Part II — Probability and Measure
Theorems

Based on lectures by J. Miller
Notes taken by Dexter Chua

Michaelmas 2016

These notes are not endorsed by the lecturers, and I have modified them (often significantly) after lectures. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

*Analysis II is essential*


Chebyshev’s inequality, tail estimates. Jensen’s inequality. Completeness of \( L^p \) for \( 1 \leq p \leq \infty \). The Hölder and Minkowski inequalities, uniform integrability. [4]

\( L^2 \) as a Hilbert space. Orthogonal projection, relation with elementary conditional probability. Variance and covariance. Gaussian random variables, the multivariate normal distribution. [2]

The strong law of large numbers, proof for independent random variables with bounded fourth moments. Measure preserving transformations, Bernoulli shifts. Statements *and proofs* of maximal ergodic theorem and Birkhoff’s almost everywhere ergodic theorem, proof of the strong law. [4]

The Fourier transform of a finite measure, characteristic functions, uniqueness and inversion. Weak convergence, statement of Lévy’s convergence theorem for characteristic functions. The central limit theorem. [2]
# Contents

0 **Introduction** 3

1 **Measures** 4  
  1.1 Measures 4  
  1.2 Probability measures 4

2 **Measurable functions and random variables** 5  
  2.1 Measurable functions 5  
  2.2 Constructing new measures 5  
  2.3 Random variables 6  
  2.4 Convergence of measurable functions 6  
  2.5 Tail events 7

3 **Integration** 8  
  3.1 Definition and basic properties 8  
  3.2 Integrals and limits 8  
  3.3 New measures from old 9  
  3.4 Integration and differentiation 9  
  3.5 Product measures and Fubini’s theorem 10

4 **Inequalities and \( L^p \) spaces** 12  
  4.1 Four inequalities 12  
  4.2 \( L^p \) spaces 12  
  4.3 Orthogonal projection in \( L^2 \) 13  
  4.4 Convergence in \( L^1(\mathbb{F}) \) and uniform integrability 13

5 **Fourier transform** 14  
  5.1 The Fourier transform 14  
  5.2 Convolutions 14  
  5.3 Fourier inversion formula 14  
  5.4 Fourier transform in \( L^2 \) 15  
  5.5 Properties of characteristic functions 15  
  5.6 Gaussian random variables 15

6 **Ergodic theory** 17  
  6.1 Ergodic theorems 17

7 **Big theorems** 18  
  7.1 The strong law of large numbers 18  
  7.2 Central limit theorem 18
0 Introduction
1 Measures

1.1 Measures

**Proposition.** A collection $\mathcal{A}$ is a $\sigma$-algebra if and only if it is both a $\pi$-system and a $d$-system.

**Lemma (Dynkin’s $\pi$-system lemma).** Let $\mathcal{A}$ be a $\pi$-system. Then any $d$-system which contains $\mathcal{A}$ contains $\sigma(\mathcal{A})$.

**Theorem (Caratheodory extension theorem).** Let $\mathcal{A}$ be a ring on $E$, and $\mu$ a countably additive set function on $\mathcal{A}$. Then $\mu$ extends to a measure on the $\sigma$-algebra generated by $\mathcal{A}$.

**Theorem.** Suppose that $\mu_1, \mu_2$ are measures on $(E, \mathcal{E})$ with $\mu_1(E) = \mu_2(E) < \infty$. If $\mathcal{A}$ is a $\pi$-system with $\sigma(\mathcal{A}) = \mathcal{E}$, and $\mu_1$ agrees with $\mu_2$ on $\mathcal{A}$, then $\mu_1 = \mu_2$.

**Theorem.** There exists a unique Borel measure $\mu$ on $\mathbb{R}$ with $\mu([a,b]) = b - a$.

**Proposition.** The Lebesgue measure is translation invariant, i.e.

$$\mu(A + x) = \mu(A)$$

for all $A \in \mathcal{B}$ and $x \in \mathbb{R}$, where

$$A + x = \{y + x, y \in A\}.$$

**Proposition.** Let $\tilde{\mu}$ be a Borel measure on $\mathbb{R}$ that is translation invariant and $\mu([0,1]) = 1$. Then $\tilde{\mu}$ is the Lebesgue measure.

