\mu\).

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

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 \(\nu\) 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 \(g: \mathbb{R} \rightarrow [0, \infty)\) measurable, \(E(g(x)) = \int_{\mathbb{R}} g \, d\mu_x = \int_{\mathbb{R}} g(x) \, f(x) \, dx\).

3.4 Fundamental Theorem of Calculus

Theorem 3.4.1

Let \(f: [a, b] \rightarrow \mathbb{R}\) be a continuous function.

Define \(F_a(t) = \int_a^t f(x) \, dx\), \(t \in [a, b]\). Then \(F_a\) is differentiable on \([a, b]\) with \(F'_a = f\).
b) Suppose \( F: [a, b] \rightarrow \mathbb{R} \) is differentiable 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 \( 0 < h \leq \delta \),

\[
\left| \frac{F(a + h) - F(a)}{h} - f(t) \right| = \left| \int_{t}^{t+h} \frac{f(x) - f(t)}{h} \, dx \right| \leq \frac{1}{h} \int_{t}^{t+h} |f(x) - f(t)| \, dx < \varepsilon.
\]

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 \( f: U \times E \rightarrow \mathbb{R} \) satisfies

\begin{enumerate}[(I)]
\item \( x \mapsto f(t, x): E \rightarrow \mathbb{R} \) is integrable for all \( t \in U \).
\item \( t \mapsto f(t, x): U \rightarrow \mathbb{R} \) is differentiable for all \( x \in E \).
\end{enumerate}

Then, there exists an integrable function \( g \) on \( E \) such that

\[
\left| \frac{df}{dt}(t, x) \right| \leq g(x) \text{ for all } t \in U, x \in E.
\]
(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 \( \mathbb{R} \) \( \{0\} \) such that \( t + h_n \in U \)

for all \( n \). Define \( g_n(x) = \frac{1}{h_n} \left( f(t + h_n, x) - f(t, x) \right) - \frac{\partial f}{\partial t}(t, x) \)

Then \( g_n(x) \to 0 \) as \( n \to \infty \) for all \( x \) by (II). By the Mean Value Theorem \( |g_n(x)| = \left| \frac{\partial f}{\partial t}(t + h_n, \xi_n, x) - \frac{\partial f}{\partial t}(t, x) \right| \)

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 \( \frac{1}{h_n} \left( F(t + h_n) - F(t) \right) - \int_E \frac{\partial f}{\partial t}(t, x) \mu(dx) = \int_E g_n(x) \mu(dx) \to 0 \)

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 $\nu$ is Lebesgue by uniqueness of extension \qed
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 \(\sigma\)-system on \(E\) satisfying \(E \in \mathcal{A}\) and \(\sigma(\mathcal{A}) = \mathcal{E}\). Suppose that \(V\) is a vector space of bounded \(E\)-measurable functions such that

i) \(I_A \in V\) 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 \(E\)-measurable functions.

Proof

Consider \(D = \{ A \in \mathcal{E} : I_A \in V \}\). Then \(A \in D \iff E \in D\). For \(A = B\), \(I_{B \setminus A} = I_B - I_A\), so if \(A, B \in D\) then \(B \setminus A \in D\).

For \(A_n \uparrow A\), \(I_A \leq \limsup I_{A_n}\), so if \(A_n \in D\) for all \(n\), then \(A \in D\). Hence \(D\) is a \(\sigma\)-system, so \(D = \sigma(\mathcal{A}) = \mathcal{E}\) by Dynkin's Lemma.

Now for \(f \geq 0\), measurable, bounded, consider \(f_n = 2^{-n} I_{f \leq 2^n}\). 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, \(f^+ \in V\) so \(f = f^+ - f^- \in V\).

Let \((E_1, \mathcal{E}_1, \mu_1)\) and \((E_2, \mathcal{E}_2, \mu_2)\) be finite \((for \; now)\) measurable spaces. Write Set \(A = \{ A_1 \times A_2 : A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2 \}\)

Write \(E = E_1 \times E_2\), \(E = \mathcal{E}_1 \otimes \mathcal{E}_2 = \sigma(\mathcal{A})\). Note that \(A\) is a \(\sigma\)-system on \(E\) with \(E \in A\) and \(\sigma(A) = \mathcal{E}\)
Lemma 3.6.1

Let \( f : E \rightarrow \mathbb{R} \) be an \( E \)-measurable function.

Then for \( x_1 \in E_1 \), the function \( x_2 \mapsto f(x_1, x_2) : E_2 \rightarrow \mathbb{R} \) is

\( E_2 \)-measurable.

Proof

Apply the monotone class theorem: with \( V \) the set of all bounded, \( E \)-measurable functions \( f \) for which the conclusion holds.

Lemma 3.6.2

Let \( f : E \rightarrow \mathbb{R} \) be a bounded, \( E \)-measurable function and define

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

Then \( f_1 \) is an \( E_1 \)-measurable function.

Proof

Use the monotone class theorem.

Theorem 3.6.3 (Product Measure)

There exists a unique measure \( \mu = (\mu_1 \otimes \mu_2) \) on \( E \) such that

\( \mu(A_1 \times A_2) = \mu_1(A_1) \mu_2(A_2) \) for all \( A_1 \in E_1, A_2 \in E_2 \).

Proof

For uniqueness, since \( A \) is a \( \sigma \)-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 \( A \in E \)

\[ \mu(A) = \int_{E_1} \left( \int_{E_2} 1_A(x_1, x_2) \, \mu_2(dx_2) \right) \mu_1(dx_1). \]
This is well-defined by the lemma, and \( \mu(A \times A') = \mu_1(A) \mu_2(A') \) for all \( A, A' \in E_1 \). For \( (A_n : n \in \mathbb{N}) \) disjoint in \( E_1 \),

\[
\mu \left( \bigcup_n A_n \right) = \int_{E_1} \left( \sum_n 1_{A_n}(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1) = \sum \mu(A_n) \text{ by the monotone convergence theorem.}
\]

Proposition 3.6.4

Define \( E = E_2 \times E_1 \), \( E = E_2 \times E_1 \), \( \hat{\mu} = \mu_2 \otimes \mu_1 \). For \( f : E \to \mathbb{R} \), define \( \hat{f} : E \to \mathbb{R} \) by \( \hat{f}(x_2, x_1) = f(x_1, x_2) \). Then suppose that \( f \) is non-negative and \( E \)-measurable. \( \hat{f} \) is then \( \hat{\mu} \)-measurable and \( \hat{\mu}(\hat{f}) = \mu(f) \).

Proof

(If \( f_n \) bounded, then \( \hat{\mu}(f_n) = \lim \mu(f_n) \) and \( \mu(f) = \lim \mu(f_n) \text{ by monotone convergence} \).)

Use the monotone class theorem.

Theorem 3.6.3 (Fubini's Theorem)

a) Let \( f \) be a non-negative \( E \)-measurable function. Then

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

Prop 3.6.4

b) Let \( f \) be a \( \mu \)-integrable (\( E \)-measurable) function, then

i) \( x_2 \mapsto f(x_1, x_2) : E_2 \to \mathbb{R} \) is \( \mu_2 \)-integrable for almost all 

\( \mu_1 \) \( x_1 \in E_1 \).

ii) \( x_1 \mapsto \int_{E_2} f(x_1, x_2) \mu_2(dx_2) \) is \( \mu_1 \)-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 \( f_n = (2^{-n} L^{2^{n-1}} + 1) \cdot n \).

Then \( \mu(f) = \lim_n \mu(f_n) = \lim_n \int_{E_1} f_n(x, x_2) \, d\mu_2(x_2) \, d\mu_1(x_1) \).

b) For \( f \) a \( \mu \)-integrable function, we have

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

So \( \int_{E_2} f(x_1, x_2) \, d\mu_2(x_2) < \infty \) for almost everywhere \( x_1 \) with

and \( x_1 \mapsto \int_{E_2} f(x_1, x_2) \, d\mu_2(x_2) \) is \( \mu \)-integrable.
Recall $(E_1, E_1, \mu_1), (E_2, E_2, \mu_2)$ finite measure spaces
$E = E_1 \times E_2, \mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2 = \sigma(A_i \times A_2, A_i \in \mathcal{E}_1, A_2 \in \mathcal{E}_2)$
$\mu = \mu_1 \otimes \mu_2$, product measure on $E : \mu(A_i \times A_2) = \mu(A_i) \mu(A_2)$

Theorem 3.6.5

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

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

$f_1(x_1) = \int_{E_2} f(x_1, x_2) \mu_2(dx_2) \quad \text{for } x_1 \in A_1$

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

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

Proof

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 $\int_{E_1} \left( \int_{E_2} |f(x_1, x_2)| \mu_2(dx_2) \right) \mu_1(dx_1)$ is finite.

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

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

Then $\int_{E_1} f_1(x_1) dx_1 < \infty$ on $A_1$ and $f_1 = f^+ - f^-$ on $A_1$.

