Chapter 0
Measures, Integration, Convergence

1. Measures
   Definitions
   Basic Examples
   Extension Theorem
   Completion

2. Measurable Functions and Integration
   Simple functions
   Monotone convergence theorem (MCT)
   Fatou’s lemma
   Dominated convergence theorem (DCT)
   Absolute continuity of the integral
   Induced measures
   Theorem of the unconscious statistician

3. Absolute Continuity, Radon-Nikodym Theorem, Fubini’s Theorem
   Absolute continuity
   Radon-Nikodym Theorem
   Fubini-Tonelli Theorem
Chapter 0

Measures, Integration, Convergence

1 Measures

Let $\Omega$ be a fixed non-void set.

Definition 1.1 (fields, $\sigma$-fields, monotone classes) A non-void class $\mathcal{A}$ of subsets of $\Omega$ is called a:

(i) **field or algebra** if $A, B \in \mathcal{A}$ implies $A \cup B \in \mathcal{A}$ and $A^c \in \mathcal{A}$.

(ii) **$\sigma$-field** or **$\sigma$-algebra** if $A, A_1, A_2, \ldots \in \mathcal{A}$ implies $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ and $A^c \in \mathcal{A}$.

(iii) **monotone class** if $A_n$ is a monotone $\nearrow$ ($\searrow$) sequence in $\mathcal{A}$ implies $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ ($\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$).

(iv) $(\Omega, \mathcal{A})$ with $\mathcal{A}$ a $\sigma$-field of subsets of $\Omega$ is called a **measurable space**.

Remark 1.1 (i) $A, B \in \mathcal{A}$ imply $A \cap B \in \mathcal{A}$ for a field.

(ii) $A_1, \ldots, A_n, \ldots \in \mathcal{A}$ implies $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$ for a $\sigma$-field.

(iii) $\emptyset, \Omega \in \mathcal{A}$ for both a field and $\sigma$-field.

(iv) To prove that $\mathcal{A}$ is a field ($\sigma$-field) it suffices to show that $\mathcal{A}$ is closed under complements and finite (countable) intersections.

Proposition 1.1 (i) Arbitrary intersections of fields ($\sigma$-fields) ((monotone classes)) are fields ($\sigma$-fields) ((monotone classes)).

(ii) There exists a minimal field ($\sigma$-field) ((monotone class)) $\sigma(\mathcal{C})$ generated by any class of subsets of $\Omega$.

(iii) a $\sigma$-field is a monotone class and conversely if it is a field.

Proof. (iii) $(\leftarrow)$ $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \left( \bigcup_{k=1}^{n} A_k \right) \equiv \bigcup_{i=1}^{\infty} B_n$ where $B_n$ $\nearrow$. $\square$

Notation 1.1 If $\Omega$ is a set, $2^\Omega$ is the family of all subsets of $\Omega$.

$2^\Omega$ is always a $\sigma$-field.

Example 1.1 If $\Omega = R$, let $\mathcal{B}_0$ consist of $\emptyset$ together with all finite unions of disjoint intervals of the form $\bigcup_{i=1}^{n} (a_i, b_i]$, or $\bigcup_{i=1}^{n} (a_i, b_i] \cup (a_{n+1}, \infty)$, $(-\infty, b_{n+1}] \cup \bigcup_{i=1}^{n} (a_i, b_i]$, with $a_i, b_i \in R$. Then $\mathcal{B}_0$ is a field.
Example 1.2 If $\Omega = (0, 1]$, let $B_0$ consist of $\emptyset$ together with all finite unions of disjoint intervals of the form $\bigcup_{i=1}^{k} (a_i, b_i], 0 \leq a_i \leq b_i \leq 1$. Then $B_0$ is a field. But note that $B_0$ does not contain intervals of the form $[a, b]$ or $(a, b)$; however $(a, b) = \bigcup_{n=1}^{\infty} (a, b - 1/n]$.

Example 1.3 If $\Omega = R$, let $\mathcal{C} = B_0$ of example 1.1, and let $B$ be the $\sigma$-field generated by $B_0$; $B = \sigma(B_0)$. $B$ is a $\sigma$-field which contains all intervals, open, closed or half-open. From real analysis, any open set $O \subset R$ can be written as a countable union of (disjoint) open intervals:

$$O = \bigcup_{n=1}^{\infty} (a_n, b_n).$$

Thus $B$ contains all open sets in $R$. This particular $B \equiv B_1$ is called the family of Borel sets. In fact, $B = \sigma(O)$, where $O$ is the collection of all open sets in $R$.

Example 1.4 Suppose that $\Omega$ is a metric space with metric $\rho$. Let $\mathcal{O}$ be the collection of open subsets of $\Omega$. The the $\sigma$-field $B = \sigma(\mathcal{O})$ is called the Borel $\sigma$-field. In particular, for $\Omega = R^k$ with the Euclidean metric $\rho(x, y) = |x - y| = \left\{ \sum_{i=1}^{k} |x_i - y_i|^2 \right\}^{1/2}$, $B = B_k \equiv \sigma(\mathcal{O})$ is the $\sigma$-field of Borel sets.

