Content

1 THEORY OF INTEGRATION ...................................... 1
  1.1 Measure Spaces ........................................... 1
  1.2 Measurable Functions and Mappings ...................... 6
  1.3 Integration of Measurable Functions .................... 10
  1.4 $L_p$-Spaces .................................................. 17
  1.5 Convergence of Functions .................................. 23
  1.6 The Theorem of Radon-Nikodym .......................... 29
  1.7 Dual Spaces and Conditional Expectation ............... 35
  1.8 Fubini’s Theorem ............................................ 44

2 MARTINGALE THEORY ........................................... 51
  2.1 Random Variables ........................................... 51
  2.2 Martingales ................................................... 55
  2.3 A.S. Convergence of Martingales ......................... 63

3 DISTRIBUTIONAL CONVERGENCE .............................. 71
  3.1 Vague and Weak Convergence ............................... 71
  3.2 Fourier-Transform ........................................... 79
  3.3 Central Limit Theorems ..................................... 85
  3.4 Donsker’s Invariance Principle ............................ 95

Index ................................................................. 99
1 THEORY OF INTEGRATION

1.1 Measure Spaces

This section contains the basic facts from measure theory, in particular the definition of probability spaces and the construction of such spaces.

**Definition 1.** A collection \( \mathcal{F} \) of subsets of a given (non-empty) set \( \Omega \neq \emptyset \) is called a \( \sigma \)-Algebra (or \( \sigma \)-field) on \( \Omega \), if the following properties hold:

(Sa) \( \Omega \in \mathcal{F} \).
(Sb) \( F \in \mathcal{F} \implies F^c \in \mathcal{F} \).
(Sc) \( F_n \in \mathcal{F}, n = 1, 2, 3, \ldots \implies \bigcup_{n \geq 1} F_n \in \mathcal{F} \).

Sets in \( \mathcal{F} \) are called measurable.

**Remark 1.** The following facts are immediate from the definition. A \( \sigma \)-algebra \( \mathcal{F} \) has the following properties:

(Sd) \( \mathcal{F} \neq \emptyset \).
(Se) \( \emptyset \in \mathcal{F} \).
(Sf) \( F_1, F_2, \ldots, F_n \in \mathcal{F}, n \in \mathbb{N} \implies \bigcup_{i=1}^n F_i \in \mathcal{F} \).
(Sg) \( F_1, F_2, \ldots, F_n \in \mathcal{F}, n \in \mathbb{N} \implies \bigcap_{i=1}^n F_i \in \mathcal{F} \).
(Sh) \( F_1, F_2 \in \mathcal{F} \implies F_1 \setminus F_2 \in \mathcal{F} \).
(Si) \( F_1, F_2 \in \mathcal{F} \implies F_1 \Delta F_2 \in \mathcal{F} \).
(Sj) \( F_1, F_2, F_3, \ldots \in \mathcal{F} \implies \bigcap_{n=1}^\infty F_n \in \mathcal{F} \).

**Theorem 1.** Let \( \mathcal{F}_i \ (i \in I) \) be \( \sigma \)-algebras on \( \Omega \). Then

\[
\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i
\]

is also a \( \sigma \)-algebra.

**Definition 2.** Let \( \Sigma \subset \mathcal{P}(\Omega) \) (\( \mathcal{P}(\Omega) \) denotes the set of all subsets of \( \Omega \), the power set) be a collection of subsets from \( \Omega \neq \emptyset \). Then there is a smallest \( \sigma \)-algebra \( \sigma(\Sigma) \) containing \( \Sigma \), defined by
Theorem of Integration

\[ \sigma(\Sigma) := \bigcap_{\mathcal{F}} \mathcal{F} \]

\( \sigma(\Sigma) \) is called the \( \sigma \)-algebra generated by \( \Sigma \).

Example 1. a. The power set of a nonempty set is always a \( \sigma \)-algebra.
b. Let \( \Omega = \mathbb{R} \) and let \( \mathcal{F} \) be defined by

\[ F \in \mathcal{F} \iff F \text{ or } F^c \text{ is at most countable} \ . \]

\( \mathcal{F} \) is a \( \sigma \)-algebra.
c. Let \( \Omega = \mathbb{R}^d, d \geq 1 \), denote the \( d \)-dimensional Euclidean space and let

\[ \Sigma^d = \left\{ \prod_{i=1}^{d} (a_i, b_i] : a_i \leq b_i \in \mathbb{R}, 1 \leq i \leq d \right\} \]

be the collection of all rectangles. Then

\[ B^d := \sigma(\Sigma^d) \]

is called the (\( d \)-dimensional) Borel \( \sigma \)-algebra (on \( \mathbb{R}^d \)).
d. Let \( \Omega \neq \emptyset \) be a topological space and \( T \) be its system of open sets. Then

\[ B = B(T) = \sigma(T) \]

is called the Borel \( \sigma \)-algebra on the topological space \( \Omega \). It is easy to show that this notion agrees with c., if \( \Omega \) is the Euclidean space.

Definition 3. For a nonempty set \( \Omega \) the collection \( \mathcal{R} \) of subsets from \( \Omega \) is called a ring (on \( \Omega \)), if the following properties hold:

(Ra) \( \emptyset \in \mathcal{R} \).
(Rb) \( R_1, R_2 \in \mathcal{R} \implies R_1 \cup R_2 \in \mathcal{R} \).
(Rc) \( R_1, R_2 \in \mathcal{R}, R_2 \subset R_1 \implies R_1 \setminus R_2 \in \mathcal{R} \).

Example 2. Let \( \mathcal{R}^d \) be the system of all finite unions of sets from \( \Sigma^d \). Then \( \mathcal{R}^d \) is a ring, which even contains the intersection of two of its elements. In general: Let \( \Omega_i \ (i \in I) \) be nonempty sets with \( \sigma \)-algebras \( \mathcal{F}_i \). Let us define \( \mathcal{R} \) as the system of all finite collections of sets of the form (which are called cylindersets)

\[ [F_1, ..., F_n; i_1, ..., i_n] = \{(x_i)_{i \in I} \in \prod_{i \in I} \Omega_i : x_{i_l} \in F_{i_l}, l = 1, ..., n\} \]

where \( F_l \in \mathcal{F}_{i_l}, l = 1, ..., n, n \in \mathbb{N} \). Then \( \mathcal{R} \) is a ring.

Definition 4. Let \( \mathcal{F} \) be a \( \sigma \)-algebra (resp. a ring). A mapping

\[ m : \mathcal{F} \to \mathbb{R} \cup \{\pm \infty\} \]

is called
1.1 Measure Spaces

(Ma) \( \sigma \)-additive, if for all pairwise disjoint sets \( F_1, F_2, \ldots \in \mathcal{F} \) with the property \( \bigcup_{n \geq 1} F_n \in \mathcal{F} \) one has (the values \( \pm \infty \) are included if they make sense):

\[
m\left( \bigcup_{n \geq 1} F_n \right) = \sum_{n \geq 1} m(F_n).
\]

(The convergence of the infinite series is an absolute convergence!)

(Mb) a pre-measure, if \( m \) \( \sigma \)-additive and positive, if \( m(\emptyset) = 0 \) holds and if \( \mathcal{F} \) is a ring.

(Mc) a measure, if \( m \) is \( \sigma \)-additive and positive and if \( \mathcal{F} \) is a \( \sigma \)-algebra.

(Md) a signed measure, if \( m \) is \( \sigma \)-additive, if \( m(\emptyset) = 0 \) holds and if \( \mathcal{F} \) is a \( \sigma \)-algebra.

**Definition 5.** A measure \( m \) is called finite, if \( m(\Omega) < \infty \) holds, and it is called \( \sigma \)-finite, if there are sets \( \Omega_n \ (n \geq 1) \), such that \( \Omega = \bigcup_{n \geq 1} \Omega_n \) and \( m(\Omega_n) < \infty \) for all \( n \in \mathbb{N} \). Finally, it is called a probability measure, if \( m(\Omega) = 1 \).

**Remark 2.** The following properties hold for measures: Let \( E, F, F_n \ (n \in \mathbb{N}) \) measurable sets. Then:

(Me) \( \exists F \) with \( m(F) < \infty \Rightarrow m(\emptyset) = 0 \).

(Mf) \( E, F \) disjoint \( \Rightarrow m(E \cup F) = m(E) + m(F) \) (infinity is permitted!).

(Mg) \( E \subset F, m(F) < \infty \Rightarrow m(F \setminus E) = m(F) - m(E) \), in particular, \( m(E) \leq m(F) \).

(Mh) (Continuity from below) \( \forall n \ F_n \subset F_{n+1} \Rightarrow \)

\[
m\left( \bigcup_{n \geq 1} F_n \right) = \lim_{n \to \infty} m(F_n).
\]

(Mi) (Continuity from above) \( \forall n \ F_n \supset F_{n+1}, \exists n \geq 1 \) mit \( m(F_n) < \infty \Rightarrow \)

\[
m\left( \bigcap_{n \geq 1} F_n \right) = \lim_{n \to \infty} m(F_n).
\]

(Mj) \( m(E), m(F) < \infty \Rightarrow m(E \cup F) = m(E) + m(F) - m(E \cap F) \).

**Theorem 2.** (Extension Theorem of Caratheodory) Let \( m_0 \) be a pre-measure on the ring \( \mathcal{R} \). Then there exists a measure \( m \) on the \( \sigma \)-algebra \( \sigma(\mathcal{R}) \) generated by \( \mathcal{R} \), which extends \( m_0 \).

**Theorem 3.** (Uniqueness Theorem) Let \( \Sigma \) be a generator of the \( \sigma \)-algebra \( \mathcal{F} \) (i.e. \( \sigma(\Sigma) = \mathcal{F} \)) having the following two properties:

1. \( \forall F_1, F_2 \in \Sigma \Rightarrow F_1 \cap F_2 \in \Sigma \). (\( \Sigma \) is called intersection-stable or closed under intersections.)
2. \( \exists E_n \in \Sigma, n \geq 1, E_n \subset E_{n+1} \) and \( \bigcup_{n \geq 1} E_n = \Omega \).
If two measures $m_1$ and $m_2$ agree on $\Sigma$, and if $m_i(E_n) < \infty$ $(n \geq 1, i = 1, 2)$, then $m_1$ and $m_2$ agree on all of $\mathcal{F}$.

**Example 3.**
1. Let $\widehat{m}$ be set function defined on $\Sigma^d$, which is positive, finite and $\sigma$-additive (this is defined analogously). Then there is a unique, $\sigma$-additive extension $m_0$ on the ring $\mathcal{R}^d$, generated by $\Sigma^d$, that is

$$R \in \mathcal{R}^d \Rightarrow \exists R_1, \ldots, R_n \in \Sigma^d, R_i \cap R_j = \emptyset (i \neq j), R = \bigcup_{j=1}^{n} R_j;$$

and $m_0(R) = \sum_{j=1}^{n} \widehat{m}(R_j)$.

Consequently, there is only one extension $m$ of $\widehat{m}$ on the Borel $\sigma$-algebra $\mathcal{B}^d$.

In particular, for the density $f : \mathbb{R}^d \to \mathbb{R}_+$, the set function defined by

$$\widehat{m}(\prod_{i=1}^{d} (a_i, b_i]) = \int_{\prod_{i=1}^{d} (a_i, b_i]} f(x) \, dx$$

is positive, finite and $\sigma$-additive. The resulting measure on $\mathbb{R}^d$ is the measure with density $f$. If $f = 1$, this measure is called the Lebesgue measure on $\mathbb{R}^d$.

2. Let $m$ be a measure on the real line, satisfying $m(E) < \infty$ for every upper bounded set $E$. Defining

$$F(t) = m((\infty, t]) \quad (t \in \mathbb{R}),$$

$F$ is a right continuous, monotonically increasing function with $\lim_{t \to -\infty} F(t) = 0$, such that the left-sided limits exist. Conversely, if $F : \mathbb{R} \to \mathbb{R}_+$ is such a monotonically increasing function,

$$\widehat{m}((s, t]) = F(t) - F(s) \quad (s, t \in \mathbb{R}, s \leq t)$$

defines a set function on the ring of finite unions of intervals and 1. guarantees the extension to a measure on $\mathbb{R}$. This measure is called a Stieltjes measure. In particular, if $\lim_{t \to \infty} F(t) = 1$ this fact is known from probability classes.

3. If $\omega_i$ $(i \in \mathbb{N})$ are at most countably many points in a non-empty set $\Omega$ with $\sigma$-algebra $\mathcal{F}$, and if $\alpha_i$ $(i \in \mathbb{N})$ are positive reals, then

$$m(F) = \sum_{i=1}^{\infty} \alpha_i 1_F(\omega_i) \quad (F \in \mathcal{F})$$

defines a measure on $(\Omega, \mathcal{F})$. Measures of this type are called discrete. In case $\alpha_i = \delta_{i, i_0}$, $i_0$ fixed, the measure is called a point mass in $\omega_{i_0}$.
1.1 Measure Spaces

Definition 6. A tuple $(\Omega, \mathcal{F})$ is called a measurable space, if $\mathcal{F}$ is a $\sigma$-algebra on the non-empty set $\Omega$, and a triple $(\Omega, \mathcal{F}, m)$ is called a measure space, if $(\Omega, \mathcal{F})$ is a measurable space, and $m$ is a measure on the $\sigma$-algebra $\mathcal{F}$.

If $(\Omega, \mathcal{F})$ is a measurable space, we let $\mathcal{M}$ denote the set of all measures on $(\Omega, \mathcal{F})$. Moreover we set

\[ M_s = \{ m \in \mathcal{M} : m(\Omega) \leq s \} \quad (s \in \mathbb{R}_+) \]
\[ M_\infty = \{ m \in \mathcal{M} : m(\Omega) < \infty \}. \]

$\mathcal{M}_s$ is a positive cone in the vector space

\[ \{ m_1 - m_2 : m_1, m_2 \in \mathcal{M}_s \}. \]
1.2 Measurable Functions and Mappings

We shall assume throughout that $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}')$ and $(\Omega_i, \mathcal{F}_i) \ (i \in I)$ ... are measurable spaces.

**Definition 7.** A mapping $T : \Omega \to \Omega'$ is called measurable, if

$$T^{-1} \mathcal{F}' \subset \mathcal{F}.$$  

In case of $\Omega' = \mathbb{R}$ (resp. $\Omega' = \mathbb{R} = \mathbb{R} \cup \{\pm \infty\}$) and $\mathcal{F}'$ the Borel $\sigma$-algebra, $T$ is called a measurable (non-numerical) function (resp. measurable numerical function, or simply a measurable function). In case of $m(\Omega) = 1$ and $\Omega' = \mathbb{R}^d$ or of a Banach space resp. $\Omega'$ a function space or arbitrary), $T$ is called a random variable ((a $d$-dimensional) random vector resp. random function or random element).

**Lemma 1.** If $\Sigma \subset \mathcal{F}'$ generates, then $T : \Omega \to \Omega'$ is measurable, if and only if $T^{-1} \Sigma \subset \mathcal{F}$.

**Proof.** $\Rightarrow$ $T^{-1} \Sigma \subset T^{-1} \mathcal{F}' \subset \mathcal{F}$.  

$\Leftarrow$ The system $\mathcal{A} := \{E \subset \Omega' : T^{-1} E \in \mathcal{F}\}$ is a $\sigma$-algebra, as is easily checked. Since $\Sigma \subset \mathcal{A}$, and since $\mathcal{F}'$ is the smallest $\sigma$-algebra with this property, it follows that $\mathcal{F}' \subset \mathcal{A}$, which proves the claim.

**Lemma 2.** 1.) The concatenation of two measurable maps is also measurable.  

2.) A continuous mapping is measurable with respect to the Borel $\sigma$-algebras (Borel fields).

**Proof.** 1.) Let $T_i : \Omega_i \to \Omega_{i+1}, \ i = 1, 2,$ be measurable mappings. Then

$$(T_2 \circ T_1)^{-1} \mathcal{F}_3 = T_1^{-1} \circ T_2^{-1} \mathcal{F}_3 \subset T_1^{-1} \mathcal{F}_2 \subset \mathcal{F}_1.$$  

2.) Let $\Omega_i, \ i = 1, 2,$ be topological spaces with Borel $\sigma$-algebras $\mathcal{F}_i$. Let $T : \Omega_1 \to \Omega_2$ be continuous (meaning that preimages of open sets are open). Let $\mathcal{T}_i$ denote the systems of open sets, which, by definition, generate the corresponding Borel fields. Continuity of $T$ implies $T^{-1} \mathcal{T}_2 \subset \mathcal{T}_1$, and the claim follows now from the last lemma.

**Definition 8.** The product of measurable spaces $(\Omega_i, \mathcal{F}_i) \ (i \in I)$ is defined as the set-theoretic product

$$\Omega = \prod_{i \in I} \Omega_i = \{(\omega_i)_{i \in I} : \omega_i \in \Omega_i\}$$.
equipped with the product-$\sigma$-algebra $F = \prod_{i \in I} F_i$, which is defined to be the smallest $\sigma$-algebra, which contains all sets of the form (cylinder or cylinder set)

$$[F,j] = \{(\omega_i)_{i \in I} \in \Omega : \omega_j \in F\}$$

where $j \in I$ and $F \in F_j$. This is obviously equivalent of saying that $F$ is the smallest $\sigma$-algebra, such that all projections onto finitely many coordinates are measurable.

**Lemma 3.** Let $T_i : \Omega_i \to \Omega_i'$ be measurable maps ($i = 1, 2$). Then the map

$$T_1 \times T_2 : \Omega_1 \times \Omega_2 \to \Omega_1' \times \Omega_2'$$

is measurable with respect to the product-$\sigma$-algebras.

**Proof.** Let $F$ und $F'$ denote the respective product-$\sigma$-algebras. For $F \in F_1'$ we have

$$(T_1 \times T_2)^{-1} F \times \Omega_2' = T_1^{-1} F \times \Omega_2 \in F$$

and likewise the same relation interchanging the indices. Therefore, the generating system

$$\Sigma = \{F_1 \times \Omega_2', \Omega_1' \times F_2 : F_i \in F_i', i = 1, 2\}$$

of $F'$ is mapped under $(T_1 \times T_2)^{-1}$ into $F$, and the claim follows from the first lemma as before.

**Remark 3.** In the situation of the last lemma one easily calculates that for $\Omega_1 = \Omega_2$ the map

$$\Omega_1 \to \Omega_1 \times \Omega_1$$

$$\omega \to (\omega, \omega)$$

is measurable. Thus also maps of the form

$$T_1 \otimes T_2 : \Omega_1 \to \Omega_1' \times \Omega_2'$$

$$T_1 \otimes T_2(\omega) = (T_1(\omega), T_2(\omega))$$

are measurable.

What has been proved so far, provides basic properties of measurable functions:

**Theorem 4.** Let $f, g, f_n$ ($n \in \mathbb{N}$) be measurable (numerical) functions on $(\Omega, F)$. The following properties are true:

(M1) Every constant function is measurable.
8

1 THEORY OF INTEGRATION

(M2) Every indicator-function of a measurable set in $\mathcal{F}$ is measurable.

(M3) $f + g$ and $f \cdot g$ are measurable, as long as $f$ and $g$ are non-numerical.

(M4) If $h : \mathbb{R} \to \mathbb{R}$ is measurable, so is $h \circ f$.

(M5) $\max(f, g)$ and $\min(f, g)$ are measurable.

(M6) $|f|$ is measurable.

(M7) $\{f < g\}$, $\{f \leq g\}$ and $\{f = g\}$ are measurable sets.

(M8) $\sup_{n \in \mathbb{N}} f_n$, $\inf_{n \in \mathbb{N}} f_n$, $\limsup_{n \to \infty} f_n$, $\liminf_{n \to \infty} f_n$ and $\lim_{n \to \infty} f_n$ (if exists) are measurable (numerical) functions.

(M9) $\{\omega \in \Omega : \lim_{n \to \infty} f_n \text{ exists}\}$ is a measurable set.

Definition 9. A function $f : \Omega \to \mathbb{R}$ is called an elementary function (or a step function), if $f$ is measurable and if it attains only finitely many real values, i.e. it has a representation of the form

$$f(\omega) = \sum_{k=1}^{s} \alpha_k 1_{A_k}(\omega),$$

where $\alpha_k \in \mathbb{R}$, $s \in \mathbb{N}$ and $A_k \in \mathcal{F}$.

Remark 4. An immediate consequence of the last theorem is this: the sum and the product of two step functions is again a step function. Moreover, every constant function is a step function.

Proposition 1. (Approximation theorem) Every positive measurable (non-numerical) function $f$ can be approximated by an increasing sequence of step functions. If $f$ is bounded, the approximation can be chosen uniformly.

Proof. Put

$$f_n = \sum_{k=1}^{2^n} \frac{k-1}{2^n} 1_{\{(k-1)2^{-n} \leq f < k2^{-n}\}}.$$

Proposition 2. (Decomposition theorem) Every measurable (numerical) function $f$ can be written in a unique way as the difference of two positive measurable functions $f^+$ and $f^-$, such that $\inf(f^+, f^-) = 0$. One has

$$f = f^+ - f^-,$$

$$f^+ = \max(f, 0),$$

$$f^- = \max(-f, 0).$$

Proof. By (M1) and (M5) $f^\pm$ are measurable functions. By definition we have $\min(f^+, f^-) = 0$.

Uniqueness follows by this argument: If $f = \phi - \psi$ is another representation with positive measurable functions $\phi$ and $\psi$ satisfying $\min(\phi, \psi) = 0$, it follows that $f^+ + \psi = f^- + \phi$. Is $f(x) > 0$ one has $f^-(x) = \psi(x) = 0$ and also $f^+(x) = \phi(x)$. In all other cases the argument is similar.
Theorem 5. (Transport of measures)
Let $T : \Omega \to \Omega'$ be measurable. If $m$ is a measure on $\mathcal{F}$, then

$$m'(F') := m(T^{-1}F') \quad (F' \in \mathcal{F}')$$

defines a measure on $\mathcal{F}'$. It will be written as $m' = m \circ T^{-1}$ and is called the
measure transported by $T$. 
1.3 Integration of Measurable Functions

Consider a measure space \((\Omega, \mathcal{F}, m)\). We shall use the convention that \(0 \cdot \infty = 0\) and start the investigation with some simple properties:

**Lemma 4.** Let \(f \geq 0\) be a step function with representation

\[
f = \sum_{i=1}^{n} \alpha_i 1_{F_i}.
\]

Then the value

\[
\sum_{i=1}^{n} \alpha_i m(F_i)
\]

is independent of the representation.

**Proof.** Let \(f = \sum_{k=1}^{p} \beta_k 1_{E_k}\) another representation of \(f\). We may assume that the sets \(F_i\), and the sets \(E_k\) as well, are pairwise disjoint, so we have \(f = \alpha_i \) on \(F_i\). It follows that \(f\) is constant on \(F_i \cap E_k\), hence \(\alpha_i = \beta_k\), provided \(F_i \cap E_k \neq \emptyset\). We can assume, that the sets in the representation decompose the support of \(f\) (i.e. the set \(\{\omega \in \Omega : f(\omega) \neq 0\}\)). It follows that

\[
\sum_{i=1}^{n} \alpha_i m(F_i) = \sum_{i=1}^{n} \beta_i m(F_i) = \sum_{i=1}^{n} \beta_i m(F_i \cap E_k) = \sum_{k=1}^{p} \beta_k m(E_k).
\]

**Definition 10.** Let \(f = \sum_{i=1}^{n} \alpha_i 1_{F_i} \geq 0\) be a step function. Then

\[
I_0(f) := \sum_{i=1}^{n} \alpha_i m(F_i)
\]

is called the integral of \(f\).

If \(f \geq 0\) is measurable, the integral of \(f\) is defined by

\[
\int f \, dm = I(f) := \sup \{I_0(g) : 0 \leq g \leq f, \ g \text{ step function}\}.
\]

Finally, if \(f\) is a measurable function, such that \(I(f^+) < \infty\) or \(I(f^-) < \infty\), then its integral is

\[
\int f \, dm = I(f) := I(f^+) - I(f^-).
\]

\(f\) is called integrable, if both, its positive and its negative part have finite integral. This is equivalent of saying that

\[
\int |f| \, dm < \infty.
\]
Remark 5. A random variable $X$ is just a measurable function on a probability space. Its expectation is its integral

$$E(X) = \int X(\omega)m(d\omega) = I(X).$$

Infinite expectation means that $I(|X|) = \infty$. The variance of a random variable is the integral of the function $(X - E(X))^2$, which is also a measurable function by section 1.2.

The integral $I_0$ is obviously monotone and linear, as can be deduced easily using the first lemma. The important property, however, is the following fact. The next lemma shows that integrals can be calculated by arbitrary countable approximations from below.

**Lemma 5.** Let $f \geq 0$ be measurable, and let $0 \leq f_n \leq f$ be step functions, which approximate $f$ pointwise and monotone. Then for any step function $0 \leq g \leq f$:

$$I_0(g) \leq \lim_{n \to \infty} I_0(f_n),$$

in particular, the limit exists and

$$I(f) = \lim_{n \to \infty} I_0(f_n).$$

**Proof.** Let $g = \sum_{k=1}^{s} \gamma_k 1_{C_k}$ be a step function, $g \leq f$.

1. In case of $I_0(g) = \infty$, there is a $k$ such that $\gamma_k > 0$ and $m(C_k) = \infty$. Let $\epsilon > 0$ and $A_n = \{f_n + \epsilon > g\}$. Because of the monotonicity of $I_0$

$$I_0(f_n) \geq I_0(f_n 1_{A_n}) \geq I_0(f_n 1_{A_n \cap C_k}) \geq (\gamma_k - \epsilon)m(A_n \cap C_k) \to \infty,$$

(because of (Mh)) also $\lim_{n \to \infty} I_0(f_n) = \infty$.

2. In case of $I_0(g) < \infty$, put $T = \{g > 0\}$. Then $m(T) < \infty$. Let $0 < \epsilon < \min_{1 \leq k \leq s} \gamma_k$ (w.l.o.g. $\gamma_1 > 0$). Denoting $A_n = \{f_n > g - \epsilon\}$ it follows as before that

$$I_0(f_n) \geq I_0(f_n 1_{A_n \cap T}) \geq I_0((g - \epsilon) 1_{A_n \cap T}) \geq I_0((g - \epsilon) 1_{A_n \cap T}) - \epsilon m(A_n \cap T) - I_0(1_{A_n \cap T}).$$

For $n \to \infty$ it follows that $T \subset \lim_{n \to \infty} A_n$, hence $m(A_n \cap T) \to m(T) < \infty$ and $m(A_n^c \cap T) \to 0$. Moreover, if $\epsilon \to 0$, the claim follows.

**Definition 11.** A property $E$ for points in $\Omega$ is said to hold almost everywhere (a.e.) or almost surely (a.s.), if there is a measurable set $N$ of measure 0 such that $\{\omega \in \Omega : E \text{ does not hold for } \omega\} \subset N$. A measurable set of measure 0 is called a nullset.
Lemma 6. Let \( f = g \) a.e. If \( \int f \, dm \) exists, so does \( \int g \, dm \) and both integrals are equal.

Proof. First of all one observes that \( f = g \) a.e. also implies that \( f^+ = g^+ \) and \( f^- = g^- \) a.e. Let \( 0 \leq f_n^* \leq f^+ \) be step functions which approximate \( f^+ \) monotonically, and let \( 0 \leq g_n^* \leq g^+ \) be step functions which approximate \( g^+ \) monotonically. Set \( F = \{ f = g \} \). Then \( f_n = f_n^*1_F + f_n^*1_{F^c} \) and \( g_n = g_n^*1_F + g_n^*1_{F^c} \) are step functions, which approximate \( f^+ \) and \( g^+ \) from below. Since \( m(F^c) = 0 \) we have that \( I_0(f_n) = I_0(g_n) \) for every \( n \geq 1 \), and the claim follows from the last lemma.

Elementary properties of the integral are contained in the next theorem.

Theorem 6. Let \( f, g \) be measurable functions, for which \( I(f) \) and \( I(g) \) exist, and let \( \alpha \) denote a constant.

(I1) \( f \geq 0 \) a step function \( \implies I(f) = I_0(f) \)

(I2) \( f \) integrable \( \implies |f| < \infty \) a.e.

(I3) \( \int \alpha f \, dm = \alpha \int f \, dm \).

(I4) If \( f \) and \( g \) are integrable, then \( f + g \) is a.e. well defined and one has

\[
\int (f + g) \, dm = \int f \, dm + \int g \, dm.
\]

(I5) \( f \leq g \) a.e. \( \implies \int f \, dm \leq \int g \, dm \).

(I6) \( f = 0 \) a.e. \( \iff \int |f| \, dm = 0 \).

(I7) \( |f| \leq g \) a.e., \( g \) integrable \( \implies f \) integrable.

(I8) \( \int f \, dm \leq \int |f| \, dm \).

(I9) \( f, g \) integrable, \( \int_A f \, dm = \int_A g \, dm \) for all \( A \in \mathcal{F} \) \( \implies f = g \) a.e.

Proof.

(I1): follows from the definition of \( I(f) \), since obviously for a step function \( g \leq f \) we have that \( I_0(g) \leq I_0(f) \).

(I2): If \( |f| = \infty \) on a measurable set \( F \) and \( m(F) > 0 \), then \( 0 \leq N1_F \leq |f| \)

for every \( N \) and \( |f| \) is not integrable. But by definition \( f \) is integrable, if

and only if \( |f| \) is integrable. A contradiction!

(I3): Let \( \alpha \geq 0 \) and \( f \geq 0 \). Then it follows that

\[
I(\alpha f) = \sup \{ I_0(\alpha g) : 0 \leq g \leq f \} = \alpha I(f).
\]

Then it also holds for any \( \alpha \) and \( f \) that

\[
\int \alpha f \, dm = \int \alpha f^+ \, dm - \int \alpha f^- \, dm = \alpha \int f \, dm \quad (\alpha \geq 0)
\]

\[
\int \alpha f \, dm = \int (-\alpha) f^- \, dm - \int (-\alpha) f^+ \, dm = \alpha \int f \, dm \quad (\alpha < 0).
\]
1.3 Integration of Measurable Functions

(I4): By (I2) $f + g$ is a.e. well defined.