Now $f_1$ are $\mu$-measurable and by part (a)

$\mu_1(f_1^+) = \mu(f^+) < \infty$

Hence $f_1$ is $\mu_1$-integrable and

$\mu_1(f_1) = \mu_1(f_1^+) - \mu_1(f_1^-) = \mu(f^+) - \mu(f^-) = \mu(f)$
Extensions and non-extensions

- Fubini's Theorem extends easily to σ-finite measures:

\[ \mu_k = \sum \mu_k^{(n)} \quad \text{with} \quad \mu_k^{(n)} \text{ finite for all } n, \quad k = 1, 2. \]

\[ \mu = \mu_1 \otimes \mu_2 = \sum \mu_1^{(n)} \otimes \mu_2^{(m)} \quad f \geq 0, \text{ measurable.} \]

\[ \mu(f) = \sum \mu_1^{(n)}(f) \otimes \mu_2^{(m)}(f) = \sum \text{iterated integrals} \]

= iterated integrals (using monotone convergence)

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

Take \( E_1 = E_2 = [0, 1] \), \( \mu_1 = \text{Lebesgue} \), \( \mu_2 = \text{counting} \).

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

\[ \int_{E_1} \left( \int_{E_2} f(x_1, x_2) \, \mu_2(dx_2) \right) \, \mu_1(dx_1) = \frac{1}{2} \]

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

Now consider:

\[ \int_0^1 \left( \int_0^1 f(x_1, x_2) \, dx_1 \right) \, dx_2 = \frac{1}{2} \neq \frac{1}{2} \]

- 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:

\[ \mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}(\mathbb{R}) \]

Write \( \lambda \) for Lebesgue measure on \( \mathbb{R} \). Then the \( n \)-fold product measure \( \lambda_n = \lambda \otimes \ldots \otimes \lambda \) is a measure on \( \mathbb{R}^n \).
and is the unique measure such that $\lambda_n \left( \bigcap_{k=1}^{n} A_k \right) = \prod_{k=1}^{n} \lambda_k(A_k)$.

We will write $\lambda_n(F) = \int_F f(x) \, dx$.

### 3.7 Independent Random Variables

Recall that random variables $X_1, \ldots, X_n$ on $(\Omega, \mathcal{F}, \mathbb{P})$ are independent if the $\sigma$-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 $E = E_1 \times \cdots \times E_n$ and $E = E_1 \otimes \cdots \otimes E_n$. 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) $\mu_X = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$

c) For all non-negative measurable functions $f_k : E_k \to \mathbb{R}$, we have

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

**Proof**

Consider $A = \bigcap_{k=1}^{n} A_k : A_k \in \mathcal{E}_k$. Then for $A = \bigcap_{k=1}^{n} A_k \in \mathcal{A}$,

$$X^{-1}(A) = \bigcap_{k=1}^{n} X_k^{-1}(A_k) \in \mathcal{F}.$$ But $\sigma(A) = E \otimes X$, $E$ measurable.
Suppose that a) holds. Take $A = \bigcap_{k=1}^{n} A_k \in \mathcal{A}$. Then

$$\mu_x (A) = \prod_{k=1}^{n} \mu (x \in A) = \prod_{k=1}^{n} \mu (x \in A_k) = \prod_{k=1}^{n} \mu_x (A_k) = \mu_x (\bigcap_{k=1}^{n} A_k) = \upnu (A)$$

But $\mathcal{A}$ is a $\sigma$-system generating $\mathcal{E}$. So $\mu_x = \upnu$ on $\mathcal{E}$. Hence b) holds.

Suppose now that b) holds. Then

$$E \left( \bigcap_{k=1}^{n} f_k (x_k) \right) = \int_{E \in \mathcal{E}} \bigcap_{k=1}^{n} f_k (x_k) \mu_x (dx_k) = \prod_{k=1}^{n} \int_{E_k} f_k (x_k) \mu_x (dx_k)$$

Finally, if c) holds, then we can deduce a) by taking $f_k = 1_{A_k}$ for $\mathcal{E}_k$-measurable sets $A_k$. $\square$
Let $(E, E, \mu)$ be a measure space. Define, for $1 \leq p < \infty$, $L^p = L^p(E, E, \mu)$ as the set of all measurable functions for $E$ having finite $L^p$-norm. Here:

$$\|f\|_p = \left( \int |f|^p \, d\mu \right)^{1/p}$$

for $1 \leq p < \infty$

$$\|f\|_\infty = \inf \{ A > 0, |f| \leq A \text{ \, almost everywhere} \}$$

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

Note $\|f\|_p \leq \mu(E)^{1/p} \|f\|_\infty$ 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 Chebychev's Inequality

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

$$\lambda \leq \mu([f \geq \lambda]) \leq \mu(E)$$

Chebychev's Inequality

For any measurable function $g$ on $E$ and any non-negative Borel-function $\Phi$ on $\mathbb{R}$ we can apply this to $f = \Phi \circ g$ 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 $\lambda \to \infty$. 

4.3 Jensen's Inequality

Let $I$ be an interval in $\mathbb{R}$. Say that a function $f: I \rightarrow \mathbb{R}$ is convex if for all $x, y \in I$, and all $t \in [0, 1]$, $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$.

**Lemma 4.3.1**

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

**Proof**

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, if $a \in \mathbb{R}$ 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 $a(x-m) + f(m) \leq f(x)$ for $x \in I$.

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 $f: I \rightarrow \mathbb{R}$ be a convex function. Then $E(f(X))$ is well defined and $f(E(X)) \leq E(f(X))$. 
Proof
The case where $X = c$ almost surely for some constant $c$ is obvious. We exclude it. Then $m = E(X)$ must be an interior point of $I$. $(X \neq E(X) \text{ almost surely means } X - E(X) > 0, \mathbb{P}(X - E(X) > 0) = 0)$. We will see $f(E(X)) = f(m) = am + b = E(ax + b)$.

For all $x \in I$

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

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

$f(E(X)) = f(m) = am + b = E(ax + b) \leq E(f(X))$.

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

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

$(E(|X|^p))^{\frac{1}{p}} \leq E(|X|^q)^{\frac{1}{q}}$.

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

Hence $L^p \supseteq L^q$ and Minkowski's

4.4 Hölder'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 $p, q \in [1, \infty]$ be conjugate indices. Then $\mu (|fg|) \leq \|f\|^p \|g\|^q$

Proof

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

The case where $\|f\|^p = 0, \infty$ is clear. We exclude these too.

(N.B. $\|f\|^p = \|Af\|^p$)

By scaling, we reduce to the case $\|f\|^p = 1$. Define $P(A) = \mu (1_{A} |f|^p)$, $A \in E$.

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

So $\mu (|fg|) = \mu \left( \frac{1}{|f|^p} |f|^p \right) = E(X) \leq E(X^q)^{\frac{1}{q}}$

(set $X = \frac{|f|^q}{|f|^p}$)

$m(m |f|^q)^{\frac{1}{q}} = \mu (|f|^q)^{\frac{1}{q}}$

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

$\leq \mu (|f|^q)^{\frac{1}{q}} = \|f\|^q = \|g\|^q \|f\|^p$ as required \qed
\[ \|F\|_p = \mu(\{F \neq 0\})^{\frac{1}{p}} \quad p \in [1, \infty) \]

\[ \|F\|_\infty = \inf \{ \lambda > 0 : \|F\|_\lambda \text{ is finite} \} \]

Hölder's inequality:
\[ \mu(\{f \neq 0\}) \leq \|F\|_p \|g\|_{q'} \quad \left( \frac{1}{p} + \frac{1}{q'} = 1 \right) \]

\[ \|af\|_p = |a| \|f\|_p \]

**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
\[ |f + g| \leq |f| + |g| \leq 2 \max\{\|f\|_p, \|g\|_p\} \]
\[ |f + g|^p \leq 2^p \max\{\|f\|_p, \|g\|_p\}^p \leq 2^p (\|f\|_p^p + \|g\|_p^p) \]

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
\[ \mu(\{f + g\}^p) \leq \|f\|_p \|f + g\|^{p-1} \|g\|_{p'} + \|g\|_p \|f + g\|^{p-1} \|f\|_p \]

But
\[ \|f + g\|_{p'} = \mu(\{f + g\}^{(p-1)q'})^{\frac{1}{p}} \quad (\text{but } (p-1)q' = p) \]

\[ = \|f + g\|^{1 - \frac{1}{p}} \] and the result follows. \( \square \)
5. Completeness of $L^p$ and orthogonal projection

5.1 $L^p$ as a Banach Space

Let $V$ be a real vector space. A norm on $V$ 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 functions.

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 may say that $(V, \| \cdot \|)$ is a Banach space.
Theorem 5.1.1 (Completeness of $L^p$)

Let $p \in [1, \infty)$, and $(f_n : n \in \mathbb{N})$ be a sequence in $L^p$ such that

$$\|f_n - f_m\|_p \to 0 \quad \text{as} \quad n, m \to \infty.$$ Then there exists $f \in L^p$ such that

$$\|f_n - f\|_p \to 0 \quad \text{as} \quad n \to \infty. \quad \text{(This says exactly that $(L^p, \|\cdot\|_p)$ is complete.)}$$

Proof (for $p \in [1, \infty)$)

There exists a subsequence $(f_{n_k})$ such that

$$S = \sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p < \infty \quad \text{(fast Cauchy subsequence)}.$$ By Minkowski's inequality, for any $E \in \mathcal{E}$,

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

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

$$S^p \geq \mu\left(\sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p^p\right)^{1/p} \geq \mu\left(\sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p^p\right)^{1/p}$$

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

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

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

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

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

Then $\mu(f_n - f)^p = \mu(\liminf_{k} f_n - f_{n_k})^p$.
\[ \mu \left( \liminf_{k \to \infty} \| f_n - f \|_p \right) \leq \liminf_{k \to \infty} \mu \left( \| f_n - f_k \|_p \right) \leq \varepsilon \]

by Fatou's Lemma, for all \( n \geq \mathbb{N} \). Hence \( f_n \to f \) in \( L^p \).

