A Brief Introduction to Measure Theory and Integration

Richard F. Bass
Department of Mathematics
University of Connecticut

September 18, 1998

1. Measures.

Let X be a set. We will use the notation: $A^c = \{x \in X : x \notin A\}$ and $A - B = A \cap B^c$.

Definition. An *algebra* or a *field* is a collection $\mathcal{A}$ of subsets of X such that
(a) $\emptyset, X \in \mathcal{A}$;
(b) if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$;
(c) if $A_1, \ldots, A_n \in \mathcal{A}$, then $\cup_{i=1}^n A_i$ and $\cap_{i=1}^n A_i$ are in $\mathcal{A}$.

$\mathcal{A}$ is a *σ-algebra* or *σ-field* if in addition
(d) if $A_1, A_2, \ldots$ are in $\mathcal{A}$, then $\cup_{i=1}^\infty A_i$ and $\cap_{i=1}^\infty A_i$ are in $\mathcal{A}$.

In (d) we allow countable unions and intersections only; we do not allow uncountable unions and intersections.

Example. Let $X = \mathbb{R}$ and $\mathcal{A}$ be the collection of all subsets of $\mathbb{R}$.

Example. Let $X = \mathbb{R}$ and let $\mathcal{A} = \{A \subset \mathbb{R} : A \text{ is countable or } A^c \text{ is countable}\}$.

Definition. A *measure* on $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to [0, \infty]$ such that
(a) $\mu(A) \geq 0$ for all $A \in \mathcal{A}$;
(b) $\mu(\emptyset) = 0$;
(c) if $A_i \in \mathcal{A}$ are disjoint, then

$$\mu(\cup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i).$$

Example. X is any set, $\mathcal{A}$ is the collection of all subsets, and $\mu(A)$ is the number of elements in A.

Example. $X = \mathbb{R}$, $\mathcal{A}$ the collection of all subsets, $x_1, x_2, \ldots \in \mathbb{R}$, $a_1, a_2, \ldots > 0$, and $\mu(A) = \sum_{\{i : x_i \in A\}} a_i$.

Example. $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise. This measure is called *point mass* at x.

Proposition 1.1. *The following hold:*
(a) If $A, B \in \mathcal{A}$ with $A \subset B$, then $\mu(A) \leq \mu(B)$.
(b) If $A_i \in \mathcal{A}$ and $A = \cup_{i=1}^\infty A_i$, then $\mu(A) \leq \sum_{i=1}^\infty \mu(A_i)$.
(c) If $A_i \in \mathcal{A}$, $A_1 \subset A_2 \subset \cdots$, and $A = \cup_{i=1}^\infty A_i$, then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.
(d) If $A_i \in \mathcal{A}$, $A_1 \supset A_2 \supset \cdots$, $\mu(A_1) < \infty$, and $A = \cap_{i=1}^\infty A_i$, then we have $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.

Proof. (a) Let $A_1 = A$, $A_2 = B - A$, and $A_3 = A_4 = \cdots = \emptyset$. Now use part (c) of the definition of measure.

(b) Let $B_1 = A_1$, $B_2 = A_2 - B_1$, $B_3 = A_3 - (B_1 \cup B_2)$, and so on. The B_i are disjoint and $\cup_{i=1}^\infty B_i = \cup_{i=1}^\infty A_i$. So $\mu(A) = \sum \mu(B_i) \le \sum \mu(A_i)$.

(c) Define the B_i as in (b). Since $\cup_{i=1}^n B_i = \cup_{i=1}^n A_i$, then

$$\begin{aligned}\mu(A) = \mu(\cup_{i=1}^\infty A_i) = \mu(\cup_{i=1}^\infty B_i) &= \sum_{i=1}^\infty \mu(B_i) \\ &= \lim_{n\to\infty} \sum_{i=1}^n \mu(B_i) = \lim_{n\to\infty} \mu(\cup_{i=1}^n B_i) = \lim_{n\to\infty} \mu(\cup_{i=1}^n A_i).\end{aligned}$$

(d) Apply (c) to the sets $A_1 - A_i$, $i = 1, 2, \ldots$. □

Definition. A *probability* or *probability measure* is a measure such that $\mu(X) = 1$. In this case we usually write $(\Omega, \mathcal{F}, \mathbb{P})$ instead of $(X, \mathcal{A}, \mu)$.