We begin letting $f, g \geq 0$. There are (the above lemma and section 1.2) step functions $0 \leq f_n \leq f$ and $0 \leq g_r \leq g$, which approximate $f$ and $g$, and satisfy $I(f) = \lim_{n \to \infty} I_0(f_n)$ and $I(g) = \lim_{r \to \infty} I_0(g_r)$. It follows that the sequence $f_n + g_n$ $(n \geq 1)$ approximates the sum $f + g$ from below, and hence

$$I(f + g) = \lim_{n \to \infty} I_0(f_n + g_n) = \lim_{n \to \infty} I_0(f_n) + \lim_{r \to \infty} I_0(g_r) = I(f) + I(g).$$

Let now $h = \phi - \psi$ be integrable and be represented as the difference of two non-negative integrable functions. Then $h^+ + \psi = h^- + \phi$ and by the relation proved so far $\int h^+ \, dm + \int \psi \, dm = \int h^- \, dm + \int \phi \, dm$. This means

$$\int h \, dm = \int h^+ \, dm - \int h^- \, dm = \int \phi \, dm - \int \psi \, dm.$$

Now consider $f$ and $g$ general integrable functions. Then $f + g = (f^+ + g^+) - (f^- + g^-)$, so

$$\int f + g \, dm = \int f^+ + g^+ \, dm - \int f^- + g^- \, dm$$
$$= \int f^+ \, dm - \int f^- \, dm + \int g^+ \, dm - \int g^- \, dm$$
$$= \int f \, dm + \int g \, dm.$$

(I5): According to the last lemma we may assume, that $f \leq g$. Moreover one has $f^+ \leq g^+$ and $g^- \leq f^-$. For positive functions the claim follows by definition. Hence

$$\int f \, dm = \int f^+ \, dm - \int f^- \, dm \leq \int g^+ \, dm - \int g^- \, dm = \int g \, dm.$$

(I6): Let $f = 0$ a.e. Then also $f^+ = 0$ and $f^- = 0$ a.e. It is immediate that $I(f^+) = 0$ holds, and since $|f| = f^+ + f^-$ and because of property (I4) it follows that $\int |f| \, dm = 0$. Conversely, if $\int |f| \, dm = 0$, it follows immediately that $\int f^+ \, dm = 0$. Let $0 \leq g \leq f^+$ be a step function. By definition of the integral $I_0$ one has $m(\{g > 0\}) = 0$. In case that $0 \leq g_n \leq f^+$ is an approximating monotone sequence of step functions, it follows that $\{f^+ > 0\} = \lim_{n \to \infty} \{g_n > 0\}$ and using (Mh) $m(\{f^+ > 0\}) = 0$. In the same way one concludes that $m(\{f^- > 0\}) = 0$, and this implies the claim.

(I7): Follows immediately from (I4) and (I5) since $|f| = f^+ + f^-$.  

(I8): $|\int f \, dm| = |\int f^+ \, dm - \int f^- \, dm| \leq \int f^+ \, dm + \int f^- \, dm = \int |f| \, dm$ because of (I4).
(19): Let $F_\eta = \{ f > g + \eta \}$ and $G_\eta = \{ g > f + \eta \}$. In case of $m(F_\eta) = 0$ and $m(G_\eta) = 0$ for every $\eta > 0$, $\bigcup_{k \geq 1} F_{1/k} \cup G_{1/k}$ is a nullset and hence on its complement one has $f = g$. Assuming that $f \neq g$ on the set $F$ satisfying $m(F) > 0$, one finds $\eta > 0$ such that $m(F_\eta) > 0$ (w.l.o.g.). This implies $g1_{F_\eta} \leq (g + \eta)1_{F_\eta} \leq f1_{F_\eta}$ and, using (14) and (15),

$$
\int_{F_\eta} g \, dm \leq \int_{F_\eta} (g + \eta) \, dm = \int_{F_\eta} g \, dm + \eta m(F_\eta)
\leq \int_{F_\eta} f \, dm = \int_{F_\eta} g \, dm.
$$

A contradiction!

**Theorem 7.** (Monotone Convergence Theorem, Theorem of B. Levi)

Let $f_n$ ($n \geq 1$) be measurable functions and let $h$ be an integrable function, such that $h \leq f_n \leq f_{n+1}$ a.e. ($n \geq 1$). Then there exists

$$
\lim_{n \to \infty} f_n \quad \text{a.e.}
$$

and for a function $f$, which agrees a.e. with this limit, it follows that

$$
\lim_{n \to \infty} \int f_n \, dm = \int f \, dm.
$$

**Proof.** W.l.o.g. let $h = 0$. Otherwise replace $f_n$ by $f_n - h$ and use linearity of the integral.

Let

$$
F = \{ \omega \in \Omega : \exists n \geq 1 \, \exists \, f_n(\omega) > f_{n+1}(\omega) \text{ or } f_n(\omega) < 0 \}.
$$

It follows that $m(F) = 0$ and we may assume (because of the last lemma), that $F = \emptyset$. Since $0 \leq f_n \leq f_{n+1}$, it follows immediately from the definition, that $\lim_{n \to \infty} \int f_n \, dm$ exists.

First assume that $\int f \, dm < \infty$. Let $\epsilon > 0$ and $g \leq f$ be a step function with $\int g \, dm \geq \int f \, dm - \epsilon$. Then $g$ is bounded and we can approximate the function $\min(g, f_n)$ uniformly by step functions from below. Consequently, there is a step function $0 \leq g_n \leq \min(g, f_n)$ satisfying

$$
\sup_{\omega \in \Omega} \min(g(\omega), f_n(\omega)) - g_n(\omega) < 1/n.
$$

Define $h_n = \max\{g_1, ..., g_n\}$. Then $0 \leq h_n \leq g$ are step functions as well, which approximate $g$ (since $f_n$ approximate $f$ and $g \leq f$). It also follows using the lemma before last, that

$$
\int g \, dm = \lim_{n \to \infty} I_0(h_n) \leq \lim_{n \to \infty} \int f_n \, dm.
$$

Letting $\epsilon \to 0$ the claim follows.

The case of $\int f \, dm = \infty$ is proven in a similar way.
Remark 6. (Monotone Convergence Theorem for Series)
Let \( f_n \geq 0 \) a.e. Then
\[
\int \liminf_{n \to \infty} f_n \, dm \leq \liminf_{n \to \infty} \int f_n \, dm.
\]

Theorem 8. (Lemma of Fatou)
(a) Let \( h \) be integrable and let \( f_n \geq h \) a.e. Then
\[
\limsup_{n \to \infty} \int f_n \, dm \leq \limsup_{n \to \infty} \int f_n \, dm.
\]
(b) Let \( h \) be integrable and let \( f_n \leq h \) a.e. Then
\[
\liminf_{n \to \infty} \int f_n \, dm \leq \int \liminf_{n \to \infty} f_n \, dm.
\]

Proof. Setting \( g_k = \inf_{n \geq k} f_n \), it follows that \( h \leq g_k \leq f = \liminf_{n \to \infty} f_n = \lim_{k \to \infty} g_k \). (consequently \( I(f^-) \) and \( I(g_k^-) \) are finite.) Moreover \( g_k \leq g_{k+1} \), so
\[
\int f \, dm = \lim_{k \to \infty} \int g_k \, dm \leq \liminf_{k \to \infty} \int f_k \, dm.
\]
(b) follows from (a) by replacing \( f_n \) by \(-f_n\).

Theorem 9. (Dominated Convergence Theorem, Theorem of Lebesgue)
Let \( f_n \) \((n \geq 1)\) be measurable and assume that \( f = \lim_{n \to \infty} f_n \) exists a.e. If there is an integrable function \( h \) such that \( |f_n| \leq h \) a.e., then
\[
\lim_{n \to \infty} \int f_n \, dm = \int f \, dm.
\]

Proof. Using Fatou’s lemma one obtains:
\[
\limsup_{n \to \infty} \int f_n \, dm \leq \liminf_{n \to \infty} \int f_n \, dm \leq \limsup_{n \to \infty} \int f_n \, dm.
\]

Remark 7. (1) it is sufficient to require stochastic convergence instead of the a.e. convergence (see section 1.5).
(2) Dominated Convergence Theorem for Series:
Let \( f_n \) be measurable functions, such that \( \sum_{k=1}^\infty f_k = f \) converges a.e. If \( \sum_{k=1}^\infty |f_k| \) is integrable, then
\[
\int \sum_{k=1}^\infty f_k \, dm = \sum_{k=1}^\infty \int f_k \, dm.
\]
(3) Let \( f_t \) \((t \in (a, b) \subset \mathbb{R})\) be measurable functions, which in absolute value are bounded by the integrable function \( g \) (i.e. \( |f_t| \leq g \forall t \)). Moreover, assume
there is a nullset $N$, such that for $\omega \notin N$ the function $t \mapsto f_t(\omega)$ is continuous in $t_0 \in (a, b)$. Then the function $t \mapsto \int_A f_t \, dm$ is continuous in $t_0$ for every $A \in \mathcal{F}$.

(4) Let $f_t$ ($t \in (a, b) \subseteq \mathbb{R}$) be functions, which a.e. differentiable in $t_0 \in (a, b)$ and such that the difference quotients are absolutely bounded by an integrable function $g$ ($|f_t - f_{t_0}| \leq g|t - t_0|$). Then

$$\frac{d}{dt} \int_A f_t \, dm \bigg|_{t=t_0} = \int_A \frac{d}{dt} f_t \bigg|_{t=t_0} \, dm$$

for every $A \in \mathcal{F}$.

The same result holds in case the derivatives exist in a neighborhood of $t_0$ and are bounded by $g$.

(5) Let $f_t$ ($t \in (a, b) \subseteq \mathbb{R}$) be measurable functions, which are absolutely bounded by the integrable function $g$, and assume $(t, \omega) \mapsto f_t(\omega)$ is measurable. Then for all $A \in \mathcal{F}$

$$\int_a^t \int_A f_{\tau} \, d\tau \, dm = \int_A \int_a^t f_{\tau} \, d\tau \, dm.$$

This follows from (4) with $F(t) = \int_a^t f_{\tau} \, d\tau$. 
1.4 \( L_p \)-Spaces

We begin stating some facts on Banach spaces.

**Definition 12.** A map \( \| \cdot \| : V \to \mathbb{R}_+ \)
of a vector space \( V \) into the positive real numbers is called a norm, if thefollowing properties hold:

\[
\begin{align*}
\| x \| = 0 & \iff x = 0 \\
\| x + y \| & \leq \| x \| + \| y \| \quad \forall x, y \in V \quad \text{(triangle inequality)} \\
\| ax \| & = |a| \cdot \| x \| \quad \forall x \in V, a \in \mathbb{R} \quad \text{(homogeneity)}
\end{align*}
\]

**Definition 13.** A vector space \( B \) (over \( \mathbb{R} \) or \( \mathbb{C} \)) with norm \( \| \cdot \| \) is called aBanach space, if it is complete with respect to the metric \( (x, y) \to \| x - y \| \),i.e. if \( x_n \in B \) and \( \lim_{n,m \to \infty} \| x_n - x_m \| = 0 \) implies the existence of a point \( x \in B \) such that \( \lim_{n \to \infty} \| x_n - x \| = 0 \).

**Theorem 10.** The dual space of a Banach space \( (B, \| \cdot \|) \) is the vector space\( B^* \) of all continuous linear forms on \( B \), which itself is a Banach space withnorm

\[
\| L \| := \sup_{\| x \| \leq 1} |L(x)| \quad (L \in B^*).
\]

**Remark 8.** A continuous linear form on \( B \) is consequently a linear map \( L : B \to \mathbb{R} \) (\( \mathbb{C} \)), which satisfies \( |L(x) - L(y)| \leq \| L \| \cdot \| x - y \| \).

**Definition 14.** A real Hilbert space \( H \) is vector space, which has an innerproduct (scalar product) and is complete with respect to the norm defined byit: If

\[
\langle \cdot, \cdot \rangle : H \times H \to \mathbb{R}
\]
is the inner product, then \( H \) and the norm \( \| x \| := \sqrt{\langle x, x \rangle} \) form a Banachspace.

**Remark 9.** In a Hilbert space \( H \) the Cauchy-Schwarz-inequality holds:

\[
|\langle x, y \rangle| \leq \| x \| \cdot \| y \| \quad \forall x, y \in H
\]

and equality holds if and only if \( x \) and \( y \) are colinear (\( x = \text{const.} \cdot y \)). Theproof of this fact is simple: The inequality \( 0 \leq \| ax - by \|^2 \) implies

\[
|\langle x, y \rangle| \leq \frac{1}{2} (a\| x \|^2/b + b\| y \|^2/a),
\]

and setting \( a\| x \| = b\| y \| \) the claim follows.
In the sequel, let \((\Omega, \mathcal{F}, m)\) be a measure space.

For \(p > 0\) define
\[
L_p(m) := \{ f : \Omega \to \mathbb{R} \text{ measurable : } \int |f|^p dm < \infty \}
\]
and
\[
L_\infty(m) := \{ f : \Omega \to \mathbb{R} \text{ measurable : } \exists N \in \mathcal{F}, m(N) = 0, \sup_{\omega \in \Omega \setminus N} |f(\omega)| < \infty \}
\]
\[
= \{ f : \Omega \to \mathbb{R} \text{ measurable : } \exists K \in \mathbb{R}_+ \ |f| \leq K \text{ a.e.} \}.
\]

Restricted to non-numerical functions these sets are vector spaces, as is deduced easily. Moreover, for \(0 < p \leq \infty\)
\[
N^0_p := \{ f \in L_p(m) : f = 0 \text{ a.e.} \}
\]
is also a vector space, when restricted to non-numerical functions. It defines an equivalence relation on \(L_p(m)\) by \(f \sim g \iff f - g \in N^0_p\) (or \(f = g\) a.e.).

We define finally
\[
L_p(m) := L_p(m)/N^0_p,
\]
and identify measurable functions on \(\Omega\) with their equivalence classes in \(L_p(m)\).

**Theorem 11.** For every \(0 < p \leq \infty\) \(L_p(m)\) is a vector space over \(\mathbb{R}\).

**Proof.** The claim is proved, once it is shown that every equivalence class has an a.e. finite representative. But this follows already from (I2), section I.3, for \(0 < p < \infty\) and from the definition for \(p = \infty\). Indeed, if \(f, g \in L_p(m)\), and \(\tilde{f}\) and \(\tilde{g}\) are representatives, define \(f + g\) to be the equivalence class belonging to \(\tilde{f} + \tilde{g}\). This definition is obviously independent of the choice of finite representatives. The multiplication is defined in a similar way.

**Theorem 12.** For \(1 \leq p < \infty\)
\[
\|f\|_p := \left( \int |\tilde{f}|^p \, dm \right)^{1/p} \quad (\tilde{f} \in f)
\]
is independent of the choice of representatives and defines a norm on \(L_p(m)\).

Moreover,
\[
\|f\|_\infty := \inf_{N \in \mathcal{F} : m(N) = 0} \sup_{\omega \in \Omega \setminus N} |\tilde{f}(\omega)|
\]
is as well independent from the choice of representatives and defines a norm on \(L_\infty(m)\).
Proof. Note that $\|f\|_p$ is well defined, which follows by definition in case $p = \infty$ and follows from the section I.3 for all other cases of $p$.

Likewise the equivalence $\|f\|_p = 0 \iff f = 0$ is immediately clear in case $p = \infty$ and follows from section I.3, (16), otherwise.

Homogeneity is easy to show and the triangle inequality follows from Minkowski’s inequality which will be proven below.

Remark 10. Because of independence of the representing functions an equivalence class is usually identified with any representative. This means $f \in L_p(m)$ means a function $f$, for which its equivalence class belongs to $L_p(m)$.

Theorem 13. (Hölder Inequality)

Let

$$1 \leq p, q \leq \infty \quad \text{and} \quad \frac{1}{p} + \frac{1}{q} = 1.$$ 

Then for $f \in L_p(m)$ and $g \in L_q(m)$ it follows that $f \cdot g \in L_1(m)$ and for all $F \in \mathcal{F}$ one has

$$\int_F |fg| \, dm \leq \left( \int_F |f|^p \, dm \right)^{1/p} \left( \int_F |g|^q \, dm \right)^{1/q}.$$

In short:

$$\|I_F fg\|_1 \leq \|I_F f\|_p \cdot \|I_F g\|_q.$$ 

In case of $1 < p < \infty$ equality holds if and only if $|f|^p/|g|^q = \text{const.}$ a.e. on $F$.

Proof. For $p = \infty$ (or $q = \infty$) it holds that

$$\|f\|_p \leq \|f\|_\infty \quad \text{a.e.},$$ 

hence by section I.3, (15),

$$\int_F |fg| \, dm \leq \int_F \|I_F f\|_\infty \cdot |g| \, dm \leq \|I_F f\|_\infty \|I_F g\|_1.$$ 

Now let $p, q > 1$. Using

$$- \log(\alpha x + \beta y) \leq -\alpha \log x - \beta \log y$$

where $\alpha + \beta = 1$ it follows that

$$x^\alpha y^\beta \leq \alpha x + \beta y,$$

where equality holds if and only if $x = y$.

If $\|I_f f\|_p = 0$ or $\|I_f g\|_q = 0$, then by section I.3, (16), nothing has to be shown. Hence it is left to consider the case when both integrals $\neq 0$. In this case the above inequality gives
THEORY OF INTEGRATION

\[
\frac{|f(\omega)| \cdot |g(\omega)|}{\|I_Ff\|_p \|I_Fg\|_q} \leq \frac{1}{p} \frac{|f|^p(\omega)}{\|I_Ff\|_p^p} + \frac{1}{q} \frac{|g|^q(\omega)}{\|I_Fg\|_q^q}
\]

and thus the claim follows by integration over \(F\), in particular, the integrability of \(fg\).

In order to show the equality statement, consider first the case when \(c|f|^p = |g|^q\) on \(F\). Then \(\int |I_Fg|^q dm = c \int |I_Ff|^p dm\) and

\[
\frac{|g(\omega)|^q}{\int |I_Fg|^q dm} = \frac{|f(\omega)|^p}{\int |I_Ff|^p dm} \quad \forall \omega \in F.
\]

Hence we have \(\leq\) in (*) and equality in Hölder’s inequality.

On the other hand, if \(c|f|^p = |g|^q\) does not hold a.e. on \(F\), there exists a subset \(G \subset F\) of positive measure, on which (**) does not hold (if (**) would hold on \(F\) then \(c|f|^p = |g|^q\) for some \(c\)), hence also strict inequality in (*).

Integration of (*) over \(G\) yields

\[
\int_G |fg| dm < \frac{1}{p} \|I_Ff\|_p^{1-p} \|I_Fg\|_q \int_G |f|^p dm + \frac{1}{q} \|I_Fg\|_q^{1-q} \|I_Ff\|_p \int_G |g|^q dm.
\]

On \(F \setminus G\) this inequality holds with \(\leq\), whence

\[
\int_F |fg| dm < \left(\int_F |f|^p dm\right)^{1/p} \left(\int_F |g|^q dm\right)^{1/q}.
\]

Theorem 14. (Minkowski Inequality)

Let \(p \geq 1\) and \(f, g \in L_p(m), F \in \mathcal{F}\). Then

\[
\|I_F(f + g)\|_p \leq \|I_Ff\|_p + \|I_Fg\|_p.
\]

Proof. The cases \(p = 1\) and \(p = \infty\) are trivial.

Hence we need to consider the case \(1 < p < \infty\). An application of Hölder’s inequality yields

\[
\int_F |f + g|^p dm = \int_F |f + g| \cdot |f + g|^{p-1} dm
\]

\[
\leq \int_F |f| \cdot |f + g|^{p-1} dm + \int_F |g| \cdot |f + g|^{p-1} dm
\]

\[
\leq \left(\int_F |f|^p dm\right)^{1/p} \cdot \left(\int_F |f + g|^p dm\right)^{1-1/p} + \left(\int_F |g|^p dm\right)^{1/p} \cdot \left(\int_F |f + g|^p dm\right)^{1-1/p}.
\]

This implies

\[
\left(\int_F |f + g|^p dm\right)^{1/p} \leq \left(\int_F |f|^p dm\right)^{1/p} + \left(\int_F |g|^p dm\right)^{1/p}.
\]
Remark 11. (1) Minkowski’s inequality shows the triangle inequality for $\|\cdot\|_p$ completing the proof that it is a norm.

(2) In case $m(\Omega) < \infty$ one obtains as a special case of the Hölder inequality that

$$\int |f|^q dm \leq \left( \int |f|^p dm \right)^{1/r} m(\Omega)^{1/s}$$

for any $q > 0$ and $1/r + 1/s = 1$. Choosing $r = p/q$ with $p > q$, it follows that

$$\int |f|^q dm \leq \left( \left( \int |f|^p dm \right)^{1/p/q} \right)^{q} m(\Omega)^{(r-1)/r} < \infty.$$ 

Consequently

$L_p(m) \subset L_q(m)$.

This is no longer true in case of an infinite measure space.

Theorem 15. (Fischer-Riesz)
The normed vector spaces $L_p(m)$ ($1 \leq p \leq \infty$) are Banach spaces and $L_2(m)$ is a Hilbert space with inner product

$$\langle f, g \rangle := \int f \cdot \overline{g} \, dm \quad f, g \in L_2(m).$$

Proof. First of all note that $\langle \cdot, \cdot \rangle$ is indeed an inner product, and

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

for any $f \in L_2(m)$. Hence we only need to show completeness in view of the preceding discussions in this section.

Let $f_n \in L_p(m)$ be a Cauchy sequence, i.e.

$$\lim_{r,n \to \infty} \|f_n - f_r\|_p = 0.$$ 

Note that we do not distinguish between equivalence classes and representatives! We need to show that some subsequence converges to a function in the vector space.

(1) $p = \infty$.

Let

$$F_{r,n} := \{ \omega \in \Omega : |f_n(\omega) - f_r(\omega)| \leq \|f_n - f_r\|_\infty \},$$

and let

$$N := \bigcup_{r,n=1}^\infty F_{r,n}^c.$$ 

Then $m(N) = 0$, and for $\omega \notin N$ the sequence $\{f_n(\omega) : n \geq 1\}$ is Cauchy in $\mathbb{R}$. Therefore there exists a function $f$ with
Theorem of Integration

\[ f(\omega) = \begin{cases} 
\lim_{n \to \infty} f_n(\omega) & \text{if } \omega \notin N \\
0 & \text{if } \omega \in N.
\end{cases} \]

Clearly \( f \) is measurable by section I.2, (M7), and it follows that

\[ \|f_n - f\|_\infty \leq \sup_{\omega \notin N} |f_n(\omega) - f(\omega)| \leq \lim_{r \to \infty} \|f_n - f_r\|_\infty \to 0 \]

by the choice of \( N \).

(2) \( 1 \leq p < \infty \).

W.l.o.g., choosing a subsequence if necessary, we may assume that \( \|f_n - f_{n+1}\|_p \leq 2^{-n} \) holds. Let

\[ h := |f_1| + \sum_{k \geq 2} |f_k - f_{k-1}|. \]

Using the theorem of monotone convergence and Minkowski’s inequality one gets \( \|h\|_p \leq \|f_1\|_p + \sum_{k \geq 2} 2^{-k} < \infty \), and hence \( h \) is a \( L_p(m) \)-function, in particular it is a.e. finite. Hence there is a set \( N \) of measure 0, such that \( \sum_{k \geq 2} |f_k - f_{k-1}| \) is convergent on \( N^c \), and the sequence \( \{f_n(\omega) : n \geq 1\} \) is Cauchy for \( \omega \notin N \). Like in the case \( p = \infty \) there exists a measurable function \( f = \lim_{n \to \infty} f_n \) a.e., which satisfies \( |f| \leq h \). The theorem of dominated convergence (observe that \( |f - f_n|^p \leq 2^p h^p \) and \( |f - f_n| \to 0 \) a.e.)

\[ \lim_{n \to \infty} \int |f - f_n|^p \, dm = \int \lim_{n \to \infty} |f - f_n|^p \, dm = 0. \]

Hence \( \|f - f_n\|_p \to 0 \).

Remark 12. The Cauchy-Schwarz-inequality in \( L_2(m) \) reads:

\[ \int |fg| \, dm \leq \left( \int f^2 \, dm \right)^{1/2} \left( \int g^2 \, dm \right)^{1/2} \]

for square-integrable functions \( f \) and \( g \).
1.5 Convergence of Functions

Let \((\Omega, \mathcal{F}, m)\) be a measure space with \(m(\emptyset) = 0\).

**Definition 15.** A sequence of measurable (numerical) functions \(f_n, n \geq 1\), converges almost surely (almost everywhere), if there is a nullset \(N \in \mathcal{F}\), such that for \(\omega \notin N\)

\[
f(\omega) = \lim_{n \to \infty} f_n(\omega)
\]

exists (i.e. is finite, \(\infty\) or \(-\infty\)). Then \(f\) can be extended to \(N\) as a (numerical) function, and then \(f_n\) is called to be convergent a.e. (a.s.) to \(f\).

**Definition 16.** A sequence of measurable functions \(f_n (n \geq 1)\) converges (locally) stochastically to \(f\), if for all \(A \in \mathcal{F}\) with \(m(A) < \infty\) and for all \(\epsilon > 0\) one has:

\[
\lim_{n \to \infty} m(A \cap \{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \epsilon\}) = 0.
\]

If this holds for all measurable sets \(A \in \mathcal{F}\) (equivalently for \(A = \Omega\)), the convergence is called stochastic convergence in the general sense.

**Definition 17.** A sequence of functions \(f_n \in L^p(m) (n \geq 1)\) converges in the \(p\)-th mean towards function \(f \in L^p(m)\), if

\[
\lim_{n \to \infty} \int_{\Omega} |f_n - f|^p dm = 0
\]

holds. If \(p \geq 1\), this is equivalent of saying that \(\lim_{n \to \infty} \|f_n - f\|_p = 0\). For \(p = \infty\) one speaks of \(f_n\) converging to \(f\) in sup-norm.

**Definition 18.** A family \(I\) of measurable (numerical) functions is called uniformly integrable, if for any \(\epsilon > 0\) there is a non-negative, integrable function \(h\), such that

\[
\int_{\{|f| \geq h\}} |f| dm < \epsilon
\]

for every \(f \in I\).

**Remark 13.** If \(f\) is a function of a uniformly integrable family, then it is integrable itself, since

\[
\int |f| dm = \int_{\{|f| \geq h\}} |f| dm + \int_{\{|f| < h\}} |f| dm \leq \int h dm < \infty.
\]

If \(I\) is a uniformly integrable family and if \(h\) is integrable, then \(\{f + h : f \in I\}\) is uniformly integrable.

**Example 4.** Let \([0,1]\), equipped with Lebesgue-measure. Consider
(a) $f_n = 1_{[0,1/n]}$. Then $f_n \to 0$ a.e., stochastically and in the $p$-th mean for $p < \infty$. For $p = \infty$ one has $\|f_n\|_\infty = 1$.

(b) $g_n = nf_n$. Then $g_n \to 0$ a.e. and stochastically. However, there is no convergence in mean (1-st mean), and the sequence is not uniformly integrable.

(c) For $2^{k-1} \leq n < 2^k$ let

$$h_n = 1_{[n2^{-k},(n+1)2^{-k}]}.$$ 

Then $h_n$ converges to 0 stochastically and in the $p$-th mean for $p < \infty$. convergence is not true.

(d) Let $u_n = nh_n$. Then only stochastic convergence holds.

**Theorem 16.** Let $f_n (n \geq 1)$ be measurable functions. Consider the following statements:

(A) $f_n$ converges (locally) stochastically to $f$ and is uniformly integrable.

(B) $f_n$ converges to $f$ in mean.

(C) $f_n$ converges to $f$ stochastically in the general sense.

(D) $f_n$ converges to $f$ (locally) stochastically.

(E) $f_n$ converges to $f$ almost everywhere.

Then the following implications hold:

$$ (A) \iff (B) \Rightarrow (C) \Rightarrow (D), $$

$$ (E) \Rightarrow (D).$$

The converse implications are not true in general.

If $m(\Omega) < \infty$, convergence in the $p$-th mean implies that one in the $q$-th mean for $1 \leq q \leq p \leq \infty$.

**Proof.** The last assertion follows from Hölder’s inequality as in section I.4. The implication ‘(C)$\iff$(D)’ is trivial. Counterexamples to the converse implications are contained in the above example.

‘(E)$\Rightarrow$(D)’. Let $\epsilon > 0$ and $A \in \mathcal{F}$ with $m(A) < \infty$, $N = \{\omega \in \Omega : \lim_{n \to \infty} f_n(\omega) \neq f(\omega)\}$ and

$$A_n := \{\omega \in \Omega \setminus N : \exists m \geq n : |f_m(\omega) - f(\omega)| > \epsilon\} \cap A.$$ 

The sets $A_n$ are monotonically decreasing and because of the convergence on $\Omega^c$ it follows that $\bigcap_{n \geq 1} A_n = \emptyset$, whence

$$\lim_{n \to \infty} m(A_n) = 0,$$

because it is a set of finite measure. This yields

$$\lim_{n \to \infty} m(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \epsilon\} \cap A) = \lim_{n \to \infty} m(N) + m(A_n) = 0.$$
Convergence of Functions

\[ m(\{ \omega \in \Omega : g(\omega) \geq a \}) = \int I_{\{g \geq a\}} \, dm \]
\[ \leq \int I_{\{g \geq a\}} h(g) \frac{1}{h(a)} \, dm \]
\[ \leq \frac{1}{h(a)} \int I_{\{g > a\}} h(g) \, dm. \]

Therefore
\[ m(\{|f_n - f| > \epsilon\}) \leq \epsilon^{-1} \|f_n - f\|_1 \to 0. \]

'(B)⇒(C)’. Use the Markov-inequality: Let \( h \) be a monotone function. Then
\[ m(\{|f_n - f| > \epsilon\}) \leq \epsilon^{-1} \|f_n - f\|_1 \to 0. \]

'(B)⇒(A)’. It suffices to show that the functions \( f_n \) form a uniformly integrable family. Let \( \epsilon > 0 \). Since
\[ \int (|f_n| - |f|)^+ \, dm \leq \int |f_n - f| \, dm, \]
there is \( N \in \mathbb{N} \) with \( \int (|f_n| - |f|)^+ \, dm < \epsilon \) for all \( n > N \). Put \( h = |f| + |f_1| + \ldots + |f_N| \). It follows for every \( n \geq 1 \)
\[ \int (|f_n| - h)^+ \, dm < \epsilon, \]
and furthermore
\[ \int I_{\{|f_n| \geq 2h\}} |f_n| \, dm \]
\[ = \int I_{\{|f_n| \geq 2h\}} h \, dm + \int I_{\{|f_n| \geq 2h\}} (|f_n| - h) \, dm \]
\[ \leq 2 \int I_{\{|f_n| \geq 2h\}} (|f_n| - h)^+ \, dm \quad (h \leq |f_n| - h \text{ on } \{|f_n| \geq 2h\}) \]
\[ \leq 2 \int (|f_n| - h)^+ \, dm \leq 2\epsilon. \]

'(A)⇒(B)’. Use the uniform continuity at the \( \emptyset \) of the maps
\[ B \in \mathcal{F} \to \int_B g \, dm \]
for \( g \) in a uniformly integrable family of positive functions. This can be seen as follows: Let \( \epsilon > 0 \). There is a positive integrable function \( h \) with
\[ \int (g \geq h) \, g \, dm \leq \epsilon/3. \]
Next there exists a constant \( c \) satisfying \( \int (h - c)^+ \, dm \leq \epsilon/3. \) Let \( B \in \mathcal{F} \) such that \( m(B) \leq \epsilon/3c. \) Then
\[ \int_B g \, dm \leq \int_{B \cap \{g \geq h\}} g \, dm + \int_B h \, dm \]
\[ \leq \frac{\epsilon}{3} + cm(B) + \int_B (h - c)^+ \, dm \leq \epsilon. \]
shows the claim.
After this preparation assume w.l.o.g., that $f_n \geq 0$ and $f = 0$ (otherwise consider $\|f_n - f\|$). Let $\eta > 0$. Choose $h \geq 0$ integrable with $\int_{\{f_n \geq h\}} f_n \, dm \leq \eta$. Now choose $A \in \mathcal{F}$ with $m(A) < \infty$ and $\int_{\Omega \setminus A} h \, dm \leq \eta$ (this set exists, since $h$ is integrable). It follows for $\xi > 0$:

$$\int f_n \, dm = \int_A f_n \, dm + \int_{A^c} f_n \, dm \leq \int_A f_n \, dm + \int_{A^c} h \, dm + \int_{A^c} (f_n - h)^+ \, dm$$

$$\leq \int_{A \cap \{f_n \geq \xi\}} f_n \, dm + \int_{A \cap \{f_n < \xi\}} f_n \, dm + \int_{A^c} h \, dm + \int_{\{f_n \geq h\}} f_n \, dm.$$ 

By stochastic convergence $m(A \cap \{f_n \geq \xi\}) \to 0$, hence by the uniform continuity of the integrals

$$\int_{A \cap \{f_n \geq \xi\}} f_n \, dm \leq \eta$$

for all sufficiently large $n$.