Note that if \( \| f_n - f \|_p \to 0 \), \( \| f_n \|_p - \| f \|_p \leq \| f_n - f \|_p \to 0 \).

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

\[ \langle v, v \rangle \geq 0 \quad \text{for all } v \in V \quad \text{with equality only if } v = 0. \]

Given an inner product, we can define a norm by

\[ \| v \| = \sqrt{\langle v, v \rangle} \quad \text{(Proof: exercise, Cauchy-Schwarz)} \]

If the resulting normed space is a Banach space, then we call \( (V, \langle \cdot, \cdot \rangle) \) a Hilbert space.
Theorem 5.12 (Approximation in $L^p$)

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

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

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 $D$ denote the set of all $A \in \mathcal{E}$ such that $1_A \in V$. Then $\mathcal{A} \subseteq D$ and ...

(to be continued)

5.2 $L^2$ as a Hilbert Space

Let $V$ be a real vector space. A bilinear map

$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is an inner product if $\langle v, v \rangle > 0$ for all $v \in V$ with equality only if $v = 0$. Given an inner product, we can define a norm on $V$ by $\|v\| = \langle v, v \rangle^{1/2}$ (Cauchy-Schwarz). Then $\|v + u\|^2 = \|v\|^2 + 2\langle v, u \rangle + \|u\|^2$ (Pythagorean) 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 $\langle f, g \rangle = \mu(fg) = \int_{\mathcal{E}} fg \, d\mu$. 

Then \( <\cdot, \cdot> : L^2 \times L^2 \to \mathbb{R} \) (\( \mu (|fg|) \leq \|f\|_2 \|g\|_2 \)) is bilinear and \( <f, f> \geq 0 \) but equality holds iff \( f = 0 \) almost everywhere. Note \( <f, f> = \|f\|_2^2 \). Hence, since \( L^2 \) is complete, we see that \( L^2 = \frac{1}{\sqrt{\mu}} \) in a Hilbert space.

For \( f, g \in L^2 \), we say \( f, g \) are orthogonal if \( <f, g> = 0 \). Define for \( V \subseteq L^2 \), \( V^\perp = \{ f \in L^2 : <f, g> = 0 \text{ for all } g \in V \} \) \( \forall n \in \mathbb{N} \).

We say that \( V \) is closed if for any sequence 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.

**Theorem 5.2.1 (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 \( v \), \( f - v \in V^\perp \).

**Proof**

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,

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

\[
\|f - \frac{v_n + v_m}{2}\|_2^2.
\]

Since \( \frac{v_n + v_m}{2} \in V \), this is no less than

\[
4d(f, V)^2.
\]

Hence \( \|v_n - v_m\|_2 \to 0 \) as \( n, m \to \infty \).

By completeness of \( L^2 \), there exists \( g \in L^2 \) such that
Probability and Measure

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

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

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

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

So $\langle f - v, g \rangle = 0$. Hence $f - v \subseteq V$.

5.3 Conditional Expectation $(\Omega, \mathcal{F}, P)$

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

$E(X|\mathcal{G}) = \left\{ \begin{array}{ll} E(X|\mathcal{A}) & \text{if } P(A) > 0 \\ 0 & \text{otherwise} \end{array} \right.$

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

Theorem

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

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

Proof

It is clear that $E(X, \mathcal{G}) \subseteq L^2(\Omega, \mathcal{G}, P)$. Then it suffices to show that $E((X - X)Z) = 0$ for all $Z \subseteq L^2(\Omega, \mathcal{G}, P)$.
Any such \( Z \) has the form \( \sum_{A \in \mathcal{A}} q_A 1_A \), so we reduce to the case
\( Z = 1_A, \ A \in \mathcal{A} \). Then \( E(X 1_A) = E(X 1_A) P(A) = E(X 1_A) \)
as required.

Note that if \( f \in L^2 \) and \( v \in V \) with \( f - v \in V^\perp \) then
\( d(f, v) = \| f - v \| \) no \( v \) in the orthogonal projection.
For, if \( g \in V \), then
\[
\| f - g \|^2 = \| f - v + (v - g) \|^2 + \| f - v \|^2 + 2 \langle f - v, v - g \rangle + \| v - g \|^2.
\]
\( = 0 \)
6. Convergence in $L^1(P)$

6.1 Bounded Convergence

**Theorem 6.1**

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

Suppose there exists a constant $C < \infty$ such that $|X_n| \leq C$ almost surely for all $n$. Then $X_n \rightarrow X$ in $L^1(E(|X_n - X|) \rightarrow 0)$.

**Proof**

There is a subsequence $(X_{n_k})$ such that $X_{n_k} \rightarrow \tilde{X}$ almost surely. So $|X_1| \leq C$ almost surely. Note that $|X_{n_k} - X| \leq \frac{\varepsilon}{3} + 2C |X_{n_k} - X| \geq \frac{\varepsilon}{3}$

There exists $N$ such that for all $n > N$, $P(|X_n - X| > \frac{\varepsilon}{3}) \leq \frac{\varepsilon}{4C}$

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

**6.2 Uniform Integrability**

**Lemma 6.2.1**

Let $X$ be an integrable random variable. Define

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

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

**Proof**

Suppose not. Then there exists $\varepsilon > 0$ and $A_n \in \mathcal{F}$ with $P(A_n) \leq \frac{\varepsilon}{n}$ and $E(|X|1_{A_n}) > \varepsilon$. By the First Borel-Cantelli Lemma,

$P(A_n \text{ infinitely often}) = 0$ (As infinitely often)

$\Rightarrow P(\cap_{n=1}^\infty A_n) = 0$. 

$\Rightarrow P(A_n \text{ infinitely often}) = 0$.
Then by dominated convergence (using (1x1))

\[ \varepsilon \leq E(1x1 \mathbb{1}_{\max A_n}) \rightarrow E(1x1 \mathbb{1}_{\text{an infin. set}}) = 0 \]  

Let \( X \) be a family of random variables. We say that \( X \) is bounded in \( L^p \) if \( \sup_{x \in X} \| x \|_p < \infty \). We say that \( X \) is uniformly integrable (UI) if \( X \) is bounded in \( L^1 \) and \( I_x(\delta) \rightarrow 0 \) as \( \delta \rightarrow 0 \). Here, \( I_x(\delta) = \sup \{ E(x^2) : A \in \mathcal{F}, P(A) \leq \delta, \ x \in X \} \)

Thus, for \( X \in L^1 \), \( \{ x \} \) is UI.

Suppose that \( X \) is bounded in \( L^p \) for some \( p > 1 \). By Hölder's inequality,

\[ E(1x1^A) \leq \| x \|_p \| 1^A \|_q = \| x \|_p P(A)^{\frac{1}{q}} \]  

(here \( p,q \) conjugate indices)

Hence, bounded \( X \) is uniformly integrable.

Example

On \([0,1]\) consider \( X_n = n \mathbb{1}_{[0,\frac{1}{n}]} \). Then the sequence

\( X = \{ X_n : n \in \mathbb{N} \} \) is bounded in \( L^1 \), but \( I_x(\delta) = 1 \) for all \( \delta > 0 \),

so \( X \) is not uniformly integrable.

Suppose that \( |x| \leq Y \) for all \( x \in X \), for some (integrable) \( Y \in L^1 \). Then \( I_x(\delta) \leq I_Y(\delta) \rightarrow 0 \) as \( \delta \rightarrow 0 \), so such a family is uniformly integrable.

Lemma 6.2.2

Let \( X \) be a family of random variables. Then \( X \) is UI

if and only if \( \sup_{x \in X} E(1x1^1x_k) \rightarrow 0 \) as \( k \rightarrow \infty \). (X)
Proof

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

\[ P(1_{X} > k) \leq \frac{\delta}{k} E(1_{X}) \leq \delta, \quad \Rightarrow \quad E(1_{X} 1_{1_{X} > k}) \leq \varepsilon. \]

Hence the condition holds.

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

So $I(1_{X}) \leq k_0 + 1$. Given $\varepsilon > 0$, note that for $A \in \mathcal{F}$ with $P(A) \leq \delta$, $E(1_{X} 1_{1_{X} \leq k_0}) \leq \delta$ for all $X \in \mathcal{F}$. Then for $\delta \leq \varepsilon/2$, we have $E(1_{X} 1_{A}) \leq k_0 P(A) + E(1_{X} 1_{1_{X} > k_0}) \leq \delta/2 + \delta/2 = \varepsilon.$

Hence $X$ is UI.

\[\]

Theorem 5.2.3

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

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

b) $X_n \to X$ in probability and $\sum_{n \in \mathbb{N}} P(1_{X_n \neq X}) < \infty$.