2. Construction of Lebesgue measure.

Define $m((a, b)) = b - a$. If G is an open set and $G \subset \mathbb{R}$, then $G = \cup_{i=1}^\infty (a_i, b_i)$ with the intervals disjoint. Define $m(G) = \sum_{i=1}^\infty (b_i - a_i)$. If $A \subset \mathbb{R}$, define

$$m^*(A) = \inf\{m(G) : G \text{ open}, A \subset G\}.$$

We will show the following.

(1) m^* is not a measure on the collection of all subsets of $\mathbb{R}$.

(2) m^* is a measure on the σ-algebra consisting of what are known as m^*-measurable sets.

(3) Let $\mathcal{A}_0$ be the algebra (not σ-algebra) consisting of all finite unions of sets of the form $[a_i, b_i)$. If $\mathcal{A}$ is the smallest σ-algebra containing $\mathcal{A}_0$, then m^* is a measure on $(\mathbb{R}, \mathcal{A})$.

We will prove these three facts (and a bit more) in a moment, but let's first make some remarks about the consequences of (1)-(3).

If you take any collection of σ-algebras and take their intersection, it is easy to see that this will again be a σ-algebra. The smallest σ-algebra containing $\mathcal{A}_0$ will be the intersection of all σ-algebras containing $\mathcal{A}_0$.

Since $(a, b]$ is in $\mathcal{A}_0$ for all a and b, then $(a, b) = \cup_{i=i_0}^\infty (a, b - 1/i] \in \mathcal{A}$, where we choose i_0 so that $1/i_0 < b - a$. Then sets of the form $\cup_{i=1}^\infty (a_i, b_i)$ will be in $\mathcal{A}$, hence all open sets. Therefore all closed sets are in $\mathcal{A}$ as well.

The smallest σ-algebra containing the open sets is called the *Borel σ-algebra*. It is often written $\mathcal{B}$.

A set N is a *null set* if $m^*(N) = 0$. Let $\mathcal{L}$ be the smallest σ-algebra containing $\mathcal{B}$ and all the null sets. $\mathcal{L}$ is called the *Lebesgue σ-algebra*, and sets in $\mathcal{L}$ are called *Lebesgue measurable*.

As part of our proofs of (2) and (3) we will show that m^* is a measure on $\mathcal{L}$. *Lebesgue measure* is the measure m^* on $\mathcal{L}$. (1) shows that $\mathcal{L}$ is strictly smaller than the collection of all subsets of $\mathbb{R}$.

Proof of (1). Define $x \sim y$ if $x - y$ is rational. This is an equivalence relationship on $[0, 1]$. For each equivalence class, pick an element out of that class (by the axiom of choice) Call the collection of such points A. Given a set B, define $B + x = \{y + x : y \in B\}$. Note $m^*(A + q) = m^*(A)$ since this translation invariance holds for intervals, hence for open sets, hence for all sets. Moreover, the sets $A + q$ are disjoint for different rationals q.

Now

$$[0,1] \subset \cup_{q\in[-2,2]}(A+q),$$

where the sum is only over rational q, so $1 \le \sum_{q\in[-2,2]} m^*(A+q)$, and therefore $m^*(A) > 0$. But

$$\cup_{q\in[-2,2]}(A+q) \subset [-6,6],$$

where again the sum is only over rational q, so $12 \ge \sum_{q\in[-2,2]} m^*(A+q)$, which implies $m^*(A) = 0$, a contradiction. $\square$

Proposition 2.1. *The following hold:*
(a) $m^*(\emptyset) = 0$;
(b) if $A \subset B$, *then* $m^*(A) \le m^*(B)$;
(c) $m^*(\cup_{i=1}^\infty A_i) \le \sum_{i=1}^\infty m^*(A_i)$.

Proof. (a) and (b) are obvious. To prove (c), let $\varepsilon > 0$. For each i there exist intervals $I_{i1}, I_{i2}, \ldots$ such that $A_i \subset \cup_{j=1}^\infty I_{ij}$ and $\sum_j m(I_{ij}) \le m^*(A_i) + \varepsilon/2^i$. Then $\cup_{i=1}^\infty A_i \subset \cup_{i,j} I_{ij}$ and

$$\sum_{i,j} m(I_{ij}) \le \sum_i m^*(A_i) + \sum_i \varepsilon/2^i = \sum_i m^*(A_i) + \varepsilon.$$

Since ε is arbitrary, $m^*(\cup_{i=1}^\infty A_i) \le \sum_{i=1}^\infty m^*(A_i)$. $\square$

A function on the collection of all subsets satisfying (a), (b), and (c) is called an *outer measure*.