1.2 Probability measures

**Proposition.** Events $(A_n)$ are independent iff the $\sigma$-algebras $\sigma(A_n)$ are independent.

**Theorem.** Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are $\pi$-systems in $\mathcal{F}$. If

$$P[A_1 \cap A_2] = P[A_1]P[A_2]$$

for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$, then $\sigma(A_1)$ and $\sigma(A_2)$ are independent.

**Lemma (Borel–Cantelli lemma).** If

$$\sum_{n} P[A_n] < \infty,$$

then

$$P[A_n \ i.o.] = 0.$$

**Lemma (Borel–Cantelli lemma II).** Let $(A_n)$ be independent events. If

$$\sum_{n} P[A_n] = \infty,$$

then

$$P[A_n \ i.o.] = 1.$$
2 Measurable functions and random variables

2.1 Measurable functions

Lemma. Let \((E, \mathcal{E})\) and \((G, \mathcal{G})\) be measurable spaces, and \(\mathcal{G} = \sigma(\mathcal{Q})\) for some \(\mathcal{Q}\). If \(f^{-1}(A) \in \mathcal{E}\) for all \(A \in \mathcal{Q}\), then \(f\) is measurable.

Proposition. Let \(f_i : E \to F_i\) be functions. Then \(\{f_i\}\) are all measurable iff \((f_i) : E \to \prod F_i\) is measurable, where the function \((f_i)\) is defined by setting the \(i\)th component of \((f_i)(x)\) to be \(f_i(x)\).

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

\[ f_1 + f_2, \quad f_1 f_2, \quad \max\{f_1, f_2\}, \quad \min\{f_1, f_2\}, \]
\[ \inf_n f_n, \quad \sup_n f_n, \quad \liminf_n f_n, \quad \limsup_n f_n. \]

The same is true with “real” replaced with “non-negative”, provided the new functions are real (i.e. not infinity).

Theorem (Monotone class theorem). Let \((E, \mathcal{E})\) be a measurable space, and \(A \subseteq E\) be a \(\pi\)-system with \(\sigma(A) = \mathcal{E}\). Let \(V\) be a vector space of functions such that

(i) The constant function \(1 = 1_E\) is in \(V\).
(ii) The indicator functions \(1_A \in V\) for all \(A \in A\)
(iii) \(V\) is closed under bounded, monotone limits.

More explicitly, if \((f_n)\) is a bounded non-negative sequence in \(V\), \(f_n \nearrow f\) (pointwise) and \(f\) is also bounded, then \(f \in V\).

Then \(V\) contains all bounded measurable functions.

2.2 Constructing new measures

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

\[ g(\pm \infty) = \lim_{x \to \pm \infty} g(x). \]

We set \(I = (g(-\infty), g(\infty))\). Since \(g\) is non-constant, this is non-empty.

Then there is a non-decreasing, left continuous function \(f : I \to \mathbb{R}\) such that for all \(x \in I\) and \(y \in \mathbb{R}\), we have

\[ x \leq g(y) \iff f(x) \leq y. \]

Thus, taking the negation of this, we have

\[ x > g(y) \iff f(x) > y. \]

Explicitly, for \(x \in I\), we define

\[ f(x) = \inf\{y \in \mathbb{R} : x \leq g(y)\}. \]

Theorem. Let \(g : \mathbb{R} \to \mathbb{R}\) be non-constant, non-decreasing and right continuous. Then there exists a unique Radon measure \(dg\) on \(\mathcal{B}\) such that

\[ dg((a, b]) = g(b) - g(a). \]

Moreover, we obtain all non-zero Radon measures on \(\mathbb{R}\) in this way.
2.3 Random variables

**Proposition.** We have

\[ F_X(x) \to \begin{cases} 
0 & x \to -\infty \\
1 & x \to +\infty 
\end{cases}. \]

Also, \( F_X(x) \) is non-decreasing and right-continuous.

**Proposition.** Let \( F \) be any distribution function. Then there exists a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) and a random variable \( X \) such that \( F_X = F \).