Proofs

Suppose that a) holds. By Chebyshev's Inequality, for all $\varepsilon > 0$
\[ P(\{x_n - x > \varepsilon\}) \leq \varepsilon \Rightarrow E(x_n - x) \to 0 \text{ as } n \to \infty. \]

Hence, \( x_n \to x \) in probability. Given \( \varepsilon > 0 \), there exists \( N \) such that for \( n > N \), \( E(x_n - x) \leq \varepsilon_2 \). Also, there exists \( \delta > 0 \) such that for \( n = 1, \ldots, N \) and any \( A \in \mathcal{F} \) with \( P(A) \leq \delta \),

\[ E(1_{x_n \leq 1 | A}) \leq \varepsilon, \text{ and } E(1_{x > 1 | A}) \leq \varepsilon_3 \]

But for \( n > N \) we have for each \( A \), \( 1_{x_n \leq 1} \leq 1_{x_n - x > x} \implies E(1_{x_n \leq 1 | A}) \leq E(1_{x_n - x} + E(1_{x > 1} | A)) \leq E \).

Hence, \( \{x_n : n \in \mathbb{N}\} \text{ is UI. so b) holds.} \)

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

\[ 1_{x_n - x > 1} \leq \varepsilon_4 + 2k \cdot 1_{x_n - x > \varepsilon_4} + 1_{x_n \leq k} + 1_{x > 1} \cdot 1_{x \geq k} \]

This is because either \( 1_{x_n - x > \varepsilon_4} \), or \( (1_{x_n - x > \varepsilon_4}, 1_{x_n \leq k}, 1_{x > k}) \)

or \( 1_{x_n > k} \) or \( 1_x > k \).

There exists \( k \) such that \( E(1_{x_n \leq 1, x_n \leq k}) \leq \varepsilon_4 \) and \( E(1_{x > 1, x > k}) \leq \varepsilon_4 \). Then there exists \( N \) such that for all \( n > N \), \( P(\{x_n - x > \varepsilon_4\}) \leq \delta \cdot \varepsilon \). Then for \( n > N \) we have \( E(x_n - x) \leq \varepsilon_4 + 2k \cdot \varepsilon + \varepsilon_4 + \varepsilon_4 = \varepsilon \)

(We should have noted that there is a subsequence \( \{x_n\} \) such that \( x_n \to x \) almost surely, so that \) \[ E(x) = E(\liminf_{k \to \infty} x_{n_k}) \leq \liminf_{k \to \infty} E(x_{n_k}) < \infty \text{ by UI.} \]
Probability and Measure

\[ |x_n - x| \leq \varepsilon + 2k \left( |x_n - x| + |x| \right) + |x_n - x| + 2k \]

\[ x_k = (-1)^k x \wedge k \]

\[ |x_k - y|^k \leq |x - y| N(k) \]

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

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

For \( f \in L^p \), \( f = u + iv \) say, for \( \lambda = \) Lebesgue Measure on \( \mathbb{R}^d \),

\[ \lambda(f) = \lambda(u) + i\lambda(v) \]

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

\[ = \lambda(\tilde{f}) = \lambda(\tilde{u}) \leq \lambda(4|f|) \quad |\tilde{u}| \leq |f| \]

For \( f \in L^p \), we define the Fourier transform \( \hat{f} : \mathbb{R}^d \to \mathbb{C} \) by

\[ \hat{f}(u) = \int_{\mathbb{R}^d} e^{i\langle u, x \rangle} f(x) \, dx, \quad u \in \mathbb{R}^d \]

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

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

Let \( f \in L^p \). If \( f \in L^1 \), we say that the Fourier inversion formula 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 \( f \in L^1 \).
If \( f \in L^2 \), we say the Plancherel identity holds for \( f \) if
\[
\| f \|_2 = (2\pi)^{d/2} \| \hat{f} \|_2. \]
We will show that this holds whenever \( f \in L^2 \).

Let \( \mu \) be a finite Borel measure on \( \mathbb{R}^d \). Define \( \hat{\mu} : \mathbb{R}^d \to \mathbb{C} \) by
\[
\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{i u \cdot x} \mu(dx), \quad \mu \in \mathbb{R}^d.
\]
If \( \mu \) 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 characteristic function \( \phi_X : \mathbb{R}^d \to \mathbb{C} \) by
\[
\phi_X(u) = \hat{\mu}_X(u) = \mathbb{E}(e^{i u \cdot X})
\]

7.2 Convolutions

Let \( f \in L^p \) and let \( \nu \) be a Borel probability measure on \( \mathbb{R}^d \).

We define the convolution \( f * \nu \) in \( L^p \) by
\[
f * \nu (x) = \int_{\mathbb{R}^d} f(x-y) \nu(dy) \quad \text{if} \quad f(x-\cdot) \in L^1(\nu)
\]
0 otherwise

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,
\[
\int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} |f(x-y)| \nu(dy) \right) dx \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x-y)| \nu(dy) dx
\]
\[
= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x)|^p dx \nu(dy) = \| f \|_p^p < \infty
\]
So \( f(x-\cdot) \in L^1(\nu) \) for almost all \( x \) and...
\[ \|f \ast \nu\|_p = \left( \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} f(x-y) \nu(dy) \right)^p dx \right)^{1/p} \leq \|f\|_p \]

If \( \nu \) has a density \( g \) then we write \( f \ast g \) for \( f \ast \nu \).

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

Thus, if \( \mu \) has a density \( f \in L^1(\mathbb{R}^d) \) then
\[
\mu \ast \nu(A) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1_A(x+y) f(x) dx \nu(dy)
= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1_A(x+y) f(x-y) dx \nu(dy) = \text{Fubini: Swap order of integration then bring } 1_A(x)
\]
outside the middle integral.

So \( \mu \ast \nu \) has density \( f \ast \nu \).

7.3 Gaussian

Consider for \( t \in (0, \infty) \) the centered Gaussian probability density function of variance \( t \) on \( \mathbb{R}^d \) given by
\( g_t(x) = (2\pi t)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2t}}, \quad x \in \mathbb{R}^d \). \( N(0, tI) \)

We will compute \( g_t \). 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 \( \Phi_Z \) is differentiable
(because \( Z \) is integrable and we can differentiate under \( \mathbb{E} \))
\[
\Phi_Z'(u) = \mathbb{E}(\text{ie}^{iuZ}) = \int_{\mathbb{R}} \text{ie}^{iux} \frac{1}{\sqrt{2\pi}} xe^{-\frac{x^2}{2}} dx = -u \Phi_Z(u)
\]
So \( \frac{d}{du} (e^{-\frac{u^2}{2}} \Phi_Z(u)) = 0 \), so \( \Phi_Z(u) = e^{-\frac{u^2}{2}} \Phi_Z(0) = e^{-\frac{u^2}{2}} \).
4.5 Approximation in $L^p$

Theorem 4.5.1
Let $\mathcal{A}$ be a $\tau$-system on $E$ generating $E$, 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

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

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 $A \in \mathcal{A}$, $\|1_A\|_p = \mu(A)^{\frac{1}{p}} < \infty$, 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 $D = \{ A \in E : 1_A \in V \}$
Then $A \subseteq D$, $1_D \in D$; if $A, B \in D$ with $A \subseteq B$, then
$$1_B - 1_A = 1_B - 1_A \in V$$
for $B \setminus A \in D$; for $A, B \in D$, $A \cap A^c \subseteq B \setminus A$, hence $D$ is a $\delta$-system, so $d = E$ by Dynkin's Lemma.

Since $V$ is a vector space it must contain every simple function.
For $f \in L^p$ with $\|f\|_p > 0$, consider the simple functions
\[ f_n = (2^{-n} L^{2^nf}) \wedge n, \text{ then } f_n \to f, \infty \]
\[ \|f - f_n\|_p \to 0, \text{ pointwise, } \infty \|f - f_n\|_p \to 0 \]
by dominated convergence, \( f \in V \).

Then \( V \) is a vector space.

Returning to the general case (where \( E \) may not be in \( \mathcal{A} \)), the above argument shows that \( f 1_{E_n} \to f \) for any \( f \in L^p \). Then
\[ \|f - f 1_{E_n}\|_p \to 0 \text{ pointwise. Hence } \|f - f 1_{E_n}\|_p \to 0 \]
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 Gaussian (continued)

We showed that
\[ \int_{\mathbb{R}} e^{imx - \frac{x^2}{12\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \, dx = e^{-\frac{u^2}{2\sigma^2}}, u \in \mathbb{R} \]

By a change of variable, \( x = \frac{y}{\sigma} \), we obtain
\[ \int_{\mathbb{R}} e^{i\frac{uy}{\sigma^2} - \frac{y^2}{2\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \, dy = e^{-\frac{u^2}{2\sigma^2}} \]

Then by Fubini, for \( g_t(x) = (2\pi\sigma^2)^{-\frac{d}{2}} e^{-\frac{x^2}{2\sigma^2}} \)
\[ g_t(u) = \int_{\mathbb{R}^d} e^{-i \langle u, x \rangle} g_t(x) \, dx = \int_{\mathbb{R}^d} e^{-i \frac{uy}{\sigma^2} + \frac{y^2}{2\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \, dy \]
\[ = \int_{\mathbb{R}^d} e^{-\frac{u^2}{2\sigma^2}} = e^{-\frac{u^2}{2\sigma^2}} \]

So \( g_{\sigma^2} = (2\pi)^{\frac{d}{2}} e^{-\frac{\sigma^2}{2}} g_{\frac{1}{\sigma^2}} \), so \( g_{\sigma^2} = (2\pi)^{\frac{d}{2}} g_{\sigma^2} \).