Definition. Let m^* be an outer measure. A set $A \subset X$ is m^**-measurable* if

$$m^*(E) = m^*(E \cap A) + m^*(E \cap A^c) \tag{2.1}$$

for all $E \subset X$.

Theorem 2.2. *If m^* is an outer measure on X, then the collection $\mathcal{A}$ of m^* measurable sets is a σ-algebra and the restriction of m^* to $\mathcal{A}$ is a measure. Moreover, $\mathcal{A}$ contains all the null sets.*

Proof. By Proposition 2.1(c),

$$m^*(E) \le m^*(E \cap A) + m^*(E \cap A^c)$$

for all $E \subset X$. So to check (2.1) it is enough to show $m^*(E) \ge m^*(E \cap A) + m^*(E \cap A^c)$. This will be trivial in the case $m^*(E) = \infty$.

If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$ by symmetry and the definition of $\mathcal{A}$. Suppose $A, B \in \mathcal{A}$ and $E \subset X$. Then

$$\begin{aligned} m^*(E) &= m^*(E \cap A) + m^*(E \cap A^c) \\ &= (m^*(E \cap A \cap B) + m^*(E \cap A \cap B^c)) + (m^*(E \cap A^c \cap B) + m^*(E \cap A^c \cap B^c)) \end{aligned}$$

The first three terms on the right have a sum greater than or equal to $m^*(E \cap (A \cup B))$ because $A \cup B \subset (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$. Therefore

$$m^*(E) \ge m^*(E \cap (A \cup B)) + m^*(E \cap (A \cup B)^c),$$

which shows $A \cup B \in \mathcal{A}$. Therefore $\mathcal{A}$ is an algebra.

Let A_i be disjoint sets in $\mathcal{A}$, let $B_n = \cup_{i=1}^n A_i$, and $B = \cup_{i=1}^\infty A_i$. If $E \subset X$,

$$\begin{aligned} m^*(E \cap B_n) &= m^*(E \cap B_n \cap A_n) + m^*(E \cap B_n \cap A_n^c) \\ &= m^*(E \cap A_n) + m^*(E \cap B_{n-1}). \end{aligned}$$

Repeating for $m^*(E \cap B_{n-1})$, we obtain

$$m^*(E \cap B_n) = \sum_{i=1}^n m^*(E \cap A_i).$$

So

$$m^*(E) = m^*(E \cap B_n) + m^*(E \cap B_n^c) \geq \sum_{i=1}^n m^*(E \cap A_i) + m^*(E \cap B^c).$$

Let $n \to \infty$. Then

$$\begin{aligned} m^*(E) &\geq \sum_{i=1}^\infty m^*(E \cap A_i) + m^*(E \cap B^c) \\ &\geq m^*(\cup_{i=1}^\infty (E \cap A_i)) + m^*(E \cap B^c) \\ &= m^*(E \cap B) + m(E \cap B^c) \\ &\geq m^*(E). \end{aligned}$$

This shows $B \in \mathcal{A}$.

If we set $E = B$ in this last equation, we obtain

$$m^*(B) = \sum_{i=1}^\infty m^*(A_i),$$

or m^* is countably additive on $\mathcal{A}$.

If $m^*(A) = 0$ and $E \subset X$, then

$$m^*(E \cap A) + m^*(E \cap A^c) = m^*(E \cap A^c) \leq m^*(E),$$

which shows $\mathcal{A}$ contains all null sets. $\square$

None of this is useful if $\mathcal{A}$ does not contain the intervals. There are two main steps in showing this. Let $\mathcal{A}_0$ be the algebra consisting of all finite unions of intervals of the form $(a, b]$. The first step is

Proposition 2.3. *If $A_i \in \mathcal{A}_0$ are disjoint and $\cup_{i=1}^\infty A_i \in \mathcal{A}_0$, then we have $m(\cup_{i=1}^\infty A_i) = \sum_{i=1}^\infty m(A_i)$.*

Proof. Since $\cup_{i=1}^\infty A_i$ is a finite union of intervals $(a_k, b_k]$, we may look at $A_i \cap (a_k, b_k]$ for each k. So we may assume that $A = \cup_{i=1}^\infty A_i = (a, b]$.