**Proposition.** Two real-valued random variables \( X, Y \) are independent iff

\[ P[X \leq x, Y \leq y] = P[X \leq x]P[Y \leq y]. \]

More generally, if \( (X_n) \) is a sequence of real-valued random variables, then they are independent iff

\[ P[x_1 \leq x_1, \ldots, x_n \leq x_n] = \prod_{j=1}^{n} P[X_j \leq x_j] \]

for all \( n \) and \( x_j \).

**Proposition.** Let

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

be our probability space. Then there exists as sequence \( R_n \) of independent Bernoulli(1/2) random variables.

**Proposition.** Let

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

Given any sequence \( (F_n) \) of distribution functions, there is a sequence \( (X_n) \) of independent random variables with \( F_{X_n} = F_n \) for all \( n \).

2.4 Convergence of measurable functions

**Theorem.**

(i) If \( \mu(E) < \infty \), then \( f_n \to f \) a.e. implies \( f_n \to f \) in measure.

(ii) For any \( E \), if \( f_n \to f \) in measure, then there exists a subsequence \( (f_{n_k}) \) such that \( f_{n_k} \to f \) a.e.

**Theorem** (Skorokhod representation theorem of weak convergence).

(i) If \( (X_n), X \) are defined on the same probability space, and \( X_n \to X \) in probability. Then \( X_n \to X \) in distribution.

(ii) If \( X_n \to X \) in distribution, then there exists random variables \( (\tilde{X}_n) \) and \( \tilde{X} \) defined on a common probability space with \( F_{\tilde{X}_n} = F_{X_n} \) and \( F_{\tilde{X}} = F_X \) such that \( \tilde{X}_n \to \tilde{X} \) a.s.
2.5 Tail events

**Theorem** (Kolmogorov 0-1 law). Let \((X_n)\) be a sequence of independent (real-valued) random variables. If \(A \in \mathcal{T}\), then \(P[A] = 0\) or \(1\).

Moreover, if \(X\) is a \(\mathcal{T}\)-measurable random variable, then there exists a constant \(c\) such that

\[ P[X = c] = 1. \]
3 Integration

3.1 Definition and basic properties

Proposition. A function is simple iff it is measurable, non-negative, and takes on only finitely many values.

Proposition. Let \( f : [0, 1] \to \mathbb{R} \) be Riemann integrable. Then it is also Lebesgue integrable, and the two integrals agree.

Theorem (Monotone convergence theorem). Suppose that \((f_n), f\) are non-negative measurable with \(f_n \nearrow f\). Then \(\mu(f_n) \nearrow \mu(f)\).

Theorem. Let \( f, g \) be non-negative measurable, and \(\alpha, \beta \geq 0\). We have that

(i) \(\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)\).

(ii) \( f \leq g \) implies \(\mu(f) \leq \mu(g)\).

(iii) \( f = 0 \) a.e. iff \(\mu(f) = 0\).

Theorem. Let \( f, g \) be integrable, and \(\alpha, \beta \geq 0\). We have that

(i) \(\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)\).

(ii) \( f \leq g \) implies \(\mu(f) \leq \mu(g)\).

(iii) \( f = 0 \) a.e. implies \(\mu(f) = 0\).

Proposition. If \(\mathcal{A}\) is a \(\pi\)-system with \(E \in \mathcal{A}\) and \(\sigma(\mathcal{A}) = \mathcal{E}\), and \(f\) is an integrable function that \(\mu(f1_A) = 0\) for all \(A \in \mathcal{A}\). Then \(\mu(f) = 0\) a.e.

Proposition. Suppose that \((g_n)\) is a sequence of non-negative measurable functions. Then we have

\[
\mu \left( \sum_{n=1}^{\infty} g_n \right) = \sum_{n=1}^{\infty} \mu(g_n).
\]

3.2 Integrals and limits

Theorem (Fatou’s lemma). Let \((f_n)\) be a sequence of non-negative measurable functions. Then

\[
\mu(\lim \inf f_n) \leq \lim \inf \mu(f_n).
\]

Theorem (Dominated convergence theorem). Let \((f_n), f\) be measurable with \(f_n(x) \to f(x)\) for all \(x \in E\). Suppose that there is an integrable function \(g\) such that \(|f_n| \leq g\) for all \(n\), then we have

\[
\mu(f_n) \to \mu(f)
\]
as \(n \to \infty\).
3.3 New measures from old

**Lemma.** For \((E, \mathcal{E}, \mu)\) a measure space and \(A \in \mathcal{E}\), the restriction to \(A\) is a measure space. □

**Proposition.** Let \((E, \mathcal{E}, \mu)\) and \((F, \mathcal{F}, \mu')\) be measure spaces and \(A \in \mathcal{E}\). Let \(f : E \to F\) be a measurable function. Then \(f|_A\) is \(\mathcal{E}_A\)-measurable.