So \( g_t(x) = g_{\sigma^2}(-x) = (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} e^{-i \langle u, x \rangle} g_{\sigma^2}(u) \, dt \)

Hence the Fourier Inversion Formula holds for \( g_t \).
7.4 Gaussian Convolutions

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

Note that $f * g_t (u) = \hat{f}(u) e^{-i\omega t}$.

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

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

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

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

$\|f * g_t\|_{L^\infty} \leq \|f\|_{L^\infty} \|g_t\|_{L^\infty} \leq \|f\|_1$.

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,

$$(2\pi)^d f * g_t (x) = (2\pi)^d \int_{\mathbb{R}^d} f(x-y) g_t(y) \, dy = \int_{\mathbb{R}^d} f(x-y) e^{-i\omega y} \hat{g_t}(u) \, du \, dy$$

$$= \int_{\mathbb{R}^d} \hat{f}(x-y) \int_{\mathbb{R}^d} e^{-i\omega y} \hat{g_t}(u) \, du \, dy$$

$$= \int_{\mathbb{R}^d} \hat{g_t}(u) \int_{\mathbb{R}^d} f(x-y) e^{i\omega y} \, dy \, du$$

$(\text{change of variable } x-y \rightarrow y)$

$$= \int_{\mathbb{R}^d} \hat{g_t}(u) \int_{\mathbb{R}^d} f(x) e^{-i\omega x} \, dx \, du$$

$$= \int_{\mathbb{R}^d} \hat{g_t}(u) \int_{\mathbb{R}^d} f(x) e^{-i\omega x} \, dx \, du = \int_{\mathbb{R}^d} \hat{f}(u) e^{-i\omega x} \, du$$
Lemma 7.4.2

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

Proofs

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

$$\|f^* g_t - h^* g_t\|_p \leq \|f - h\|_p \leq \frac{\varepsilon}{3}.$$  

Consider

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

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  

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

by Jensen's inequality in a probability measure $\mu$

$$\int_{\mathbb{R}} g_t(y) e(y) \, dy = \int_{\mathbb{R}} g_t(y) e(1E_y) \, dy \to 0$$

Hence

$$\|f - f^* g_t\|_p \leq \|f - h\|_p + \|h - h^* g_t\|_p + \|h^* g_t - f^* g_t\|_p \leq 3\|f - h\|_p \leq \varepsilon$$

for sufficiently large $t$.

\[ \square \]

7.5 Uniqueness and Inversion

Theorem 7.5.1

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

$$f_t(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{-i \langle u, x \rangle} f(u) e^{-\frac{|u|^2}{2t}} \, du.$$  

Then $f_t \to f$ in $L^1$. Moreover, if $f \in L^1$, the Fourier Inversion Formula holds.
Since \( f * g_t(u) = \hat{f}(u) e^{-i u t / 2} \), by Lemma 7.4.1, we have \( f_t = f * g_t \). Hence \( f_t \rightarrow f \) in \( L^1' \) by Lemma 7.4.2.

Then if \( f \in L^1' \), we can apply dominated convergence with \( |f| \) as dominating function to see that
\[
\hat{f}_t(x) \rightarrow \int_{\mathbb{R}^d} e^{-i \langle u, x \rangle} \hat{f}(u) \, du \quad \text{as} \quad t \rightarrow 0.
\]

But there is a sequence \( t_n \rightarrow 0 \) such that \( f_{t_n} \rightarrow f \) a.e.

Hence the inversion formula holds for \( f \). \( \square \)
Theorem 7.6.1

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

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

Moreover, there is a unique Hilbert space automorphism \( F \) on \( L^2 \) such that for all \( f \in L^1 \cap L^2 (\mathbb{R}^d) \),

\[
F[f] = \left[(2\pi)^{-d/2} \hat{f}\right]
\]

(Thus \( F : L^2 \to L^2 \) is a linear bijection with \( \| Ff \| = \| f \| \forall f \in L^2 \)).

Proof

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

\[
(2\pi)^d \| \hat{f} \|_2^2 = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{-i(x,y)} f(x) \hat{f}(y) \, dx \, dy
\]

\[
= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{-i(x,y)} f(x) \, dx \, \hat{f}(y) \, dy \quad \text{by the inversion formula}
\]

\[
= \int_{\mathbb{R}^d} \hat{f}(w) \int_{\mathbb{R}^d} e^{-i(x,y)} f(x) \, dx \, dy \quad \text{by Fubini, since } (x,y) \mapsto \hat{f}(w)\text{ is integrable}
\]

\[
= \int_{\mathbb{R}^d} \hat{f}(w) \int_{\mathbb{R}^d} e^{-i(x,y)} f(x) \, dx \, dw
\]

\[
= \int_{\mathbb{R}^d} \hat{f}(w) F(w) \, dw = \| \hat{f} \|_2^2
\]

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 \( L^1 \), so \( \| f_t \|_2 \to \| f \|_2 \) as \( t \to 0 \).

Now \( \hat{f_t}(w) = \hat{f}(w) e^{-|w|^2 t} \), so \( \| \hat{f_t} \|_2 \to \| f \|_2 \) as \( t \to 0 \).

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

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

Define \( F_0 \) on \( L^1 \cap L^2 \) by \( F_0[f] = \left[(2\pi)^{-d/2} \hat{f}\right] \) for \( f \in L^1 \cap L^2 \).
Then $F_0$ is a well-defined linear map $L^2 \rightarrow L^2$ with $\|F_0 h\|_2 = \|h\|_2$ for all $h \in L^1 \cap L^2$. Now $L^1 \cap L^2$ is dense in $L^2$ (it contains all step functions), so $F_0$ extends uniquely to a linear isometry from $L^2$ into $L^2$.

It remains to show that $F$ is injective, $L^2 \rightarrow L^2$. Then $V \subseteq L^1 \cap L^2$. Consider $V = \{ Ff : f \in L^1 \}$ with $F \in L^1 \cap L^2$. Then $V$ is dense in $L^2$ (it contains all Gaussian convolutions) and by the inversion formula $F^*h = h$. So $F$ is onto $V$ and hence $F$ is onto $L^2$ as required.

7.7 Weak Convergence and Characteristic Functions

Let $\mu_n, (n \in \mathbb{N})$ be Borel probability measures on $\mathbb{R}^d$.

We say that $\mu_n \rightharpoonup \mu$ weakly if $\mu_n (f) \rightarrow \mu (f)$ 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-supported) or even to $C_c^1 (\mathbb{R}^d)$ without affecting whether $\mu_n (f) \rightarrow \mu (f)$ for all $f$.

For random variables $X, (X_n : n \in \mathbb{N})$ in $\mathbb{R}^d$, say $X_n \rightharpoonup X$ weakly if $X_n \rightarrow X$ weakly (in the sense of probability 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 $\mu_X$ of $X$ is uniquely determined by its characteristic function $\Phi_X (t) = \mathbb{E} e^{it^T X}$.
Moreover, if \( \Phi_x \) is integrable, then \( X \) has a continuous bounded density (with respect to \( dx \)) given by
\[
\Phi_x(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i<x,u>} \Phi_x(u) \, du, \quad x \in \mathbb{R}^d.
\]
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.

**Proof**

Let \( Z \) be a random variable in \( \mathbb{R}^d \) independent of \( X \) with density \( g \) (that is, \( Z \sim N(0, I) \)). Then \( \Phi X \) has density \( g_0 \) and
\[
X + \Phi X \text{ has density function } \hat{f}_z = \mu_x \ast g_0.
\]
Now \( f_z \in \mathcal{L}' \) and \( \hat{f}_z = \Phi_x \hat{g}_x \), \( \hat{g}_x(u) = e^{-i\langle u, x \rangle - \frac{1}{2} \| x \|^2} \), \( \| \hat{g}_x \|_2 \leq 1 \), so \( \hat{f}_z \in \mathcal{L}' \).

So by inversion
\[
\Phi_x(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i<x,u>} e^{-\frac{1}{2} \| x \|^2} \Phi_x(u) \, du \quad (\ast)
\]

So \( \Phi_x \) determines the distribution of \( X + \Phi X \) uniquely. By bounded convergence, for all \( g \in C_b(\mathbb{R}^d) \)
\[
\mu_x(g) = E(g(X)) = \lim_{\varepsilon \to 0} E\left( g\left(X + \varepsilon \Phi X\right)\right)
\]
So \( \Phi_x \) determines \( \mu_x(g) \) uniquely and hence \( \mu_x \).