First,

$$m(A) = m(\cup_{i=1}^n A_i) + m(A - \cup_{i=1}^n A_i) \geq m(\cup_{i=1}^n A_i) = \sum_{i=1}^n m(A_i).$$

Letting $n \to \infty$,

$$m(A) \geq \sum_{i=1}^\infty m(A_i).$$

Let us assume a and b are finite, the other case being similar. By linearity, we may assume $A_i = (a_i, b_i]$. Let $\varepsilon > 0$. The collection $\{(a_i, b_i + \varepsilon/2^i)\}$ covers $[a + \varepsilon, b]$, and so there exists a finite subcover.

Discarding any interval contained in another one, and relabeling, we may assume $a_1 < a_2 < \cdots a_N$ and $b_i + \varepsilon/2^i \in (a_{i+1}, b_{i+1} + \varepsilon/2^{i+1})$. Then

$$\begin{aligned} m(A) &= b - a = b - (a+\varepsilon) + \varepsilon \\ &\le \sum_{i=1}^{N} (b_i + \varepsilon/2^i - a_i) + \varepsilon \\ &\le \sum_{i=1}^{\infty} m(A_i) + 2\varepsilon. \end{aligned}$$

Since ε is arbitrary, $m(A) \le \sum_{i=1}^\infty m(A_i)$. □

The second step is the Carathéodory extension theorem. We say that a measure m is *σ-finite* if there exist $E_1, E_2, \ldots$, such that $m(E_i) < \infty$ for all i and $X \subset \cup_{i=1}^\infty E_i$.

Theorem 2.4. *Suppose $\mathcal{A}_0$ is an algebra and m restricted to $\mathcal{A}_0$ is a measure. Define*

$$m^*(E) = \inf\Big\{ \sum_{i=1}^\infty m(A_i) : A_i \in \mathcal{A}_0, E \subset \cup_{i=1}^\infty A_i \Big\}.$$

Then
(a) $m^*(A) = m(A)$ if $A \in \mathcal{A}_0$;
(b) every set in $\mathcal{A}_0$ is m^*-measurable;
(c) if m is σ-finite, then there is a unique extension to the smallest σ-field containing $\mathcal{A}_0$.

Proof. We start with (a). Suppose $E \in \mathcal{A}_0$. We know $m^*(E) \le m(E)$ since we can take $A_1 = E$ and $A_2, A_3, \ldots$ empty in the definition of m^*. If $E \subset \cup_{i=1}^\infty A_i$ with $A_i \in \mathcal{A}_0$, let $B_n = E \cap (A_n - \cup_{i=1}^{n-1} A_i)$. The the B_n are disjoint, they are each in $\mathcal{A}_0$, and their union is E. Therefore

$$m(E) = \sum_{i=1}^\infty m(B_i) \le \sum_{i=1}^\infty m(A_i).$$

Thus $m(E) \le m^*(E)$.

Next we look at (b). Suppose $A \in \mathcal{A}_0$. Let $\varepsilon > 0$ and let $E \subset X$. Pick $B_i \in \mathcal{A}_0$ such that $E \subset \cup_{i=1}^\infty B_i$ and $\sum_i m(B_i) \le m^*(E) + \varepsilon$. Then

$$\begin{aligned} m^*(E) + \varepsilon &\ge \sum_{i=1}^\infty m(B_i) = \sum_{i=1}^\infty m(B_i \cap A) + \sum_{i=1}^\infty m(B_i \cap A^c) \\ &\ge m^*(E \cap A) + m^*(E \cap A^c). \end{aligned}$$

Since ε is arbitrary, $m^*(E) \ge m^*(E \cap A) + m^*(E \cap A^c)$. So A is m^*-measurable.

Finally, suppose we have two extensions to the smallest σ-field containing $\mathcal{A}_0$; let the other extension be called n. We will show that if E is in this smallest σ-field, then $m^*(E) = n(E)$.

Since E must be m^*-measurable, $m^*(E) = \inf\{\sum_{i=1}^\infty m(A_i) : E \subset \cup_{i=1}^\infty A_i, A_i \in \mathcal{A}_0\}$. But $m = n$ on $\mathcal{A}_0$, so $\sum_i m(A_i) = \sum_i n(A_i)$. Therefore $n(E) \le \sum_i n(A_i)$, which implies $n(E) \le m^*(E)$.