**Proposition.** If \(f\) is integrable, then \(f|_A\) is \(\mu_A\)-integrable and \(\mu_A(f|_A) = \mu(f 1_A)\). □

**Proposition.** If \(g\) is a non-negative measurable function on \(G\), then
\[
\nu(g) = \mu(g \circ f).
\]

**Proposition.** The \(\nu\) defined above is indeed a measure.

3.4 Integration and differentiation

**Proposition** (Change of variables formula). Let \(\phi : [a, b] \to \mathbb{R}\) be continuously differentiable and increasing. Then for any bounded Borel function \(g\), we have
\[
\int_{\phi(a)}^{\phi(b)} g(y) \, dy = \int_a^b g(\phi(x)) \phi'(x) \, dx.
\] (*)&nbsp;

**Theorem** (Differentiation under the integral sign). Let \((E, \mathcal{E}, \mu)\) be a space, and \(U \subseteq \mathbb{R}\) be an open set, and \(f : U \times E \to \mathbb{R}\). We assume that

(i) For any \(t \in U\) fixed, the map \(x \mapsto f(t, x)\) is integrable;

(ii) For any \(x \in E\) fixed, the map \(t \mapsto f(t, x)\) is differentiable;

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

Then the map
\[
x \mapsto \frac{\partial f}{\partial t}(t, x)
\]
is integrable for all \(t\), and also the function
\[
F(t) = \int_E f(t, x) \, d\mu
\]
is differentiable, and
\[
F'(t) = \int_E \frac{\partial f}{\partial t}(t, x) \, d\mu.
\]
3.5 Product measures and Fubini’s theorem

**Lemma.** Let $E = E_1 \times E_2$ be a product of $\sigma$-algebras. Suppose $f : E \to \mathbb{R}$ is $\mathcal{E}$-measurable function. Then

(i) For each $x_2 \in E_2$, the function $x_1 \mapsto f(x_1, x_2)$ is $\mathcal{E}_1$-measurable.

(ii) If $f$ is bounded or non-negative measurable, then

$$f_2(x_2) = \int_{E_1} f(x_1, x_2) \, \mu_1(dx_1)$$

is $\mathcal{E}_2$-measurable.

**Theorem.** There exists a unique measurable function $\mu = \mu_1 \otimes \mu_2$ on $\mathcal{E}$ such that

$$\mu(A_1 \times A_2) = \mu_1(A_1) \mu_2(A_2)$$

for all $A_1 \times A_2 \in \mathcal{A}$.

**Theorem** (Fubini’s theorem).

(i) If $f$ is non-negative measurable, then

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

In particular, we have

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

This is sometimes known as *Tonelli’s theorem*.

(ii) If $f$ is integrable, and

$$A = \left\{ x_1 \in E : \int_{E_2} |f(x_1, x_2)| \mu_2(dx_2) < \infty \right\},$$

then

$$\mu_1(E_1 \setminus A) = 0.$$ 

If we set

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

then $f_1$ is a $\mu_1$ integrable function and

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

**Proposition.** Let $X_1, \ldots, X_n$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $(E_1, \mathcal{E}_1), \ldots, (E_n, \mathcal{E}_n)$ respectively. We define

$$E = E_1 \times \cdots \times E_n, \quad \mathcal{E} = \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n.$$ 

Then $X = (X_1, \ldots, X_n)$ is $\mathcal{E}$-measurable and the following are equivalent:
(i) $X_1, \ldots, X_n$ are independent.

(ii) $\mu_X = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$.