This yields

$$\int f_n \, dm \leq 3\eta + \xi m(A).$$

Letting $\eta$ and $\xi$ tend to zero, the claim follows.

In the last proof the following results have been proven as a by-product:

**Corollary 1.** (Markov-inequality)

Let $h$ be a monotone function, such that $I(h \circ f)$ exists. Then

$$m(\{\omega \in \Omega : f(\omega) \geq a\}) \leq \frac{1}{h(a)} \int I_{\{f > a\}} h(f) \, dm.$$ 

**Corollary 2.** ($\emptyset$-continuity of integrals)

Let $\mathcal{I}$ be a uniformly integrable family of positive functions. Then for every $\epsilon > 0$ there is $\delta > 0$, such that $m(B) \leq \delta$ implies

$$\int_B g \, dm \leq \epsilon \quad \forall g \in \mathcal{I}.$$ 

**Corollary 3.** Let $f_n$ converge in the mean to $f$ and $1_{A_n}$ stochastically to $1_A$. Then

$$\lim 1_{A_n} f_n = 1_A f \quad \text{in mean}.$$ 

**Proof.** $f_n \to f$ in mean implies, that $\{f_n\}$ is uniformly integrable and stochastic convergence. But then the sequence $1_{A_n} f_n$ is uniformly integrable and stochastically convergent to $1_A f$.
1.5 Convergence of Functions

Theorem 17. Let $m$ be $\sigma$-finite. The following two assertions are equivalent:

(a) The sequence $f_n$ ($n \geq 1$) converges stochastically to $f$.
(b) Every subsequence of $f_n$ has a further subsequence which converges to $f$.

Proof. 

(a)$\Rightarrow$ (b). W.L.O.G. let $f_n \geq 0$ and $f = 0$. Fix $E_k \in \mathcal{F}$ with $E_k \uparrow \Omega$ and $m(E_k) < \infty$. W.L.O.G. let $f_n$ denote the arbitrary subsequence (renumbering the subsequence). Then there is a subsequence $f_{n_k}$ satisfying

$$m(E_k \cap \{f_{n_k} \geq 2^{-k}\}) \leq 2^{-k}.$$ 

Let $F_k := E_k \cap \{f_{n_k} \geq 2^{-k}\}$. By the theorem of monotone convergence this implies

$$\int \sum_{k=1}^{\infty} I_{F_k} \, dm < \infty,$$

hence also $\sum_{k=1}^{\infty} I_{F_k} < \infty$ a.e.. Thus there is a nullset $N$, such that $\omega \not\in N$ does not lie in any $F_k$ for all large $k$, i.e.

$$\lim_{k \to \infty} f_{n_k}(\omega) = 0.$$

(b)$\Rightarrow$ (a). Suppose, the conclusion is wrong. Then there is some $\eta > 0$, $A \in \mathcal{F}$ and a subsequence $n_k$ with $m(A) < \infty$ and

$$m(A \cap \{f_{n_k} \geq \eta\}) \geq \eta$$

for every $k \geq 1$. Using (b) it follows that this subsequence has a further subsequence which converges a.e., hence also stochastically. A contradiction.

Theorem 18. (General Fatou Lemma)

Let $f_n$ be measurable functions, such that the family $\{f_n\}$ is uniformly integrable and

$$\liminf_{n \to \infty} \int f_n \, dm < \infty.$$

Then

$$\int \liminf_{n \to \infty} f_n \, dm \leq \liminf_{n \to \infty} \int f_n \, dm.$$

Proof. Let $\eta > 0$. Choose $g \geq 0$ integrable with $\int_{\{f_n \geq g\}} f_n^- \, dm \leq \eta$. Let

$$g_n := \max(f_n, -g).$$

Then $g_n \geq -g$ and $-g$ are integrable. Fatou lemma yields
\[ \int \lim \inf g_n \, dm \leq \lim \inf \int g_n \, dm \]
\[ = \lim \inf \int_{\{g_n \geq -g\}} f_n \, dm + \int_{\{g_n < -g\}} -g \, dm \]
\[ \leq \lim \inf \int f_n \, dm + \lim \sup \int_{\{f_n > g\}} f_n^- \, dm \]
\[ \leq \lim \inf \int f_n \, dm + \eta. \]

Since \( \lim \inf f_n \leq \lim \inf g_n \), it follows that also \( \lim \inf f_n \) is integrable, and the theorem follows from the monotonicity of the integrals.

**Theorem 19. (Theorem of Egorov)**

Assume that \( m(\Omega) < \infty \) and let \( f_n \ (n \geq 1) \) be a towards \( f \) a.e. convergent sequence of measurable functions. For any \( \epsilon > 0 \) there exists a set \( A \in \mathcal{F} \) with \( m(\Omega \setminus A) < \epsilon \), such that \( \{f_n\} \) converges to \( f \) uniformly on \( A \).

**Proof.** Let \( \eta_k \to 0 \). There is \( N_k \), such that

\[ E_k = \{ \omega \in \Omega : \forall n \geq N_k \ |f_n(\omega) - f(\omega)| < \eta_k \} \]

\[ m(E_k) = m(\Omega) - 2^{-k} \]. Choose \( k \) mit \( \sum_{j \geq k} 2^{-j} \leq \epsilon \), \( A = \bigcap_{j \geq k} E_j \). It follows that uniform convergence holds on \( A \).
1.6 The Theorem of Radon-Nikodym

In this section \((\Omega, \mathcal{F})\) is a measurable space (fixed).

**Definition 19.** Let \(\Lambda\) be a signed measure on \((\Omega, \mathcal{F})\). A measurable set \(A \in \mathcal{F}\) is called

(a) negative, if for all \(F \in \mathcal{F} \cap A\) \(\Lambda(F) \leq 0\).

(b) positive, if for all \(F \in \mathcal{F} \cap A\) \(\Lambda(F) \geq 0\).

**Theorem 20.** (Hahn Decomposition)

Let \(\Lambda\) be a signed measure on \((\Omega, \mathcal{F})\) satisfying \(\Lambda(F) > -\infty\) \(\forall F \in \mathcal{F}\).

Then there is a negative set \(N \in \mathcal{F}\), such that \(\Omega \setminus N\) is positive. This splitting of \(\Omega\) into a positive and a negative set is called the Hahn decomposition of \(\Lambda\).

In particular, every signed measure \(\Lambda\) can be written as the difference of two finite measures. More precisely:

\[ \Lambda = \Lambda^+ - \Lambda^- \]

\[ \Lambda^+(F) = \Lambda(F \cap (\Omega \setminus N)) \quad (F \in \mathcal{F}) \]

\[ \Lambda^-(F) = -\Lambda(F \cap N) \quad (F \in \mathcal{F}). \]

**Proof.** Define \(\mathcal{D} := \{F \in \mathcal{F} : F \text{ is negative}\}\).

Let \(F_k \in \mathcal{D}, k \geq 1\). Put \(E_1 = F_1, E_2 = F_2 \setminus (E_1 \cap F_2)\) and in general

\[ E_n = F_n \setminus [(E_1 \cup E_2 \cup \ldots \cup E_{n-1}) \cap F_n]. \]

Then every \(E_n\) is negative, so \(E_n \in \mathcal{D}\), and these sets are disjoint. Their (disjoint) union is also negative, since for \(A \subset \bigcup_{n \geq 1} E_n\)

\[ \Lambda(A) = \sum_{n \geq 1} \Lambda(A \cap E_n) \leq 0. \]

It follows that

\[ \bigcup_{n \geq 1} F_n \in \mathcal{D}. \]

Now let \(\alpha := \inf \{\Lambda(F) : F \in \mathcal{D}\} \in \mathbb{R} \cup \{-\infty\}\).

Choosing \(F_k \in \mathcal{D}\) such that \(\alpha = \inf \Lambda(F_k)\) we just showed that \(N := \bigcup_{n \geq 1} F_n\) belongs to \(\mathcal{D}\) and satisfies

\[ \alpha \leq \Lambda(N) = \sum_{n \geq 1} \Lambda(E_n) \leq \sum_{1 \leq k \leq n} \Lambda(E_k) = \Lambda(F_n) \to \alpha \quad \text{for } n \to \infty. \]
It follows that $\Lambda(N) = \alpha > -\infty$, and we have to show that $P = \Omega \setminus N$ is positive.

Assume that $P$ is not positive. Then there is a measurable set $N' \subset P$ with $\Lambda(N') < 0$. We define sets $P_k$ as follows:

\[
\beta_1 := \sup\{\Lambda(A) : A \subset N'\}; \quad \Lambda(P_1) > \frac{\min(\beta_1, 1)}{2} \\
\beta_2 := \sup\{\Lambda(A) : A \subset N' \setminus P_1\}; \quad \Lambda(P_2) > \frac{\min(\beta_2, 1)}{2} \\
\vdots \\
\beta_n := \sup\{\Lambda(A) : A \subset N' \setminus (P_1 \cup \ldots \cup P_{n-1})\}; \quad \Lambda(P_n) > \frac{\min(\beta_n, 1)}{2} \\
\vdots
\]

Note that $\beta_1 > 0$ since if $N'$ is negative, $\Lambda(N \cup N') < \alpha$ a contradiction. Moreover, $\beta_k \geq 0$.

Define $P_0 = \bigcup_{k \geq 1} P_k$. Then $0 < \Lambda(P_0) < \infty$ since $\Lambda(N') = \Lambda(N' \setminus P_0) + \Lambda(P_0)$ and since $\Lambda(P_0) = \infty$ implies that $\Lambda(N' \setminus P_0) = -\infty$. It follows from this that $\Lambda(N' \setminus P_0) < 0$ and $\lim_{k \to \infty} \beta_k = 0$ since

\[
\infty > \Lambda(P_0) > \frac{1}{2} \sum \min(\beta_k, 1).
\]

Let $F \subset N' \setminus P_0 \subset N' \setminus P_k$ (for all $k \geq 1$). By definition of $P_k$

\[
\Lambda(F) \leq \Lambda(N' \setminus P_k) \leq \beta_{k+1},
\]

so $\Lambda(F) \leq 0$. We have shown that $N' \setminus P_0$ is negative. Hence the set

\[
N_0 = N \cup (N' \setminus P_0)
\]

is negative and satisfies $\Lambda(N_0) < \alpha$. A contradiction.

**Definition 20.** Let $m$ be a measure on $(\Omega, \mathcal{F})$. A signed measure $\Lambda$ on $(\Omega, \mathcal{F})$ is called

(a) non-singular (with respect to $m$) ($h \ll m$), if for every $F \in \mathcal{F}$ with $m(F) = 0$ it follows that $\Lambda(F) = 0$.

(b) absolutely continuous (with respect to $m$) if

\[
\forall \epsilon > 0 \exists \delta > 0 \text{ such that } \forall F \in \mathcal{F} \text{ with } m(F) < \delta \implies |\Lambda(F)| < \epsilon.
\]

(c) orthogonal to $m$ ( $\Lambda \perp m$), if there is a measurable set $T$ with $m(T) = 0$ and $\Lambda(F) = 0$ for all $F \subset \Omega \setminus T$. A set $T$ with $\Lambda(F) = 0$ for all $F \subset \Omega \setminus T$ is called a support of $\Lambda$.

**Remark 14.** Let $m$ be a measure on $(\Omega, \mathcal{F})$. 

1.6 The Theorem of Radon-Nikodym

1.) If the signed measure $\Lambda$ is finite, then $\Lambda$ is absolutely continuous with respect to $m$, if and only if it is non-singular with respect to $m$. We only have to show the implication $\Leftarrow$. Assume $\Lambda$ is non-singular. In case the claim is wrong, there are $\varepsilon > 0$ and sets $F_k$ satisfying $m(F_k) < 2^{-k}$ and $|\Lambda(F_k)| > \varepsilon$. W.l.o.g. we may assume that $\Lambda(F_k) \geq 0$, since otherwise we can take a subsequence. Setting

$$ G_k := \bigcup_{j \geq k} F_k \quad \text{and} \quad G := \bigcap_{j \geq 1} G_j, $$

it follows from section I.1 (Mi) that $m(G) = 0$ and $\Lambda(G) \geq \varepsilon$, a contradiction.

If $\Lambda$ is not finite, the claim is wrong. One can choose $\Lambda(\{k\}) = 1$, $m(\{k\}) = 2^{-k}$ for $k \geq 1$.

2.) Let $\mathcal{E}$ denote the set of all measurable functions $f$, for which the integral

$$ \int f^- \, dm < \infty. $$

then

$$ \nu_f(F) = \int 1_F \cdot f \, dm \quad (F \in \mathcal{F}) $$

defines a signed measure on $(\Omega, \mathcal{F})$. $\sigma$-additivity follows from the theorem of monotone convergence (B. Levi). If two functions are equal a.e., then the corresponding signed measures are also equal. By definition $\nu_f \ll m$.

Restricted to integrable functions ($L^1(m)$) the operator

$$ \nu : L_1(m) \to \{ \nu \ll m : \nu \text{ is a finite signed measure } \ll m \} $$

defined by $\nu(f) = \nu_f$ is linear and positive. $\nu_{f+g} = \nu_f + \nu_g$ is always true. If $f \geq 0$, then we have $g \in L_1(\nu_f) \iff g \cdot f \in L_1(m)$.

Theorem 21. (Theorem of Radon-Nikodym)

Let $m$ be a $\sigma$-finite measure on $(\Omega, \mathcal{F})$. The map

$$ \nu : \mathcal{E} \to \{ A : A \text{ is a signed measure } \ll m, \ A(F) > -\infty \forall F \in \mathcal{F} \} $$

defined by

$$ \nu(f)(A) = \nu_f(A) = \int_A f \, dm \quad (A \in \mathcal{F}) $$

is a bijection (1-1 and onto). In particular, $\nu$ defines a bijection of $L_1(m)$ into the vector space of all non-singular finite signed measures.

Proof. Injectivity of $\nu$ is simple. Let $f, g \in \mathcal{E}$, such that $m(\{f > g\}) > 0$ (so $f$ and $g$ are not equal a.e.). Then there are $\delta > 0$ and a measurable set $A \subset \{ f \geq g + \delta \}$ with $0 < m(A) < \infty$ and $\int_A g \, dm < \infty$. (if this integral
equals $\infty$ for all such subsets $A$, then $g = \infty$ a.e. on $A$, which is impossible!

It follows immediately that

$$\nu_f(A) = \int_A f \, dm \geq \int_A (g + \delta) \, dm = \int_A g \, dm + \delta m(A) > \nu_g(A).$$

In order to show that the map is onto, assume first that the measure is finite.

(1) \textbf{Lemma:} Let $\Lambda \ll m$ be a positive, finite signed measure (a finite measure), not equal to 0. Then there is a function $g \in L_1(m)$, $g \geq 0$, $g \neq 0$, such

that $\int_A g \, dm \leq \Lambda(A)$ for every $A \in \mathcal{F}$.

This can be seen as follows. There is $\alpha > 0$ with $\alpha m(\Omega) < \Lambda(\Omega)$. Let $\Omega = N \cup P$ the Hahn decomposition of the finite signed measure $\Lambda - \alpha m$. Defining $g = \alpha \chi_P$, it follows that

$$A \in \mathcal{F} \cap P \implies \int_A g \, dm = \alpha \int_{A \cap P} dm = \alpha m(\emptyset) = 0 \leq \Lambda(A)$$

$$A \in \mathcal{F} \implies \int_A g \, dm = \int_{A \cap P} g \, dm + \int_{A \cap N} g \, dm \leq \Lambda(A).$$

(2) Let $\Lambda$ be a fixed finite measure, which is non-singular with respect to $m$. Let $\mathcal{G} = \{g \in L_1(m) : \nu(g) \leq \Lambda\}$. First observe that $\mathcal{G}$ is closed under forming the maximum of two functions $g_1, g_2 \in \mathcal{G}$:

$$\nu_{g_1 \vee g_2}(A) = \int_A g_1 \vee g_2 \, dm = \int_{A \cap \{g_1 > g_2\}} g_1 \, dm + \int_{A \cap \{g_1 \leq g_2\}} g_2 \, dm$$

$$\leq \Lambda(A \cap \{g_1 > g_2\}) + \Lambda(A \cap \{g_1 \leq g_2\}) = \Lambda(A)$$

holds for any measurable set $A$.

Let $u = \sup\{ \int g \, dm : g \in \mathcal{G} \}$, and let $g_n \in \mathcal{G}$ satisfy $\int g_n \, dm \to u$. Since $\mathcal{G}$ is closed under taking maxima we may assume that the $g_n$ increase monotonically (and converge to a possibly numerical function $g$), and because of the theorem of monotone convergence we obtain

$$\nu_g(A) = \int_A g \, dm = \lim_{n \to \infty} \int_A g_n \, dm \leq \Lambda(A) \quad (A \in \mathcal{F}),$$

in particular $\int g \, dm < \infty$. This function $g$ is the one we need: Assume there is some $A \in \mathcal{F}$ satisfying $\nu_g(A) < \Lambda(A)$. Then $\Lambda - \nu_g$ is a finite measure, not equal to 0, and using the lemma one obtains a function $g^* \geq 0$, $\neq 0$ with

$$\int_A g^* \, dm \leq \Lambda(A) - \int_A g \, dm \quad (A \in \mathcal{F}).$$

It follows that $g + g^* \in \mathcal{G}$ and $\int g + g^* \, dm = u + \int g^* \, dm > u$, a contradiction!
(3) In a next step let $A$ be an arbitrary measure, which is non-singular with respect to $m$. For $n \in \mathbb{N}$ let $N_n$ denote the negative set for the signed measure $A - nm$. Then $A_n(A) = A_N(A) \cap A$ $(A \in \mathcal{F})$ is a finite measure, and there is a function $g_n$ with
\[ \int_A g_n \, dm = A(A \cap N_n) \quad (A \in \mathcal{F}). \]
By definition the sets $N_n$ are monotonically increasing, hence also the functions $g_n$. Let $N = \bigcup_{n \geq 1} N_n$ and $g = \lim_{n \to \infty} g_n$ on $N$, $g = \infty$ on $\Omega \setminus N$. Obviously $A(F) = \infty$ for $F \subset \Omega \setminus N$, hence by the theorem of monotone convergence
\[ A(F) = A(F \cap (\Omega \setminus N)) + A(F \cap N) = \int_{F \cap (\Omega \setminus N)} g \, dm + \lim_{n \to \infty} \int_{F \cap N_n} g_n \, dm = \int_F g \, dm. \]
(4) To conclude the proof, let $A$ be an arbitrary non-singular signed measure which satisfies $A(F) > -\infty$ $(F \in \mathcal{F})$. Then by (3) there are (considering the positive signed measures $A^\pm$) functions $g^\pm$ with $A^\pm(F) = \int_F g^\pm \, dm$ for $F \in \mathcal{F}$. In particular $\int g^- \, dm < \infty$. It follows that $A(F) = A^+(F) - A^-(F) = \int_F g^+ - g^- \, dm$, and $g = g^+ - g^-$ is the desired function in $\mathcal{E}$.

Finally we need to extend the proof to any $\sigma$-finite measure $m$. Fix $E_n \in \mathcal{F}$ satisfying $E_n \uparrow \Omega$ and $m(E_n) < \infty$. Then $m_n(F) = m(F \cap E_n)$ $(F \in \mathcal{F})$ are finite measures and $A_n(F) = A(F \cap E_n)$ are signed measures with $A_n \ll m_n$. By the above proof there are functions $g_n$ with
\[ A(E_n) = A_n(F) = \int_{F \cap E_n} g_n \, dm_n = \int_{F \cap E_n} g_n \, dm. \]
The functions $\|E_n g_n$ can be chosen monotonically increasing, and thus their limit $g$ (which can be infinite) satisfies
\[ A(F) = \lim_{n \to \infty} A_n(F) = \lim_{n \to \infty} \int_F \|E_n g_n \, dm = \int_F g \, dm. \]
This concludes the proof.

**Definition 21.** Let $m$ be a measure and $A$ a signed measure on $(\Omega, \mathcal{F})$. If there is a measurable function $f$ satisfying
\[ A(F) = \int_F f \, dm \quad (F \in \mathcal{F}), \]
then $f$ is called the density of $A$ with respect to $m$, or the Radon-Nikodym derivative of $A$ with respect to $m$. It is denoted by
\[ f = \frac{dA}{dm}. \]
If $m$ is $\sigma$-finite and $A \ll m$, $A > -\infty$, the Radon-Nikodym derivative always exist.
Example 5. If \( m = \lambda \) is Lebesgue measure, \( \Lambda([a,b]) = b^3 - a^3 \), then

\[
\frac{d\Lambda}{d\lambda}(x) = 3x^2 \quad (x \in \mathbb{R}).
\]

Remark 15. (1) Let \( m \) and \( \mu \) be two measures, \( m \sigma \)-finite. Then the following statements are equivalent:

(i) \( \mu \) has a density with respect to \( m \).

(ii) \( \mu \ll m \).

(2) If \( f \) denotes the Radon-Nikodym derivative of a measure \( \mu \) with respect to \( m \), then for every function \( g \in L_1(\mu) \):

\[
\int g \, d\mu = \int g \cdot f \, dm.
\]

This follows by approximation with step functions from below.

(3) The Radon-Nikodym derivative satisfies the chain rule: If \( m, \mu \) are \( \sigma \)-finite, and if \( \Lambda \) is a signed measure such that \( \Lambda > -\infty \) and

\[
\Lambda \ll \mu \ll m,
\]

then it follows using (2) that

\[
\int_F \frac{d\Lambda}{dm} \, dm = \Lambda(F) = \int_F \frac{d\Lambda}{d\mu} \, d\mu = \int \left( \int_{F} \frac{d\mu}{d\lambda} \right) \frac{d\mu}{dm} \, dm = \int \frac{d\mu}{d\mu} \, dm.
\]

This is true for any \( F \in \mathcal{F} \), hence we also have that

\[
\frac{d\Lambda}{dm} = \frac{d\Lambda}{d\mu} \frac{d\mu}{dm} \quad \text{a.e..}
\]

(4) Let \( m \) and \( \mu \) be finite measures such that \( m \ll \mu \ll m \). Then we have that

\[
0 < \frac{dm}{d\mu} \frac{d\mu}{dm} < \infty \quad \text{and} \quad \frac{dm}{d\mu} = \left( \frac{d\mu}{dm} \right)^{-1}.
\]

This can be seen as follows:

\[
m(F) = \int (1_F \frac{dm}{d\mu}) d\mu = \int 1_F \frac{dm}{d\mu} \frac{d\mu}{dm} dm,
\]

whence \( \frac{dm}{d\mu} \frac{d\mu}{dm} = 1 \) a.e.
1.7 Dual Spaces and Conditional Expectation

Let \((\Omega, \mathcal{F}, m)\) be a measure space. In this section we discuss the two most important applications of the theorem of Radon-Nikodym, the characterization of the dual spaces of \(L_p(m)\) and the definition of conditional expectation.

Recall that the dual space \(L_p(m)^*\) of \(L_p(m)\) (for \(p \geq 1\)) is defined as the Banach space of all continuous linear forms \(L : L_p(m) \rightarrow \mathbb{R}\) equipped with the norm

\[
\|L\| := \sup_{f \in L_p(m)} \frac{|L(f)|}{\|f\|}.
\]

The set of all linear forms \(L\) satisfying \(L(f) \geq 0\) for every non-negative \(f \in L_p(m)\) makes up the positive cone \(L_p(m)^*_+\). For every linear form \(L\) there exist two linear forms \(L^+\) and \(L^-\) in this cone such that

\[
L = L^+ - L^-.
\]

The following theorem provides a different representation of this space. Recall that two positive real numbers \(p\) and \(q\) are called dual if

\[
\frac{1}{p} + \frac{1}{q} = 1.
\]

If \(p = 1\) (or \(q = 1\)), this relation is still meaningful with \(q = \infty\) (or \(p = \infty\)), and we shall make use of it. Dual pairs satisfy the condition \(1 \leq p, q \leq \infty\).

**Theorem 22.** Let \(p\) and \(q\) be dual.

(1) The map

\[
F : L_q(m) \rightarrow L_p(m)^*, \quad g \rightarrow F_g
\]

defined by

\[
F_g(f) = \int f \cdot g \, dm \quad (g \in L_q(m), f \in L_p(m)),
\]

is linear and order preserving. For \(1 < p \leq \infty\) \(F\) is an isometry. If \(m\) is \(\sigma\)-finite, Then \(F\) is also an isometry for \(p = 1\).

(2) If \(m\) is \(\sigma\)-finite and \(p < \infty\), then \(F\) is an isomorphism, hence

\[
L_p(m)^* \cong L_q(m).
\]
Remark 16. Order preserving of the map $F$ means:

$$g \leq g' \implies Fg \leq Fg'.$$

The order in $L_p(m)^*$ is defined as follows:

$$L \leq L' \iff \forall f \in L_p(m), f \geq 0 \implies L(f) \leq L'(f).$$

Consequently this means, that $L' - L$ is a positive operator (in the positive cone $L_p(m)^*$).

Isometry means, that

$$\|Fg\| = \|g\|_q$$

for every $g \in L_q(m)$. Isometry implies injectivity, as can easily be seen. Consequently, in general $L_q(m)$ is isometrically embedded in the dual space $L_p(m)^*$. Equality holds for $p < \infty$ and for $\sigma$-finite measures. In this case every linear functional (linear form) $L : L_p(m) \rightarrow \mathbb{R}$ can be represented by some $Fg$ where $g \in L_q(m)$.

Proof. (1) Let $g \in L_q(m)$. The Hölder inequality shows for $f \in L_p(m)$

$$|Fg(f)| = |\int f \cdot g \, dm| \leq \|g\|_q \cdot \|f\|_p.$$

Therefore $Fg(f)$ is well defined for $f \in L_p(m)$. $Fg$ is obviously linear on $L_p(m)$, since the integral is linear. The continuity of $Fg$ follows as well, and even more precisely

$$\|Fg\| \leq \|g\|_q.$$

Hence $Fg \in L_p(m)^*$ for every $g \in L_q(m)$, and the map $F$ is well defined. By definition one gets immediately that $F$ itself is linear, and Hölder’s inequality shows, that $F$ is continuous, more precisely $\|F\| \leq 1$.

The map $F$ is order preserving: let $g' \leq g$ be functions in $L_q(m)$ and let $0 \leq f \in L_p(m)$. Then $f \cdot g' \leq f \cdot g$ and hence, using section I 3, (I5), it follows that $Fg'(f) \leq Fg(f)$.

It is left to show the isometry property. Since $\|Fg\| \leq \|g\|_q$ for every $g \in L_q(m)$ it is sufficient, to show for fixed $g \in L_q(m)$, that there exists a sequence $f_n \in L_p(m)$ satisfying

$$\lim_{n \to \infty} \frac{|Fg(f_n)|}{\|f_n\|_p} = \|g\|_q.$$

If $1 < p < \infty$, so $q = p(q - 1)$, put $f_n = f = \text{sign}(g) \cdot |g|^{q-1}$. Then

$$|Fg(f)| = \left| \int \text{sign}(g) \cdot |g|^{q-1} \cdot \text{sign}(g) \cdot |g| \, dm \right|$$

$$= \int |g|^q \, dm = \|g\|_q \left( \int |g|^q \, dm \right)^{(q-1)/q}$$

$$= \|g\|_q \left( \int |f|^q/(q-1) \, dm \right)^{(q-1)/q} = \|g\|_q \|f\|_p.$$
If $p = \infty$, then we get $q = 1$. Set $f_n = f = \text{sign}(g)$. It follows that
\[
|F_g(f)| = \int \text{sign}(g) \cdot g \, dm = \int |g| \, dm = \|g\|_1.
\]

If $p = 1$, so $q = \infty$, and we have to assume in addition that, that $m$ is $\sigma$-finite in addition. Let $A_n := \{\|g\|_\infty - |g| < n^{-1}\}$. There exist sets $B_n \subset A_n$ with $0 < m(B_n) < \infty$, so $\|g\|_1 \in L_1(m)$. it follows setting $f_n = \text{sign}(g) \cdot \mathbb{1}_{B_n}$, that
\[
|F_g(f_n)| = \int_{B_n} \text{sign}(g) \cdot g \, dm = \int_{B_n} |g| \, dm \\
\geq \int_{B_n} (\|g\|_\infty - n^{-1}) \, dm = (\|g\|_\infty - n^{-1})\|f_n\|_1.
\]
Letting $n \to \infty$ the claim follows.