Definition 1.2 (i) A measure (finitely additive measure) is a function $\mu : \mathcal{A} \rightarrow [0, \infty]$ such that $\mu(\emptyset) = 0$ and $\mu(\bigcup A_n) = \sum \mu(A_n)$ for countable (finite) disjoint sequences $A_n$ in $\mathcal{A}$.
(ii) A measure space is a triple $(\Omega, \mathcal{A}, \mu)$ with $\mathcal{A}$ a $\sigma$-field and $\mu$ a measure.

Definition 1.3 (i) $\mu$ is a finite measure if $\mu(\Omega) < \infty$.
(ii) $\mu$ is a probability measure if $\mu(\Omega) = 1$.
(iii) $\mu$ is an infinite measure if $\mu(\Omega) = \infty$.
(iv) A measure $\mu$ on a field (\sigma-field) $\mathcal{A}$ is called $\sigma$-finite if there exists a partition $\{F_n\}_{n \geq 1} \subset \mathcal{A}$ such that $\Omega = \bigcup_{n=1}^{\infty} F_n$ and $\mu(F_n) < \infty$ for all $n \geq 1$.
(v) A probability measure is a measure space $(\Omega, \mathcal{A}, \mu)$ with $\mu$ a probability measure.

Definition 1.4 (i) A measure $\mu$ on $(\Omega, \mathcal{A})$ is discrete if there are finitely or countably many points $\omega_i$ in $\Omega$ and masses $m_i \in [0, \infty)$ such that

$$\mu(A) = \sum_{\omega_i \in A} m_i \quad \text{for} \quad A \in \mathcal{A}.$$ 

(ii) If $\mu$ is defined on $(\Omega, 2^\Omega)$, $\Omega$ arbitrary, by $\mu(A) = \sharp$ of points in $A$, $\mu(A) = \infty$ if $A$ is not finite, then $\mu$ is called counting measure.

Example 1.5 (i) A discrete measure $\mu$ on $(\Omega, \mathcal{A}) = (R^1, B_1)$: $x_i = i, m_i = 2^i$.
(ii) A discrete measure $\mu$ on $(\Omega, \mathcal{A}) = (\mathbb{Z}^+, 2^{\mathbb{Z}^+})$: $x_i = 2i, m_i = 1/i$. ($\mathbb{Z}^+ = \{1, 2, \ldots\}$).
(iii) Counting measure on $(R^1, B_1)$: not a $\sigma$-finite measure!
(iv) Counting measure on $(\mathbb{Z}^+, 2^{\mathbb{Z}^+})$.
(v) A probability measure on $\mathbb{Q}$, the rationals: With $\{x_i\}$ an enumeration of the rationals, let $m_i = 6/(\pi^2 i^2)$.

Proposition 1.2 Let $(\Omega, \mathcal{A}, \mu)$ be a measure space.
(i) If $\{A_n\}_{n \geq 1} \subset \mathcal{A}$ with $A_n \subset A_{n+1}$ for all $n$, then $\mu(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \to \infty} \mu(A_n)$.
(ii) If $\mu(A_1) < \infty$ and $A_n \supset A_{n+1}$ for all $n$, then $\mu(\bigcap_{n=1}^{\infty} A_n) = \lim_{n \to \infty} \mu(A_n)$. 

1. MEASURES

Proof. (i) 
\[
\mu(\bigcup_{n=1}^{\infty} A_n) = \mu(\bigcup_{n=1}^{\infty} (A_n \setminus A_{n-1})) \quad \text{where} \quad A_0 = \emptyset \\
= \sum_{n=1}^{\infty} \mu(A_n \setminus A_{n-1}) \quad \text{by countable additivity} \\
= \lim_{n \to \infty} \sum_{i=1}^{n} \mu(A_n \setminus A_{n-1}) \\
= \lim_{n \to \infty} \mu\left(\bigcup_{i=1}^{n} (A_n \setminus A_{n-1})\right) \quad \text{by finite additivity} \\
= \lim_{n \to \infty} \mu(A_n). 
\]

(ii) Let \( B_n = A_1 \setminus A_n = A_1 \cap A_n^c \) so that \( B_n \nrightarrow \). Thus, on the one hand we have 
\[
\lim_{n \to \infty} \mu(B_n) = \mu(\bigcup_{n=1}^{\infty} B_n) \quad \text{by part (i)} \\
= \mu(\bigcup_{n=1}^{\infty} (A_1 \cap A_n^c)) \\
= \mu(A_1 \cap \bigcup_{n=1}^{\infty} A_n^c) \\
= \mu(A_1 \cap (\bigcap_{n=1}^{\infty} A_n)^c) \\
= \mu(A_1) - \mu(\bigcap_{n=1}^{\infty} A_n) \quad \text{by finite additivity,}
\]
while on the other hand,
\[
\lim_{n \to \infty} \mu(B_n) = \lim_{n \to \infty} \mu(A_1 \setminus A_n) = \lim_{n \to \infty} \{\mu(A_1) - \mu(A_n)\} \quad \text{by finite additivity} \\
= \mu(A_1) - \lim_{n \to \infty} \mu(A_n).
\]
Combining these two equalities yield the conclusion of (ii). □

Definition 1.5

(i) \( \lim A_n \equiv \bigcup_{n=1}^{\infty} (\cap_{n=1}^{\infty} A_k) \equiv \{\omega \in \Omega : \omega \in A_n \text{ but a finite number of } A_k^c \} \equiv [A_n \text{ a.a.}] \); \\
(ii) \( \overline{\lim} A_n \equiv \cap_{n=1}^{\infty} (\cup_{k=1}^{\infty} A_k) \equiv \{\omega \in \Omega : \omega \in \text{infinitely many } A^c \text{'s} \} \equiv [A_n \text{ i.o.}] \).