Suppose now that \( \Phi_x \in \mathcal{L}' \). We use dominated convergence with dominating function \( \| \Phi_x \|_1 \) to see that
\[
\hat{f}_z(x) \to \Phi_x(x) \text{ for all } x \in \mathbb{R}^d, \quad \text{and } \| \hat{f}_z(x) \|_2 \leq (2\pi)^{-d} \| \Phi_x \|_2.
\]

Then for any \( g \in C_c(\mathbb{R}^d) \),
\[
E(g(X)) = \lim_{\varepsilon \to 0} E\left( g\left(X + \varepsilon \Phi X\right)\right)
= \int g(y) \hat{f}_z(y) \, dy \to \int g(y) \Phi_x(x) \, dx \quad \text{by bounded convergence.}
\]
Note that \( f_\epsilon(x) = \lim_{\epsilon \to 0} f_\epsilon(x) \geq 0 \) \( \forall x \in \mathbb{R}^d \). It follows that 
\[ m_\epsilon(dx) = f_\epsilon(x) \, dx \quad \text{(exercise)} \]

Suppose now that \( g \in C_c^1(\mathbb{R}^d) \) and \( \Phi_{x_n}(u) \to \Phi_x(u) \) \( \forall u \in \mathbb{R}^d \).

\[ |g(x) - g(x + \epsilon z)| \leq ||\nabla g||_{\infty} \epsilon |z| \quad \text{(by the Mean Value Theorem)} \]

so \[ |E(g(x)) - E(g(x + \epsilon z))| \leq ||\nabla g||_{\infty} \epsilon E(|z|) < \infty \]

Similarly, \[ |E(g(x_n)) - E(g(x_n + \epsilon z))| \leq ||\nabla g||_{\infty} \epsilon E(|z|) \]

Now, \[ E(g(x_n + \epsilon z)) = \int_{\mathbb{R}^d} g(x) \int_{\mathbb{R}^d} e^{-i \langle x, u \rangle} e^{-i \mu z / 2} \Phi_{x_n}(u) \, du \, dx \]

\[ \to \int_{\mathbb{R}^d} g(x) \int_{\mathbb{R}^d} e^{-i \langle x, u \rangle} e^{-i \mu z / 2} \Phi_x(u) \, du \, dx = (2\pi)^d E(g(x + \epsilon z)) \]

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

Given \( \epsilon > 0 \), choose \( \mu > 0 \) so that \( ||\nabla g||_{\infty} \epsilon E(|z|) \leq \epsilon^2 / 3 \).

Then for sufficiently large \( n \), we have

\[ |E(g(x)) - E(g(x_n))| \leq \epsilon \]

\( \square \)
Lemma 7.7.2

Let $\mu$ be a Borel probability measure on $\mathbb{R}^d$ and let $(\mu_n : n \in \mathbb{N})$ be a sequence of such measures. Suppose $\mu_n (A) \to \mu (A)$ for all $C^\infty_c (\mathbb{R}^d)$. Then $\mu_n \rightharpoonup \mu$ weakly on $\mathbb{R}^d$ (i.e., $\mu_n (A) \to \mu (A)$ for all $f \in C_b (\mathbb{R}^d)$).

Proof:

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

$$
\rho (x) = \frac{c}{x} \exp \left[ - (1 - |x|^2)^{1/3} \right] \quad \text{if } |x| < 1
$$

$$
= 0 \quad \text{if } |x| \geq 1
$$

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

Set $\rho_\varepsilon (x) = \varepsilon^{-d} \rho (\varepsilon x)$ for $\varepsilon \in (0, \infty)$ ("$\varepsilon = \varepsilon^{-1}$") and set $f_\varepsilon = f \ast \rho_\varepsilon$. Then $f_\varepsilon \in C^\infty_c (\mathbb{R}^d)$ and

$$
\|f_\varepsilon - f\|_\infty \to 0 \quad \text{as } \varepsilon \to 0.
$$

Now $\mu_n (f_\varepsilon) \to \mu (f_\varepsilon)$ as $n \to \infty$ and $|\mu_n (f_\varepsilon) - \mu (f_\varepsilon)| \leq \|f_\varepsilon - f\|_\infty$ as $\varepsilon \to 0$. So $\mu_n (f) \rightharpoonup \mu (f)$ as $n \to \infty$.

Now suppose $f \in C_b (\mathbb{R}^d)$. Consider $g (x) = \varepsilon \sqrt{(R^2 - |x|^2)} 1_{N_1} (x)$. 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 $R > 0$ such that

$$
\mu (1 - g) < \frac{\varepsilon}{3},
$$

so $\mu_n (1 - g) \leq \frac{\varepsilon}{3} \|f\|_\infty$, for all sufficiently large $n$. Hence $|\mu_n (E) - \mu (E)| = |\mu_n (E \setminus g) + 1 \mu_n (f ) - \mu (f) - \mu (E \setminus g) + \mu (f )|$

$$
+ |\mu (f - g) + |\mu (f) - \mu (g)|
$$
\[ \leq \|f\|_{\infty} \mu_n(-g) + 1 \cdot 1 + \|f\|_{\infty} \mu(1-g) \]
\[ \leq \frac{\epsilon_3 + \epsilon_3 + \epsilon_3}{3} \text{ for all sufficiently large } n \]
(because \( f g \in C_c(\mathbb{R}^d) \))

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 \( f_X(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \).

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

b) \( \var(X) = \sigma^2 \),

c) \( \phi_X(u) = \mathbb{E}(e^{iuX}) = e^{i\mu u - \frac{1}{2} \sigma^2 u^2} \)

d) For \( a, b \in \mathbb{R} \), \( aX + b \sim N(\mu a + b, \sigma^2 a^2) \) (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 \( Y_1, \ldots, Y_d \) independent \( N(0,1) \). Set \( Y = (Y_1, \ldots, Y_d) \).

Then for \( u \in \mathbb{R}^d \), \( \mathbb{E}(e^{i\langle u, Y \rangle}) = \prod_{k=1}^d \mathbb{E}(e^{i u_k X_k}) = \prod_{k=1}^d e^{-\frac{1}{2} u_k^2} = e^{-\frac{1}{2} u^2} \).

So \( \langle u, Y \rangle \sim N(0, u^2) \) by uniqueness of characteristic function.
Theorem 8.2.1

Let $X$ be a Gaussian random variable in $\mathbb{R}^d$. Then $X \in L^2(\mathbb{P})$

Set $E(X) = \mu \in \mathbb{R}^d$, $V = \text{var}(X) = (\text{cov}(X_i, X_k))_{i,k} \in \mathbb{R}^{d \times d}$.

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 $\mu$ and $V$ (we write $X \sim \mathcal{N}(\mu, V)$).

c) If $A$ is a $d \times d$ 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

$$f_X(x) = \frac{1}{(2\pi)^{\frac{d}{2}} (\det V)^{\frac{1}{2}}} e^{-\frac{1}{2} \langle x - \mu, V^{-1}(x - \mu) \rangle}, x \in \mathbb{R}^d$$

e) If $X = (x_1, x_2)$ with $x_1 \in \mathbb{R}^{d_1}$, $x_2 \in \mathbb{R}^{d_2}$, and if

$\text{cov}(x_1, x_2) = 0$, then $x_1, x_2$ are independent.

Proof:

Each component $X_k \in L^2(\mathbb{P})$, so $X$ is too. Note that

$E(\langle u, X \rangle) = \langle u, \mu \rangle$, $\text{var}(\langle u, X \rangle) = \langle u, Vu \rangle$

Now $\langle u, X \rangle$ is Gaussian so $E(e^{i\langle u, X \rangle}) = e^{i\langle u, \mu \rangle - \frac{1}{2} \langle u, Vu \rangle}$

Hence a) and b) hold by uniqueness of characteristic functions.

c) is left as an exercise.

Consider $Y = (Y_1, \ldots, Y_d)$ as above, and since $V$ is non-negative

definite, there exists a symmetric $d \times d$ matrix $\Sigma$ such that

$$\Sigma^2 = V \quad (V = U \text{diag}(\lambda_k) U^*, \Sigma = U \text{diag}(1^2) U^*)$$

Consider $\tilde{X} = \Sigma Y + \mu$. Then $\tilde{X}$ is Gaussian and

$E(\tilde{X}) = \mu$, $\text{var}(\tilde{X}) = V$, so $\tilde{X}$ has the same distribution as $X$. 

3
Then, by a linear change of variables in $\mathbb{R}^d$, $X$ has density

$$f_X(x) = f_X(x) = (2\pi)^{-d/2} (\det V)^{-1/2} e^{-\frac{1}{2} \langle x - \mu, V^{-1}(x - \mu) \rangle}$$

Finally, if $\text{cov}(X_1, X_2) = 0$, then $V = \begin{pmatrix} V_{11} & 0 \\ 0 & V_{22} \end{pmatrix}$

$$\phi_X(u) = e^{-\frac{1}{2} \langle u, V^{-1} u \rangle} = \phi_{X_1} \phi_{X_2}, \quad u = (u_1, u_2) \in \mathbb{R}^d$$

### 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 $\mu = \mathbb{E}(X)$, $\sigma^2 = \text{Var}(X)$. Assume that $\sigma^2 > 0$. Set $S_n = X_1 + \ldots + X_n$

$$Z_n = \frac{S_n - n\mu}{\sigma \sqrt{n}}$$

Then $Z_n \Rightarrow Z$ weakly in $\mathbb{R}$ where $Z \sim \mathcal{N}(0, 1)$ (that is, $\mathbb{P}(Z_n \leq a) \to \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx$ as $n \to \infty$)

**Proof**

Set $\phi(u) = \mathbb{E}(e^{iuX})$ and $\phi_n(u) = \mathbb{E}(e^{iuZ_n})$. Since $X_1 \in L^2$, we can differentiate $\phi$ twice under $\mathbb{E}$ and obtain

$\phi(0) = 1$, $\phi'(0) = \mu$, $\phi''(0) = 1$ 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

$$\phi_n(u) = \left( \phi \left( \frac{u}{\sqrt{n}} \right) \right)^n = \left( 1 - \frac{u^2}{2n} + o \left( \frac{u^2}{n} \right) \right)^n \quad \text{as} \quad \frac{u^2}{n} \to 0$$

The complex logarithm satisfies $\log(1 + Z) = Z + o(1|Z|)$ as $|Z| \to 0$.