(2) Let $1 \leq p < \infty$ and let $L \in L_p(m)^*$ be a fixed continuous linear form on $L_p(m)$.
(A) We show the claim first when $m(\Omega) < \infty$.
In this case $\mathbb{1}_A \in L_p(m)$ for every $A \in \mathcal{F}$, and
\[
\Lambda(A) := L(\mathbb{1}_A) \quad (A \in \mathcal{F})
\]
defines a non-singular finite signed measure $(\Omega, \mathcal{F})$. This can be seen as follows:

(i) $\sigma$-additivity: Let $A_n \in \mathcal{F}$ be pairwise disjoint, $B_n = A_1 \cup \ldots \cup A_n$ and $A = \bigcup_n A_n$. Then
\[
\|\mathbb{1}_A - \mathbb{1}_{B_n}\|_p = \left(\int |\mathbb{1}_A - \mathbb{1}_{B_n}|^p \, dm\right)^{1/p} = m(A \setminus B_n)^{1/p} \to 0 \quad (n \to \infty).
\]
Because of linearity and continuity of $L$ it follows that
\[
\Lambda(B_n) = L(\mathbb{1}_{B_n}) = \sum_{k=1}^{n} L(\mathbb{1}_{A_k}) = \sum_{k=1}^{n} \Lambda(A_k) \to L(\mathbb{1}_A) = \Lambda(A) \quad (n \to \infty).
\]

Absolute convergence of this series is a by-product, since the convergence to $L(\mathbb{1}_A)$ is independent of the numeration of the sets $A_k$. Another argument of this can be given using the decomposition of $L$ in $L = L^+ - L^-$ (see the beginning of this section), hence
\[
\sum_{k=1}^{\infty} |\Lambda(A_k)| = \sum_{k=1}^{\infty} |L(\mathbb{1}_{A_k})| = \sum_{k=1}^{\infty} (L^+(\mathbb{1}_{A_k}) + L^-(\mathbb{1}_{A_k})) = L^+(\mathbb{1}_A) + L^-(\mathbb{1}_A).
\]
Altogether it follows now that
\[
\sum_{n=1}^{\infty} \Lambda(A_n) = \Lambda(A).
\]
(ii) \( A(\emptyset) = 0 \): \( A(\emptyset) = L(I_\emptyset) = L(0) = 0 \).

(iii) \( A \ll m \): Let \( m(A) = 0 \). Then \( I_A = 0 \ a.e. \), hence \( \mathbb{I}_A = 0 \) considered as an element of \( L_p(m) \) and thus \( A(A) = L(I_A) = 0 \).

(iv) \( A \) is finite: Since for \( A \in F \) one has \( I_A \in L_p(m) \), it follows that \( 0 \leq |L(I_A)| = |A(A)| < \infty \).

The Radon-Nikodym theorem provides a function \( g \in L_1(m) \) with

\[
A(A) = L(I_A) = \int I_A \cdot g \ dm \quad (A \in F).
\]

First we show that \( g \in L_\infty(m) \). First consider \( 1 < p < \infty \). Let \( g_n, (n \geq 1) \), be non-negative step functions monotonically convergent to \( |g| \). We have

\[
\int (g_n)^q dm = \int (g_n)^{q-1} g_n dm \leq \int (g_n)^{q-1} g dm = \int (g_n)^{q-1} \text{sign}(g) \cdot \text{sign}(g) \cdot |g| dm.
\]

Now \( g_n^{q-1} \text{sign}(g) \) is a step function as well, so it has a form

\[
g_n^{q-1} \text{sign}(g) = \sum_{k=1}^{N_n} \alpha_k^n \mathbb{I}_{A_k} \quad (n \geq 1),
\]

and together with the representation of \( L(I_A) \) and since \( q = p(q-1) \) it follows that

\[
\|g_n\|_q^q = \int (g_n)^q dm \leq \int \sum_{k=1}^{N_n} \alpha_k^n L(I_{A_k}) g dm
\]

\[
= \sum_{k=1}^{N_n} \alpha_k^n L(I_{A_k}) = L \left( \sum_{k=1}^{N_n} \alpha_k^n I_{A_k} \right)
\]

\[
\leq \|L\| \left\| \sum_{k=1}^{N_n} \alpha_k^n I_{A_k} \right\|_p = \|L\| \left( \int |g_n^{q-1} \text{sign}(g)|^p dm \right)^{1/p}
\]

\[
= \|L\| \left( \int |g_n|^{p(q-1)} dm \right)^{1/p} = \|F\| \|g_n\|^{q-1}_q.
\]

Dividing by \( \|g_n\|_{q-1}^q \) (w.l.o.g. \( \|g\|_q \neq 0 \)) we obtain

\[
\|g_n\|_q^q = \int g_n^q dm \leq \|L\|^q \quad (n \geq 1),
\]

and hence also \( \int g^q dm \leq \|L\|^q \).

In case \( p = 1 \) we have to show that \( g \in L_\infty(m) \). Assume, this is not correct. Then it follows for every \( n \in \mathbb{N} \) that \( m(\{|g| \geq n\}) > 0 \). Let \( A_n = \{|g| \geq n\} \) and \( f_n = \text{sign}(g) \cdot I_{A_n} \). It follows that
\[ L(f_n) = L(I_{A_n \cap \{ g \geq 0 \}}) - L(I_{A_n \cap \{ g \leq 0 \}}) \]

\[ = \int_{A_n} \text{sign}(g) \cdot g \, dm \geq nm(A_n) = n \| I_{A_n} \|_1 = n \| f_n \|_1, \]

and further that

\[ \| L \| = \sup_{f \in L_1(m)} \frac{|L(f)|}{\| f \|_1} \geq \sup_{n \geq 1} \frac{L(f_n)}{\| f_n \|_1} = \infty, \]

a contradiction to the continuity of \( L \).

We are now in the position to use \( F_g \). It has been shown up to this point that \( L = F_g \) holds for indicator functions, i.e. \( L(I_A) = \int F_g(I_A) \). Because of linearity this equality extends to step functions. Step functions are dense in \( L_p(m) \) and the functionals \( L \) and \( F_g \) are continuous, hence \( L = F_g \).

(B) the case \( m(\Omega) = \infty \).

By \( \sigma \)-finiteness there are sets \( E_n \in \mathcal{F}, m(E_n) < \infty \) and \( E_n \uparrow \Omega \). Put \( m_n(A) = m(A \cap E_n) \) for \( A \in \mathcal{F} \).

It is possible to embed the space \( L_p(m_n) \) into the space \( L_p(m) \) continuously, linearly and isometrically: Just use the map

\[ f \rightarrow \overline{f} \]

\[ \overline{f} = \begin{cases} f & \text{auf } E_n \\ 0 & \text{auf } \Omega \setminus E_n. \end{cases} \]

This also defines a continuous, linear functional by restriction

\[ L_n = L|_{L_p(m_n)}. \]

Since each \( m_n \) is a finite measure, there are functions \( g_n \in L_p(m) \) with \( g_n = 0 \) on \( \Omega \setminus E_n \), such that

\[ L(f \mathbb{1}_{E_n}) = \int f \mathbb{1}_{E_n} \cdot g_n \, dm \]

for \( f \in L_p(m) \) and

\[ \| g_n \|_q \leq \| L \| \]

holds.

For measurable sets \( A \subset E_n \) one obtains

\[ \int_A g_{n+1} \, dm = L(\mathbb{1}_A \mathbb{1}_{E_n}) = L(\mathbb{1}_A \mathbb{1}_{E_{n+1}}) = \int_A g_{n+1} \, dm, \]

since \( E_n \subset E_{n+1} \). By section I.3, (19), it follows that \( g_n \mathbb{1}_{E_n} = g_n = g_{n+1} \mathbb{1}_{E_n} \), and thus there exists

\[ g := \lim_{n \to \infty} g_n \quad \text{a.e.} \]
We show first that $g \in L^q(m)$. Since $\|g_n\|_q$ converges monotonically to $\|g\|_q$, the theorem of monotone convergence gives

$$\lim_{n \to \infty} \|g_n\|_q^q = \lim_{n \to \infty} \|g_n\|_q^q dm = \int |g|^q dm,$$

and by uniform boundedness of the $\|g_n\|_q$ by $\|L\|$ it follows that $g \in L^q(m)$.

Consider $f \in L^p(m)$. Then $f 1_{E_n}$ converges pointwise to $f$ a.e. and this sequence is bounded by the function $|f| \in L^p(m)$. The theorem of dominated convergence yields

$$\|f - f 1_{E_n}\|_p^p = \int |f - f 1_{E_n}|^p dm \to 0 \quad (n \to \infty).$$

The continuity of $L$ implies moreover that

$$L(f) = \lim_{n \to \infty} L(f 1_{E_n}) = \lim_{n \to \infty} \int f 1_{E_n} g_n dm = \lim_{n \to \infty} \int f \cdot g_n dm.$$

We obtain $\lim_{n \to \infty} fg_n = fg$ a.e.. Hölder’s inequality implies

$$\|fg\|_1 \leq \|f\|_p \|g\|_q < \infty,$$

that is: $fg_n$ is bounded by the integrable function $|fg|$, and the theorem of dominated convergence shows that

$$L(f) = \lim_{n \to \infty} \int f \cdot g_n dm = \int f \cdot g dm.$$

This ends the proof.

Let us now turn to the second application.

**Theorem 23. (Existence of conditional expectation)**

Let $A \subset F$ be a $\sigma$-subalgebra of $F$, such that the restriction of $m$ on $A$ is $\sigma$-finite. Then for every function $f \in L^1(m)$ there exists a $A$-measurable function $g \in L^1(m)$, which satisfies

$$\int_A f dm = \int_A g dm \quad (BE1)$$

for every $A \in A$. The function $g$ is a.e. uniquely determined by this equality and by the measurability condition.

**Definition 22.** The function $g$ is called the conditional expectation of $f$ given $A$ and will be written as

$$g = E(f|A).$$

In case the $\sigma$-algebra $A$ is the smallest $\sigma$-algebra for which the functions $h_i$ or random variables $X_i$ are measurable, or which contains a collection $\Sigma$ of subsets, these generating quantities appear instead of $A$ in the notation: e.g. $E(X|Y)$, $E(f|h_1, ..., h_n), ...$)
Proof. Let $m_0$ be the restriction of the set function $m$ on $\mathcal{F}$ to $\mathcal{A}$. ($m_0$ is, of course, again a measure.) Define

$$A(A) = \int_A f \, dm \quad (A \in \mathcal{A}).$$

$A$ is a finite signed measure on $(\Omega, \mathcal{A})$, and by the theorem of Radon Nikodym there exists a function $g \in L_1(m_0)$ satisfying

$$\Lambda(A) = \int_A g \, dm_0.$$

$g$ is a.e. uniquely determined, since for another function $g'$, $\mathcal{A}$-measurable and having this property, it follows for any $A \in \mathcal{A}$ that

$$\int_A g \, dm = \int_A g' \, dm.$$

Therefore $g = g'$ a.e. by section I.3 (19). We also may consider $g$ as a function in $L_1(m)$ and hence we have

$$A(A) = \int_A g \, dm.$$

Remark 17. The conditional expectation can be defined in the same manner for positive functions.

Theorem 24. The conditional expectation (for $L_1$-functions) has the following properties:

(BE2) Linearity:

$$E(\alpha f + \beta g|A) = \alpha E(f|A) + \beta E(g|A) \quad \text{a.e.}$$

for $f, g \in L_1(m)$ and $\alpha, \beta \in \mathbb{R}$.

(BE3) Positivity: $f \geq 0 \implies E(f|A) \geq 0$.

(BE4) Monotonicity: If $f \leq g$, then $E(f|A) \leq E(g|A)$ a.e.

(BE5) Convergence: $f_n \to f$ a.e., $|f_n| \leq h \in L_1(m)$ implies that

$$\lim_{n \to \infty} E(f_n|A) = E(f|A)$$

a.e. and in mean.

(BE6) Connection to integrals: If $\mathcal{A} = \{\emptyset, \Omega\}$ mod $m$, then the restriction of $m$ is only $\sigma$-finite, if $m$ is finite itself, and it holds that

$$E(f|A) = \frac{1}{m(\Omega)} \int f \, dm \quad (f \in L_1(m)).$$

(BE7) If $f$ is $\mathcal{A}$-measurable, then $E(f|A) = f$ a.e.
1 THEORY OF INTEGRATION

Proof. (BE2) \(\alpha E(f|A) + \beta E(g|A)\) is \(\mathcal{A}\)-measurable and for \(A \in \mathcal{A}\) it follows that
\[
\int_A (\alpha E(f|A) + \beta E(g|A)) \, dm = \int_A \alpha f + \beta g \, dm.
\]
Consequently
\[
\alpha E(f|A) + \beta E(g|A) = E(\alpha f + \beta g|A)
\]
by the uniqueness of the conditional expectation.

(BE3) For \(A \in \mathcal{A}\) one has
\[
\int_A E(f|A) \, dm = \int_A f \, dm \geq 0.
\]
Therefore \(E(f|A) \geq 0\) holds by section I.3 (I5).

(BE4) Let \(f \leq g\). It follows for \(A \in \mathcal{A}\) that
\[
\int_A E(f|A) \, dm = \int_A f \, dm \leq \int_A g \, dm \leq \int_A E(g|A) \, dm.
\]

(BE5) let \(A \in \mathcal{A}\). The theorem of dominated convergence implies that
\[
\lim_{n \to \infty} \int_A E(f_n|A) \, dm = \lim_{n \to \infty} \int_A f_n \, dm = \int_A f \, dm = \int_A E(f|A) \, dm.
\]
First let \(0 \leq f_n \uparrow f\). By (BE4) also \(E(f_n|A)\) is increasing and
\[
g := \lim_{n \to \infty} E(f_n|A)
\]
exists. \(g\) is \(\mathcal{A}\)-measurable and the theorem of monotone convergence shows that for \(A \in \mathcal{A}\)
\[
\int_A g \, dm = \lim_{n \to \infty} \int_A E(f_n|A) \, dm = \int_A E(f|A) \, dm.
\]
Because of the uniqueness of the conditional expectation we have that \(g = E(f|A)\).

Finally let the functions \(f_n\) be arbitrary. Define
\[
g_n = \sup_{r \geq n} |f_r - f|.
\]
Then \(g_n \downarrow 0\) a.e. and \(g_n \leq |f| + |h|\). Thus we obtain
\[
0 \leq |f| + |h| - g_n \uparrow |f| + |h|
\]
and using the first part of proof it follows that
\[
\lim_{n \to \infty} E(|f| + |h| - g_n|A) = E(|f| + |h||A) - \lim_{n \to \infty} E(g_n|A) = E(|f| + |h||A).
\]
Therefore \(\lim_{n \to \infty} E(g_n | \mathcal{A}) = 0\) a.e. Using (BE2) and (BE4), we obtain that
\[
\lim_{n \to \infty} |E(f_n | \mathcal{A}) - E(f | \mathcal{A})| = \lim_{n \to \infty} |E(f_n - f | \mathcal{A})| \\
\leq \lim_{n \to \infty} E(|f_n - f||\mathcal{A}) \leq \lim_{n \to \infty} E(g_n | \mathcal{A}) = 0.
\]

Finally, the convergence in mean follows from
\[
\lim_{n \to \infty} \int |E(f | \mathcal{A}) - E(f_n | \mathcal{A})| \, dm \\
= \lim_{n \to \infty} \int_{\{E(f | \mathcal{A}) - E(f_n | \mathcal{A}) \geq 0\}} E(f | \mathcal{A}) - E(f_n | \mathcal{A}) \, dm \\
- \lim_{n \to \infty} \int_{\{E(f | \mathcal{A}) - E(f_n | \mathcal{A}) < 0\}} E(f | \mathcal{A}) - E(f_n | \mathcal{A}) \, dm \\
= \lim_{n \to \infty} \int_{\{E(f | \mathcal{A}) - E(f_n | \mathcal{A}) \geq 0\}} f - f_n \, dm - \lim_{n \to \infty} \int_{\{E(f | \mathcal{A}) - E(f_n | \mathcal{A}) < 0\}} f - f_n \, dm \\
\leq \int |f - f_n| \, dm = 0
\]
using the theorem of dominated convergence once more.

(BE6) If \(m(\Omega) = \infty\), the restriction to the trivial \(\sigma\)-subalgebra \(\mathcal{A}\) is not \(\sigma\)-finite. Hence \(m\) is finite. Since \(\mathcal{A}\) is trivial, \(1_A\) is as a \(\mathcal{A}\)-measurable function a.e. constant. Thus
\[
m(\Omega) E(f | \mathcal{A}) = \int E(f | \mathcal{A}) \, dm = \int f \, dm.
\]
(BE7) follows from the definition of the conditional expectation and the uniqueness in the last theorem.
1.8 Fubini’s Theorem

In this section \((\Omega_i, \mathcal{F}_i, m_i)\) are \(\sigma\)-finite measure spaces \((i = 1, 2)\), \(\Omega = \Omega_1 \times \Omega_2\) and \(\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2\) their measure theoretic product equipped with the product-\(\sigma\)-algebra \(\mathcal{F}\). Let us define the maps

\[
T_{\omega_1} : \Omega_2 \to \Omega \quad (\omega_1 \in \Omega_1)
T_{\omega_1}(\eta) = (\omega_1, \eta)
\]

\[
T_{\omega_2} : \Omega_1 \to \Omega \quad (\omega_2 \in \Omega_2)
T_{\omega_2}(\eta) = (\eta, \omega_2).
\]

Note that

\[
T_{\omega_1}^{-1}(F_1 \times F_2) = \begin{cases} F_2 & \text{if } \omega_1 \in F_1 \\ \emptyset & \text{if } \omega_1 \notin F_1 \end{cases}
\]

hence \(T_{\omega_1}\) is measurable for every \(\omega_1 \in \Omega_1\). The same holds for the maps \(T_{\omega_2}\).

**Definition 23.** Let \(F \in \mathcal{F}\). For every \(\omega_1 \in \Omega_1\) the set

\[
F_{\omega_1} := T_{\omega_1}^{-1}F
\]

is called the \(\omega_1\)-section of \(F\), and for every \(\omega_2 \in \Omega_2\) the set

\[
F_{\omega_2} := T_{\omega_2}^{-1}F
\]

is called the \(\omega_2\)-section of \(F\).

**Lemma 7.** For every \(\omega_2 \in \Omega_2\) (resp. \(\omega_1 \in \Omega_1\))

\[
m_{\omega_2}^1(F) := m_1(F_{\omega_2}) \quad F \in \mathcal{F}
\]

(resp.

\[
m_{\omega_1}^2(F) := m_2(F_{\omega_1}) \quad F \in \mathcal{F}
\]

defines a measure on \((\Omega, \mathcal{F})\).

**Proof.** Since we have that \(m_{\omega_2}^1 \geq 0\), only \(\sigma\)-additivity has to be shown. Let \(F_k \in \mathcal{F}\) be pairwise disjoint and \(F = \bigcup_{k \geq 1} F_k\). It is easy to see that the sets \((F_k)_{\omega_2}\) are pairwise disjoint and that \(F_{\omega_2} = \bigcup_{k \geq 1}(F_k)_{\omega_2}\). This implies \(\sigma\)-additivity by definition.

The following theorem is the key point in the theorem of Fubini:

**Theorem 25.** The maps

\[
\phi_{2,F} : \Omega_2 \to \mathbb{R}^+ \cup \{\infty\} \\
\omega_2 \mapsto m_{\omega_2}^2(F)
\]

\[
\phi_{1,F} : \Omega_1 \to \mathbb{R}^+ \cup \{\infty\} \\
\omega_1 \mapsto m_{\omega_1}^1(F)
\]
(for all \( F \in \mathcal{F} \)) are positive and measurable functions satisfying
\[
\int_{\Omega_2} m_2^*(F) m_2(d\omega_2) = \int_{\Omega_1} m_2^*(F) m_1(d\omega_1).
\]

**Proof.** Let
\[
\Sigma := \{ F \in \mathcal{F} : \phi_{1,F}, \phi_{2,F} \text{ measurable, } \int \phi_{1,F} dm_1 = \int \phi_{2,F} dm_2 \}.
\]

It is sufficient to show the following:

(i) \( \Sigma \) contains the generating class \( \Gamma \) for \( \mathcal{F} \), which consists of all sets of form \( F_1 \times F_2 \) (\( F_i \in \mathcal{F}_i \)).

(ii) \( \Sigma \) is closed under union of two disjoint sets.

(iii) \( \Sigma \) is a monotone class, i.e.
\[
\Sigma \ni F_n \uparrow F \implies F \in \Sigma
\]
and
\[
\Sigma \ni F_n \downarrow F \implies F \in \Sigma.
\]

By (i) and (ii) and using section I.1 it follows that \( \Sigma \) contains the ring generated by \( \Gamma \). Since \( \Omega \in \Gamma \), this ring is closed under forming complements. Thus the smallest monotone class \( \mathcal{M}(\Gamma) \), which contains this ring, is a \( \sigma \)-algebra, whence \( \mathcal{F} \). On the other hand this monotone class \( \mathcal{M}(\Gamma) \) is contained in \( \Sigma \subset \mathcal{F} \). Therefore \( \Sigma \) is itself a \( \sigma \)-algebra, which contains a generating system.

(i): By the preceding it follows for \( F = F_1 \times F_2 \) that
\[
\phi_{1,F}(\omega_1) = m_2(F_{\omega_1}) = m_2(F) 1_{F_1}(\omega_1)
\]
\[
\phi_{2,F}(\omega_2) = m_1(F_{\omega_2}) = m_1(F_1) 1_{F_2}(\omega_2)
\]
are measurable and satisfy
\[
\int \phi_{1,F} dm_1 = \int m_2(F) 1_{F_1} dm_1 = m_1(F_1) m_2(F_2) = \int \phi_{2,F} dm_2.
\]

(ii): Let \( F, G \in \Sigma \) be pairwise disjoint. It has been seen in the proof of the above lemma that the sections are also pairwise disjoint, hence one has
\[
\phi_{1,F \cup G}(\omega_1) = m_2((F \cup G)_{\omega_1}) = m_2(F_{\omega_1}) + m_2(G_{\omega_1}) = \phi_{1,F} + \phi_{1,G}.
\]
This implies the measurability and, moreover, that
\[
\int \phi_{1,F \cup G} dm_1 = \int \phi_{1,F} dm_1 + \int \phi_{1,G} dm_1 = \int \phi_{2,F \cup G} dm_2.
\]

(iii): First, let \( F \subset G \). Since \( F_{\omega_1} \subset G_{\omega_1} \), it follows that
\[ \phi_{1,F}(\omega_1) = m_2(F_{\omega_1}) \leq m_2(G_{\omega_1}) = \phi_{1,G}. \]

Similarly it follows that \( \phi_{2,F} \leq \phi_{2,G} \).

For \( \Sigma \ni F_n \uparrow F \) the functions \( \phi_{i,F_n} \) (i = 1, 2) are positive and converges to \( \phi_{i,F} \); therefore \( \phi_{i,F} \) is measurable. Using the theorem of monotone convergence it follows in addition that

\[
\int \phi_{2,F} dm_2 = \lim_{n \to \infty} \int \phi_{2,F_n} dm_2 = \lim_{n \to \infty} \int \phi_{1,F_n} dm_1 = \int \phi_{1,F} dm_1.
\]

If \( \Sigma \ni F_n \downarrow F \), then one concludes similarly that \( F \in \Sigma \) using the theorem of dominated convergence. This theorem is only applicable for finite measures \( m_i \) (because the constant function 1 has to be integrable). In case the measures are \( \sigma \)-finite choose sets \( E_{t_i} \in \mathcal{F}_i \) such that \( E_{t_i} \uparrow \Omega_i \) and \( m_i(E_{t_i}) < \infty \) (i = 1, 2). For fixed \( t \in \mathbb{N} \) die the sets \( E_{t_1} \times E_{t_2} \) belong to \( \Gamma \subset \Sigma \), hence \( G_{n,t} := F_n \cap (E_{t_1} \times E_{t_2}) \in \Sigma \) by (ii). According to the finite case which has been proven so far, for \( G^t = F \cap (E_{t_1} \times E_{t_2}) \) the functions \( \phi_{t,G^t} \) are measurable, and one has

\[
\int \phi_{1,G^t} dm_1 = \int \phi_{2,G^t} dm_2.
\]

Since \( G^t \uparrow F \), it follows that

\[
\int \phi_{1,F} dm_1 = \int \phi_{2,F} dm_2
\]

using the theorem of monotone convergence (as above).

This finishes the proof.

**Theorem 26.** There is exactly one measure \( m \) on \((\Omega, \mathcal{F})\) satisfying

\[ m(F_1 \times F_2) = m_1(F_1) \cdot m_2(F_2) \quad (F_i \in \mathcal{F}_i). \]

\( m \) has the property that

\[ m(F) = \int m_1(F_{\omega_1}) m_2(d\omega_2) = \int m_1(F_{\omega_2}) m_2(d\omega_2) \quad (F \in \mathcal{F}) \]

and is called the product measure of \( m_1 \) and \( m_2 \), written as

\[ m = m_1 \times m_2. \]

**Proof.** It is known already that

\[ m(F) = \int m_1(F_{\omega_1}) m_2(d\omega_2) = \int m_1(F_{\omega_2}) m_2(d\omega_2) \quad (F \in \mathcal{F}) \]
is well defined and satisfies
\[ m(F_1 \times F_2) = m_1(F_1) \cdot m_2(F_2) \quad (F_i \in \mathcal{F}_i). \]

Therefore it suffices to show that \( m \) is a measure and uniquely determined.

Obviously, \( m \) is positive. Let \( F_n \in \mathcal{F} \) be pairwise disjoint sets. Since \( m_i^{\omega} \) are measures, it follows from the theorem of monotone convergence (series form) that
\[ m(F) = \int m_1^{\omega}(F)m_2(d\omega_2) = \int \sum_{n \geq 1} m_1^{\omega_n}(F_n)m_2(d\omega_2) = \sum_{n \geq 1} m(F_n). \]

Hence \( m \) is \( \sigma \)-additive.

In particular, \( m_0(F_1 \times F_2) = m(F_1 \times F_2) \) defines a positive, \( \sigma \)-additive set function on \( \mathcal{F} \), which by the extension- and uniqueness theorems in section I.1 permits only one extension. Thus \( m \) is unique.

For a function \( f : \Omega \to \mathbb{R} \cup \{\pm \infty\} \) define
\[ f_{\omega_1} := f \circ T_{\omega_1} : \Omega_2 \to \mathbb{R} \cup \{\pm \infty\} \]
and
\[ f_{\omega_2} := f \circ T_{\omega_2} : \Omega_1 \to \mathbb{R} \cup \{\pm \infty\}. \]

Obviously these functions are measurable.

**Corollary 4.** Let \( f : \Omega \to \mathbb{R}_+ \cup \{\infty\} \) be a positive measurable function. Then the functions
\[ \omega_1 \rightarrow \phi_{1,f}(\omega_1) := \int f_{\omega_1} dm_2 \]
and
\[ \omega_2 \rightarrow \phi_{2,f}(\omega_2) := \int f_{\omega_2} dm_1 \]
are positive and measurable. Moreover,
\[ \int f \, d(m_1 \times m_2) = \int \left[ \int f_{\eta} dm_2 \right] m_1(d\eta) = \int \left[ \int f_{\eta} dm_1 \right] m_2(d\eta). \]

**Proof.** The assertions hold for indicator functions \( f \) because of the last two theorems. It is immediate that the assertions then also hold for step functions. By the theorem of monotone convergence the assertions follow by monotone approximation through step functions: Let \( f_n \uparrow f \). Then \( (f_n)_{\omega_1} \uparrow f_{\omega_1} \) and hence
\[ \int (f_n)_{\omega_1} \, dm_2 \uparrow \int f_{\omega_1} \, dm_2. \]

This yields
\[ \int f_{(m_1 \times m_2)} \uparrow \int f_n \, dm_1 \times dm_2 = \int [\int (f_n)_{\omega_1} \, dm_2] \, m_1(d\omega_1) \uparrow \int [\int f_{\omega_1} \, dm_2] \, m_1(d\omega_1). \]

**Theorem 27.** (Fubini)

(A1) \( f \in L^1(m_1 \times m_2) \Rightarrow f_{\omega_1} \in L^1(m_2) \) for \( m_1 \) almost all \( \omega_1 \in \Omega_1 \), and

\[ \phi_1^* (\omega_1) = \begin{cases} \int f_{\omega_1} \, dm_2 & \text{if } f_{\omega_1} \in L^1(m_2) \\ 0 & \text{else} \end{cases} \]

the function \( \phi_1^* \in L^1(m_1) \).

(A2) If \( f \geq 0 \), then \( f_{\omega_1} \in L^1_+(m_2) \) for \( m_1 \) a.e. \( \omega_1 \in \Omega_1 \); and if \( \phi_{f,1} \in L^1(m_1 \times m_2) \).

(A3) If the assumptions in (A1) or (A2) hold, then

\[ \int fd(m_1 \times m_2) = \int \left[ \int f(\omega_1, \omega_2) \, dm_2(d\omega_2) \right] \, m_1(d\omega_1) \]

\[ = \int \left[ \int f(\omega_1, \omega_2) \, m_1(d\omega_1) \right] \, m_2(d\omega_2) =: \int \int f(\omega_1, \omega_2) \, m_1(d\omega_1) \, m_2(d\omega_2). \]

(B) If \( f \in L^1(m_1 \times m_2) \), then \( \int f_{\omega_1} \, dm_1 \) is independent of the choice of the representative in \( f \), and \( f_{\omega_1} \in L^1(m_2) \) for \( m_1 \) a.e. \( \omega_1 \in \Omega_1 \). Moreover, the function

\[ \eta \rightarrow \int f_\eta \, dm_2 \quad \eta \in \Omega_1 \]

belongs to \( L^1(m_1) \), and

\[ \int fd(m_1 \times m_2) = \int \left[ \int f(\omega_1, \omega_2) \, dm_2(d\omega_2) \right] \, m_1(d\omega_1) \]

\[ = \int \left[ \int f(\omega_1, \omega_2) \, m_1(d\omega_1) \right] \, m_2(d\omega_2) =: \int \int f(\omega_1, \omega_2) \, m_1(d\omega_1) \, m_2(d\omega_2). \]

**Remark 18.** In (A1), (A2) and (B), the corresponding statements for the \( \omega_2 \)-sections hold as well.

**Proof.** (A1) Let \( f \geq 0 \). By the last corollary \( \phi_{1,f} \) is positive and measurable. It satisfies

\[ \int f \, dm_1 \times m_2 = \int \phi_{1,f} \, dm_1. \]

Since the left hand side is finite by assumption, it follows that

\[ \int f_{\omega_1} \, dm_2 = \phi_{1,f}(\omega_1) < \infty \]
$m_1$ a.e., so $f_{\omega_1} \in L_1(m_2)$ $m_1$ a.e.