Let $\varepsilon > 0$ and choose $A_i \in \mathcal{A}_0$ such that $m^*(E) + \varepsilon \ge \sum_i m(A_i)$ and $E \subset \cup_i A_i$. Let $A = \cup_i A_i$ and $B_k = \cup_{i=1}^k A_i$. Observe $m^*(E) + \varepsilon \ge m^*(A)$, hence $m^*(A - E) < \varepsilon$. We have

$$m^*(A) = \lim_{k\to\infty} m^*(B_k) = \lim_{k\to\infty} n(B_k) = n(A).$$

Then
$$m^*(E) \le m^*(A) = n(A) = n(E) + n(A - E) \le n(E) + m(A - E) \le n(E) + \varepsilon.$$
Since ε is arbitrary, this completes the proof. $\square$

We now drop the $*$ from m^* and call m Lebesgue measure.

3. Lebesgue-Stieltjes measures. Let $\alpha : \mathbb{R} \to \mathbb{R}$ be nondecreasing and right continuous (i.e., $\alpha(x+) = \alpha(x)$ for all x). Suppose we define $m_\alpha((a, b)) = \alpha(b) - \alpha(a)$, define $m_\alpha(\cup_{i=1}^\infty (a_i, b_i)) = \sum_i (\alpha(b_i) - \alpha(a_i))$ when the intervals (a_i, b_i) are disjoint, and define $m_\alpha^*(A) = \inf\{m_\alpha(G) : A \subset G, G \text{ open}\}$. Very much as in the previous section we can show that m_α^* is a measure on the Borel σ-algebra. The only differences in the proof are that where we had $a+\varepsilon$, we replace this by a', where a' is chosen so that $a' > a$ and $\alpha(a') \le \alpha(a)+\varepsilon$ and we replace $b_i + \varepsilon/2^i$ by b_i', where b_i' is chosen so that $b_i' > b_i$ and $\alpha(b_i') \le \alpha(b_i) + \varepsilon/2^i$. These choices are possible because α is right continuous.

Lebesgue measure is the special case of m_α when $\alpha(x) = x$.

Given a measure μ on $\mathbb{R}$ such that $\mu(K) < \infty$ whenever K is compact, define $\alpha(x) = \mu((0, x])$ if $x \ge 0$ and $\alpha(x) = -\mu((x, 0])$ if $x < 0$. Then α is nondecreasing, right continuous, and it is not hard to see that $\mu = m_\alpha$.

4. Measurable functions. Suppose we have a set X together with a σ-algebra $\mathcal{A}$.

Definition. $f : X \to \mathbb{R}$ is *measurable* if $\{x : f(x) > a\} \in \mathcal{A}$ for all $a \in \mathbb{R}$.

Proposition 4.1. *The following are equivalent.*
(a) $\{x : f(x) > a\} \in \mathcal{A}$ *for all* a*;*
(b) $\{x : f(x) \le a\} \in \mathcal{A}$ *for all* a*;*
(c) $\{x : f(x) < a\} \in \mathcal{A}$ *for all* a*;*
(d) $\{x : f(x) \ge a\} \in \mathcal{A}$ *for all* a*.*

Proof. The equivalence of (a) and (b) and of (c) and (d) follow from taking complements. The remaining equivalences follow from the equations
$$\{x : f(x) \ge a\} = \cap_{n=1}^\infty \{x : f(x) > a - 1/n\},$$
$$\{x : f(x) > a\} = \cup_{n=1}^\infty \{x : f(x) \ge a + 1/n\}.$$
$\square$

Proposition 4.2. *If X is a metric space, $\mathcal{A}$ contains all the open sets, and f is continuous, then f is measurable.*

Proof. $\{x : f(x) > a\} = f^{-1}(a, \infty)$ is open. $\square$

Proposition 4.3. *If f and g are measurable, so are $f + g$, cf, fg,* $\max(f, g)$*, and* $\min(f, g)$*.*