$\log \phi_n(u) = n \log (1 - \frac{u^2}{2n} + o \left( \frac{u^2}{n} \right)) = -\frac{u^2}{2} + o(1) \quad \text{as} \quad n \to \infty$

So $\phi_n(u) \equiv e^{-u^2/2} \quad \text{as} \quad n \to \infty$ for all $u \in \mathbb{R}$. But $e^{-u^2/2} = \phi(u)$

$Z_n \Rightarrow Z$ 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.
9. Ergodic Theory: Definition

Let \((E, \mathcal{E}, \mu)\) be a measure space. A measurable function \(\Theta: E \to E\) is a measure preserving transformation 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 invariant if \(\Theta^{-1}(A) = A\). The set \(\mathcal{E}_0\) of all invariant sets is a \(\sigma\)-algebra. A measurable function \(F\) on \(E\) is invariant if \(F = F \circ \Theta\). Then (exercise) \(F\) is invariant if and only if \(F\) is \(\mathcal{E}_0\) measurable.

We say that \(\Theta\) is ergodic if \(\mathcal{E}_0\) contains only sets of measure 0 and their complements (we say that \(\mathcal{E}_0\) is trivial).

Examples

i) Shift Map on the torus. Fix \(a \in E\).

\[E = (0, 1)^d, \quad E = \mathcal{B}(E), \quad \mu = \text{Lebesgue}\]

Define \(\Theta: E \to E\) by \(\Theta(x_1, \ldots, x_d) = (x_1 + a, \ldots, x_1 + ad)\)

It is easy to check that for \(A = (b_1, c_1] \times \ldots \times (b_d, c_d]\) we have \(\mu(\Theta^{-1}(A)) = \mu(A)\). So \(\mu\) and \(\mu \circ \Theta^{-1}\) agree on the \(\sigma\)-system of such sets, which generates \(E\). So \(\Theta\) is a measure preserving transformation.

ii) Bakers Map. \(E = (0, 1], \mathcal{B}(E), \text{Lebesgue}\)

\[\Theta(x) = 2x - \lfloor 2x \rfloor\]

It is easy to see that \(\mu(\Theta^{-1}(a, b]) = b - a\) so \(\mu\) is a measure preserving transformation by the same \(\sigma\)-system argument.
Proposition 9.1.1
Let $f$ be an integrable function, $\Theta$ 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 $\Theta$ is ergodic. Then there exists a constant $c \in \mathbb{R}$ such that $f = c$ almost everywhere.

Idea: $\{x \in E : f(x) \leq y\} \in \mathcal{E}_a$, so $\mu(f^{-1}y) = 0$ 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 $S_n = \sum_{k=0}^{n-1} f \circ \Theta^k$.

Thus for $x \in E$, $S_n(x) = \frac{1}{n} (f(x) + f(\Theta(x)) + \ldots + f(\Theta^{n-1}(x)))$ which is the empirical average of $f$ along the point orbit of $\Theta$ from $x$.

\[ Birkhoff's \ Theorem \]
\[ \sigma \text{-finite } \mu, \Theta \text{ a measure preserving transformation} \]
\[ \Rightarrow S_n \text{ converges almost everywhere.} \]

9.3.1 Lemma (Maximal Ergodic Lemma)
Let $\Theta$ be a measure-preserving transformation, and $F$ be an integrable function. Define $S^*_n(x) = \sup_{n \geq 0} S_n(x)$.

Then $\int_{\{S_n^* > 0\}} f d\mu \geq 0$.

Proof
Consider $S_n^* = \max_{0 \leq m \leq n} S_m$ and set $A_n = \{S_n^* > 0\}$.
For \(m = 1, \ldots, n\), 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 \(A_n \cap \{S_n^{*} > 0\}\) as \(n \to \infty\) so by dominated convergence with dominating function \(1f_1\).

\[
\int_{\{S_n^{*} > 0\}} f \, d\mu = \lim_{n \to \infty} \int_{A_n} f \, d\mu \geq 0
\]

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 \(E(X_n) = 0\), \(E(X_n^4) \leq M\) for all \(n\) for some \(M < \alpha_4\). Set \(S_n = X_1 + \ldots + X_n\). Then \(\frac{S_n}{n} \to 0\) almost surely.

**Proof**

Observe that

\[
E(S_n^4) = \sum_{k \leq n} E(X_k^4) + \binom{4}{2} \sum_{\substack{i < j < k \leq n}} E(X_i^2 X_j^2)
\]

because \(E(X_i X_j X_k X_l) = E(X_i X_j^3) = E(X_i X_j X_k^2)\) for all distinct indices \(i, j, k, l\).
Now \( E\left(x_i^2x_k^2\right) \leq E\left(x_i^4\right)^{\frac{1}{2}} E\left(x_k^4\right)^{\frac{1}{2}} \leq M \)

So \( E(S_n^4) \leq nM + \frac{6}{n} \sqrt{n} (n-1)M \leq 3n^2M \)

So \( E\left(\sum_{\lambda=1}^{\infty} \left(\frac{S_n}{n}\right)^4\right) = \sum_{\lambda=1}^{\infty} E\left(\left(\frac{S_n}{n}\right)^4\right) \leq \sum_{\lambda=1}^{\infty} n^{-4} 3n^2M \)

\[ = 3M \sum_{\lambda=1}^{\infty} \frac{1}{\lambda^2} < \infty \]

So \( \sum_{\lambda=1}^{\infty} \left(\frac{S_n}{n}\right)^4 < \infty \) almost surely \( \Rightarrow \frac{S_n}{n} \rightarrow 0 \) almost surely
Recap

\((E, E, \mu)\) a measure space. We say that a measurable function \(\Theta : E \to E\) is a measure-preserving transformation if
\[\mu = \mu \circ \Theta^{-1}.
\]

We say that a measurable function \(f : E \to \mathbb{R}\) is invariant if
\[f = f \circ \Theta^{n}.
\]

We are interested in the averages of functions on the orbits of \(\Theta\).

\[S_n(x) = S_n^+(x) = f(x) + f(\Theta(x)) + \ldots + f(\Theta^{n-1}(x))\]

**Maximal Ergodic Lemma**

Let \(\Theta\) be a measure-preserving transformation and let \(f\) be an integrable function on \(E\). Define \(S^* = \sup_n S_n\).

Then
\[\int_{\{s^* > 0\}} f \, d\mu \geq 0.
\]

**Theorem 9.3.2** (Birkhoff's almost-everywhere ergodic theorem)

Assume that \(\mu\) is \(\sigma\)-finite. Let \(\Theta\) be a measure-preserving transformation and let \(f\) be an integrable function on \(E\). Then there exists an invariant function \(F\) on \(E\) such that
\[\mu(|F|) \leq \mu(|f|)\text{ and } \frac{S_n(x)}{n} \to F(x)\text{ almost everywhere.}
\]

**Proof**

Consider for \(a, b \in \mathbb{R}\), with \(a < b\),
\[D = D(a, b) = \left\{ \liminf_{n \to \infty} \frac{S_n}{n} < a, \quad b < \limsup_{n \to \infty} \frac{S_n}{n} \right\}
\]

The functions \(\liminf_{n \to \infty} \frac{S_n}{n}\), \(\limsup_{n \to \infty} \frac{S_n}{n}\) are invariant.
\[ S_{n+1} = S_n \Theta + f \]  
So D is invariant
\[ \frac{S_{n+1}}{n+1} = \frac{1}{n+1} \left( \frac{S_n}{n} \right) \Theta + \frac{f}{n+1} \]  
Since \( \lim_{n \to \infty} S_n/n \), \( \limsup \frac{S_n}{n} = \left( \limsup \frac{S_n}{n} \right) \Theta \) \( \liminf \frac{S_n}{n} \) are.

We shall show that \( \mu(D) = 0 \).

By restricting \( \mu, \Theta, f \) 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 \( B \in E \) with \( \mu(B) < \infty \) and consider \( g = f - b \chi_B \).
Then \( S_n^g(x) = S_n^f(x) - b S_n^{\chi_B}(x) \geq S_n^f(x) - nb > 0 \) for some \( n \in \mathbb{N} \) for all \( x \). That is, \( S_n^g > 0 \) 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 \( B_n \in E \) 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 \( b > a \).
Consider $\Delta = \{ \liminf_{n \to \infty} \frac{S_n}{n} < \limsup_{n \to \infty} \frac{S_n}{n} \}$. Then

$$\Delta = \bigcup_{a,b \in \mathbb{Q}} \mathcal{D}(a,b),$$
so $\Delta$ is invariant and $\mu(\Delta) = 0$.

So we can define an invariant function $\tilde{F} : E \to [-\infty, \infty]$ by

$$\tilde{F}(x) = \begin{cases} 
\lim_{n \to \infty} \frac{S_n}{n}(x) & x \in \Delta \\
0 & x \notin \Delta 
\end{cases}$$

Note that $\mu(\{1\} \circ \Theta^n) = \mu(\{1\})$ so

$$\mu(\{S_n\}) \leq \sum_{k=0}^{n-1} \mu(\{1\} \circ \Theta^k) \leq n \mu(\{1\})$$

By Fatou's Lemma, $\mu(\{\tilde{F}\}) = \mu(\liminf_{n \to \infty} \frac{S_n}{n}) \leq \liminf_{n \to \infty} \mu(\frac{S_n}{n})$.