Remark 1.2 \( \lim A_n \subset \overline{\lim} A_n; \lim A_n = \lim A_n \) provided \( \lim A_n = \lim_{n \to \infty} A_n \).

Proposition 1.3 Monotone \( \nrightarrow (\\setminus \\setminus) \) \( A_n \)'s have \( \lim A_n = \bigcup_{n=1}^{\infty} A_n = (\cap_{n=1}^{\infty} A_n) \).

Example 1.6 Let \( A = B = \sigma(B_0) \) as in example 1.3. For \( B \in B_0 \), let \( \mu(B) \equiv \) the sum of the lengths of intervals \( A \in B_0 \) composing \( B \). Then \( \mu \) is a countably additive measure on \( B_0 \). Can \( \mu \) be extended to \( B \)? The answers is yes, and depends on the following:

Theorem 1.1 (Caratheodory Extension Theorem) A measure \( \mu \) on a field \( C \) can be extended to a measure on the minimal \( \sigma \)-field \( \sigma \)-field \( \sigma(C) \) over \( C \). If \( \mu \) is \( \sigma \)-finite on \( C \), then the extension is unique and is also \( \sigma \)-finite.

Proof. See Billingsley (1986), pages 29 - 35 and 137 - 139. □
Example 1.7 (example 1.3, continued.) The extension of the countably additive measure \( \mu \) on \( B_0 \) to \( B_1 = \sigma(B_0) \), the Borel \( \sigma \)-field, is called Lebesgue measure; thus \((R^1, B_1, \mu)\) where \( \mu \) is the extension of the Caratheodory extension theorem, is a measure space. The usual procedure is to complete \( B_1 \) as follows.

**Definition 1.6** If \((\Omega, A, \mu)\) is a measure space such that \( B \subset A \) with \( A \in A \) and \( \mu(A) = 0 \) implies \( B \in A \), then \((\Omega, A, \mu)\) is a complete measure space. If \( \mu(A) = 0 \), then \( A \) is called a null set. (Of course there can be non-empty null sets.)

**Exercise 1.1** Let \((\Omega, A, \mu)\) be a measure space. Define
\[
\overline{A} \equiv \{ A \cup N : A \in A, \, N \subset B \text{ for some } B \in A \text{ such that } \mu(B) = 0 \}
\]
and let \( \overline{\mu}(A \cup N) \equiv \mu(A) \). Then \((\Omega, \overline{A}, \overline{\mu})\) is a complete measure space.

Example 1.8 (example 1.3, continued.) Completing \((R^1, B_1, \mu)\) where \( \mu = \text{Lebesgue measure} \) yields the complete measure space \((R^1, \overline{B_1}, \overline{\mu})\). \( \overline{B_1} \) is called the \( \sigma \)-field of Lebesgue sets.

So far we know only a few measures. But we will now construct a whole batch of them; and they are just the ones most useful for probability theory.

**Definition 1.7** A measure \( \mu \) on \( R \) assigning finite values to finite intervals is called a Lebesgue-Stieltjes measure.

**Definition 1.8** A function \( F \) on \( R \) which is finite, increasing, and right continuous is called a generalized distribution function (generalized df).
\[
F(a, b] \equiv F(b) - F(a)
\]
for \(-\infty < a \leq b < \infty\) is called the increment function of the generalized df \( F \). We identify generalized df's having the same increment function.

**Theorem 1.2** (Correspondence theorem.) The relation
\[
\mu((a, b]) = F(a, b] \quad \text{for} \quad -\infty < a \leq b < \infty
\]
establishes a one-to-one correspondence between Lebesgue-Stieltjes measures \( \mu \) on \( B = B_1 \) and equivalence classes of generalized df's.

**Proof.** See Billingsley (1986), pages 147, 149 - 151. \( \square \)

**Definition 1.9** (Probability measures on \( R \).) If \( \mu(\Omega) = 1 \), then \( \mu \) is called a probability distribution or probability measure and is denoted by \( P \).

**Definition 1.10** An \( \mathcal{F} \)-right-continuous function \( F \) on \( R \) such that \( F(-\infty) = 0 \) and \( F(\infty) = 1 \) is a distribution function (df).

**Corollary 1** The relation
\[
P((a, b]] = F(b) - F(a) \quad \text{for} \quad -\infty < a \leq b < \infty
\]
establishes a one-to-one correspondence between probability measures on \( R \) and df's.
2. MEASURABLE FUNCTIONS AND INTEGRATION

2. Measurable Functions and Integration

Let \((\Omega, \mathcal{A})\) be a measurable space.
Let \(X\) denote a function, \(X: \Omega \to R\).