Proof. If $f(x)+g(x) < \alpha$, then $f(x) < \alpha - g(x)$, and there exists a rational r such that $f(x) < r < \alpha - g(x)$. So
$$\{x : f(x) + g(x) < \alpha\} = \bigcup_{r \text{ rational}} (\{x : f(x) < r\} \cap \{x : g(x) < \alpha - r\}).$$

f^2 is measurable since $\{x : f(x)^2 > a) = \{x : f(x) > \sqrt{a}\} \cup \{x : f(x) < -\sqrt{a}\}$. The measurability of fg follows since $fg = \frac{1}{2}[(f + g)^2 - f^2 - g^2]$.

$$\{x : \max(f(x), g(x)) > a\} = \{x : f(x) > a\} \cup \{x : g(x) > a\}. \qquad \square$$

Proposition 4.4. *If f_i is measurable for each i, then so is $\sup_i f_i$, $\inf_i f_i$, $\limsup_{i\to\infty} f_i$, and $\liminf_{i\to\infty} f_i$.*

Proof. The result will follow for lim sup and lim inf once we have the result for the sup and inf by using the definitions. We have $\{x : \sup_i f_i > a\} = \cap_{i=1}^{\infty}\{x : f_i(x) > a\}$, and the proof for $\inf f_i$ is similar. $\square$

Definition. We say $f = g$ *almost everywhere*, written $f = g$ a.e., if $\{x : f(x) \neq g(x)\}$ has measure zero. Similarly, we say $f_i \to f$ a.e., if the set of x where this fails has measure zero.

5. Integration. In this section we introduce the Lebesgue integral.

Definition. If $E \subset X$, define the *characteristic function* of E by

$$\chi_E(x) = \begin{cases} 1 & x \in E; \\ 0 & x \notin E. \end{cases}$$

A *simple function* s is one of the form

$$s(x) = \sum_{i=1}^{n} a_i \chi_{E_i}(x)$$

for reals a_i and sets E_i.

Proposition 5.1. *Suppose $f \geq 0$ is measurable. Then there exists a sequence of nonnegative measurable simple functions increasing to f.*

Proof. Let $E_{ni} = \{x : (i-1)/2^n \leq f(x) < i/2^n\}$ and $F_n = \{x : f(x) \geq n\}$ for $n = 1, 2, \ldots,$ and $i = 1, 2, \ldots, n2^n$. Then define

$$s_n = \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \chi_{E_{ni}} + n\chi_{F_n}.$$

It is easy to see that s_n has the desired properties. $\square$

Definition. If $s = \sum_{i=1}^n a_i \chi_{E_i}$ is a nonnegative measurable simple function, define the *Lebesgue integral* of s to be

$$\int s\, d\mu = \sum_{i=1}^{n} a_i \mu(E_i). \tag{5.1}$$

If $f \geq 0$ is measurable function, define

$$\int f\, d\mu = \sup\Big\{ \int s\, d\mu : 0 \leq s \leq f, s \text{ simple}\Big\}. \tag{5.2}$$

If f is measurable and at least one of the integrals $\int f^+\, d\mu$, $\int f^-\, d\mu$ is finite, where $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$, define

$$\int f\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu. \tag{5.3}$$

A few remarks are in order. A function s might be written as a simple function in more than one way. For example $\chi_{A\cup B} = \chi_A + \chi_B$ is A and B are disjoint. It is clear that the definition of $\int s\, d\mu$ is unaffected by how s is written. Secondly, if s is a simple function, one has to think a moment to verify that the definition of $\int s\, d\mu$ by means of (5.1) agrees with its definition by means of (5.2).

Definition. If $\int |f|\, d\mu < \infty$, we say f is *integrable*.

The proof of the next proposition follows from the definitions.

Proposition 5.2. *(a) If f is measurable, $a \le f(x) \le b$ for all x, and $\mu(X) < \infty$, then $a\mu(X) \le \int f\,d\mu \le b\mu(X)$;*

(b) If $f(x) \le g(x)$ for all x and f and g are measurable and integrable, then $\int f\,d\mu \le \int g\,d\mu$.

(c) If f is integrable, then $\int cf\,d\mu = c\int f\,d\mu$ for all real c.

(d) If $\mu(A) = 0$ and f is measurable, then $\int f\chi_A\,d\mu = 0$.

The integral $\int f\chi_A\,d\mu$ is often written $\int_A f\,d\mu$. Other notation for the integral is to omit the μ if it is clear which measure is being used, to write $\int f(x)\,\mu(dx)$, or to write $\int f(x)\,d\mu(x)$.