If $f$ is arbitrary, decompose it into positive and negative parts: First one has

$$(f^+)_{\omega_1}(\eta) = f^+(\omega_1,\eta) = \max(f(\omega_1,\eta);0) = \max(f_{\omega_1}(\eta);0) = (f_{\omega_1})^+(\eta).$$

Analogously one obtains $(f^-)_{\omega_1} = (f_{\omega_1})^-$. Consequently $f_{\omega_1} = (f^+)_{\omega_1} - (f^-)_{\omega_1}$ belongs to $L_1(m_2)$ $m_1$ and

$$\omega_1 \to \int (f^+)_{\omega_1} - (f^-)_{\omega_1} \, dm_2$$

(and thus $\phi^*_1$) in $L_1(m_1)$.

(A2) This follows immediately from the last corollary.
(A3) follows from the last corollary by decomposing into positive and negative parts.
(B) It suffices to show the independence of $f_{\omega_1}$ from the choice of a representative. Let $h$ and $g$ be two different functions in $f$. Observe that

$$\int \int |h_\eta - g_\eta| \, dm_2 \, m_1(\eta) = \int \int |h_\eta - g_\eta| \eta \, dm_2 \, m_1(\eta) = \int |h_\eta - g_\eta| \, d(m_1 \times m_2) = 0,$$

hence $h_\eta = g_\eta$ $m_2$ a.e. for almost all $\eta$ (with respect to $m_1$).
2 MARTINGALE THEORY

2.1 Random Variables

In this section let \((\Omega, \mathcal{F}, P)\) be a fixed probability space. Recall that a measurable function defined on a probability space is called a random variable and denoted by \(X, Y, Z, \ldots\) More generally, a measurable map \(X : \Omega \rightarrow \Omega'\) with values in a measurable space \((\Omega', \mathcal{F}')\) is called a random element.

Definition 24. A family of random elements \((X_t)_{t \in T}\) defined on the same probability space and taking values in the same measurable space \((E, \mathcal{B})\) is called a random process or stochastic process. The set \(T\) is called the time parameter of the process, and the space \((E, \mathcal{B})\) is called the state space of the process.

Definition 25. Let \(X\) and \(Y\) be random variables. Then

1. \(E(X) = \int XdP\) is called the expectation of \(X\). \(X\) has finite expectation if \(X \in L_1(P)\).
2. \(E(X^p) = \int X^p dP\) is called the \(p\)-th moment of \(X\). It is finite if and only if \(X \in L_p(P)\). \(E((X - E(X))^p)\) is called the \(p\)-th centralized moment. If the second centralized moment exists, it is called the variance of \(X\) and is denoted by \(\text{Var}(X)\).
3. \(X\) is bounded if \(X \in L_\infty(P)\).
4. The covariance of \(X\) and \(Y\) is defined by

   \[
   \text{Cov}(X, Y) = \int (X - E(X))(Y - E(Y))dP,
   \]

   whenever the integral exists and is finite.
5. \(F_X(t) = \frac{1}{2}(P(X \leq t^+) + P(X < t)) \) (resp. \(P(X \leq t)\), resp. \(P(X < t)\)) is called the normalized (resp. left continuous, resp. right continuous) version of the distribution function of \(X\). Here

   \[
   P(X \leq t^+) = \lim_{s \downarrow t} P(X \leq s).
   \]

Definition 26. A family \(\mathcal{F}\) of sub-\(\sigma\)-algebras of \(\mathcal{F}\) is called independent if for all choices \(A_1, \ldots, A_n\) of \(\sigma\)-algebras in \(\mathcal{F}\) and for all choices of events \(A_i \in A_i\),
\[ P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2)\ldots P(A_n) \]

holds. A family of measurable sets or random variables is called independent if the \(\sigma\)-algebras generated by these quantities are independent.

Remark 19. It can easily be shown, that two events \(E\) and \(F\) are independent if and only if
\[ P(E \cap F) = P(E)P(F). \]
Likewise, finitely many events \(E_1, \ldots, E_n\) are independent if for all choices \(i_1, \ldots, i_s\)
\[ P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_s}) = \prod_{l=1}^{s} P(E_{i_l}). \]
Random variables \(X_n (n \in \mathbb{N})\) are independent if for all \(n \geq 1\) and all choices of Borel sets \(B_1, \ldots, B_n \subset \mathbb{R}\)
\[ P \left( \bigcap_{i=1}^{n} \{X_i \in B_i\} \right) = \prod_{i=1}^{n} P(X_i \in B_i). \]

We collect some elementary facts on random variables and probability spaces.

**Theorem 28. (Borel-Cantelli Lemma)**
1. Let \(E_n \in \mathcal{F}, n = 1, 2, \ldots\). Then
\[ \sum_{n=1}^{\infty} P(E_n) < \infty \implies P \left( \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} E_m \right) = 0. \]
This means that with probability one, only finitely many \(E_n\) occur.

2. Let \(E_n \in \mathcal{F}, n = 1, 2, \ldots\), be independent. Then
\[ \sum_{n=1}^{\infty} P(E_n) = \infty \implies P \left( \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} E_m \right) = 1. \]
This means that with probability one, infinitely many \(E_n\) occur.

**Proof.**
1. \[ P \left( \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} E_m \right) \leq \lim_{n \to \infty} P \left( \bigcup_{m=n}^{\infty} E_m \right) \leq \lim_{n \to \infty} \sum_{m=n}^{\infty} P(E_m) = 0. \]

2.
2.1 Random Variables

$$1 - P \left( \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} E_m \right) = P \left( \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} E_m^c \right)$$

$$= \lim_{n \to \infty} P \left( \bigcap_{m=n}^{\infty} E_m^c \right)$$

$$= \lim_{n \to \infty} \lim_{N \to \infty} P \left( \bigcap_{m=n}^{N} E_m^c \right)$$

$$= \lim_{n \to \infty} \prod_{m=n}^{\infty} (1 - P(E_m))$$

$$= \lim_{n \to \infty} \prod_{m=n}^{\infty} (1 - P(E_m))$$

$$= 0,$$

since for a sequence $a_k$ of positive numbers, $\sum_{k=1}^{\infty} a_k = \infty$ if and only if $\prod_{k=1}^{\infty} 1 - a_k = 0$.

The next result is a reformulation of properties of integrals.

**Theorem 29.** Let $X, Y$ be random variables.

1. $\text{Var}(X) = E(X^2) - (E(X))^2$ if $X \in L_2(P)$.
2. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ if $X, Y \in L_2(P)$.
3. If $X$ and $Y$ are independent and in $L_1(P)$ then

   $$E(XY) = E(X)E(Y),$$

   in particular $\text{Cov}(X, Y) = 0$.
4. If $X$ and a $\sigma$-algebra $A$ are independent, then

   $$E(X|A) = E(X)$$

5. $E(aX + bY) = aE(X) + bE(Y)$ if $X, Y \in L_1(P)$ and $a, b \in \mathbb{R}$.
6. If $X$ and $Y$ are independent and in $L_2(P)$ then for $a, b \in \mathbb{R}$

   $$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y).$$

**Theorem 30.** (Weak Law of Large Numbers.) Let $X_n$, $n = 1, 2, 3, \ldots$, be uncorrelated random variables (which means that $\text{Cov}(X_n, X_m) = 0$ for $n \neq m$) in $L_2(P)$. If

$$a_n = \sum_{k=1}^{n} \text{Var}(X_k) \to \infty,$$

then for any $\eta > 0$
\[
\lim_{n \to \infty} P \left( \left| \frac{1}{a_n} \sum_{k=1}^{n} (X_k - E(X_k)) \right| \geq \eta \right) = 0
\]

**Proof.** By Markov’s inequality (using the function \( h(x) = x^2 \))

\[
\lim_{n \to \infty} P \left( \left| \frac{1}{a_n} \sum_{k=1}^{n} (X_k - E(X_k)) \right| \geq \eta \right) \leq \lim_{n \to \infty} \eta^{-2} a_n^{-2} \text{Var}(X_1 + \ldots + X_n) = 0,
\]

since

\[
\text{Var}(X_1 + \ldots + X_n) = \sum_{k,l=1}^{n} \text{Cov}(X_k, X_l) = a_n,
\]

and since \( a_n \to \infty \).

**Theorem 31. (Strong Law of Large Numbers)** Let \( X, X_1, X_2, \ldots \) be independent, identically distributed and square integrable random variables. Then

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X_k = E(X) \quad \text{a.s.}
\]

**Proof.** Define \( S_n = X_1 + \ldots + X_n \). We first show that

\[
\lim_{n \to \infty} \frac{1}{n^2} S_n^2 = E(X).
\]

Fix a sequence of positive numbers \( \eta_l \to 0 \). Let

\[
\Omega_l = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{ \left| \frac{1}{n^2} S_m^2 - E(X) \right| \leq \eta_l \}.
\]

It is easy to see that

\[
\lim_{n \to \infty} \frac{1}{n^2} S_n^2(\omega) = E(X)
\]

for \( \omega \in \bigcap_{l=1}^{\infty} \Omega_l \). Hence we need to show that the latter set has probability one, equivalently it needs to be shown that \( P(\Omega_l) = 1 \) for all \( l \). But

\[
1 - P(\Omega_l) = P \left( \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{ \left| \frac{1}{n^2} S_m^2 - E(X) \right| > \eta_l \} \right)
\]

\[
\leq \lim_{n \to \infty} \sum_{m=n}^{\infty} \frac{\text{Var}(X)}{\eta_l^2 m^2} = 0.
\]

(By Markov’s inequality and observing that we consider the subsequence \( n^2 \), so that \( \sum \frac{1}{m^2} \) is summable!)
Now consider the full sequence $S_n$. Let $n = k^2 + l$ with $l = 0, 1, ..., (k + 1)^2 - 1$ and

$$E_n = \{|\frac{1}{n}S_n - \frac{1}{k^2}S_{k^2}| \geq \eta_l\}.$$  

It suffices to show that the event $E_n$ occurs only finitely often. Since

$$\text{Var}\left(\frac{1}{n}S_n - \frac{1}{k^2}S_{k^2}\right) = \text{Var}(\frac{1}{n} - \frac{1}{k^2})S_{k^2} + \frac{1}{n}(S_n - S_{k^2})$$

$$= (k^2(\frac{1}{n} - \frac{1}{k^2})^2)\text{Var}(X) + \frac{n - k^2}{n^2}\text{Var}(X))$$

$$\leq \frac{k^2(k^2 - n)^2}{(k^2n)^2}\text{Var}(X) + \frac{1}{n^2}\text{Var}(X).$$

Since $l \leq (k + 1)^2 - k^2 = 2k + 1 \leq 3k \leq 3\sqrt{n}$ it follows that

$$\text{Var}\left(\frac{1}{n}S_n - \frac{1}{k^2}S_{k^2}\right) \leq \left(\frac{9}{n^2} + \frac{3}{n^{3/2}}\right)\text{Var}(X)$$

and so by Markov’s inequality again

$$\sum_{n=1}^{\infty} P(E_n) \leq \text{Var}(X) \left(9 \sum_{n=1}^{\infty} \frac{1}{n^2} + 3 \sum_{n=1}^{\infty} \frac{1}{n^{3/2}}\right).$$

Using the Borel-Cantelli lemma, the claim follows.

### 2.2 Martingales

Let $(\Omega, \mathcal{F}, P)$ be a fixed probability space. Recall that a random variable $X$ is a measurable function defined on the space $(\Omega, \mathcal{F}, P)$. We have seen in chapter 1 that for any sub-$\sigma$-algebra $\mathcal{A}$ of $\mathcal{F}$ there is a random variable $E(X|\mathcal{A})$, which is $\mathcal{A}$-measurable and satisfies

$$\int_A E(X|\mathcal{A}) \, dP = \int_A X \, dP \quad (A \in \mathcal{A}).$$

It is called the conditional expectation of $X$ given $\mathcal{A}$. In case $Y_i$ ($i \in I$) are further random variables on the same probability space, then $E(X|Y_i, i \in I)$ stands for the conditional expectation of $X$ given the $\sigma$-algebra generated by all $Y_i$. Properties of conditional expectations were listed before in section 1.7. This will be used now.

The following example motivates the definition of a martingale.
Example 6. Let $X_1, X_2, \ldots$ be independent, identically distributed random variables satisfying $P(\{X_1 = \pm 1\}) = 1/2$ (fair coin tossing). We bet the amount of $b_n$ on the outcome of the $n+1$-st trial $X_{n+1}$. You win the amount of $b_n$ if the outcome is 1, otherwise you lose this amount. If $S_n$ denotes the capital after the $n$-th trial, we have

$$S_{n+1} = S_n + b_n X_{n+1}$$

after the next trial. It is reasonable to make the amount $b_n$ depending on the outcomes of the trials $X_1, \ldots, X_n$, meaning that $b_n$ is a function of these variables:

$$b_n = b_n(X_1, \ldots, X_n).$$

This implies

$$E(S_{n+1}|X_1, \ldots, X_n) = E(S_n + b_n(X_1, \ldots, X_n)X_{n+1}|X_1, \ldots, X_n)$$
$$= E(S_n|X_1, \ldots, X_n) + b_n(X_1, \ldots, X_n)E(X_{n+1}|X_1, \ldots, X_n)$$
$$= S_n + b_n(X_1, \ldots, X_n)E(X_{n+1}) = S_n.$$

In its derivation we made use of several properties of conditional expectations as was shown before. In particular: If $Z$ and $A$ are independent, then

$$\int_A E(Z|A) \, dP = \int_A Z \, dP = P(A)E(Z) \quad (A \in \mathcal{A}),$$

so $E(Z|A) = E(Z)$ holds. We find that we are not able to make an expected win on the basis of previous outcomes of the gambling procedure.

Definition 27. Let $I$ be an ordered index set, the order relation will be denoted by $\leq$. Moreover, let $\{X_i : i \in I\}$ be a family of integrable random variables (defined on the probability space $(\Omega, \mathcal{F}, P)$) and let $\{\mathcal{A}_i : i \in I\}$ be a family of sub-$\sigma$-algebras with $\mathcal{A}_i \subset \mathcal{A}_j$ for $i \leq j$.

1. $(X_i)$ is called adapted to $(\mathcal{A}_i)$, if $X_i$ is $\mathcal{A}_i$ measurable.
2. $(X_i)$ is called a supermartingale (SMG) with respect to $(\mathcal{A}_i)$, if $(X_i)$ is adapted to $(\mathcal{A}_i)$ and for $i \leq j$ the inequality

$$E(X_j|\mathcal{A}_i) \leq X_i$$

holds.
3. $(X_i)$ is called a submartingale with respect to $(\mathcal{A}_i)$, if $(-X_i)$ is a supermartingale with respect to $(\mathcal{A}_i)$.
4. $(X_i)$ is called a martingale (MG) with respect to $(\mathcal{A}_i)$, if $(X_i)$ is both, a super- and a submartingale with respect to $(\mathcal{A}_i)$.

Remark 20. (1) If the connection to the $\sigma$-algebras $\mathcal{A}_i$ is omitted, it is understood that the $\sigma$-algebras
\[ A_i = \sigma(\{X_j : j \leq i\}) \]

are used in the definition of the (super-, sub-) martingale properties. The submartingale property is this
\[
E(X_j | A_i) \geq X_i \quad (\forall i \leq j),
\]
and the martingale property is
\[
E(X_j | A_i) = X_i \quad (\forall i \leq j).
\]

(2) The elementary properties of martingales and submartingales are summarized in the sequel. They are immediate consequences of the definitions and the properties of conditional expectations.

- (M1) \((X_i)\) MG (SMG) w.r.t. \((A_i)\) \iff \(X_i\) is \(A_i\)-measurable
  \[
  \forall i \leq j; \forall A \in A_i : \int_A X_j dP = (\leq) \int_A X_i dP.
  \]

- (M2) \(\{X_i : i \in \mathbb{N}\}\) MG (SMG) w.r.t. \(\{A_i : i \in \mathbb{N}\}\) \iff \(X_i\) is \(A_i\)-measurable
  \[
  \forall n \in \mathbb{N} : E(X_{n+1} | A_n) = (\leq) X_n.
  \]

- (M3) \(\{X_i : i \in \mathbb{N}\}\) MG (SMG) \implies
  \[
  \forall i \leq j : E(X_j) = (\leq) E(X_i).
  \]

- (M4) \(\{X_i : i \in \mathbb{N}\}\) independent, \(E(X_n) = 0 \implies S_n = X_1 + X_2 + \ldots + X_n\) is a martingale.

- (M5) \((X_i), (Y_i)\) MG w.r.t. \((A_i), \alpha, \beta \in \mathbb{R} \implies (\alpha X_i + \beta Y_i)\) is a martingale w.r.t. \((A_i)\).

- (M6) \((X_i), (Y_i)\) SMG w.r.t. \((A_i), \alpha, \beta \in \mathbb{R}_+ \implies (\alpha X_i + \beta Y_i)\) is a supermartingale w.r.t. \((A_i)\).

- (M7) \((X_i), (Y_i)\) SMG w.r.t. \((A_i)\) \implies (\min\{X_i, Y_i\}) is a supermartingale w.r.t. \((A_i)\).

- (M8) \((X_i)\) SMG w.r.t. \((A_i)\) \implies (\text{\textit{\textminus}} X_i\text{\textsuperscript{\textdegree}}) is a supermartingale w.r.t. \((A_i)\).

- (M9) \((X_i)\) MG w.r.t. \((A_i)\), \(q \cup\text{-}\textit{convex} (\text{resp.} \cap\text{-}\textit{convex}) \implies (q(X_i))\) is a submartingale (resp. supermartingale) w.r.t. \((A_i)\).

\textit{Example 7.} Let \((\Omega, \mathcal{F}, P)\) be a probability space. Let

\[ I = \{\beta = \{B_1, \ldots, B_s\} : \beta\text{ is a finite partition of } \Omega \text{ into measurable sets}\}. \]

The order on \(I\) is determined by inclusion: \(\beta \leq \beta' \iff\) every \(B \in \beta\) is a union of sets in \(\beta' \iff\) \(\sigma(\beta) \subset \sigma(\beta')\). Let \(Q\) be a finitely additive, finite set function on \(\mathcal{A}\). Define
\[ X_\beta := \sum_{B \in \beta} \frac{Q(B)}{P(B)} 1_B, \]

and \( X_\beta = 0 \) for every \( B \in \beta \) with \( P(B) = 0 \). Then

(a) \( A_\beta := \sigma(\beta) \) is increasing and \((X_\beta)\) is adapted to \( \sigma(\beta) \).

(b) \( \beta \leq \beta', B \in \beta \Rightarrow \int_B X_{\beta'} dP = \int_B \sum_{C \in \beta', P(C) > 0} \frac{Q(C)}{P(C)} P(B \cap C) \)

\[ = \sum_{B \supseteq C \in \beta', P(C) > 0} Q(C) \leq Q(B) \]

\[ = \int_B X_\beta dP. \]

Thus we have a supermartingale.

**Definition 28.** Let \( I \) be an ordered index set and \( \{A_i : i \in I\} \) a family of sub-\( \sigma \)-algebras of \( \mathcal{F} \) such that \( A_i \subset A_j \) for \( i \leq j \). A mapping

\[ T : \Omega \rightarrow I \]

is called a stopping time with respect to \((A_i)_i\), if

\[ \{\omega \in \Omega : T(\omega) \leq t\} \in A_t \]

for every \( t \in I \).

The set system

\[ \mathcal{A}_T := \{A \subset \Omega : A \cap \{\omega \in \Omega : T(\omega) \leq t\} \in A_t \ (t \in I)\} \]

is a \( \sigma \)-algebra and is called the \( \sigma \)-algebra of the \( T \)-past.

**Example 8.**

(1) Let \( I = \mathbb{R} \) and \( T(\omega) = \alpha \) for some \( \alpha \in \mathbb{R} \). Then \( T \) is a stopping time, because

\[ \{T \leq t\} = \begin{cases} \Omega & \text{if } \alpha \leq t \\ \emptyset & \text{if } \alpha > t. \end{cases} \]

Moreover, one has

\[ \mathcal{A}_T = \{A \subset \Omega : A \cap \{\alpha \leq t\} \in A_t \ (t \in \mathbb{R})\} = \mathcal{A}_\alpha. \]

(2) Let \( I = \mathbb{N} \cup \{\infty\} \) and let \( X_1, X_2, \ldots \) be random variables. For a measurable set \( A \subset \mathbb{R} \) let

\[ T(\omega) = \inf\{n \in \mathbb{N} : X_n(\omega) \in A\}. \]

If no such \( n \) exists, the infimum taken over the empty set is defined to be \( = \infty \). Then \( T \) is a stopping time w.r.t. \( \mathcal{A}_n = \sigma(X_i : 1 \leq i \leq n) \ (n = \infty \) is included).
Remark 21. Elementary properties of stopping times are these: Let $I$ be an ordered index set and let $S, T$ be stopping times w.r.t. $(A_i)$.

(S1) $\inf\{S, T\}$ and $\max\{S, T\}$ are stopping times w.r.t. $(A_i)$.
(S2) If $I = \mathbb{N}$ then $S + T$ is a stopping time w.r.t. $(A_i)$.
(S3) $T \leq S \Rightarrow A_T \subset A_S$.
(S4) Assume there exists a countable subset $I_0 \subset I$, such that $\forall t \in I \exists i_0 \in I_0$ with $t \leq i_0$. Then $A_T \subset \mathcal{F}$.
(S5) $\forall t \in I$ let $T(\{\omega \in \Omega : T(\omega) \leq t\})$ be countable. Then the stopped random variable

$$X_T(\omega) := X_{T(\omega)}(\omega)$$

is $A_T$ measurable.

We sketch the proof of these facts.

(1) $\{\min\{S, T\} \leq t\} = \{T \leq t\} \cup \{S \leq t\}$ and $\{\max\{S, T\} \leq t\} = \{T \leq t\} \cap \{S \leq t\}$.
(2) $\{S + T \leq t\} = \bigcup_{1 \leq k \leq t} \{S = k, T \leq t - k\} \in A_t$.
(3) $T \leq S$ implies $\{S \leq t\} \subset \{T \leq t\}$. Moreover,

$$E \in A_T \Rightarrow E \cap \{S \leq t\} = E \cap \{T \leq t\} \cap \{S \leq t\} \in A_t,$$

i.e. $E \in A_S$.
(4) $E \in A_T \Rightarrow E = \bigcup_{i_0 \in I_0} E \cap \{T \leq i_0\} \in \mathcal{F}$.
(5) Let $B$ be a Borel set and $A = \{\omega \in \Omega : X_T(\omega) \in B\}$. Let $t \in I$. Then

$$A \cap \{T \leq t\} = \bigcup_{j \in T((T \leq t))} A \cap \{T = j\}$$

$$= \bigcup_{j \in T((T \leq t))} \{X_j \in B\} \cap \{T = j\} \in A_t,$$

because $\{T = j\} = \{T \leq j\} \setminus \{T < j\}$.

Theorem 22. (Optional Sampling Theorem)
Let $\{X_n : n \in \mathbb{N}\}$ be a martingale (supermartingale) w.r.t. $\{A_n : n \in \mathbb{N}\}$ and $\{T_n : n \geq 1\}$ be stopping times w.r.t. $\{A_n : n \in \mathbb{N}\}$. Assume the following conditions are satisfied:

(1) $T_n \leq T_{n+1}$ ($n \geq 1$).
(2) $\liminf_{n \to \infty} \mathbb{E}_{T_n} |X_{n+1}| dP = 0$ for every $n \in \mathbb{N}$.
(3) $\limsup_{n \to \infty} \mathbb{E}_{T_n} X_n < \infty$ or $\mathbb{E}(X_{T_n}) < \infty$ for every $n \in \mathbb{N}$.

Then $\{X_{T_n} : n \geq 1\}$ is a martingale (supermartingale) w.r.t. the family $\{A_{T_n} : n \geq 1\}$.

Proof. It follows immediately from the previous remark that $A_{T_n} \subset A_{T_{n+1}} \subset \mathcal{F}$ and that $X_{T_n}$ is $A_{T_n}$ measurable. The proof is carried out in three steps.
We shall show the result for supermartingales, the proof for martingales is just an adaption.

(A) Assume first that every $X_{T_n}$ ($n \geq 1$) is integrable. For $n \geq 1$ and $A \in A_{T_n}$ we show that

$$\int_{A} X_{T_{n+1}} dP \leq \int_{A} X_{T_{n}} dP,$$

i.e. the supermartingale property.

Let $A_j = A \cap \{T_n = j\} \in A_j \cap A_{T_n} \subset A_j \cap A_{T_{n+1}}$. Then

$$\int_{A} X_{T_{n+1}} dP = \sum_{k=1}^{\infty} \int_{A_k} X_{T_{n+1}} dP$$

and

$$\int_{A} X_{T_{n}} dP = \sum_{k=1}^{\infty} \int_{A_k} X_{T_{n}} dP = \sum_{k=1}^{\infty} \int_{A_k} X_{k} dP.$$

It is left to show

$$\int_{A_k} X_{T_{n+1}} dP \leq \int_{A_k} X_{k} dP$$

for every $k \in \mathbb{N}$.

Fix $k \in \mathbb{N}$ and consider any $N > k$. Using the relation

$A_k \cap \{T_{n+1} = j\} \in A_j \cap A_{T_{n+1}}$ ($j \geq k$)

$A_k \cap \{T_{n+1} \geq N\} = A_k \cap (\Omega \setminus \{T_{n+1} \leq N - 1\}) \in A_{N-1} \cap A_{T_{n+1}}$ ($j \geq k$)

$A_k \cap \{T_{n+1} > N\} \in A_N \cap A_{T_{n+1}}$ ($j \geq k$)

it follows that

$$\int_{A_k} X_{T_{n+1}} dP = \sum_{j=k}^{N} \int_{A_k \cap \{T_{n+1} = j\}} X_{T_{n+1}} dP + \int_{A_k \cap \{T_{n+1} > N\}} X_{T_{n+1}} dP$$

$\leq \sum_{j=k}^{N-1} \int_{A_k \cap \{T_{n+1} = j\}} X_{k} dP + \int_{A_k \cap \{T_{n+1} \geq N\}} X_{N} dP - \int_{A_k \cap \{T_{n+1} > N\}} \int_{A_k \cap \{T_{n+1} > N\}} (X_{N} - X_{T_{n+1}}) dP$

$\leq \sum_{j=k}^{N-1} \int_{A_k \cap \{T_{n+1} = j\}} X_{k} dP + \int_{A_k \cap \{T_{n+1} \geq N\}} X_{N-1} dP - \int_{A_k \cap \{T_{n+1} > N\}} \int_{A_k \cap \{T_{n+1} > N\}} (X_{N} - X_{T_{n+1}}) dP$

$\leq ... \leq \int_{A_k \cap \{T_{n+1} = k\}} X_{k} dP - \int_{A_k \cap \{T_{n+1} > N\}} \int_{A_k \cap \{T_{n+1} > N\}} (X_{N} - X_{T_{n+1}}) dP$.

The last integral converges to 0, if $N$ runs through a suitable subsequence of integers, and this shows the claim for integrable $X_{T_n}$. 
2.2 Martingales

(B) Again we assume that the $X_{T_n}$ are integrable. We show:

\[
E(X_1) \geq E(X_{T_n}) \geq \liminf_{N \to \infty} E(X_N)
\]

\[
E(|X_{T_n}|) \leq E(X_1) + 2 \limsup_{N \to \infty} E(X_N^+).
\]

Using part (A) it follows for $S_1 = 1$, $S_m = T_n (m \geq 2)$, that $X_{S_m}$ is a supermartingale, whence

\[
E(X_{S_m}) = E(X_{T_n}) \leq E(X_1).
\]

On the other hand

\[
E(X_{T_n}) = \lim_{N \to \infty} \sum_{k=1}^{N} \int_{\{T_n = k\}} X_k \, dP
\]

\[
\geq \limsup_{N \to \infty} \sum_{k=1}^{N} \int_{\{T_n = k\}} X_N \, dP
\]

\[
= \limsup_{N \to \infty} \int_{\{T_n \leq N\}} X_N \, dP
\]

\[
= \limsup_{N \to \infty} E(X_N) - \int_{\{T_n > N\}} X_N \, dP
\]

\[
\geq \liminf_{N \to \infty} E(X_N).
\]

The second inequality is derived from the first one:

\[
E(|X_{T_n}|) = E(X_{T_n}^+) + E(X_{T_n}^-) = E(X_{T_n}) + 2E(X_{T_n}^-).
\]

Since $-X_{T_n}^-$ is a supermartingale, it follows that

\[
E(-X_{T_n}^-) \geq \liminf_{N \to \infty} E(-X_N^-),
\]

and together with $E(X_{T_n}) \leq E(X_1)$ also that

\[
E(|X_{T_n}|) \leq E(X_1) + 2 \limsup_{N \to \infty} E(X_N^-).
\]

(C) We assume finally that $X_{T_n}$ is not integrable. We shall show that this cannot hold. Let $M \in \mathbb{N}$ be fixed but arbitrary. Define

\[
S_n^M = \begin{cases} T_n & \text{if } T_n \leq M \\ M & \text{else,} \end{cases}
\]

then $S_n^M$ ($n \geq 1$) is a sequence of stopping times, and one gets
\[ E(|X_{SM}|) = \sum_{k=1}^{M} \int_{S_n = k} |X_k| \, dP < \infty. \]

By part (B) and by assumption \( E(|X_{SM}|) \leq \alpha < \infty \), where \( \alpha = E(X_1) + 2 \limsup_{N \to \infty} E(X_N) \). By construction
\[ \lim_{M \to \infty} X_{SM} = X_T, \]
and using the lemma of Fatou, one obtains that
\[ E(|X_T|) = E(\lim_{M \to \infty} |X_{SM}|) \leq \liminf_{M \to \infty} E(|X_{SM}|) \leq \alpha < \infty. \]

Corollary 5.
\[ E(X_1) \geq E(X_T) \geq \liminf_{N \to \infty} E(X_N) \]
\[ E(|X_T|) \leq E(X_1) + 2 \limsup_{N \to \infty} E(X_N). \]

Corollary 6. Let \((X_i)_{1 \leq i \leq n}\) be a SMG (MG) and let \((T_j)_{1 \leq j \leq m}\) stopping times (w.r.t. \((A_i)_{i=1}^{n}\)) satisfying \( T_j \leq T_{j+1} \). Then \((X_{T_j})_{1 \leq j \leq m}\) is a SMG (MG) as well.

Corollary 7. Let \((X_i)_{1 \leq i \leq n}\) be a SMG (MG) and \( T \) a stopping time. Then
\[ Y_n(\omega) := \begin{cases} X_n(\omega) & \text{if } n \leq T(\omega) \\ X_T(\omega) & \text{if } n > T(\omega) \end{cases} \]
is a SMG (MG).