(iii) For any $f_1, \cdots, f_n$ bounded and measurable, we have

$$E \left[ \prod_{k=1}^{n} f_k(X_k) \right] = \prod_{k=1}^{n} E[f_k(X_k)].$$
4 Inequalities and \( L^p \) spaces

4.1 Four inequalities

**Proposition** (Chebyshev’s/Markov’s inequality). Let \( f \) be non-negative measurable and \( \lambda > 0 \). Then
\[
\mu(\{f \geq \lambda\}) \leq \frac{1}{\lambda} \mu(f).
\]

**Proposition** (Jensen’s inequality). Let \( X \) be an integrable random variable with values in \( I \). If \( c : I \to \mathbb{R} \) is convex, then we have
\[
E[c(X)] \geq c(E[X]).
\]

**Lemma.** If \( c : I \to \mathbb{R} \) is a convex function and \( m \) is in the interior of \( I \), then there exists real numbers \( a, b \) such that
\[
c(x) \geq ax + b
\]
for all \( x \in I \), with equality at \( x = m \).

**Proposition** (Hölder’s inequality). Let \( p, q \in (1, \infty) \) be conjugate. Then for \( f, g \) measurable, we have
\[
\mu(|fg|) = \|fg\|_1 \leq \|f\|_p \|g\|_q.
\]
When \( p = q = 2 \), then this is the Cauchy-Schwarz inequality.

**Lemma.** Let \( a, b \geq 0 \) and \( p \geq 1 \). Then
\[
(a + b)^p \leq 2^p(a^p + b^p).
\]

**Theorem** (Minkowski inequality). Let \( p \in [1, \infty] \) and \( f, g \) measurable. Then
\[
\|f + g\|_p \leq \|f\|_p + \|g\|_p.
\]

4.2 \( L^p \) spaces

**Theorem.** Let \( 1 \leq p \leq \infty \). Then \( L^p \) is a Banach space. In other words, if \( (f_n) \) is a sequence in \( L^p \), with the property that \( \|f_n - f_m\|_p \to 0 \) as \( n, m \to \infty \), then there is some \( f \in L^p \) such that \( \|f_n - f\|_p \to 0 \) as \( n \to \infty \).
4.3 Orthogonal projection in $L^2$

**Theorem.** Let $V$ be a closed subspace of $L^2$. Then each $f \in L^2$ has an orthogonal decomposition

$$f = u + v,$$

where $v \in V$ and $u \in V^\perp$. Moreover,

$$\|f - v\|_2 \leq \|f - g\|_2$$

for all $g \in V$ with equality iff $g \sim v$.

**Lemma** (Pythagoras identity).

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

**Lemma** (Parallelogram law).

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

**Proposition.** The conditional expectation of $X$ given $G$ is the projection of $X$ onto the subspace $L^2(G, \mathbb{P})$ of $G$-measurable $L^2$ random variables in the ambient space $L^2(\mathbb{P})$.

4.4 Convergence in $L^1(\mathbb{P})$ and uniform integrability

**Theorem** (Bounded convergence theorem). Suppose $X, (X_n)$ are random variables. Assume that there exists a (non-random) constant $C > 0$ such that $|X_n| \leq C$. If $X_n \to X$ in probability, then $X_n \to X$ in $L^1$.

**Proposition.** Finite unions of uniformly integrable sets are uniformly integrable.

**Proposition.** Let $\mathcal{X}$ be an $L^p$-bounded family for some $p > 1$. Then $\mathcal{X}$ is uniformly integrable.

**Lemma.** Let $\mathcal{X}$ be a family of random variables. Then $\mathcal{X}$ is uniformly integrable if and only if

$$\sup\{E[|X|1_{|X| > k}] : X \in \mathcal{X}\} \to 0$$

as $k \to \infty$.

**Corollary.** Let $\mathcal{X} = \{X\}$, where $X \in L^1(\mathbb{P})$. Then $\mathcal{X}$ is uniformly integrable. Hence, a finite collection of $L^1$ functions is uniformly integrable.

**Theorem.** Let $X, (X_n)$ be random variables. Then the following are equivalent:

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

(ii) $\{X_n\}$ is uniformly integrable and $X_n \to X$ in probability.
5 Fourier transform

5.1 The Fourier transform

Proposition. \[ \|\hat{f}\|_\infty \leq \|f\|_1, \quad \|\hat{\mu}\|_\infty \leq \mu(\mathbb{R}^d). \]

Proposition. The functions \( \hat{f}, \hat{\mu} \) are continuous.