**Definition 2.1** \(X: \Omega \to R\) is measurable if \([X \in B] \equiv X^{-1}(B) = \{\omega \in \Omega: X(\omega) \in B\} \in \mathcal{A}\) for all \(B \in \mathcal{B}_1\).

**Definition 2.2**
(i) For \(A \in \mathcal{A}\) the **indicator function** of \(A\) is the function

\[
1_A(\omega) = \begin{cases} 
1 & \text{if } \omega \in A \\
0 & \text{if } \omega \in A^c 
\end{cases}
\]

(ii) A **simple function** is \(X(\omega) \equiv \sum_{i=1}^n x_i 1_{A_i}(\omega)\) for \(\sum_{i=1}^n A_i = \Omega, \ A_i \in \mathcal{A}, \ x_i \in R\).

(iii) An **elementary function** is \(X(\omega) \equiv \sum_{i=1}^\infty x_i 1_{A_i}(\omega)\) for \(\sum_{i=1}^\infty A_i = \Omega, \ A_i \in \mathcal{A}, \ x_i \in R\).

**Proposition 2.1** \(X\) is measurable if and only if \(X^{-1}(C) \equiv \{X^{-1}(C) : C \in \mathcal{C}\} \subset \mathcal{A}\) where \(\sigma(\mathcal{C}) = \mathcal{B}\). Hence \(X\) is measurable if and only if \(X^{-1}((x, \infty)) \equiv [X > x] \in \mathcal{A}\) for all \(x \in R\).

**Proof.** \((\Rightarrow)\) This direction is trivial.
\((\Leftarrow)\) \(X^{-1}(B) = X^{-1}(\sigma(C)) = \sigma(X^{-1}(C))\) since \(X^{-1}\) preserves all set operations and since \(X^{-1}(C) \subset \mathcal{A}\) with \(\mathcal{A}\) a \(\sigma\)-field by hypothesis.

Further, \(\sigma([x, \infty) : x \in R]) = \mathcal{B}_1\) since \((a, b] = (a, \infty) \cap (b, \infty)\) and \(\mathcal{B}_1\) is generated by intervals of the form \([a, b]\). \(\square\) Note that the assertion of the proposition would work with \((x, \infty)\) replaced by any of \([x, \infty), (-\infty, x], (-\infty, x)\).

**Proposition 2.2** Suppose that \(\{X_n\}\) are measurable. Then so are \(\sup_n X_n\), \(-X_n\), \(\inf_n X_n\), \(\underline{\lim} X_n\), \(\lim X_n\), and \(\liminf X_n\).

**Proof.** \(\sup_n X_n > x = \bigcup_n [X_n > x]\);
\([-X_n > x] = [X_n < -x]\);
\(\inf_n X_n = -\sup_n (-X_n)\);
\(\lim X_n = \inf_n (\sup_{k>n} X_k)\);
\(\lim X_n = -\lim(-X_n)\);
\(\liminf X_n = \liminf X_n\) when \(\lim X_n\) exists. \(\square\)

**Proposition 2.3** \(X\) is measurable if and only if it is the limit of a sequence of simple functions:

\[X_n = n1_{X \leq n} + \sum_{k=-n2^n + 1}^{n2^n} \frac{k - 1}{2^n} 1_{[(k-1)/2^n \leq X < k/2^n]} + n1_{X > n}\].

**Proof.** \((\Rightarrow)\) The \(X_n\)'s exhibited above have \(|X_n(\omega) - X(\omega)| < 2^{-n}\) for \(|X(\omega)| < n\).
\((\Leftarrow)\) The exhibited \(X_n\)'s are simple, converge to \(X\), and \(\lim X_n\) is measurable by prop 2.2. \(\square\)

**Remark 2.1** If \(X \geq 0\), then \(0 \leq X_n \leq X\).
Proposition 2.4 Let \( X, Y \) be measurable. Then \( X \pm Y, XY, X/Y, X^+ \equiv X1_{[X \geq 0]}, X^- \equiv -X1_{[X \leq 0]}, |X|, g(X) \) for measurable \( g \) are all measurable.

Proof. Let \( X_n, Y_n \) be simple functions, \( X_n \to X, Y_n \to Y \). Then \( X_n \pm Y_n, X_nY_n, X_n/Y_n \) are simple functions converging to \( X \pm Y, XY, \) and \( X/Y, \) and hence the limits are measurable by prop 2.3. \( X^+ \) and \( X^- \) are easy by prop 2.3, and \( |X| = X^+ + X^- \). For \( g : \mathbb{R} \to \mathbb{R} \) measurable we have, for \( B \in B_1, \)

\[
(gX)^{-1}(B) = X^{-1}(g^{-1}(B)) = X^{-1}(\text{a Borel set}) \quad \text{since} \quad g \text{ is measurable}
\]

\[
\in A \quad \text{since} \quad X^{-1} \text{ is measurable}.
\]

\[\square\]

Remark 2.2 Any continuous function \( g \) is measurable since

\[g^{-1}(B) = g^{-1}(\sigma(O)) = \sigma(g^{-1}(O)) = \sigma(\text{a subcollection of open sets} \subset B).\]

Now let \((\Omega, \mathcal{A}, \mu)\) be a measure space, and let \( X, Y \) denote measurable functions from \((\Omega, \mathcal{A})\) to \((\mathbb{R}, \mathcal{B}), \mathcal{B} \equiv R \cup \{\pm \infty\}, \mathcal{B} \equiv \sigma(B \cup \{\infty\} \cup \{-\infty\}).\)

CONVENTIONS: \(0 \cdot \infty = 0 = \infty \cdot 0, x \cdot \infty = \infty \cdot x = \infty \text{ if } 0 < x < \infty; \infty \cdot \infty = \infty.\)

Definition 2.3 (i) For \( X = \sum^m_i x_i A_i \) with \( x_i \geq 0, \sum^m_i A_i = \Omega, \) then \( \int X d\mu = \sum^m_i x_i \mu(A_i)\).