Hence $\tilde{F}(x) \in \mathbb{R}$ for $\mu$-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, $\mu(\{1\}) = \mu(\{\tilde{F}\}) \leq \mu(\{1\})$ and $\frac{S_n}{n} \to \bar{F}$ almost everywhere.

**Theorem 9.3.3 (von Neumann's $L^p$-ergodic theorem)**

Assume $\mu(E) < \infty$. Let $p \in [1, \infty)$. Let $\Theta$ be a measure-preserving transformation and let $f \in L^p$. Then

$$\frac{S_n}{n} \to \bar{F}.$$

**Proof**

We have $\mu(\{1\} \circ \Theta^n) = \mu(\{1\})$. So by Mikowski's Inequality,

$$\|S_n\|_p \leq \sum_{k=0}^{n-1} \|1\|_p \circ \Theta^k \|_p = n \|1\|_p$$

Hence $\|\frac{S_n}{n}\|_p \leq \|1\|_p$. By Fatou's Lemma,

$$\mu(\{\frac{S_n}{n}\}) = \mu(\liminf_{n \to \infty} \frac{S_n}{n}) \leq \liminf_{n \to \infty} \mu(\frac{S_n}{n})$$
\[
\lim_{n \to \infty} \mu \left( \frac{S_n}{n} \right) \leq \mu \left( \frac{1}{n} \right), \quad \text{so } \|F\|_p \leq \|F\|_p
\]

Fix \( k \in (0, \infty) \), and consider \( g = (-k) \cdot 1_{\{|H| \geq k\}} \). By Dominated Convergence (using \( \|F\|_p \) as dominating function)
\[
\mu \left( \frac{1}{n} \cdot g \right) \leq \mu \left( \frac{1}{n} \cdot 1_{|H| > k} \right) \to 0 \quad \text{as } k \to \infty.
\]

Given \( \varepsilon > 0 \), choose \( k \) so that \( \|F - g\|_p \leq \varepsilon/3 \). Note that \( f \in L^p(\mu) \) and \( \mu \) is a finite measure.

Now \( S_n^+ - S_n^- = S_n^+ - g \) and taking limits, \( f - g = f - g \)

by Birkhoff's Theorem.
\[
\|S_n^+ - S_n^-\|_p = \|S_n^+ - g\|_p \leq \varepsilon/3,
\]
\[
\|F - g\|_p = \|F - g\|_p \leq \varepsilon/3.
\]

Now \( S_n^+ \to g \) almost everywhere and \( \|S_n^-\|_p \leq k \), so by bounded convergence, \( \mu \left( \frac{1}{n} \cdot g \right) \to 0 \quad \text{as } n \to \infty. \)

So we can choose \( N \) so that \( \forall n \geq N, \)
\[
\|S_n^+ - g\|_p \leq \varepsilon/3
\]

So by Minkowski's Inequality,
\[
\|S_n^+ - F\|_p \leq \|S_n^+ - S_n^-\|_p + \|S_n^- - g\|_p + \|g - F\|_p \leq \varepsilon
\]

\( \square \)
9.2 Bernoulli Shifts

Let $\mathcal{M}$ 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 $(\Omega, \mathcal{F}, P)$ all having distribution $\mathcal{M}$.

Consider $E = \mathbb{R}^N = \{ x = (x_n : n \in \mathbb{N}) : x_n \in \mathbb{R} \text{ for all } n \}$

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 $\mathcal{A} = \bigwedge\{ A_n : A_n \in \mathcal{B}, A_n = \mathbb{R} \text{ for all but finitely many } n \}$

Note that $\mathcal{A}$ is a $\sigma$-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 $\mu$ on $(E, \mathcal{E})$ by $\mu = P \circ Y^{-1}$

Thus $\mu$ is the law of $Y$.

Call $(E, \mathcal{E}, \mu)$ the canonical model for i.i.d. sequences with distribution $\mathcal{M}$. Consider the shift map $\Theta : E \to E$ by

$\Theta(x_1, x_2, \ldots) = (x_2, x_3, \ldots)$

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 $\mathcal{M}$.

For $A = \bigwedge A_n \in \mathcal{A}, \quad \Theta^{-1}(A) = \mathbb{R} \times A_1 \times A_2 \times \ldots \in \mathcal{A}$.

and $\mu(A) = \bigwedge m(A_n) = \mu(\Theta^{-1}(A))$
Since \( A \) generates \( \mathcal{E} \), this shows that \( \Theta \) 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 \( Y_n = \sigma(X_{n+k} : k \in \mathbb{N}), Y = \bigcap_n Y_n \).

For \( A = \bigcap_n A_n \in A \), \( \Theta^{-n}(A) = \{ x_{nk} \in A_k \text{ for all } k \in \mathbb{N} \} \in Y \).

Since \( Y_n \) is a \( \sigma \)-algebra, and \( A \) generates \( \mathcal{E} \), this implies that \( \Theta^{-n}(A) \in Y_n \) for all \( A \in \mathcal{E} \).

Suppose that \( A \in \mathcal{E}_0 \) (\( A \) is invariant), thus \( A = \Theta^{-1}(A) = \Theta^{-n}(A) \in Y \).

Hence \( A \in Y \). But by Kolmogorov's zero-one law, for all \( A \in Y \), \( \mu(A) \in \{0, 1\} \). \( \square \)

**10.2 Strong Law of Large Numbers**

**Theorem 10.2.1**

\[ S_n(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x) \]

Assume that \( \int_{\mathbb{R}} |x| \mu(dx) < \infty \), and \( \int_{\mathbb{R}} x \mu(dx) = \bar{\mu} \).

Then \( \mu(\{x \in E : \frac{x_1 + \ldots + x_n}{n} \to \bar{\mu} \text{ as } n \to \infty \}) = 1 \).

**Proof**

We apply consider the function \( X_i \) on \( (E, \mathcal{E}, \mu) \). Then \( X_i \) is integrable \( \mu(|X_i|) = \int_{\mathbb{R}} |x| \mu(dx) < \infty \) and \( \mu(X_i) = \int_{\mathbb{R}} x \mu(dx) = \bar{\mu} \).

Consider the shift map \( \Theta \) m.i.p.t.

Note that \( S_n(x) = x_1 + \ldots + x_n \). By Birkhoff's Theorem, there is an invariant function \( \bar{X} \), such that \( \frac{S_n}{n}(x) \to \bar{X}(x) \) for \( \mu \)-almost-all \( x \).
Since $\Theta$ is ergodic, there exists a constant $c \in \mathbb{R}$ such that $\bar{X}_i = c$ almost everywhere. But $\frac{S_n}{n} \Rightarrow \bar{X}_i$ in $L^1$

(von Neumann ergodic theorem)

So $c = \mu(\bar{X}_i) = \lim_{n \to \infty} \mu\left(\frac{S_n}{n}\right) = \mu(X_i) = \nu \Rightarrow \Omega$

Theorem 10.2.2 (Strong Law of Large Numbers)

Let $(X_n : n \in \mathbb{N})$ be a sequence of integrable, i.i.d. r.v.s having mean $\nu$. Set $S = X_1 + \ldots + X_n$. Then $\frac{S_n}{n} \Rightarrow \nu$ almost surely.

Proof.

Observe $\left\{ \frac{S_n}{n} \Rightarrow \nu \right\} = \left\{ Y \in E_0 \right\}$ where $E_0 = \left\{ x \in E : \frac{x_1 + \ldots + x_n}{n} \Rightarrow \nu \right\}$. So $P\left( \frac{S_n}{n} \Rightarrow \nu \right) = \mu(E_0) = 0$

Convergence of $X_n \Rightarrow X$

A subsequence

Almost Everywhere $\Rightarrow$ In Probability $\Rightarrow$ In Distribution

In $L^p (p>1) \Rightarrow$ In $L^1$

Weak Convergence $\Rightarrow$ Characteristic Function

We can return from $X_n \Rightarrow X$ in distribution to $X_1 \Rightarrow X$ a.e. by changing the probability space.

$X_1, X+E_1, \ldots, X+E_n, \quad |E_i| \leq \infty$ i.i.d. $\Rightarrow E E_n = 0$

$X_n = \frac{1}{n}(X_1 + \ldots + X_n) \Rightarrow X$

$X, X_1, X_2, \ldots$ independent copies of $X$. 

3
\[ F_n(x) = \frac{1}{n} \sum_{k=1}^{n} I_{X_k} \leq x \Rightarrow F(x) \quad \text{a.e.} \]

\[ X_1, X_2, \ldots \text{ i.i.d. r.v.s } \mathbb{E}(X_i^2) < \infty \]

Want to estimate \( \mathbb{E}(X_i) = \mu \). Use \( \frac{S_n}{n} \approx \mu \) a.s.

\[
P\left(\left| \frac{S_n}{n} - \mu \right| \leq \varepsilon \right) = P\left(\left| \frac{\frac{S_n}{n} - \mu}{\sigma / \sqrt{n}} \right| \leq \frac{\varepsilon}{\sigma / \sqrt{n}} \right) \xrightarrow{\text{CLT}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx 
\]