Proof. \( S_n := \min\{n, T\} \) is a stopping time, bounded by \( n \). The assumptions of the theorem are fulfilled for \( X_{S_n} \), and since \( X_{S_n} = Y_n \) the claim follows.
2.3 A.S. Convergence of Martingales

The main result of this section is

**Theorem 33. (Doob’s convergence theorem for supermartingales)**

Let \( \{X_n : n \in \mathbb{N}\} \) be a supermartingale satisfying

\[
\sup_{n \in \mathbb{N}} E(X_n^-) < \infty.
\]

Then there is an integrable random variable \( X \), such that

\[
\lim_{n \to \infty} X_n = X \quad \text{a.e.}
\]

**Remark 22.** Note that the filtration \( \mathcal{A}_n \) is not specified. The supermartingale is therefore adapted to any monotone family of \( \sigma \)-algebras.

**Proof.** Let

\[
E := \{\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) =: X(\omega) \text{ exists and is finite}\},
\]

\[
F_{p,q}^+ := \{\omega \in \Omega : X_n \geq \frac{q + 1}{p} \text{ for only finitely many } n \in \mathbb{N}\} \quad (p \in \mathbb{N}, \ q \in \mathbb{Z}),
\]

\[
F_{p,q}^- := \{\omega \in \Omega : X_n \leq \frac{q}{p} \text{ for only finitely many } n \in \mathbb{N}\} \quad (p \in \mathbb{N}, \ q \in \mathbb{Z}),
\]

\[
F_{p,q} := F_{p,q}^+ \cup F_{p,q}^- \quad (p \in \mathbb{N}, \ q \in \mathbb{Z}),
\]

\[
F_0 := \{\omega \in \Omega : \lim_{n \to \infty} |X_n(\omega)| = \infty\}.
\]

One easily checks that

\[
E = F_0^c \cap \bigcap_{p=1}^{\infty} \bigcap_{q \in \mathbb{Z}} F_{p,q},
\]

since for \( \omega \in E \), it follows for \( p \in \mathbb{N}, \ q \in \mathbb{Z} \) that

\[
\lim_{n \to \infty} \sup X_n(\omega) \leq \frac{q + 1}{p}
\]

or

\[
\lim_{n \to \infty} \inf X_n(\omega) \geq \frac{q}{p}.
\]

Conversely, one of both inequalities holds for any \( p \in \mathbb{N}, \ q \in \mathbb{Z} \), then the limit exists (is finite or infinite), hence if in addition \( \omega \notin F_0 \), the limit is finite.

It is left to show that \( P(E) = 1 \). It suffices to show that \( P(F_0) = 0 \) and \( P(F_{p,q}^c) = 0 \) for every \( p \in \mathbb{N} \) and \( q \in \mathbb{Z} \).

Let \( p \in \mathbb{N} \) and \( q \in \mathbb{Z} \) be fixed and set \( a = q/p, \ b = (q + 1)/p \) and \( G = \Omega \setminus F_{p,q} \). Define recursively for \( m \in \mathbb{N} \)
By construction one has $T_k < T_{k+1}$ and
$$G = \bigcap_{k=1}^{\infty} \{ \omega \in \Omega : T_k(\omega) < \infty \}.$$  

For the moment, fix $M \in \mathbb{N}$ and define
$$S_k^M := \min\{T_k, M\} \quad k \in \mathbb{N}.$$  

$S_k^M$ are a stopping times with $S_k^M \leq S_{k+1}^M$. Moreover,
$$E(|X_{S_k^M}|) = \sum_{j=1}^{M} \int_{\{S_k^M = j\}} |X_j| \, dP < \infty \quad k \in \mathbb{N}$$
and
$$\int_{\{S_k^M \geq N\}} X_N \, dP = 0 \quad \text{for } k \in \mathbb{N}, N > M.$$  

By the optional sampling theorem $X_{S_k^M}$ ($k \in \mathbb{N}$) is a supermartingale.

Let
$$Z_M := \sum_{k=1}^{\infty} (X_{S_k^M} - X_{S_k^M}).$$  

$Z_M$ is a finite sum, since the sequence of stopping times $S_k^M$ is finally constant $= M$. Define
$$Y_M(\omega) = \min\{n \in \mathbb{N} : X_{S_n^M}(\omega) = M\}.$$  

On the one hand it follows using the SMG-property that
$$E(Z_M) = \sum_{k=1}^{\infty} E(X_{S_k^M} - X_{S_k^M}) \geq 0,$$
and on the other hand that
$$Z_M = \sum_{k=1}^{Y_M} (X_{S_{2k-1}^M} - X_{S_{2k}^M}) + (X_{S_{2Y_M+1}^M} - X_{S_{2Y_M+2}^M}) \leq Y_M(a - b) + (a - X_M)^+.$$
Consequently

\[ 0 \leq E(Z_M) \leq -(b-a)E(Y_M) + a + E(X_M^-), \]

and so

\[ E(Y_M) \leq \frac{a - E(X_M)}{b-a} \leq \frac{a}{b-a} + \sup_{n \in \mathbb{N}} E(X_n^-), \]

Assume, that \( P(G) > 0 \), the sequence \( Y_M \leq Y_{M+1} \) is monotonically increasing and converges on \( G \) towards \( \infty \). Therefore

\[ \lim_{M \to \infty} E(Y_M) = \infty \]

is a contradiction.

It will be shown now that \( P(F_0) = 0 \). If \( P(F_0) > 0 \), then the lemma of Fatou implies that

\[ \infty = \int \liminf_{n \to \infty} |X_n| dP \leq \liminf_{n \to \infty} \int X_n dP = \liminf_{n \to \infty} E(X_n) + 2 \inf_{n \to \infty} E(X_n^-) \leq E(X_1) + 2 \liminf_{n \to \infty} E(X_n^-) < \infty. \]

This calculation finishes the proof of the existence of the limit \( X \) almost surely. In particular the previous calculation also shows that

\[ \int |X| dP \leq E(X_1) + 2 \liminf_{n \to \infty} E(X_n^-) < \infty. \]

This concludes the proof.

**Theorem 34. (Convergence theorem for uniformly integrable SMG)**

Let \( \{X_n : n \geq 1\} \) be a supermartingale with respect to \( \{A_n : n \geq 1\} \) and a uniformly integrable family of random variables. Define \( I := \mathbb{N} \cup \{\infty\} \) with the order completed by setting \( n \leq \infty \) (\( n \in \mathbb{N} \)). Let \( \mathcal{A}_\infty := \sigma(\{A_n : n \in \mathbb{N}\}) \).

Then there is an integrable random variable \( X_\infty \) with the following properties:

(a) \( \lim_{n \to \infty} X_n = X_\infty \) almost surely.
(b) \( X_n \ (n \geq 1) \) converges to \( X_\infty \) in \( L_1(P) \), i.e.

\[ \lim_{n \to \infty} E|X_n - X_\infty| = 0. \]

(c) \( \{X_n : n \in I\} \) is a supermartingale.

**Proof.** We first prove (a). One has to show that the assumptions in Doob’s convergence theorem hold. Let \( \epsilon > 0 \). There \( M \in \mathbb{N} \) satisfying

\[ \sup_{n \in \mathbb{N}} \int_{\{|X_n| \geq M\}} |X_n| dP < \epsilon. \]

This implies that
\[
\sup_{n \in \mathbb{N}} E(|X_n|) < \epsilon + \sup_{n \in \mathbb{N}} \int_{\{|X_n| < M\}} |X_n| \, dP \leq M + \epsilon < \infty.
\]

We show (b). Since \( L_1(P) \) is complete, it suffices to show that the sequence \( (X_n) \) is Cauchy in \( L_1(P) \). If so the sequence has a limit in \( L_1(P) \), which – since it is also the limit in probability – equals \( X_\infty \) a.s. So we let \( \epsilon > 0 \). Since the family \( \{X_n - X_m : n, m \in \mathbb{N}\} \) is as well uniformly integrable, there is \( M \in \mathbb{N} \) with
\[
\sup_{n \in \mathbb{N}} \int_{\{|X_n - X_m| \geq M\}} |X_n - X_m| \, dP < \epsilon.
\]
Because of the almost sure convergence there exists \( n_0 \), such that for \( n,m \geq n_0 \)
\[
P(\{|X_n - X_m| \geq \epsilon\}) \leq \frac{\epsilon}{M}.
\]
It follows for \( n,m \geq n_0 \)
\[
\int_{\{|X_n - X_m| < M\}} |X_n - X_m| \, dP = \int_{\{|X_n - X_m| < \epsilon\}} |X_n - X_m| \, dP + \int_{\{\epsilon \leq |X_n - X_m| < M\}} |X_n - X_m| \, dP \\
\leq \epsilon + MP(\{|X_n - X_m| \geq \epsilon\}) \leq 2\epsilon.
\]
Combining everything yields
\[
E|X_n - X_m| \leq 3\epsilon \quad (n,m \geq n_0).
\]
It is clear (because of (a)) that \( X_\infty \) is \( \mathcal{A}_\infty \) measurable. The SMG-property in (c) needs to be proved only for \( n \in \mathbb{N} \) and \( \infty \). Let \( A \in \mathcal{A}_n \), we have that
\[
\int_A X_m \, dP \leq \int_A X_n \, dP \quad \forall m \geq n.
\]
Because of the \( L_1 \)-convergence in (b) one may pass to the limit when \( m \to \infty \), whence (c) holds.

**Theorem 35. (Convergence theorem for uniformly integrable MG)**

Let \( \{X_n : n \geq 1\} \) be a martingale with respect to \( \{\mathcal{A}_n : n \geq 1\} \) and a uniformly integrable family of random variables. Define \( I := \mathbb{N} \cup \{\infty\} \) with the order completed by setting \( n \leq \infty \) (\( n \in \mathbb{N} \)). Let \( \mathcal{A}_\infty := \sigma(\{\mathcal{A}_n : n \in \mathbb{N}\}) \). Then there exists an integrable random variable \( X_\infty \) with the following properties:

(a) \( \lim_{n \to \infty} X_n = X_\infty \) almost surely.

(b) \( X_n \ (n \geq 1) \) converges to \( X_\infty \) in \( L_1(P) \), i.e.
\[
\lim_{n \to \infty} E|X_n - X_\infty| = 0.
\]
(c) \( \{X_n : n \in I \} \) is a martingale.
(d) \( X_n = E(X_\infty | A_n) \) a.s. for every \( n \in \mathbb{N} \).

**Proof.** (a)-(c) follow from the last theorem, since \( \{X_n\} \) and \( \{-X_n\} \) are uniformly integrable SMG. It is left to show (d). let \( n \in \mathbb{N} \) and \( A \in \mathcal{A}_n \). Since \( X_n \) is \( \mathcal{A}_n \)-measurable, the assertion follows from

\[
A \in \mathcal{A}_n \implies \int_A E(X_\infty | \mathcal{A}_n) \, dP = \int_A X_\infty \, dP = \int_A X_n \, dP.
\]

**Example 9.** (A) Let \( X \) be an integrable random variable on \((\Omega, \mathcal{F}, P)\), and let \( \{A_t : t \in I\} \) be an increasing family (i.e. \( t \leq t' \implies A_t \subset A_{t'} \)) of sub-\( \sigma \)-algebras. Then

\[
\{X_t := E(X|A_t) : t \in I\}
\]

is a uniformly integrable martingale.

Only uniform integrability has to be proven. Jensen’s inequality implies

\[
|E(X|A_\infty)| \leq E(|X| | A). \quad (m \to \infty),
\]

since by Jensens inequality

\[
P(|X_t| \geq M) \leq \frac{1}{M} E(|X_t|) = \frac{1}{M} E(|E(X|A_t)|) \leq \frac{1}{M} E(E(|X| | A_t)) = \frac{1}{M} E(|X|).
\]

(B) Let \( X_1, X_2, \ldots \) be independent identically distributed random variables with \( P(\{X_1 = \pm 1\}) = 1/2 \pm \eta, \eta \leq 0 \). Denote by \( S_0 > 0 \) the capital at the beginning of the gamble and

\[
S_{n+1} = S_n + b_n(X_1, \ldots, X_n)X_{n+1} \quad (n \geq 0)
\]

the capital after the \( n+1 \)-st play. We have that \( b_n \) is \( \mathcal{A}_n \)-measurable, and \( S_n \) is a supermartingale. We assume, that the players are forced to bet a minimal amount \( \alpha \) (if he bets at all). Since the capital cannot become negative, one has that \( S_n \) is never negative, so \( ES_n = 0 \) for every \( n \in \mathbb{N} \), and \( S_n \to S \) a.s. by the convergence theorem for SMG. The event, playing infinitely often, is described by the event

\[
\{|S_{n+1} - S_n| \geq \alpha \text{ infinitely often}\}.
\]

Since \( S_n \) converges to a finite limit, this event can occur only with probability 0. Hence the player will lose his capital in finite time.

**Theorem 36.** (Strong law of large numbers of Kolmogorov)
Let \( X_1, X_2, \ldots \) be independent random variables and let \( a_n \in \mathbb{R}_+ \), such that:

1. \( M := \sum_{n=1}^{\infty} \frac{\sigma^2(X_n)}{a_n} < \infty \)
(2) \( a_n \leq a_{n+1} \rightarrow \infty \).

Then
\[
\frac{1}{a_n} \sum_{i=1}^{n} (X_i - E(X_i)) \quad (n \geq 1)
\]
converges almost surely to 0.

**Proof.** W.l.o.g. \( E(X_i) = 0 \) for every \( i \geq 1 \). The sequence
\[
Z_n = \sum_{i=1}^{n} \frac{X_i}{a_i} \quad (n \geq 1)
\]
is a martingale. One has
\[
E(|Z_n|) \leq \left( \sum_{i=1}^{n} \frac{\sigma_i^2(X_i)}{a_i} \right)^{1/2} \leq \sqrt{M}.
\]

By the convergence theorem for SMG, the sequence \((Z_n)_{n \geq 1}\) converges a.s. to a random variable \( X \). Using Kronecker’s lemma, the claim follows.

**Remark 23.** Kronecker’s lemma says: If \(\{c_n : n \in \mathbb{N}\}\) is a summable sequence of real numbers and \(\{a_n : n \in \mathbb{N}\}\) a further sequence monotonically increasing to infinity, then \(\sum_{1 \leq k \leq n} a_k c_k\) converges to zero.

**Theorem 37.** Let \( X \) be an integrable random variable and \((A_n : n \geq 1)\) a monotonically decreasing sequence of \(\sigma\)-algebras. Then
\[
\lim_{n \to \infty} E(X|A_n) = E(X|A_\infty) \quad a.s. \text{ and in } L_1(P),
\]
where \( A_\infty = \bigcap_{n \geq 1} A_n \).

**Proof.** This can be proved in the same way as Doob’s convergence theorem for SMG. Just apply the optional sampling theorem for time periods \( n, n-1, n-2, \ldots, 1 \).

**Theorem 38.** (Strong law of large numbers for identically distributed random variables)
Let \( X_1, X_2, \ldots \) be independent and identically distributed random variables with \( E(|X_1|) < \infty \). then the sequence
\[
\frac{1}{n} \sum_{k=1}^{n} X_k \quad (n \geq 1)
\]
converges almost surely to \( E(X_1) \).
Proof. Let
\[ A_n := \sigma\left( (X_{(1)}, \ldots, X_{(n)}), X_{n+1}, X_{n+2}, \ldots \right) \]
denote the σ-algebra generated by the order statistics \((X_{(1)}, \ldots, X_{(n)})\) of the first \(n\) random variables and all subsequent variables \(X_{n+j}\). This sequence of σ-algebras is monotonically decreasing, so monotonically increasing with respect to the order defined by \(n \preceq m \iff n \geq m\). Let
\[ S_n = \frac{1}{n} \sum_{k=1}^{n} X_k \quad (n \geq 1). \]
Observe that \(S_n\) is a function of the order statistic, so it is \(A_n\)-measurable. For any \(A \in A_n\) one also has that
\[ \int_A S_n \, dP = \int_A X_1 \, dP, \]
since the sets in \(A_n\) do not change if two variables \(X_i\) and \(X_j\) with \(1 \leq i, j \leq n\) are exchanges. It follows that
\[ S_n = E(X_1 | A_n) \quad (n \geq 1). \]
It follows that \((S_n)_{n \geq 1}\) is a martingale, which is uniformly integrable. Hence \(S_n\) converges almost surely to a random variable \(S_\infty\). This limiting random variable must be constant and equal to \(E(X_1)\), since the sequence \(S_n\) converges by the weak law of large numbers stochastically to \(E(X_1)\).
3 DISTRIBUTIONAL CONVERGENCE

3.1 Vague and Weak Convergence

Let $E$ be a Banach space and $E^* = \{L : E \to \mathbb{R}; L \text{ linear and continuous} \}$ its (topological) dual (space). For $E_0 \subset E$, one can introduce a notion of convergence on $E^*$ as follows:

$$L, L_n \in E^*, \lim_{n \to \infty} L_n = L \iff \forall f \in E_0 : \lim_{n \to \infty} L_n(f) = L(f).$$

Likewise one can introduce a notion of convergence on $E$ using $E_0^* \subset E^*$:

$$f, f_n \in E, \lim_{n \to \infty} f_n = f \iff \forall L \in E_0^* : \lim_{n \to \infty} L(f_n) = L(f).$$

If $\Omega$ is a topological space, we denote by $C_b(\Omega)$ the Banach space of all bounded continuous functions equipped with the supremum-norm

$$\|f\| = \sup_{\omega \in \Omega} |f(\omega)|.$$

Likewise we use $C_K(\Omega)$ to denote the subspace of $C_b(\Omega)$ to denote those functions which have a compact support.

The above idea for different types of convergence leads to

**Definition 29.** (1) Let $E$ be a Banach space. A sequence $f_n \in E$ is called weakly convergent to $f \in E$, if

$$\lim_{n \to \infty} L(f_n) = L(f) \quad \forall L \in E^*$$

holds.

(2) Let $E$ be a Banach space. A sequence $L_n \in E^*$ is called weak-* convergent to $L \in E^*$, if

$$\lim_{n \to \infty} L_n(f) = L(f) \quad \forall f \in E$$

holds.

(3) Let $\Omega$ be a topological space. A sequence $m_n$ ($n \in \mathbb{N}$) of finite measures is called weakly convergent to the finite measure $m$, if for every $f \in C_b(\Omega)$

$$\lim_{n \to \infty} \int f \, dm_n = \int f \, dm$$

holds. This means they converge, considered as elements in $C_b(\Omega)^*$, in the weak-* sense to $m$. We write $s - \lim_{n \to \infty} m_n = m$. 
DISTRIBUTIONAL CONVERGENCE

(4) Let $\Omega$ be a topological space. A sequence $m_n$ ($n \in \mathbb{N}$) of finite measures is called vaguely convergent to the finite measure $m$, if for every $f \in C_K(\Omega)$

$$\lim_{n \to \infty} \int f \, dm_n = \int f \, dm$$

holds. This means, that as elements of $C_K(\Omega)^*$, the convergence is a weak-* one towards $m$. We write $v\lim_{n \to \infty} m_n = m$.

The set $M_e(\Omega)$ of all finite measures on a topological space $\Omega$ can be turned into a topological space, such that convergence in this topology agrees with the weak convergence of measures. The topology is defined by the following open neighborhood basis of a measure $m$:

$$U(f_1, \ldots, f_n; \varepsilon) := \{\mu \in M_e(\Omega) : |\int f_i \, dm - \int f_i \, d\mu| < \varepsilon \quad (1 \leq i \leq n)\},$$

where $\varepsilon > 0$ and where $f_1, \ldots, f_n$ are continuous bounded functions. The topology is separable if the function space has a countable dense set. In this case $M_e(\Omega)$ is metrizable.

Example 10. Let $E = L_1(m)$ where $m$ is a $\sigma$-finite measure. Then $E^* = L_\infty(m)$, hence $f_n \to f$ weakly $\iff \int f_n g \, dm \to \int f g \, dm$ for every function $g \in L_\infty(m)$.

Theorem 39. (Vitali-Hahn-Sachs) Let $m$ be a $\sigma$-finite measure on a measurable space $(\Omega, \mathcal{F})$ and let $f_n$ ($n \geq 1$) be integrable functions. If for $A \in \mathcal{F}$ the integrals $\int_A f_n \, dm$ converge to a (finite) limit, then the sequence $\{f_n : n \in \mathbb{N}\}$ is uniformly integrable, and there exists a function $f \in L_1(m)$, such that $f_n$ converges weakly in $L_1(m)$ to $f$.

Proof. (1) Assume the theorem is correct for every finite measure. Since $m$ is $\sigma$-finite, there is $h \in L_1(m)$ with $h > 0$ on $\Omega$ (for example $h = \sum_{n \geq 1} 1_{E_n} 2^{-n} / m(E_n)$ where $E_n \uparrow \Omega$). Consider $g_n = f_n / h \in L_1(h \, dm)$ and the finite measure $m_h(F) = \int_F h \, dm$. For $F \in \mathcal{F}$ it follows that

$$\int_F g_n \, dm = \int_F f_n \, dm$$

converges to an absolutely continuous signed measure (evaluated at $F$), and that $\{g_n\}$ is uniformly integrable. This implies that also $\{f_n\}$ is uniformly integrable.

(2) It is left to show the theorem when $m(\Omega)$ is finite.

Consider the equivalence relation $\sim$ on $\mathcal{F}$ which is defined by

$$E \sim F \iff m(E \Delta F) = 0 \iff 1_F = 1_E \quad a.e.$$
It is easy to see (embed $F_{f^-}$ in $L_1(m)$ isometrically; completeness of $L_1(m)$), such that the space $F_{f^-}$ with metric $d(F, E) = m(E \Delta F)$ becomes a complete metric space and hence is of second category. The last statement means, that the space is not representable as a countable union of nowhere dense and closed sets. The completeness can be seen as follows: $1_{F_n} \to f$ in $L_1(m)$ means $f \in \{0, 1\}$ almost surely, hence $f$ is an indicator function.

Now let
\[
F_k(\epsilon) := \{ B \in F_{f^-} : \sup_{n \geq k} \int_B |f_n - f_k| dm \leq \epsilon \}.
\]
Since $F_k(\epsilon)$ are closed sets and their union (over $k$) covers $F$ there exist $k_0 \geq 1$, $B_0 \in F_{f^-}$ and $\delta > 0$, such that $K(B_0, \delta) \subset F_{k_0}(\epsilon)$.

Now let $B \in F_{f^-}$, $m(B) < \delta$. We have that $B = B_1 \setminus B_2$ with $B_1 = B_0 \cup B$, $B_2 = B_0 \cap (B \cap B_0)^c$ and $m(B_0 \Delta B_i) < \delta$ ($i = 1, 2$). It follows for $n \geq k_0$ that
\[
|\int_B f_n - f_{k_0} dm| \leq \sum_{i=1}^2 |\int_{B_i} f_n - f_{k_0} dm| \leq 2\epsilon.
\]
We may choose $\delta$ so small, that $m(B) < \delta$ implies $|\int_B f_{k_0} dm| < \epsilon$. Hence the following statement holds: For $\eta > 0$ there are $k_0 \in \mathbb{N}$ and $\delta(\eta) > 0$, such that $m(B) < \delta(\eta)$ implies
\[
|\nu(B)| \leq \eta \quad \text{and} \quad |\int_B f_n dm - \nu(B)| < \eta,
\]
where $\nu(B) = \lim_{n \to \infty} \int_B f_n dm$.

(3) We now show the theorem in case $m(\Omega) < \infty$. It is immediately clear from (2), that $\nu$ is a finite signed measure on $(\Omega, \mathcal{F})$ which is absolutely continuous with respect to $m$:

1. $\nu(\emptyset) = \lim_{n \to \infty} \int_B f_n dm = 0$,
2. For disjoint measurable sets $A$ and $B$ one gets $\nu(A \cup B) = \lim_{n \to \infty} \int(1_A + 1_B) f_n dm = \nu(A) + \nu(B)$,
3. $m(A) = 0$ implies that $\nu(A) = \lim_{n \to \infty} \int_A f_n dm = 0$. Hence $\nu$ absolutely continuous with respect to $m$.
4. Let $A_n$ be pairwise disjoint measurable sets ($n \geq 1$). Letting $D_N = \bigcup_{k=N+1}^{\infty} A_k$, $\lim_{N \to \infty} m(D_N) = 0$. By (2) it follows that $\nu(D_N) \to 0$, and, moreover, that
\[
\nu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \nu(A_k) + \nu(D_N) \to \sum_{k=1}^{\infty} \nu(A_k).
\]
The absolute convergence of the series follows by taking the subsequences $\nu(A_n) \geq 0$, resp. $\leq 0$. 

3.1 Vague and Weak Convergence
The theorem of Radon-Nikodym now implies that \( f = \frac{d\nu}{dm} \) is the desired function, to which \( f_n \) converges weakly: By definition
\[
\int_A f \, dm = \nu(A) = \lim_{n \to \infty} \int_A f_n \, dm \quad (A \in \mathcal{F}),
\]
and
\[
\int g \, dm = \int g \, d\nu = \lim_{n \to \infty} \int g \, f_n \, dm
\]
follows for any \( g \in L_\infty(m) \) by approximation via step functions.

Therefore it is sufficient to show: Assume the sequence is not uniformly integrable. Then there is a set \( A \in \mathcal{F} \) such that \( \lim_{n \to \infty} \int_A f_n \, dm \) does not exist. This is a contradiction.

So let \( \epsilon > 0 \) be chosen such that
\[
\lim_{N \to \infty} \sup_{n \geq 1} \int_{\{|f_n| \geq N\}} |f_n| \, dm > \epsilon.
\]
Passing to a suitable subsequence we may assume that
\[
\int_{\{|f_N| \geq N\}} |f_N| \, dm > \epsilon \quad (N \geq 1).
\]
Again, passing to a further subsequence, we may assume that
\[
\int_{\{f_N \geq N\}} f_N \, dm > \epsilon \quad (N \geq 1)
\]
(or the corresponding assertion for \( f_N \leq -N \)).

Moreover, we may assume that there are measurable sets \( A_n \subset \{f_N > N\} \) satisfying \( m(A_n) \to 0 \) and \( \int_{A_n} f_N \, dm > \epsilon \). This can be seen as follows: Let \( \Omega_1 = \{ \omega \in \Omega : m(\{\omega\}) > 0 \} \). \( \Omega_1 \) contains at most countably many points whose mass tends to zero (since \( m(\Omega) < \infty \)). Since \( \int_{\{\omega\}} f_N \, dm = f_N(\omega) \to \nu(\{\omega\}) \) for \( \omega \in \Omega_1 \) it follows that \( m(\Omega_1 \cap \{f_N > N\}) \to 0 \). The subsequence can be found immediately (for \( \epsilon/2 \)) or one considers \( \Omega \setminus \Omega_1 \). In the latter case the function \( t \to \int_{\Omega_1 \cap \{f_N > t\}} f_N \, dm \) is continuous in \( t \) and vanishes at infinity; and this gives \( A_N = \{f_N > t\} \cap \Omega_1 \) immediately by choosing \( t \) satisfying \( \int_{A_N} f_N \, dm = \epsilon/2 \).

Let \( \eta < \epsilon/6 \). We construct \( N_k \to \infty \) and \( \delta_k < \delta(\eta) \) (see (2)), such that
\begin{itemize}
  \item[(a)] \( m(A_{N_l} \cup \ldots \cup A_{N_j}) \leq \delta_l \quad (l = 1, 2, \ldots; j = l, l + 1, \ldots) \)
  \item[(b)] \( \left| \int_{A_{N_l} \cup \ldots \cup A_{N_j}} f_{N_{j+1}} \, dm - \nu(A_{N_l} \cup \ldots \cup A_{N_j}) \right| < \eta \)
  \item[(c)] \( m(B) < \delta_{j+1} \) impliziert \( |\int_B f_{N_j} \, dm| < \eta \).
\end{itemize}

For \( k = 1 \) choose \( \delta_1 < \delta(\eta) \) and \( m(A_{N_k}) < \delta_1 \). Then we choose \( \delta_2 < \delta(\eta) \) so small, that (c) for \( f_{N_k} \) as the integrable function holds.
3.1 Vague and Weak Convergence

If \( N_1, \ldots, N_k \) and \( \delta_1, \ldots, \delta_{k+1} \) are chosen, such that (a)-(c) hold, choose \( N_{k+1} \) so large, that (a) for \( l = 1, 2, \ldots, k + 1; j = l, l + 1, \ldots, k + 1 \) and (b) for \( j = k \) hold. Then choose \( \delta_{k+2} < \delta(\eta) \) so small, that (c) for \( f_{N_{k+1}} \) is true.

Put \( B_k = A_{N_1} \cup \ldots \cup A_{N_k} \) for \( k = 1, 2, \ldots, \infty \) and \( A = B_\infty \). It follows that

\[
\left| \int_A f_{N_k} dm - \nu(A) \right| \\
= \left| \int_{A_{N_k}} f_{N_k} dm + \int_{B_{k-1}} f_{N_k} dm + \int_{A \setminus B_k} f_{N_k} dm \\
- \int_{B_{k-1} \cap A_{N_k}} f_{N_k} dm - \nu(A) \right| \\
> \epsilon - 6\eta > 0,
\]

since \( |\nu(B_k)| < \eta, m(A \setminus B_k) < \delta_{k+1}, |\int_{B_{k-1}} f_{N_k} dm| \leq |\nu(B_{k-1})| + \eta < 2\eta \).

This is a contradiction.

**Corollary 8.**

(1) Let \( m_n \in \mathcal{M}_e(\Omega) \) be finite measures such that

\[
m(F) = \lim_{n \to \infty} m_n(F) \quad F \in \mathcal{F}
\]

in \( \mathbb{R} \) exists. Then \( m \) is a finite measure.

(2) If the sequence \( \{f_n : n \geq 1\} \) of measurable functions is uniformly integrable, then it is relatively compact in the weak topology on \( L_1(m) \).

**Proof.** Only (1) has to be shown. Let

\[
\nu(F) = \sum_{n \geq 1} 2^{-n} \frac{m_n(F)}{m_n(\Omega)} \quad F \in \mathcal{F}.
\]

\( \nu \) is a finite measure and every \( m_n \) is absolutely continuous with respect to \( \nu \). Let \( f_n \) denote the associated Radon-Nikodym derivative, which belongs to \( L_1(\nu) \). It follows that

\[
\int_F f_n \, d\nu = m_n(F) \to m(F) \quad F \in \mathcal{F}.
\]

By the theorem of Vitali-Hahn-Saks there exists an integrable function \( f \) with \( m(F) = \int_F f \, d\nu \). This shows the claim.

The most important theorem on weak convergence is the theorem of Helly and Bray.