(ii) For \( X \geq 0, \int X d\mu = \lim_n \int X_n d\mu \) where \( \{X_n\} \) is any \( \geq 0, / \) sequence of simple functions, \( X_n \to X.\)

(iii) For general \( X, \int X d\mu = \int X^+ d\mu - \int X^- d\mu \) if one of \( \int X^+ d\mu, \int X^- d\mu \) is finite.

(iv) If \( \int X d\mu \) is finite, then \( X \) is integrable.

JUSTIFICATION: See Loève pages 120 - 123 or Billingsley (1986), page 176.

Proposition 2.5 (Elementary properties.) Suppose that \( \int X d\mu, \int Y d\mu, \) and \( \int X d\mu + \int Y d\mu \) exist. Then:

(i) \( \int (X + Y) d\mu = \int X d\mu + \int Y d\mu, \int cX d\mu = c \int X d\mu; \)

(ii) \( X \geq 0 \) implies \( \int X d\mu \geq 0; X \geq Y \) implies \( \int X d\mu \geq \int Y d\mu; \) and \( X = Y \) a.e. implies \( \int X d\mu = \int Y d\mu.\)

(iii) \( (\text{integrability}), X \) is integrable if and only if \( |X| \) is integrable, and either implies that \( X \) is a.e. finite, \( |X| \leq Y \) with \( Y \) integrable implies \( X \) integrable; \( X \) and \( Y \) integrable implies that \( X + Y \) is integrable.

Proof. (iii) That \( X \) is integrable if and only if \( \int X^+ d\mu \) and \( \int X^- d\mu \) finite if and only if \( |X| \) integrable is easy. Now \( \int X^+ d\mu < \infty \) implies \( X^+ \) finite a.e.; if not, then \( \mu(A) > 0 \) where \( A \equiv \{\omega : X^+(\omega) = \infty\}, \) and then \( \int X^+ d\mu \geq \int X^+ 1_A d\mu = \infty \cdot \mu(A) = \infty, \) a contradiction. Now \( 0 \leq X^\leq Y, \) thus \( 0 \leq \int X^+ d\mu \leq \int Y d\mu < \infty. \) Likewise \( \int X^- d\mu < \infty. \) \[\square\]

Theorem 2.1 (Monotone convergence theorem.) If \( 0 \leq X_n \to X, \) then \( \int X_n d\mu \to \int X d\mu.\)

Corollary 1 If \( X_n \geq 0 \) then \( \int \sum_{n=1}^{\infty} X_n d\mu = \sum_{n=1}^{\infty} \int X_n d\mu.\)
2. MEASURABLE FUNCTIONS AND INTEGRATION

Proof. Note that $0 \leq \sum_{n} X_{k} \nearrow \sum_{1}^{\infty} X_{k}$ and apply the monotone convergence theorem. □

Theorem 2.2 (Fatou's lemma.) If $X_{n} \geq 0$ for all $n$, then $\liminf_{n} X_{n} d\mu \leq \lim \int X_{n} d\mu$. 

Proof. Since $X_{n} \geq \inf_{k \geq n} X_{k} \equiv Y_{n} \nearrow \lim X_{n}$, it follows from the MCT that

$$\int \liminf_{n} X_{n} d\mu = \int \lim Y_{n} d\mu = \lim \int Y_{n} d\mu \leq \lim \int X_{n} d\mu.$$ □

Definition 2.4 A sequence $X_{n}$ converges almost everywhere (or converges a.e. for short), denoted $X_{n} \to_{a.e.} X$, if $X_{n}(\omega) \to X(\omega)$ for all $\omega \in \Omega \setminus N$ where $\mu(N) = 0$ (i.e. for a.e. $\omega$). Note that \{X_{n}\}, X, are all defined on one measure space $(\Omega, A)$. If $\mu$ is a probability measure, $\mu = P$ with $P(\Omega) = 1$, we will write $\to_{a.e.}$ for $\to_{a.e.}$.

Proposition 2.6 Let \{X_{n}\}, X be finite measurable functions. Then $[X_{n} \to X] = \cap_{k=1}^{\infty} \cup_{m=n}^{\infty} [X_{m} - X] < 1/k]$, and is a measurable set.

Corollary 1 Let \{X_{n}\}, X be finite measurable functions. Then $X_{n} \to_{a.e.} X$ if and only if

$$\mu(\cap_{n=1}^{\infty} \cup_{m=n}^{\infty} [X_{m} - X] \geq \varepsilon) = 0$$

for all $\varepsilon > 0$. If $\mu(\Omega) < \infty$, $X_{n} \to_{a.e.} X$ if and only if

$$\mu(\cup_{m=n}^{\infty} [X_{m} - X] \geq \varepsilon) \to 0 \quad \text{as} \quad n \to \infty$$

for all $\varepsilon > 0$.