5.2 Convolutions

Proposition. For any \( f \in L^p \) and \( \nu \) a probability measure, we have \[ \|f \ast \nu\|_p \leq \|f\|_p. \]

Proposition. \( \hat{f} \ast \nu(u) = \hat{f}(u)\hat{\nu}(u) \).

Proposition. Let \( \mu, \nu \) be probability measures, and \( X, Y \) be independent variables with laws \( \mu, \nu \) respectively. Then \[ \hat{\mu} \ast \hat{\nu}(u) = \hat{\mu}(u)\hat{\nu}(u). \]

5.3 Fourier inversion formula

Theorem (Fourier inversion formula). Let \( f, \hat{f} \in L^1 \). Then \[ f(x) = \frac{1}{(2\pi)^d} \int \hat{f}(u)e^{-i(u,x)} \, du \text{ a.e.} \]

Proposition. Let \( Z \sim N(0,1) \). Then \[ \phi_Z(a) = e^{-a^2/2}. \]

Proposition. Let \( Z = (Z_1, \cdots, Z_d) \) with \( Z_j \sim N(0,1) \) independent. Then \( \sqrt{t}Z \) has density \[ g_t(x) = \frac{1}{(2\pi t)^{d/2}} e^{-|x|^2/(2t)}, \]
with \[ \hat{g}_t(u) = e^{-|u|^2t/2}. \]

Lemma. The Fourier inversion formula holds for the Gaussian density function.

Proposition. \[ \|f \ast g_t\|_1 \leq \|f\|_1. \]

Proposition. \[ \|f \ast g_t\|_\infty \leq (2\pi t)^{-d/2} \|f\|_1. \]

Proposition. \[ \|\hat{f} \ast g_t\|_1 = \|\hat{f}\hat{g}_t\|_1 \leq (2\pi)^{d/2} t^{-d/2} \|\hat{f}\|_1, \]
and \[ \|\hat{f} \ast g_t\|_\infty \leq \|\hat{f}\|_\infty. \]
Lemma. The Fourier inversion formula holds for Gaussian convolutions.

**Theorem (Fourier inversion formula).** Let $f \in L^1$ and

$$f_t(x) = (2\pi)^{-d} \int \hat{f}(u)e^{-|u|^2t/2}e^{-i(u,x)}du = (2\pi)^{-d} \int \hat{f}*g_t(u)e^{-i(u,x)}du.$$

Then $\|f_t - f\|_1 \to 0$, as $t \to 0$, and the Fourier inversion holds whenever $f, \hat{f} \in L^1$.

**Lemma.** Suppose that $f \in L^p$ with $p \in [1, \infty)$. Then $\|f*g_t - f\|_p \to 0$ as $t \to 0$.

### 5.4 Fourier transform in $L^2$

**Theorem (Plancherel identity).** For any function $f \in L^1 \cap L^2$, the Plancherel identity holds:

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

**Theorem.** There exists a unique Hilbert space automorphism $F : L^2 \to L^2$ such that

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

whenever $f \in L^1 \cap L^2$.

Here $[f]$ denotes the equivalence class of $f$ in $L^2$, and we say $F : L^2 \to L^2$ is a Hilbert space automorphism if it is a linear bijection that preserves the inner product.

### 5.5 Properties of characteristic functions

**Theorem.** The characteristic function $\phi_X$ of a distribution $\mu_X$ of a random variable $X$ determines $\mu_X$. In other words, if $X$ and $\tilde{X}$ are random variables and $\phi_X = \phi_{\tilde{X}}$, then $\mu_X = \mu_{\tilde{X}}$.

**Theorem.** If $\phi_X$ is integrable, then $\mu_X$ has a bounded, continuous density function

$$f_X(x) = (2\pi)^{-d} \int \phi_X(u)e^{-i(u,x)}du.$$

**Theorem.** Let $X, (X_n)$ be random variables with values in $\mathbb{R}^d$. If $\phi_{X_n}(u) \to \phi_X(u)$ for each $u \in \mathbb{R}^d$, then $\mu_{X_n} \to \mu_X$ weakly.