**Theorem 40.** (Helly-Bray)

Let \( \Omega \) be a locally compact, separable space (so it has a countable basis for the topology). Then, for any \( s \geq 0 \), the set of Borel measures \( m \) with \( m(\Omega) \leq s \) is vaguely sequentially compact.
Remark 24. Sequentially compact means the following: For every sequence of measures \( m_n \) with \( m_n(\Omega) \leq s \) there are a Borel measure \( m(\Omega) \leq s \) and a subsequence \( \{n_k : k \geq 1\} \), such that

\[
\lim_{k \to \infty} \int f \, dm_{n_k} = \int f \, dm \quad f \in C_K(\Omega).
\]

Observe also that under the assumptions of the theorem \( \Omega \) is metrizable, and hence every linear form on \( C_K(\Omega) \) is uniquely representable via integrals with respect to some Borel measure.

Proof. The space \( C_K(\Omega) \) is (as a Banach space) separable. Therefore there exists a countable sequence \( f_n (n \geq 1) \) of functions in \( C_K(\Omega) \), which is dense with respect to the metric \( d(f, g) = \|f - g\|_\Omega \). Moreover, one knows that for every \( f \in C_K(\Omega) \)

\[
|\int f \, dm_n| \leq \|f_n\|_\Omega m_n(\Omega) \leq s\|f_n\|_\Omega.
\]

By Cantor’s diagonalization procedure one can find a subsequence \( \{n_k : k \geq 1\} \), such that the limit

\[
\lim_{k \to \infty} \int f_j \, dm_{n_k} =: L(f_j)
\]

exists for every \( j \geq 1 \). This implies for arbitrary \( f \in C_K(\Omega) \):

\[
|\int f \, dm_{n_k} - \int f \, dm_{n_l}| \\
\leq |\int f \, dm_{n_k} - \int f_j \, dm_{n_k}| + |\int f_j \, dm_{n_k} - \int f_j \, dm_{n_l}| + |\int f_j \, dm_{n_l} - \int f \, dm_{n_l}| \\
\leq 2\|f - f_j\|_\Omega + |\int f_j \, dm_{n_k} - \int f_j \, dm_{n_l}|.
\]

This shows that \( L(f) = \lim_{k \to \infty} \int f \, dm_{n_k} \) exists for all \( f \).

\( L \) is an abstract \( \sigma \)-continuous abstract integral, so it has by the theorem of Riesz an integral representation using some Borel measure \( m \). Since \( \Omega \) is \( \sigma \)-compact, it also follows that

\[
m(\Omega) = \lim_{r \to \infty} m(K_r) \leq \lim_{r \to \infty} \lim_{k \to \infty} \int f_r \, dm_{n_k} \leq s,
\]

where \( K_r \uparrow \Omega \) are compact sets, and where \( 1_{K_r} \leq f_r \leq 1, f_r \in C_K(\Omega) \). This completes the proof.

Corollary 9. (Kryloff-Bogliouboff)

Let \( \Omega \) be compact and separable (= compact, metric). Let

\[
T : \Omega \to \Omega
\]
be a continuous mapping. Then there is a $T$-invariant finite Borel measure $m$, i.e.
\[ \int f \circ T \, dm = \int f \, dm \]
for every integrable function $f$.

**Proof.** Let $x_0 \in \Omega$. Define
\[ m_n(F) = \frac{1}{n} \sum_{k=0}^{n-1} 1_{F}(T^k(x_0)) \quad (F \in \mathcal{F}). \]

$m_n$ is a discrete probability measure and for $f \in C^b(\Omega)$ one obtains
\[ \int f \circ T \, dm_n = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x_0)) \]
\[ = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x_0)) + \frac{f(T^n(x_0))}{n} - \frac{f(x_0)}{n} = \int f \, dm_n + \frac{f(T^n(x_0))}{n} - \frac{f(x_0)}{n}. \]

Pick a convergent subsequence from $m_n$. Let $m$ denote the associated limit measure. Since $T$ is continuous, the function $f \circ T$ is continuous and bounded as long as the function $f$ is. It follows that
\[ \int f \circ T \, dm = \int f \, dm \]
for $f \in C^b(\Omega)$. Since $C^b(\Omega)$ is dense in $L_1(m)$, the claim follows.

**Definition 30.** A family $\mathcal{M}$ of finite measures on $(\Omega, \mathcal{B})$ is called tight, if for every $\epsilon > 0$ there is a compact set $K \subset \Omega$, such that for every $m \in \mathcal{M}$
\[ m(\Omega \setminus K) < \epsilon \]
holds.

**Remark 25.** A finite measure on a Polish space (complete and metrizable) is inner regular and hence tight.

**Definition 31.** A family $\mathcal{C} \subset C^b(\Omega)$ of bounded continuous functions on $\Omega$ is called separating, if for any two different measures $m_1$ and $m_2$ on $(\Omega, \mathcal{B})$ there is a function $f \in \mathcal{C}$ satisfying $\int f \, dm_1 \neq \int f \, dm_2$.

**Theorem 41.** Let $\Omega$ a locally compact and $\sigma$-compact topological space. The following assertions are equivalent for finite measures $m_n \ (n \geq 1)$ and $m$ on $(\Omega, \mathcal{B})$:

1. $s - \lim_{n \to \infty} m_n = m$.
2. $v - \lim_{n \to \infty} m_n = m$ and $\lim_{n \to \infty} m_n(\Omega) = m(\Omega)$. 
DISTRIBUTIONAL CONVERGENCE

\( v = \lim_{n \to \infty} m_n = m \) and \( \{m_n : n \geq 1\} \) is tight. 

(4) There is a separating set \( C \subset C_b(\Omega) \) satisfying \( \lim_{n \to \infty} \int f dm_n = \int f dm \) for every \( f \in C \) and \( \{m_n : n \geq 1\} \) is tight.

Proof. 

'(1)\( \Rightarrow \) (2)': is trivial.

'(2)\( \Rightarrow \) (3)': Let \( \epsilon > 0 \). Since \( \Omega \) is a Polish space, \( m \) is tight, and there is a compact set \( K \) with \( m(\Omega \setminus K) < \epsilon \). Since \( \Omega \) is locally compact, there is a continuous function \( g \in C_K(\Omega) \) with \( 1_K \leq g \leq 1 \). Let \( T \) denote the compact support of \( g \). It follows that

\[
m_n(\Omega \setminus T) \leq m_n(\Omega) - \int g dm_n \to m(\Omega) - \int g dm \leq m(\Omega \setminus K).
\]

Hence there exists some \( n_0 \), such that for \( n \geq n_0 \)

\[
m_n(\Omega \setminus T) \leq \epsilon.
\]

Since every \( m_n \) is tight, one can find a compact set \( K_0 \supset T \), such that for \( n \geq 1 \), \( m_n(\Omega \setminus K_0) \leq \epsilon \).

'(3)\( \Rightarrow \) (1)': Let \( f \in C_b(\Omega) \), \( \epsilon > 0 \) and \( K \) compact satisfying \( m_n(\Omega \setminus K) < \epsilon \). Choose \( g \in C_K(\Omega) \) with \( 1_K \leq g \leq 1 \). Then

\[
| \int f dm_n - \int f dm | \\
\leq | \int f \cdot g dm_n - \int f \cdot g dm | + | \int f(1-g) dm_n | + | \int f(1-g) dm | \\
\leq | \int fg dm_n - \int fg dm | + \int_{\Omega \setminus K} |f| dm_n + \int_{\Omega \setminus K} |f| dm \\
\leq | \int fg dm_n - \int fg dm | + 2\epsilon \|f\|_{\Omega}.
\]

Letting \( n \to \infty \) and then \( \epsilon \to 0 \), the claim follows.

'(1)\( \Rightarrow \) (4)': It suffices to show that \( C_b(\Omega) \) itself is separating. If there were two measures \( m_1 \) and \( m_2 \) and no function \( f \in C_b(\Omega) \) with \( \int f dm_1 \neq \int f dm_2 \), we would have a contradiction to Riesz representation theorem: The measure is uniquely determine by its integral values for functions in \( C_K(\Omega) \).

'(4)\( \Rightarrow \) (1)': Suppose, we have \( s - \lim_{n \to \infty} m_n \neq m \). Let \( m_{n_k} \) be a subsequence, for which \( m \) is no accumulation point. By the theorem of Helly-Bray there is a further vaguely convergent subsequence, which converges vaguely to \( m_0 \), say. Since the sequence \( m_{n_k} \) must be tight, it follows from what has been proven so far, that the vague convergence is in fact a weak convergence. Hence we may assume weak convergence:

\[
s - \lim_{k \to \infty} m_{n_k} = m_0
\]

For \( f \in C \) it follows that
\[
\int f\,dm = \lim_{n \to \infty} \int f\,dm_n = \lim_{k \to \infty} \int f\,dm_k = \int f\,dm_0.
\]

A contradiction!

**Corollary 10.** Let \( \Omega = \mathbb{R} \). Then the following statements are equivalent:

Let \( F_n \) and \( F \) \((n \geq 1)\) denote the distributions of measures \( m_n \) and \( m \).

1. \( s - \lim_{n \to \infty} m_n = m \).
2. \( \lim_{n \to \infty} F_n(x) = F(x) \) for all \( x \in \mathbb{R} \), such that \( F \) is continuous in \( x \).
3. \( \lim_{n \to \infty} F_n(x) = F(x) \) for all \( x \in D \subset \mathbb{R} \), such that \( D \) is dense in \( \mathbb{R} \) and \( F \) is continuous in \( x \).

**Proof.** Every indicator function \( I_x := 1_{]-\infty,x]} \) can be approximated from below and above by functions \( f, g \in C_0(\mathbb{R}) \) \((g \leq I_x \leq f)\), such that \(| \int f - g \,d\mu| \leq \mu(\{f \neq g\}) \to \mu(\{x\}) \) gilt. Therefore

\[
F(x) = m(I_x) \geq \lim_{n \to \infty} \int f\,dm_n - m(\{f \neq g\}) = \limsup_{n \to \infty} F_n(x) - m(\{f \neq g\})
\]
\[
\geq \liminf_{n \to \infty} F_n(x) - m(\{f \neq g\}) \geq \lim_{n \to \infty} \int g\,dm_n - m(\{f \neq g\})
\]
\[
\geq F(x) - 2m(\{f \neq g\}).
\]

This implies \((1) \Rightarrow (2)\). \((2) \Rightarrow (3)\) is trivial. \((3) \Rightarrow (1)\) is shown as follows: Suppose, \( m_n \) has an accumulation point \( \mu \), which is different from \( m \). The above calculation shows that

\[
\lim_{n \to \infty} F_n(x) = \mu(I_x) = F(x)
\]

for \( x \in D \). A measure is determined by its distribution function, and because of left continuity the distribution function is determined by its values on a dense subset.

### 3.2 Fourier-Transform

Let \((\Omega, \mathcal{F}, P)\) and \((\Omega_n, \mathcal{F}_n, P_n)\) \((n \geq 1)\) always denote probability spaces. Recall that measurable functions on these spaces are called random variables. Measurable mappings with values in a vector space are called random vectors, and – in the general situation of arbitrary image spaces – random elements. If the image space is a function space, a random element is called a stochastic process.

**Definition 32.** A sequence of random elements

\[
X_n : \Omega_n \to E \quad (n \geq 1)
\]
with values in a topological space $E$ converges in distribution towards the random element

$$X : \Omega \to E,$$

if the distributions $P_{X_n} := P_n \circ X_n^{-1}$ converge weakly to the distribution $P_X := P \circ X^{-1}$ of $X$. With the usual notation $\mathcal{L}(Y)$ for the distribution of a random element $Y$ one can write for short:

$$s - \lim_{n \to \infty} \mathcal{L}(X_n) = \mathcal{L}(X),$$

or

$$\mathcal{D} \quad X_n \to X.$$ 

**Remark 26.** The family

$$\mathcal{C}_d := \{ \sin <t, \cdot >, \cos <t, \cdot > : t \in \mathbb{R}^d \}$$

is a separating family of bounded continuous functions on $\mathbb{R}^d$, where $< \cdot , \cdot >$ denotes the inner product on $\mathbb{R}^d$. As has been shown in the last section convergence for such functions and tightness imply the weak convergence on locally compact spaces. The integrals over these functions are conveniently studied using the integration theory of complex function:

$$\exp[i <t, x>] = \cos <t, x> + i \sin <t, x> \quad (t, x \in \mathbb{R}^d),$$

and

$$\int \exp[i <t, x>] \mu(dx) = \int \cos <t, x> \mu(dx) + i \int \sin <t, x> \mu(dx) \quad (t \in \mathbb{R}^d).$$

**Definition 33.** Let $\mu$ be a probability distribution on $\mathbb{R}^d$.

$$\hat{\mu}(\cdot) := \int \exp[i <\cdot, x>] \mu(dx)$$

is called the Fourier-transform (characteristic function) of $\mu$. If $X$ is a $\mathbb{R}^d$-valued random vector, $\hat{X} = \mathcal{L}(X)$ is called the Fourier-transform (characteristic function) of $X$.

**Remark 27.** (1) If $\mathcal{M}$ is a tight family of probability distributions on $\mathbb{R}^d$, then the family $\{ \hat{\mu} : \mu \in \mathcal{M} \}$ is uniformly equicontinuous. In particular, the Fourier-transform of a distribution is equicontinuous.

This is shown as follows: Let $\epsilon > 0$. By tightness, there is a compact set $K$ satisfying $\mu(K) > 1 - \epsilon$ ($\mu \in \mathcal{M}$). Let $\alpha = \sup_{x \in K} \|x\|$ and $\delta > 0$ be so small, that

$$|\alpha <u| < \delta \implies |1 - \exp[i\alpha u]| < \epsilon.$$ 

It follows for $\mu \in \mathcal{M}$ and $\|t - s\| < \delta$ that
\[ |\hat{\mu}(t) - \hat{\mu}(s)| \leq \int |\exp[i < x, t >] - \exp[i < x, s >]|\mu(dx) \]
\[ \leq \int_K |\exp[i < x, t >] - \exp[i < x, s >]|\mu(dx) + 2\mu(\mathbb{R}^d \setminus K) \]
\[ = \int_K |\exp[i < x, t >]|(1 - \exp[i < x, s - t >]|\mu(dx) + 2\mu(\mathbb{R}^d \setminus K) \]
\[ \leq 3\epsilon. \]

(2) In general
\[ |\hat{\mu}(t + h) - \hat{\mu}(t)|^2 \leq 2|1 - \text{Re}\hat{\mu}(h)| \quad (t, h \in \mathbb{R}^d). \]
(\(\text{Re}z\) denotes the real part of \(z \in \mathbb{C}\) and \(\text{Im}z\) the imaginary part.) In order to prove the assertion use Cauchy-Schwarz inequality:
\[ |\hat{\mu}(t + h) - \hat{\mu}(t)|^2 \leq \int |\exp[i < x, t >]|(\exp[i < x, h >] - 1)^2\mu(dx)^2 \]
\[ \leq \int |\exp[i < x, t >]|^2\mu(dx) \int |\exp[i < x, h >] - 1|^2\mu(dx) \]
\[ = \int \exp[i < x, h >] \exp[-i < x, h >] + 1\mu(dx) \]
\[ - \int (\exp[i < x, h >] + \exp[-i < x, h >])\mu(dx) \]
\[ = 2\left(1 - \int \text{Re}\exp[i < x, h >]\mu(dx)\right) \]
\[ = 2(1 - \text{Re}\hat{\mu}(h)). \]

(3) Let \(X_1, \ldots, X_n\) be independent random vectors with values in \(\mathbb{R}^d\). Then
\[ \sum_{k=1}^n X_k = \prod_{k=1}^n \hat{X}_k. \]

Let \(\mu\) denote the distribution of the partial sum \(\sum X_i\) and \(\mu_k\) the distribution of \(X_k\). Then
\[ \int \exp[i < x, t >]\mu(dx) = \int \ldots \int \exp[i < \sum_{k=1}^n x_k, t >]\mu_1(dx_1)\ldots\mu_n(dx_n) \]
\[ = \int \ldots \int \prod_{k=1}^n \exp[i < x_k, t >]\mu_1(dx_1)\ldots\mu_n(dx_n) \]
\[ = \prod_{k=1}^n \int \exp[i < x, t >]\mu_k(dx). \]
The last equality is obtained using approximation by step functions.
Theorem 42. (Continuity theorem of Lévy)

(1) If the sequence \( \{ \mu_n : n \geq 1 \} \) of probability measures on \( \mathbb{R}^d \) converges weakly to the measure \( \mu \), then \( \{ \hat{\mu}_n : n \geq 1 \} \) converges uniformly on compact set to \( \hat{\mu} \).

(2) If the Fourier-transforms \( \hat{\mu}_n \) of probability measures \( \mu_n \) on \( \mathbb{R}^d \) \( (n \geq 1) \) converge to a function \( \phi : \mathbb{R}^d \to \mathbb{C} \), which is continuous in 0, then there exists a probability measure \( \mu \) satisfying

\[
\lim_{n \to \infty} \mu_n = \mu.
\]

Proof. (1) is simple: The weak convergence implies the convergence of

\[
\lim_{n \to \infty} \int \cos <t,x> \mu_n(dx) = \int \cos <t,x> \mu(dx),
\]

\[
\lim_{n \to \infty} \int \sin <t,x> \mu_n(dx) = \int \sin <t,x> \mu(dx),
\]

also

\[
\lim_{n \to \infty} \int \exp[i <t,x>] \mu_n(dx) = \int \exp[i <t,x>] \mu(dx).
\]

The family \( \mu_n \) is as well tight, and hence the sequence \( \hat{\mu}_n \) is uniformly equi-continuous. This implies the uniform convergence on compact sets.

(2) The assumption immediately implies that

\[
\lim_{n \to \infty} \int \cos <t,x> \mu_n(dx) = \mathbb{R}\phi(t)
\]

\[
\lim_{n \to \infty} \int \sin <t,x> \mu_n(dx) = \mathbb{I}\phi(t).
\]

Since this family of function is separating, it is sufficient to show tightness for the family \( \{ \mu_n : n \geq 1 \} \).

Let \( u > 0 \). For fixed \( 1 \leq j \leq d \) one obtains

\[
\frac{1}{u} \int_0^u \left( 1 - \int_{\mathbb{R}^d} \cos[vx_j] \mu_n(dx) \right) dv
= \frac{1}{u} \int_{\mathbb{R}^d} \left( \int_0^u (1 - \cos[vx_j]) dv \right) \mu_n(dx)
= \frac{1}{u} \int_{\mathbb{R}^d} \left( u - \frac{1}{x_j} \sin[ux_j] \right) \mu_n(dx)
\geq \alpha \mu_n(\{ x \in \mathbb{R}^d : |x_j| > 1 \}),
\]

where

\[
\alpha := \inf_{s \geq 1} \left( 1 - \frac{1}{s} \sin[s] \right).
\]

It follows, using \( e_j(v) = (0, ..., 0, v, 0, ..., 0) \) \( (j\text{-th coordinate}) \), that
\[ \limsup_{n \to \infty} \mu_n \left( \{ x \in \mathbb{R}^d : |x_j| \geq u^{-1} \} \right) \]
\[ \leq \alpha^{-1} \limsup_{n \to \infty} \frac{1}{u} \int_0^u \left( 1 - \Re \mu_n(e_j(v)) \right) dv \]
\[ \leq \alpha^{-1} \frac{1}{u} \int_0^u \left( 1 - \Re \phi(e_j(v)) \right) dv. \]

Note that one needs to apply the dominated convergence theorem. Since \( \phi \) is continuous in 0 and \( \phi(0) = 1 \), we obtain
\[ \lim_{u \to 0} \frac{1}{u} \int_0^u \left( 1 - \Re \phi(e_j(v)) \right) dv = 0 \quad (j = 1, \ldots, d, 0 < u \leq u_0) \]
whence for \( \epsilon > 0 \) there is \( u_0 \) satisfying
\[ \left| \frac{1}{u} \int_0^u \left( 1 - \Re \phi(e_j(v)) \right) dv \right| < \frac{\alpha \epsilon}{2d} \quad (j = 1, \ldots, d, u \leq u_0). \]
Furthermore, there is \( n_0 \in \mathbb{N} \), such that for \( n \geq n_0 \) and \( 0 < u \leq u_0 \)
\[ \left| \frac{1}{u} \int_0^u \left( 1 - \Re \mu_n(e_j(v)) \right) dv \right| < \frac{\alpha \epsilon}{d}. \]
Choosing \( K = \{ x \in \mathbb{R}^d : \forall j \ |x_j| \leq u^{-1} \} \), it follows for \( n \geq n_0 \) that
\[ \mu_n(\mathbb{R}^d \setminus K) \leq \sum_{j=1}^d \mu_n(\{ x \in \mathbb{R}^d : x_j \notin [-u^{-1}, u^{-1}] \}) \]
\[ \leq \sum_{j=1}^d \alpha^{-1} \frac{1}{u} \int_0^u \left( 1 - \Re \phi(e_j(v)) \right) dv \]
\[ = \sum_{j=1}^d \alpha^{-1} \frac{1}{u} \int_0^u \left( 1 - \Re \mu_n(e_j(v)) \right) dv \]
\[ \leq \epsilon. \]

**Example 11.** Let \( X \) be a normal random variable with expectation \( \mu \) and variance \( \sigma^2 > 0 \). Its Fourier-transform is given by
\[ \hat{X}(t) = \exp[i\mu t - \frac{\sigma^2 t^2}{2}] \quad (t \in \mathbb{R}). \]
This can be shown as follows: The Fourier-transform of \( X \) is
\[ \hat{X}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp[ixt - \frac{(x - \mu)^2}{2\sigma^2}] dx. \]
One integrates the complex-valued function
$z \to \exp[it - \frac{(x - \mu)^2}{2\sigma^2}] \quad (z \in \mathbb{C})$

along the closed curve which is given by circumference of the rectangle defined by the points $\pm s$ and $\pm s + it\sigma^2$. One finds (using Cauchy’s integral theorem (which says that the integral has to be zero))

$$
\int_{-s}^{s} \exp[it - \frac{(x - \mu)^2}{2\sigma^2}] \, dx + \int_{0}^{t} \exp[i(s + ix\sigma^2)t - \frac{(s + ix\sigma^2 - \mu)^2}{2\sigma^2}] \, dx
$$

$$
+ \int_{s}^{-s} \exp[i(x + it\sigma^2)t - \frac{(x + it\sigma^2 - \mu)^2}{2\sigma^2}] \, dx
$$

$$
+ \int_{t}^{0} \exp[i(-s + ix\sigma^2)t - \frac{(-s + ix\sigma^2 - \mu)^2}{2\sigma^2}] \, dx
$$

$$
= 0.
$$

For $s \to \infty$ the second and fourth integrals tend to 0. Therefore

$$
\hat{X}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp[i(x + it\sigma^2)t - \frac{(x + it\sigma^2 - \mu)^2}{2\sigma^2}] \, dx
$$

$$
= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp[itx - t^2\sigma^2 - \frac{x^2 - t^2\sigma^4 + \mu^2 + 2itx\sigma^2 - 2x\mu - 2it\mu\sigma^2}{2\sigma^2}] \, dx
$$

$$
= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp[-t^2\sigma^2 / 2 + it\mu] \, \exp[-\frac{(x - \mu)^2}{2\sigma^2}] \, dx
$$

$$
= \exp[it\mu - t^2\sigma^2 / 2].
$$
3.3 Central Limit Theorems

For a random variable $X$ let $E(X)$ denote the expectation

$$E(X) = \int X \, dP$$

and $\sigma^2(X)$ the variance

$$\sigma^2(X) = E(X - E(X))^2 = \int X^2 \, dP - \left( \int X \, dP \right)^2.$$

Recall the definition of independence:

**Definition 34.** A family of random elements

$$X_i : \Omega \rightarrow E_i \quad i \in I,$$

which is defined on the same probability space $(\Omega, \mathcal{F}, P)$ and attains values in the measurable spaces $(E_i, \mathcal{B}_i)$, is called independent, if the $\sigma$-algebras $X_i^{-1} \mathcal{B}_i$ are independent. This means that

$$\forall i_1, \ldots, i_n \in I, \forall F_j \in X_{i_j}^{-1} \mathcal{B}_{i_j} \implies P(F_1 \cap \ldots \cap F_n) = P(F_1) \cdot \ldots \cdot P(F_n).$$

**Theorem 43.** Let $X_1, \ldots, X_n$ be independent random elements with values in $E_j$ and let $\phi_j : E_j \rightarrow \mathbb{C}$ be measurable functions, such that $\phi_j \circ X_j$ are integrable. Then $\prod_{j=1}^{n} \phi_j(X_j)$ is integrable and

$$\int \prod_{j=1}^{n} \phi_j(X_j) \, dP = \prod_{j=1}^{n} \int \phi_j(X_j) \, dP.$$

**Proof.** This is easily proved by approximation via step functions.

**Definition 35.** Let $(\Omega_n, \mathcal{F}_n, P_n) \quad (n \geq 1)$ be probability spaces and $(E, \mathcal{B})$ a measurable space. A family $\{X_{jn} : 1 \leq j \leq k_n, \quad n = 1, 2, 3, \ldots\}$ of random elements

$$X_{jn} : \Omega_n \rightarrow E$$

is called an array of $E$-valued random elements. It is called independent, if for every $n \geq 1$ the random elements $X_{jn}$ ($j = 1, \ldots, k_n$) are independent.

**Definition 36.** An array $\{X_{jn} : 1 \leq j \leq k_n, n \geq 1\}$ of $\mathbb{R}^d$-valued random vectors is called asymptotically negligible, if for every $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \sup_{1 \leq j \leq k_n} P_n(\|X_{jn}\| \geq \epsilon) = 0.$$

This is equivalent to the statement, that the sequences $(X_{jn})_n$ for arbitrary choices $1 \leq j_n \leq k_n$ converge to zero uniformly in the sense of stochastic convergence.
Definition 37. Let \( \{X_{jn} : 1 \leq j \leq k_n, n \geq 1\} \) be an array of square integrable real valued random variables, i.e. \( X_{jn} \in L_2(P_n) \) \( (1 \leq j \leq k_n, n \geq 1) \).

Denote

\[
s_n^2 = \sigma^2(X_{1n} + X_{2n} + \ldots + X_{k_n}) = \sigma^2(X_{1n}) + \ldots + \sigma^2(X_{k_n}).
\]

The array is said to satisfy the Lindeberg condition, if for every \( \varepsilon > 0 \), the quantities

\[
L_n(\varepsilon) = \frac{1}{s_n^2} \sum_{j=1}^{k_n} \int_{\{|X_{jn} - E(X_{jn})| \geq \varepsilon s_n\}} (X_{jn} - E(X_{jn}))^2 \, dP_n
\]

\[
= \frac{1}{s_n^2} \sum_{j=1}^{k_n} E\left(1_{[\varepsilon s_n, \infty)}(|X_{jn} - E(X_{jn})|) \left[X_{jn} - E(X_{jn})\right]^2\right)
\]

tend to 0 as \( n \to \infty \).

Theorem 44. (Central limit theorem under the Lindeberg condition)

Let \( \{X_{jn} : 1 \leq j \leq k_n, n \geq 1\} \) be an independent array of square integrable random variables with \( \sigma^2(X_{jn}) > 0 \) for \( 1 \leq j \leq k_n, n \geq 1 \). Then the following two statements are equivalent:

(1) The array satisfies the Lindeberg condition.

(2) The array \( \{X_{jn} - E(X_{jn}) : 1 \leq j \leq k_n, n \geq 1\} \) is asymptotically negligible, and the distributions

\[
L\left(\frac{X_{1n} - E(X_{1n}) + \ldots + X_{k_n} - E(X_{k_n})}{s_n}\right)
\]

converge weakly to the standard normal distribution \( \mathcal{N}(0,1) \).

The proof of this theorem is using a few lemmas. We may assume that for every \( j \) and \( n \) \( E(X_{jn}) = 0 \). Denoting \( \sigma_{jn}^2 = \sigma^2(X_{jn}) \), we have

\[
s_n^2 = \sum_{j=1}^{k_n} \sigma_{jn}^2.
\]

Replacing \( X_{jn} \) by \( X_{jn}/s_n \) we may as well assume, that \( s_n^2 = 1 \) holds for every \( n \geq 1 \). Denote by \( \mu_{jn} \) the distribution of \( X_{jn} \) and by \( \mu_n \) the measure on \( \mathbb{R} \), which is given by

\[
\int f(x) \mu_n(dx) = \sum_{j=1}^{k_n} \int f(x) \frac{x^2}{1 + x^2} \mu_{jn}(dx)
\]

for \( f \in C_b(\mathbb{R}) \). Finally we set

\[
S_n := X_{1n} + X_{2n} + \ldots + X_{k_n}.
\]
Lemma 8. Let \( \{X_{jn} : 1 \leq j \leq k_n, n \geq 1\} \) be asymptotically negligible. For every \( t \in \mathbb{R} \) we have

\[
\lim_{n \to \infty} \max_{1 \leq j \leq k_n} |1 - \hat{\mu}_{jn}(t)| = 0.
\]

Proof. Let \( \epsilon > 0 \). Since the array is asymptotically negligible, there is \( N \geq 1 \), such that for \( n \geq N \) and \( 1 \leq j \leq k_n \)

\[
P_n(|X_j| \geq \epsilon) \leq \epsilon.
\]

Because of \( |1 - \exp[itx]| \leq |t| \epsilon \) for \( |x| < \epsilon \) it follows that

\[
|1 - \hat{\mu}_{jn}(t)| = |\int (1 - \exp[itx]) \mu_{jn}(dx)|
\]

\[
\leq \int_{|x|<\epsilon} |\exp[itx] - 1| \mu_{jn}(dx) + \int_{|x|\geq\epsilon} |1 - \exp[itx]| \mu_{jn}(dx)
\]

\[
\leq |t| \epsilon + 2P_n(|X_j| \geq \epsilon) \leq \epsilon(|t| + 2).
\]

Lemma 9.

\[
\sup_{n \geq 1} \sum_{j=1}^{k_n} |1 - \hat{\mu}_{jn}(t)| < \infty.
\]

Proof. Since \( \int itx \mu_{jn}(dx) = 0 \) for every \( t \in \mathbb{R}, 1 \leq j \leq k_n \) and \( n \geq 1 \), and since \( |\exp[itx] - 1 - itx| \leq t^2x^2 \) \((t, x \in \mathbb{R})\) it follows that

\[
\sum_{j=1}^{k_n} |1 - \hat{\mu}_{jn}(t)| = \sum_{j=1}^{k_n} |\int (\exp[itx] - 1) \mu_{jn}(dx)|
\]

\[
\leq \sum_{j=1}^{k_n} |\int (\exp[itx] - 1 - itx) \mu_{jn}(dx)| \leq \sum_{j=1}^{k_n} t^2 \int x^2 \mu_{jn}(dx)
\]

\[
= t^2 \sum_{j=1}^{k_n} \sigma_{jn}^2 = t^2.
\]

Lemma 10. If \( \max_{1 \leq j \leq k_n} |1 - \hat{\mu}_{jn}(t)| \leq 1/2 \), then

\[
|\log \hat{S}_{n}(t) - \int (\exp[itx] - 1 - itx) \frac{1 + x^2}{x^2} \mu_{n}(dx)| \leq \max_{1 \leq j \leq k_n} |1 - \hat{\mu}_{jn}(t)| \sum_{j=1}^{k_n} |1 - \hat{\mu}_{jn}(t)|.
\]

Proof. Since \( \log \hat{S}_{n}(t) = \sum_{j=1}^{k_n} \log \hat{\mu}_{jn}(t) \) and \( \int itx \mu_{jn}(dx) = 0 \) it follows that
| \log \hat{S}_n(t) - \int (\exp[itx] - 1 - itx) \frac{1 + x^2}{x^2} \mu_n(dx) |

= \left| \sum_{j=1}^{k_n} \log \hat{\mu}_{jn}(t) - \int (\exp[itx] - 1 - itx) \mu_{jn}(dx) \right|

\leq \sum_{j=1}^{k_n} | \log \hat{\mu}_{jn}(t) - \hat{\mu}_{jn}(t) + 1 |

= \sum_{j=1}^{k_n} \left| \sum_{l=1}^{\infty} (-1)^{l+1} \frac{(\hat{\mu}_{jn}(t) - 1)^l}{l} - \hat{\mu}_{jn}(t) + 1 \right|

\leq \sum_{j=1}^{k_n} \frac{1}{2} \sum_{l=2}^{\infty} |\hat{\mu}_{jn}(t) - 1|^l

= \sum_{j=1}^{k_n} \frac{1}{2} \left[ \frac{1}{1 - |\hat{\mu}_{jn}(t) - 1|} - 1 - |\hat{\mu}_{jn}(t) - 1| \right]

= \frac{1}{2} \sum_{j=1}^{k_n} |\hat{\mu}_{jn}(t) - 1|^2

\leq \max_{1 \leq j \leq k_n} |1 - \hat{\mu}_{jn}(t)| \sum_{j=1}^{k_n} |1 - \hat{\mu}_{jn}(t)|.