Proof. First note that

$$[X_{n} \to X]^{c} = \cup_{k=1}^{\infty} \cap_{n=1}^{\infty} \cup_{m=n}^{\infty} [X_{m} - X] \geq 1/k] \equiv \cup_{k=1}^{\infty} A_{k}$$

with $A_{k} \nearrow$; and $A_{k} = \cap_{n=1}^{\infty} B_{n,k}$ with $B_{n,k} \setminus$ in n. Applying prop 1.2 gives the result. □

Definition 2.5 (Convergence in measure; convergence in probability.) A sequence of finite measurable functions $X_{n}$ converge in measure to a measurable function $X$, denoted $X_{n} \to_{\mu} X$, if

$$\mu([X_{n} - X] \geq \varepsilon) \to 0$$

for all $\varepsilon > 0$. If $\mu$ is a probability measure, $\mu(\Omega) = 1$, call $\mu = P$, write $X_{n} \to_{P} X$, and say $X_{n}$ converge in probability to $X$.

Proposition 2.7 Let $X_{n}$'s be finite a.e.

(i) If $X_{n} \to_{\mu} X$ then there exist a subsequence \{n_{k}\} such that $X_{n_{k}} \to_{a.e.} X$.

(ii) If $\mu(\Omega) < \infty$ and $X_{n} \to_{a.e.} X$, then $X_{n} \to_{\mu} X$.

Theorem 2.3 (Dominated Convergence Theorem) If $|X_{n}| \leq Y$ a.e. with $Y$ integrable, and if $X_{n} \to_{\mu} X$ (or $X_{n} \to_{a.e.} X$), then $\int |X_{n} - X| d\mu \to 0$ and $\lim \int X_{n} d\mu = \int X d\mu$. 

**Proof.** We give the proof under the assumption \( X_n \rightarrow_{a.e.} X \). Then \( Z_n = |X_n - X| \rightarrow 0 \) a.e. and \( Z_n \leq |X_n| + |X| \leq 2Y \equiv Z \). Thus \( Z - Z_n \geq 0 \) and by Fatou's lemma
\[
\int Zd\mu = \int \lim_{n \to \infty} (Z - Z_n)d\mu \leq \lim_{n \to \infty} \int (Z - Z_n)d\mu = \int Zd\mu - \lim_{n \to \infty} \int Z_n d\mu,
\]
and this implies
\[
\lim_{n \to \infty} \int Z_n = \lim_{n \to \infty} \int |X_n - X|d\mu \leq 0.
\]
Thus
\[
\left| \int X_n - \int X \right| = \left| \int (X_n - X)d\mu \right| \leq \int |X_n - X|d\mu \to 0.
\]
\(\square\)

**Definition 2.6** Let \( X \) be a finite measurable function on a probability space \((\Omega, \mathcal{A}, P)\) (so that \( P(\Omega) = 1 \)). Then \( X \) is called a random variable and
\[
P_X(B) \equiv P(X \in B) = P(\{\omega \in \Omega : X(\omega) \in B\})
\]
for all \( B \in \mathcal{B} \) is called the (induced) probability distribution of \( X \) (on \( R \)). The df associated with \( P_X \) is denoted by \( F_X \) and is called the df of the random variable \( X \). Thus \((R, \mathcal{B}, P_X)\) is a probability space.

**Theorem 2.4 (Theorem of the unconscious statistician.)** If \( g \) is a finite measurable function from \( R \) to \( R \), then
\[
\int_{\Omega} g(X(\omega))dP(\omega) = \int_{R} g(x)dP_{X}(x) = \int_{R} g(x)dF_{X}(x).
\]

**Proposition 2.8 (Interchange of integral and limit or derivative.)** Suppose that \( X(\omega, t) \) is measurable for each \( t \in (a, b) \).

(i) If \( X(\omega, t) \) is a.e. continuous in \( t \) at \( t_0 \) and \( |X(\omega, t)| \leq Y(\omega) \) a.e. for \( |t - t_0| < \delta \) with \( Y \) integrable, then \( \int X(\cdot, t)d\mu \) is continuous in \( t \) at \( t_0 \).

(ii) Suppose that \( \frac{\partial}{\partial t} X(\omega, t) \) exists for a.e. \( \omega \), all \( t \in (a, b) \), and \( |\frac{\partial}{\partial t} X(\omega, t)| \leq Y(\omega) \) integrable a.e. for all \( t \in (a, b) \). Then
\[
\frac{\partial}{\partial t} \int_{\Omega} X(\omega, t)d\mu(\omega) = \int_{\Omega} \frac{\partial}{\partial t} X(\omega, t)d\mu(\omega).
\]