### 5.6 Gaussian random variables

**Proposition.** Let $X \sim N(\mu, \sigma^2)$. Then

$$\mathbb{E}[X] = \mu, \quad \text{var}(X) = \sigma^2.$$ 

Also, for any $a, b \in \mathbb{R}$, we have

$$aX + b \sim N(a\mu + b, a^2\sigma^2).$$

Lastly, we have

$$\phi_X(u) = e^{-i\mu u - u^2\sigma^2/2}.$$
Theorem. Let $X$ be Gaussian on $\mathbb{R}^n$, and let $A$ be an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then

(i) $AX + b$ is Gaussian on $\mathbb{R}^m$.
(ii) $X \in L^2$ and its law $\mu_X$ is determined by $\mu = \mathbb{E}[X]$ and $V = \text{var}(X)$, the covariance matrix.
(iii) We have
$$\phi_X(u) = e^{i(u, \mu) - (u, Vu)/2}.$$
(iv) If $V$ is invertible, then $X$ has a density of
$$f_X(x) = (2\pi)^{-n/2}(\det V)^{-1/2} \exp\left(-\frac{1}{2}(x - \mu, V^{-1}(x - \mu))\right).$$
(v) If $X = (X_1, X_2)$ where $X_i \in \mathbb{R}^{n_i}$, then $\text{cov}(X_1, X_2) = 0$ iff $X_1$ and $X_2$ are independent.
6 Ergodic theory

Proposition. If $f$ is integrable and $\Theta$ is measure-preserving. Then $f \circ \Theta$ is integrable and
\[ \int f \circ \Theta d\mu = \int_E f d\mu. \]

Proposition. If $\Theta$ is ergodic and $f$ is invariant, then there exists a constant $c$ such that $f = c$ a.e.

Theorem. The shift map $\Theta$ is an ergodic, measure preserving transformation.

6.1 Ergodic theorems

Lemma (Maximal ergodic lemma). Let $f$ be integrable, and
\[ S^* = \sup_{n \geq 0} S_n(f) \geq 0, \]
where $S_0(f) = 0$ by convention. Then
\[ \int_{\{S^* > 0\}} f \, d\mu \geq 0. \]

Theorem (Birkhoff’s ergodic theorem). Let $(E, \mathcal{E}, \mu)$ be $\sigma$-finite and $f$ be integrable. There exists an invariant function $\bar{f}$ such that
\[ \mu(|\bar{f}|) \leq \mu(|f|), \]
and
\[ \frac{S_n(f)}{n} \to \bar{f} \text{ a.e.} \]

If $\Theta$ is ergodic, then $\bar{f}$ is a constant.

Theorem (von Neumann’s ergodic theorem). Let $(E, \mathcal{E}, \mu)$ be a finite measure space. Let $p \in [1, \infty)$ and assume that $f \in L^p$. Then there is some function $\bar{f} \in L^p$ such that
\[ \frac{S_n(f)}{n} \to \bar{f} \text{ in } L^p. \]
7 Big theorems

7.1 The strong law of large numbers

**Theorem** (Strong law of large numbers assuming finite fourth moments). Let $(X_n)$ be a sequence of independent random variables such that there exists $\mu \in \mathbb{R}$ and $M > 0$ such that
\[
\mathbb{E}[X_n] = \mu, \quad \mathbb{E}[X_n^4] \leq M
\]
for all $n$. With $S_n = X_1 + \cdots + X_n$, we have that
\[
\frac{S_n}{n} \to \mu \text{ a.s. as } n \to \infty.
\]

**Theorem** (Strong law of large numbers). Let $(Y_n)$ be an iid sequence of integrable random variables with mean $\nu$. With $S_n = Y_1 + \cdots + Y_n$, we have
\[
\frac{S_n}{n} \to \nu \text{ a.s.}
\]

7.2 Central limit theorem

**Theorem.** Let $(X_n)$ be a sequence of iid random variables with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] = 1$. Then if we set
\[
S_n = X_1 + \cdots + X_n,
\]
then for all $x \in \mathbb{R}$, we have
\[
\mathbb{P}\left[\frac{S_n}{\sqrt{n}} \leq x\right] \to \int_{-\infty}^{x} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \, dy = \mathbb{P}[N(0,1) \leq x]
\]
as $n \to \infty$. 