Proof of theorem. (2)⇒ (1): Since $S_n$ converges weakly to the standard normal distribution with Fourier transform $t \mapsto \exp[-t^2/2]$, it follows from the continuity theorem of Lévy, that

$$\lim_{n \to \infty} \log \hat{S}_n(t) = -t^2/2 \quad (t \in \mathbb{R}).$$

The above lemmas then show (the array is asymptotically negligible), that

$$\lim_{n \to \infty} \int (\exp[itx] - 1 - itx) \frac{1 + x^2}{x^2} \mu_n(dx) = -t^2/2.$$

Observe next that the functions

$$x \to (\exp[itx] - 1 - itx) \frac{1 + x^2}{x^2} = -t^2 \frac{1 + x^2}{2} + o(x) \quad (x \neq 0)$$

are continuously extendable in zero by $-t^2/2$, and hence form a family of bounded and continuous functions. One easily checks that this family is as well separating, and that \{\mu_n : n \geq 1\} is a tight family. The last statement follows from
3.3 Central Limit Theorems

\[ \mu_n(\mathbb{R} \setminus [-a, a]) \leq \sum_{j=1}^{k_n} P_n(X_{jn}^2 \geq a) \leq a^{-1} \sum_{j=1}^{k_n} \sigma_{jn}^2 = a^{-1}. \]

It follows that \( s - \lim_{n \to \infty} \mu_n = \epsilon_0 \). \( \epsilon_0 \) denotes the point mass in 0. Let \( g \in C_K(\mathbb{R}) \). It follows that

\[ \int g(x)(1 + x^2)\mu_n(dx) \to g(0). \]

Moreover,

\[ \int (1 + x^2)\mu_n(dx) = \sum_{j=1}^{k_n} \int x^2 \mu_{jn}(dx) = 1, \]

yields

\[ s - \lim_{n \to \infty} \nu_n = \epsilon_0, \]

where \( \nu_n(dx) = (1 + x^2)\mu_n(dx) \). It is easy to see now that

\[ L_n(\epsilon) = \sum_{j=1}^{k_n} \int_{\{|x| \geq \epsilon\}} x^2 \mu_{jn}(dx) = \nu_n(\mathbb{R} \setminus [-\epsilon, \epsilon]) \to 0. \]

\((1) \Rightarrow (2):\) First observe that the Lindeberg condition implies asymptotic negligibility, since

\[ \sum_{j=1}^{k_n} P_n(|X_{jn}| \geq \epsilon) \leq \sum_{j=1}^{k_n} \int_{\{|x| \geq \epsilon\}} \frac{x^2}{\epsilon^2} \mu_{jn}(dx) = \frac{1}{\epsilon^2} L_n(\epsilon) \to 0. \]

Using the same notation as in the first part, it follows for any \( \epsilon > 0 \) that

\[ \lim_{n \to \infty} |\hat{\nu}_n(t) - 1| \leq \lim_{n \to \infty} \int_{\{|x| \geq \epsilon\}} (\exp[itx] - 1) \nu_n(dx) + \int_{\{|x| < \epsilon\}} (\exp[itx] - 1) \nu_n(dx) \]

\[ \leq \lim_{n \to \infty} 2L_n(\epsilon) + \lim_{n \to \infty} |\exp[it\epsilon] - 1| = |\exp[it\epsilon] - 1|. \]

Letting \( \epsilon \to 0 \), it follows that \( \lim_{n \to \infty} \hat{\nu}_n(t) = 1 \) for every \( t \in \mathbb{R} \). By the continuity theorem of Lévy, \( \nu_n \) converges weakly to \( \epsilon_0 \), because \( \epsilon_0(t) = 1 \) for every \( t \in \mathbb{R} \). If \( f \in C_b(\mathbb{R}) \), so is \( x \to f(x)/(1 + x^2) \) a bounded continuous function and it follows that

\[ \lim_{n \to \infty} \int f(x)\mu_n(dx) = \lim_{n \to \infty} \int f(x) \frac{1}{1 + x^2} \nu_n(dx) = f(0). \]

The previous lemmas imply that
DISTRIBUTIONAL CONVERGENCE

\[ \lim_{n \to \infty} \log \hat{S}_n(t) \]

\[ = \lim_{n \to \infty} \int \left( \exp[itx] - 1 - itx \right) \frac{1 + x^2}{x^2} \mu_n(dx) \]

\[ = -t^2/2. \]

By the continuity theorem of Lévy this means that the distributions of \( S_n \) converge weakly to the standard normal distribution.

**Corollary 11.** For \( l = 1, ..., d \) let \( \{ X^l_{jn} : 1 \leq j \leq k_n(l), n \geq 1 \} \) be an independent array of square integrable random variables. Assume in addition that the \( d \) arrays are independent, and each one satisfies the Lindeberg condition. Then the random vectors

\[ Z_n = (Z_{1n}, ..., Z_{dn}) \]

with

\[ Z_{ln} := \frac{X^l_{1n} - E(X^l_{1n}) + X^l_{2n} - E(X^l_{2n}) + ... + X^l_{k_n(l)n} - E(X^l_{k_n(l)n})}{\sigma^2(X^l_{1n}) + \sigma^2(X^l_{2n}) + ... + \sigma^2(X^l_{k_n(l)n})} \]

converge weakly to the standard normal distribution in \( \mathbb{R}^d \).

The standard normal distribution in \( \mathbb{R}^d \) has density

\[ f(x_1, x_2, ..., x_d) = \frac{1}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2} (x_1^2 + x_2^2 + ... + x_d^2)\right] \]

and Fourier transform

\[ (t_1, t_2, ..., t_d) \to \exp\left[-\frac{1}{2} (t_1^2 + t_2^2 + ... + t_d^2)\right]. \]

**Proof.** Let \( t = (t_1, t_2, ..., t_d) \in \mathbb{R}^d \) be fixed. It follows from the independence of the arrays and from the last theorem that

\[ E(\exp[i < t, Z_n >]) = E(\exp[i \sum_{l=1}^d t_l Z_{ln}]) \]

\[ = \prod_{l=1}^d E(\exp[i t_l Z_{ln}]) \to \prod_{l=1}^d \exp[-t_l^2/2]. \]

Then apply the continuity theorem of Lévy.

**Corollary 12.** Let \( X_j (j \in \mathbb{N}) \) be a sequence of independent random variables with the same distribution (for short: i.i.d.). If \( X_1 \in L_2(P) \) (and hence every \( X_j \)) is not constant, the central limit theorem holds:

\[ \frac{X_1 + X_2 + ... + X_n - nE(X_1)}{\sqrt{n\sigma^2(X_1)}} \xrightarrow{D} \mathcal{N}(0, 1). \]
3.3 Central Limit Theorems

Proof. Since $X_1$ is non-constant, one has $\sigma^2(X_j) > 0$, and the central limit theorem under Lindeberg’s condition is applicable. Consider the array

$$X_{jn} = X_j \quad (1 \leq j \leq n, n \geq 1)$$

and we show the Lindeberg condition. Let $\epsilon > 0$. Since the distributions are all equal,

$$L_n(\epsilon) = \frac{1}{n\sigma^2(X_1)} \sum_{j=1}^{n} \int_{\{\|X_j - E(X_1)\| \leq \epsilon \sqrt{n\sigma^2(X_1)}\}} [X_j - E(X_1)]^2 \, dP$$

$$= \frac{1}{\sigma^2(X_1)} \int_{\{\|X_1 - E(X_1)\| \leq \epsilon \sqrt{n\sigma^2(X_1)}\}} [X_1 - E(X_1)]^2 \, dP$$

$$= \frac{1}{\sigma^2(X_1)} \int_{\{|u| \leq \epsilon \sqrt{n\sigma^2(X_1)}\}} u^2 \mu(du)$$

$$= 0,$$

where $\mu$ denotes the distribution of $X_1 - E(X_1)$. Observe that by assumption $u \mapsto u^2$ is a function in $L_1(\mu)$!

We add two more results which are connected to the Lindeberg central limit theorem. The proofs are somehow following similar ideas.

Definition 38. A distribution $\mu$ (in $\mathbb{R}^d$) is called infinitely divisible, if for each $n \in \mathbb{N}$ there exist independent, identically distributed random variables (vectors) $X_1, ..., X_n$ such that the distribution of the sum $X_1 + ... + X_n$ is equal to $\mu$.

Example 12. 1. A normal distribution is infinitely divisible, since the sum of $n$ independent normally distributed random variables, each with mean $\mu/n$ and variance $\sigma^2/n$, has a normal distribution with mean $\mu$ and variance $\sigma^2$.

2. A Poisson distribution is as well infinitely divisible, since the sum of two independent Poisson distributed random variables with parameters $\lambda$ and $\kappa$ has again a Poisson distribution with parameter $\lambda + \kappa$.

3. A point mass is infinitely divisible. If $X$ is a random variable which a.s. is equal to $a$, then the variables $X_k = X/k$ ($k = 1, ..., n$) are independent and identically distributed. Their sum equals $X$, so has a distribution which equals the point mass in $a$.

4. A non-degenerate Bernoulli measure is not infinitely divisible. To see this, let $X$ and $Y$ be two independent, identically distributed random variables, such that the sum is Bernoulli with success probability $p \in (0, 1)$ (note that $p \neq 0, 1$ because otherwise it would be infinitely divisible by 3.)

First observe that the distribution must have support on $[0, 1/2]$, for otherwise it is easy to construct a set in $[0, 1/2]^c$ of positive measure for the sum. Next, each variable has positive mass on 0, since the sum has to have positive mass on this point. Then the set $\{X = 0, 0 < Y \leq 1/2\}$ has measure 0, showing that both random variables must be concentrated on 0, a contradiction.
Theorem 45. Let \( \{X_{kn} : 1 \leq k \leq k_n\} \) be an independent and asymptotically negligible array of random variables. If there is a distribution \( \mu \) on \( \mathbb{R} \) such that the distributions of
\[
S_n = X_{1n} + ... + X_{k_nn}
\]
converge weakly to \( \mu \), then \( \mu \) is infinitely divisible.

Theorem 46. (Lévy-Khintchine formula)
A distribution \( \mu \) on \( \mathbb{R} \) is infinitely divisible if and only if it has a characteristic function given by
\[
\hat{\mu}(t) = \exp \left[ i\beta t - \frac{\sigma^2 t^2}{2} + \int \left( e^{itx} - 1 - \frac{itx}{1 + x^2} \right) \frac{1 + x^2}{x^2} m(dx) \right]
\]
for some \( \beta \in \mathbb{R} \), \( \sigma^2 \geq 0 \) and some measure \( m \) on \( \mathbb{R} \) satisfying \( m(\{0\}) = 0 \).

The measure \( m \) (and also equivalent measures) is called the Lévy measure.

Another similar theorem is the Lindeberg central limit theorem for martingale differences. The extension to arrays is known as well, but we restrict to the case of a single sequence.

Definition 39. Let \( X_n \) \( (n \geq 1) \) be a sequence of random variables, adapted to the increasing sequence \( F_n \) of \( \sigma \) fields in \( F \). Then \( (X_n,F_n)_{n \geq 1} \) is called a martingale difference sequence if for each \( n \in \mathbb{N} \)
\[
E(X_{n+1}|F_n) = 0.
\]

Note that the sum \( M_n = X_1 + ... + X_n \) defines a martingale and that \( E(X_n) = 0 \) for every \( n \in \mathbb{N} \).

Theorem 47. Let \( (X_n,calF_n) \) \( (n \geq 1) \) be a martingale difference sequence, such that the array \( X_{kn} = X_k \) for \( k = 1, ..., n \) and \( n \geq 1 \) satisfies the Lindeberg condition. Moreover, assume that
\[
\sum_{j=1}^{n} E \left| E(X_j^2|F_{j-1}) - \sigma_j^2 \right| = o(s_n^2),
\]
where
\[
\sigma_j^2 = E(X_j^2) \quad \text{and} \quad s_n^2 = \sum_{j=1}^{n} \sigma_j^2.
\]
Then
\[
s - \lim_{n \to \infty} \frac{1}{s_n} \sum_{j=1}^{n} X_j = N(0,1).
\]
3.3 Central Limit Theorems

Proof. In the proof we use the following estimate (which also has been used earlier)

Lemma 11. For all $\delta \in [0, 1]$

$$|e^{itx} - \sum_{j=0}^{n} \frac{(it)^j}{j!}| \leq \frac{2^{1-\delta}n^{\delta}}{(1+\delta)(2+\delta)...(n+\delta)}.$$ 

Now let $t \in \mathbb{R}$ be fixed. Let $\epsilon > 0$. Then, using $M_n = X_1 + ... + X_n$ as before,

$$E\left(e^{itM_n/s_n}\right) - e^{t^2/2} = e^{-t^2/2} \left|\sum_{j=1}^{n} E\left(e^{itM_j/s_n + s_j^2t^2/(2s_n^2)} - e^{itM_{j-1}/s_n + s_{j-1}^2t^2/(2s_n^2)}\right)\right|.$$ 

Observe that the latter is a telescoping sum! Evaluating the expectations appearing in the sum, we obtain

$$\left|\sum_{j=1}^{n} E\left(e^{itM_j/s_n + s_j^2t^2/(2s_n^2)} - e^{itM_{j-1}/s_n + s_{j-1}^2t^2/(2s_n^2)}\right)\right| 
\leq e^{t^2/2} E\left|e^{itX_j/s_n} - e^{-\sigma_j^2t^2/(2s_n^2)}\right| |F_{j-1}|.$$ 

In these estimates we used the facts that $E(Y) = E(E(Y|A))$, and that $|e^{itM_{j-1}/s_n}| \leq 1$.

Now we use the lemma with $\delta = 1$ and $n = 1, 2$ (three times) to get

$$E\left(e^{itX_j/s_n} - e^{-\sigma_j^2t^2/(2s_n^2)}\right) |F_{j-1}| = E\left(e^{itX_j/s_n} - e^{-\sigma_j^2t^2/(2s_n^2)}\right) |F_{j-1}| 
= e^{itX_j/s_n} - e^{-\sigma_j^2t^2/(2s_n^2)} |F_{j-1}| 
= e^{itX_j/s_n} - e^{itX_j/s_n + t^2X_j^2/(2s_n^2)} |F_{j-1}| \n\quad + \frac{\sigma_j^2t^2}{2s_n^2} - E\left(e^{t^2X_j^2/(2s_n^2)} |F_{j-1}| \right)  
\leq E\left(\frac{t^2X_j^2}{s_n^2} I(|X_j| > s_n) |F_{j-1}| \right) 
\quad + E\left(\frac{|tX_j|^3}{s_n^3} I(|X_j| \leq s_n) |F_{j-1}| \right) \n\quad + \frac{\sigma_j^4t^4}{8s_n^4} + b_j,$$

where
Next observe that the Lindeberg condition implies that
\[
\max_{1 \leq j \leq n} \sigma_j^2 s_n^{-2} \leq \max_{1 \leq j \leq n} s_n^{-2} \left[ \epsilon^2 s_n^2 + \int_{\{|X_j| > \epsilon s_n\}} X_j^2 dP \right] = \epsilon^2 + L_n(\epsilon).
\]
Putting everything together we arrive at
\[
\left| E \left( e^{itM_n/s_n} \right) - e^{t^2/2} \right| \\
\leq e^{-t^2/2} \sum_{j=1}^{n} e^{t^2/2} E \left| e^{itX_j/s_n} - e^{-\sigma_j^2 t^2 / (2s_n^2)} \right| \\
\leq \sum_{j=1}^{n} E \left( e^{\left( \frac{t^2 X_j^2}{s_n^2} \right) 1_{\{|X_j| > \epsilon s_n\}} \right) + E \left( e^{-\left( \frac{tX_j^3}{s_n} \right) 1_{\{|X_j| \leq \epsilon s_n\}} \right) + \max_{1 \leq j \leq n} \sigma_j^2 s_n^{-2} \frac{\sigma_j^2 t^4}{8s_n^2} + b_j \\
\leq L_n(\epsilon) + \epsilon |t|^3 + (L_n(\epsilon) + \epsilon^2) t^4/8 + \sum_{j=1}^{n} b_j.
\]
Observe that \( \sum_{j=1}^{n} b_j \) tends to zero as \( n \) tends to infinity by assumption. Hence, letting \( n \) tend to infinity and then \( \epsilon \) to zero, shows that
\[
\lim_{n \to \infty} E \left( e^{itM_n/s_n} \right) = e^{t^2/2}
\]
for every \( t \in \mathbb{R} \). By the continuity theorem this implies the convergence to normal distribution.
3.4 Donsker’s Invariance Principle

Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with $0 < E(X_1^2) < \infty$. The space of all continuous functions on the unit interval $[0, 1]$ will be denoted by $C = C([0, 1])$. It is a separable Banach space with norm
\[ \|f\|_{[0,1]} = \|f\| = \max_{0 \leq t \leq 1} |f(t)|, \]
in particular, it is a separable, complete metric space. For $0 \leq t_1 < t_2 < \ldots, t_n \leq 1$ let
\[ \pi_{t_1 \ldots t_n} : C \to \mathbb{R}^n \]
denote the projection defined by
\[ \pi_{t_1 \ldots t_n}(f) = (f(t_1), \ldots, f(t_n)). \]

It is easy to see that each projection is continuous. Let $\mu_n, \mu$ be probability measures on $C$.

**Theorem 48.** The following properties are equivalent:
(1) $s - \lim_{n \to \infty} \mu_n = \mu$.
(2) For every projection $\pi_{t_1 \ldots t_n}$ one has
\[ s - \lim_{n \to \infty} \mu_n \circ \pi_{t_1 \ldots t_n}^{-1} = \mu \circ \pi_{t_1 \ldots t_n}^{-1}, \]
and the family $\{\mu_n : n \geq 1\}$ is tight.

This theorem is a consequence of Prohorov’s theorem:

**Theorem 49.** (Theorem of Prohorov) Let $\mathcal{M}$ be a family of finite measures on a metric space $\Omega$ such that $\sup_{\mu \in \mathcal{M}} \mu(\Omega) \leq s$ for some $s \geq 0$. Then
1. If $\mathcal{M}$ is tight then it is relatively sequentially compact.
2. If $\Omega$ is a Polish space and $\mathcal{M}$ relatively compact, then $\mathcal{M}$ is a tight family.

We need to investigate tightness in $C$. By the theorem of Arzéla-Ascoli a set of functions is relatively compact, if it is bounded and the modulus of continuity converges uniformly to 0:

**Theorem 50.** (Arzéla-Ascoli in $C$)
$A \subset C$ is relatively compact if and only if
\[ \sup_{f \in A} |f(0)| < \infty \]
and
\[ \lim_{\delta \to 0} \sup_{f \in A, |s-t|<\delta} |f(t) - f(s)| = 0. \]

This gives the following criterion for tightness:
Theorem 51. A sequence \( \{\mu_n : n \geq 1\} \) is tight if and only if
\[
\forall \eta > 0 \exists a \in \mathbb{R} \forall n \geq 1 \implies \mu_n(\{f : |f(0)| > a\}) \leq \eta
\]
and
\[
\forall \epsilon, \eta > 0 \exists \delta < 1 \exists n_0 \forall n \geq n_0 \implies \mu_n(\{f : \sup_{|s-t| < \delta} |f(t) - f(s)| \geq \epsilon\}) \leq \eta.
\]
Let us define
\[
\tilde{X}_n = X_n - E(X_n) \quad \sigma(X_n) = \sqrt{\text{Var}(X_n)},
\]
\[
S_n = \tilde{X}_1 + \tilde{X}_2 + \ldots + \tilde{X}_n,
\]
\[
Z_n(t) := \frac{1}{\sqrt{n}} \left( S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) \tilde{X}_{\lfloor nt \rfloor + 1} \right),
\]
where \( \lfloor nt \rfloor = \max\{k \in \mathbb{N} : k \leq nt\} \) denotes the Gauß-bracket. The random functions \( Z_n \) are continuous \( (Z_n \) is as well a stochastic process) and the continuity modulus of \( Z_n \) can be determined by the endpoints of the intervals, on which \( Z_n \) is linear, i.e. at the points \( t = k/n \). One gets
\[
\sup_{|t-s| < \delta} |Z_n(t) - Z_n(s)| \leq \max_{0 \leq k,j \leq n, |k-j| < \delta n + 2} \frac{1}{\sqrt{n}} |S_k - S_j|.
\]
This is done in the proof of the next theorem.

Theorem 52. (Donsker’s invariance principle)
The distributions of \( Z_n \ (n \geq 1) \) converge weakly in \( C \) to a probability measure \( W \). The measure \( W \) satisfies the following conditions:
The random variables
\[
\pi_{t_1}, \pi_{t_2} - \pi_{t_1}, \ldots, \pi_{t_n} - \pi_{t_{n-1}}
\]
(defined on \( C \) with the probability \( W \)) are for arbitrary \( 0 \leq t_1 < t_2 < \ldots < t_n \leq 1 \) independent and (jointly) normally distributed with
\[
E(\pi_{t_i} - \pi_{t_{i-1}}) = 0 \quad (i = 1, \ldots, n, t_0 = 0)
\]
\[
\sigma^2(\pi_{t_i} - \pi_{t_{i-1}}) = t_i - t_{i-1} \quad (i = 1, \ldots, n, t_0 = 0)
\]
\[
W(\pi_0 = 0) = 1.
\]

Remark 28. The probability measure \( W \) on \( C \) is called the Wiener measure. For the projections one usually writes \( W_t = \pi_t \) and considers the projections as random variables on \( (C, W) \). Then \( W := (W_t)_{t \in [0,1]} \) is a stochastic process, called the Wiener process. It is the restriction of the Brownian motion on the interval \( [0,1] \).
Corollary 13. Let $X_1, X_2, X_3, \ldots$ be i.i.d. random variables with $\sigma^2 = \sigma^2(X_1) \in (0, \infty)$. Let

$$S_n = \sum_{j=1}^{n} (X_j - E(X_1)), \quad S_0 = 0 \quad (n \geq 1).$$

Then

$$\frac{1}{\sigma \sqrt{n}} \max_{0 \leq k \leq n} S_k \quad (n \geq 1)$$

converges weakly to $|N|$, where $N$ is a standard normally distributed random variable. Hence for $\alpha \geq 0$

$$\lim_{n \to \infty} P\left(\left\{ \frac{1}{\sigma \sqrt{n}} \max_{0 \leq k \leq n} S_k \geq \alpha \right\}\right) = \frac{2}{\sqrt{2\pi}} \int_{\alpha}^{\infty} \exp\left[-\frac{1}{2} u^2\right] du.$$

Proof. Since $Z_n$ converges weakly to $W$ and $h(f) = \sup_{0 \leq t \leq 1} f(t)$ is a continuous function on $C$,

$$\frac{1}{\sigma \sqrt{n}} \max_{0 \leq k \leq n} S_k = \sup_{0 \leq t \leq 1} Z_n(t)$$

converges weakly to $\sup_{0 \leq t \leq 1} W_t$. The distribution of this random variable will be determined by a special sequence $X_1, X_2, \ldots$.

Let $P(\{X_1 = \pm 1\}) = 1/2$ the fair coin tossing. For $\alpha \geq 0$ one shows that

$$P(\{\max_{0 \leq j \leq n} S_j \geq \alpha\}) = 2P(\{S_n > a\}) + P(\{S_n = a\}).$$

If $a = 0$, both sides are equal to 1, since $P(\{S_n > 0\}) = P(\{S_n < 0\})$. So let $a > 0$. Then

$$P(\{\max_{1 \leq j \leq n} S_j \geq \alpha\}) - P(\{S_n = a\}) = P(\{\max_{1 \leq j \leq n} S_j \geq a; S_n < a\}) + P(\{\max_{1 \leq j \leq n} S_j \geq a; S_n > a\})$$

and

$$P(\{\max_{1 \leq j \leq n} S_j \geq a; S_n > a\}) = P(\{S_n > a\}).$$

It is left to show that

$$P(\{\max_{1 \leq j \leq n} S_j \geq a; S_n < a\}) = P(\{\max_{1 \leq j \leq n} S_j \geq a; S_n > a\}).$$

Let $\{X_1 = b_1, \ldots, X_n = b_n\}$ be an event, which is contained in the first set, and let $j$ denote the smallest index, for which $S_j = a$ holds. Assign the reflected event to this event which is the event $\{X_1 = b_1, \ldots, X_j = b_j, X_{j+1} = b_{j+1}', \ldots, X_n = b_n'\}$ to the second set by putting $b_k' = b_k + 1 \pmod{2}$ (so 1 replaced by 0 and 0 by 1). It is easy to see that this defines a bijection between the elementary events of both sets. Since all elementary events have probability $2^{-n}$, the claimed equality follows.
The claim of the corollary follows from this argument: Let $\alpha \geq 0$. Define $a_n = -[-\alpha \sqrt{n}]$ and one obtains from the above equality that

$$P(\{ \max_{0 \leq j \leq n} S_j \geq \alpha \sqrt{n} \}) = 2P(\{S_n > a_n\}) + P(\{S_n = a_n\}).$$

The central limit theorem (use uniform convergence of the distribution functions) gives

$$\lim_{n \to \infty} P(\{S_n \geq a_n\}) = \lim_{n \to \infty} \frac{1}{\sqrt{2\pi}} \int_{[-\frac{\alpha \sqrt{n}}{\sqrt{n}}]}^{\infty} \exp[-u^2/2] \, du = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} \exp[-u^2/2] \, du.$$

Therefore

$$\lim_{n \to \infty} P(\{ \max_{0 \leq j \leq n} S_j \geq \alpha \}) = 2P(\{N \geq \alpha\}) = P(|N| \geq \alpha),$$

and the proof is finished.

**Theorem 53.** Let $X_n$ ($n \geq 1$) be independent uniformly (on $[0,1]$) distributed random variables. Let $\hat{F}(t) = \frac{1}{n} \sum_{k=1}^{n} 1_{X_k \leq t}$ denote the empirical distribution function of $X_1, \ldots, X_n$. Then

$$\sqrt{n}(\hat{F}(t) - t) \quad (0 \leq t \leq 1)$$

converges weakly in the space of cadlag functions on $[0,1]$ to the distribution of the random variable

$$W(t) - tW(1) \quad (0 \leq t \leq 1),$$

where $W$ denotes the Wiener process.
Index

$L_p$–space, 18
$\omega$-section, 44
$\sigma$-additive, 3
$\sigma$-algebra generated by $\Sigma$, 2
$\sigma$-algebra, -field, 1
$p$-th moment, 51

absolutely continuous, 30
adapted, 56
almost sure convergence, 23
array of random elements, 85
asymptotic negligible, 86

Banach space, 17
bounded random variable, 51
Brownian motion, 96

Cauchy-Schwarz inequality, 17
central limit theorem, 86
characteristic function (Fourier-transform), 80
conditional expectation, 40
convergence in $p$-the mean, 23
convergence in sup-norm, 23
cylinder set, 7

density, 33
discrete measure, 4
distribution function, 51
Donsker’s invariance principle, 96
Doob’s convergence theorem, 63
dual Banach space, 17
dual pair, 35
dual space, 71

elementary function, 8
expectation, 51

Fourier-transform (characteristic function), 80
Fubini’s theorem, 48
Hölder inequality, 19
Hahn decomposition, 29
Hilbert space, 17

independent, 51, 85
infinitely divisible, 91
inner product, 17
integrable, 10
integral, 10
isometry, 36
isomorphism, 36

Lévy measure, 92
Lévy-Khintchine formula, 92
Lebesgue measure, 4
Lindeberg condition, 86

martingale, 56
martingale difference sequence, 92
measurable, 1, 6
measurable space, 5
measure, 3
measure finite, 3
measure space, 5
measure transport, 9
measure, $\sigma$-finite, 3
Minkowski inequality, 20

negative set, 29
non-singular, 30
norm, 17
nullset, 11

order preserving, 36
orthogonal, 30
point mass, 4
positive cone, 35
positive operator, 36
positive set, 29
pre-measure, 3
probability measure, 3
product measure, 46
Prohorov’s theorem, 95
Radon-Nikodym derivative, 33
random element, 6
random function, 6
random variable, 6, 51
random vector, 6
signed measure, 3
state space, 51
Stieltjes measure, 4
stochastic (random) process, 51
stochastic convergence, 23
stopping time, 58
strong law of large numbers, 68
supermartingale, 56
support, 30
Theorem of Fisher-Riesz, 21
tight, 77
time parameter, 51
uncorrelated random variables, 53
uniformly integrable, 23
uniformly integrable martingale, 67
vague convergence, 72
variance, 51
weak convergence, 71
weak-+ convergence, 71
Wiener measure, 96
Wiener process, 96