**Proof.** (ii). By the mean value theorem
\[
\frac{X(\omega, t+h) - X(\omega, t)}{h} = \frac{\partial}{\partial t} X(\omega, t)|_{t=s}
\]
for some \( t \leq s \leq t + h \). Also the left side of the display converges to \( \frac{\partial}{\partial t} X(\omega, t) \) as \( h \to 0 \) for a.e. \( \omega \), and by the equality of the display and the hypothesized bound, the difference quotient on the left side of the display is bounded in absolute value by \( Y \). Therefore
\[
\frac{\partial}{\partial t} \int X(\omega, t)d\mu(\omega) = \lim_{h \to 0} \frac{1}{h} \left\{ \int X(\omega, t+h)d\mu(\omega) - \int X(\omega, t)d\mu(\omega) \right\}
\]
\[
= \lim_{h \to 0} \frac{1}{h} \int \left\{ \frac{X(\omega, t+h) - X(\omega, t)}{h} \right\} d\mu(\omega)
\]
\[
= \int \frac{\partial}{\partial t} X(\omega, t)d\mu(\omega)
\]
where the last equality holds by the dominated convergence theorem. \(\square\)
3. **Absolute Continuity, Radon-Nikodym Theorem, Fubini’s Theorem**

Let \((\Omega, \mathcal{A}, \mu)\) be a measure space, and let \(X\) be a non-negative measurable function on \(\Omega\). For \(A \in \mathcal{A}\), set

\[
\nu(A) \equiv \int_A X d\mu = \int_\Omega 1_A X d\mu.
\]

Then \(\nu\) is another measure on \((\Omega, \mathcal{A})\) and \(\nu\) is finite if and only if \(X\) is integrable \((X \in L_1(\mu))\).

**Definition 3.1** The measure \(\nu\) defined by ?? is said to have density \(X\) with respect to \(\mu\).

Note that \(\mu(A) = 0\) implies that \(\nu(A) = 0\).

**Definition 3.2** If \(\mu, \nu\) are any two measures on \((\Omega, \mathcal{A})\) such that \(\mu(A) = 0\) implies \(\nu(A) = 0\) for any \(A \in \mathcal{A}\), then \(\nu\) is said to be **absolutely continuous with respect to \(\mu\)**, and we write \(\nu \ll \mu\). We also say that \(\nu\) is **dominated by \(\mu\)**.

**Theorem 3.1 (Radon-Nikodym theorem.)** Let \((\Omega, \mathcal{A}, \mu)\) be a \(\sigma\)-finite measure space, and let \(\nu\) be a measure on \((\Omega, \mathcal{A})\) with \(\nu \ll \mu\). Then there exists a measurable function \(X \geq 0\) such that \(\nu(A) = \int_A X d\mu\) for all \(A \in \mathcal{A}\). The function \(X = \frac{d\nu}{d\mu}\) is unique in the sense that if \(Y\) is another such function, then \(Y = X\) \(a.e.\) with respect to \(\mu\). \(X\) is called the **Radon-Nikodym derivative of \(\nu\) with respect to \(\mu\)**.

**Proof.** See Billingsley (1986), page 376. \(\square\)

**Corollary 1 (Change of Variable Theorem.)** Suppose that \(\nu, \mu\) are \(\sigma\)-finite measures defined on a measure space \((\Omega, \mathcal{A})\) with \(\nu \ll \mu\), and suppose that \(Z\) is a measurable function such that \(\int Z d\nu\) is well-defined. Then for all \(A \in \mathcal{A}\),

\[
\int_A Z d\nu = \int_A Z \frac{d\nu}{d\mu} d\mu.
\]

**Proof.** (i) If \(Z = 1_B\); then

\[
\int_A 1_B d\nu = \nu(A \cap B) = \int_{A \cap B} \frac{d\nu}{d\mu} d\mu = \int_A 1_B \frac{d\nu}{d\mu} d\mu.
\]

(ii) If \(Z = \sum_1^m z_i 1_{A_i}\), then

\[
\int_A Z d\nu = \sum_1^m z_i \int_A 1_{A_i} d\nu = \sum_1^m z_i \int_A 1_{A_i} \frac{d\nu}{d\mu} d\mu \quad \text{by (i)}
\]

\[
= \int_A Z \frac{d\nu}{d\mu} d\mu.
\]
(iii) If \( Z \geq 0, \) let \( Z_n \geq 0 \) be simple functions \( \nearrow Z \). Then
\[
\int_A Z d\nu = \lim \int_A Z_n d\nu \quad \text{by the monotone convergence thm.}
\]
\[
= \lim \int_A \frac{d\nu}{d\mu} d\mu \quad \text{by part (ii)}
\]
\[
= \int_A \frac{d\nu}{d\mu} d\mu \quad \text{by the monotone convergence thm.}
\]

(iv) If \( Z \) is measurable, \( Z = Z^+ - Z^- \) where one of \( Z^+, Z^- \) is \( \nu \)-integrable, then
\[
\int_A Z d\nu = \int_A Z^+ d\nu - \int_A Z^- d\nu
\]
\[
= \int_A Z^+ \frac{d\nu}{d\mu} d\mu - \int_A Z^- \frac{d\nu}{d\mu} d\mu \quad \text{by (iii)}
\]
\[
= \int_A Z \frac{d\nu}{d\mu} d\mu.
\]

Example 3.1 Let \((\Omega, \mathcal{A}, P)\) be a probability space; often this will be \((\mathbb{R}^n, \mathcal{B}_\mathbb{R}, P)\). Often in statistics we suppose that \( P \) has a density \( f \) with respect to a \( \sigma \)-finite measure \( \mu \) on \((\Omega, \mathcal{A})\) so that
\[
P(A) = \int_A f d\mu \quad \text{for} \quad A \in \mathcal{A}.
\]
If \( \mu \) is Lebesgue measure on \( \mathbb{R}^n \), then \( f \) is the density function. If \( \mu \) is counting measure on a countable set, then \( f \) is the frequency function or mass function.

Proposition 3.1 (Scheffé’s theorem) Suppose that \( \nu_n(A) = \int_A f_n d\mu \), that \( \nu(A) = \int_A f d\mu \) where \( f_n \) are densities and \( \nu_n(\Omega) = \nu(\Omega) < \infty \) for all \( n \), and that \( f_n \to f \) a.e. \( \mu \). Then
\[
\sup_{A \in \mathcal{A}} |\nu_n(A) - \nu(A)| = \frac{1}{2} \int_\Omega |f_n - f| d\mu \to 0.
\]

Proof. For \( A \in \mathcal{A} \),
\[
|\nu_n(A) - \nu(A)| = \left| \int_A (f_n - f) d\mu \right|
\]
\[
\leq \int_A |f_n - f| d\mu \leq \int_\Omega |f_n - f| d\mu,
\]
and this implies that
\[
\sup_{A \in \mathcal{A}} |\nu_n(A) - \nu(A)| \leq \int_\Omega |f_n - f| d\mu.
\]
Let \( g_n \equiv f - f_n \). Now \( g_n^+ \to 0 \) a.e. \( \mu \), and \( g_n^+ \leq f \) which is integrable. Thus by the dominated convergence theorem \( \int g_n^+ d\mu \to 0 \). But
\[
0 = \int g_n d\mu = \int_\Omega (f - f_n) d\mu = \int_\Omega (g_n^+ - g_n^-) d\mu.
\]
3. **Absolute Continuity, Radon-Nikodym Theorem, Fubini's Theorem**

so \( \int g_n^+ \, d\mu = \int g_n^- \, d\mu \), and hence

\[
\int |g_n| \, d\mu = \int g_n^+ \, d\mu + \int g_n^- \, d\mu = 2 \int g_n^+ \, d\mu \to 0,
\]

proving the claimed convergence. To prove that equality holds as claimed in the statement of the proposition, note that for the event \( B \equiv [f - f_n \geq 0] \) we have

\[
\sup_{A \in \mathcal{A}} |\nu_n(A) - \nu(A)| \geq |\nu_n(B) - \nu(B)| = \left| \int_{[f-f_n \geq 0]} (f_n - f) \, d\mu \right|
\]

\[
= \int_{[g_n^+ \geq 0]} g_n^+ \, d\mu = \int g_n^+ \, d\mu
\]

\[
= \frac{1}{2} \int |f_n - f| \, d\mu.
\]

But on the other hand

\[
|\nu_n(A) - \nu(A)| = \left| \int_A f_n \, d\mu - \int_A f \, d\mu \right|
\]

\[
= \left| \int_A (f - f_n) \, d\mu \right|
\]

\[
= \left| \int_{A \cap B} (f - f_n) \, d\mu + \int_{A \cap B^c} (f - f_n) \, d\mu \right|
\]

\[
\leq \int g_n^+ \, d\mu,
\]

so

\[
\sup_{A \in \mathcal{A}} |\nu_n(A) - \nu(A)| \leq \int g_n^+ \, d\mu = \frac{1}{2} \int |f_n - f| \, d\mu.
\]

\(\square\)

Now suppose that \((X, \mathcal{X}, \mu)\) and \((Y, \mathcal{Y}, \nu)\) are two \(\sigma\)-finite measure spaces. If \(A \in \mathcal{X}, B \in \mathcal{Y}\), a measurable rectangle is a set of the form \(A \times B \subset X \times Y\).

Let \(X \times Y \equiv \sigma(\{A \times B : A \in \mathcal{X}, B \in \mathcal{Y}\})\). Define a measure \(\pi\) on \((X \times Y, X \times Y)\) by

\[
\pi(A \times B) = \mu(A)\nu(B)
\]

for measurable rectangles \(A \times B\).

**Theorem 3.2 (Fubini - Tonelli theorem.)** Suppose that \(f : X \times Y \to R\) is \(X \times Y\)-measurable and \(f \geq 0\). Then

\[
\int_Y f(x, y) \, d\nu(y) \quad \text{is} \quad \mathcal{X} \text{- measurable},
\]

\[
\int_X f(x, y) \, d\mu(x) \quad \text{is} \quad \mathcal{Y} \text{- measurable},
\]

and

\[
(1) \quad \int_{X \times Y} f(x, y) \, d\pi(x, y) = \int_X \left( \int_Y f(x, y) \, d\nu(y) \right) \, d\mu(x) = \int_Y \left( \int_X f(x, y) \, d\mu(x) \right) \, d\nu(y).
\]

If \(f \in L_1(\pi)\) (so \(\int_{X \times Y} |f| \, d\pi < \infty\)), then (1) holds.