Proposition 5.3. *If f is integrable,*

$$\Big|\int f\Big| \le \int |f|.$$

Proof. $f \le |f|$, so $\int f \le \int |f|$. Also $-f \le |f|$, so $-\int f \le \int |f|$. Now combine these two facts. □

One of the most important results concerning Lebesgue integration is the monotone convergence theorem.

Theorem 5.4. *Suppose f_n is a sequence of nonnegative measurable functions with $f_1(x) \le f_2(x) \le \cdots$ for all x and with $\lim_{n\to\infty} f_n(x) = f(x)$ for all x. Then $\int f_n\,d\mu \to \int f\,d\mu$.*

Proof. By Proposition 5.2(b), $\int f_n$ is an increasing sequence of real numbers. Let L be the limit. Since $f_n \le f$ for all n, then $L \le \int f$. We must show $L \ge \int f$.

Let $s = \sum_{i=1}^m a_i\chi_{E_i}$ be any nonnegative simple function less than f and let $c \in (0,1)$. Let $A_n = \{x : f_n(x) \ge cs(x)\}$. Since the $f_n(x)$ increases to $f(x)$ for each x and $c < 1$, then $A_1 \subset A_2 \subset \cdots$, and the union of the A_n is all of X. For each n,

$$\begin{aligned}\int f_n \ge \int_{A_n} f_n &\ge c\int_{A_n} s_n\\ &= c\int_{A_n} \sum_{i=1}^m a_i\chi_{E_i}\\ &= c\sum_{i=1}^m a_i\mu(E_i\cap A_n).\end{aligned}$$

If we let $n\to\infty$, by Proposition 1.1(c), the right hand side converges to

$$c\sum_{i=1}^m a_i\mu(E_i) = c\int s.$$

Therefore $L \ge c\int s$. Since c is arbitrary in the interval $(0,1)$, then $L \ge \int s$. Taking the supremum over all simple $s \le f$, we obtain $L \ge \int f$. □

Once we have the monotone convergence theorem, we can prove that the Lebesgue integral is linear.

Theorem 5.5. *If f_1 and f_2 are integrable, then*

$$\int (f_1+f_2) = \int f_1 + \int f_2.$$

Proof. First suppose f_1 and f_2 are nonnegative and simple. Then it is clear from the definition that the theorem holds in this case. Next suppose f_1 and f_2 are nonnegative. Take s_n simple and increasing to f_1

and t_n simple and increasing to f_2. Then $s_n + t_n$ increases to $f_1 + f_2$, so the result follows from the monotone convergence theorem and the result for simple functions. Finally in the general case, write $f_1 = f_1^+ - f_1^-$ and similarly for f_2, and use the definitions and the result for nonnegative functions. $\square$

Suppose f_n are nonnegative measurable functions. We will frequently need the observation

$$\begin{aligned} \int \sum_{n=1}^\infty f_n &= \int \lim_{N\to\infty} \sum_{n=1}^N f_n = \lim_{N\to\infty} \int \sum_{n=1}^\infty f_n \\ &= \lim_{N\to\infty} \sum_{n=1}^N \int f_n = \sum_{n=1}^\infty \int f_n. \end{aligned} \tag{5.4}$$

We used here the monotone convergence theorem and the linearity of the integral.

The next theorem is known as Fatou's lemma.

Theorem 5.6. *Suppose the f_n are nonnegative and measurable. Then*

$$\int \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \int f_n.$$

Proof. Let $g_n = \inf_{i\ge n} f_i$. Then g_n are nonnegative and g_n increases to $\liminf f_n$. Clearly $g_n \le f_i$ for each $i \ge n$, so $\int g_n \le \int f_i$. Therefore

$$\int g_n \le \inf_{i\ge n} \int f_i.$$

If we take the supremum over n, on the left hand side we obtain $\int \liminf f_n$ by the monotone convergence theorem, while on the right hand side we obtain $\liminf_n \int f_n$. $\square$

A second very important theorem is the dominated convergence theorem.

Theorem 5.7. *Suppose f_n are measurable functions and $f_n(x) \to f(x)$. Suppose there exists an integrable function g such that $|f_n(x)| \le g(x)$ for all x. Then $\int f_n \, d\mu \to \int f \, d\mu$.*

Proof. Since $f_n + g \ge 0$, by Fatou's lemma,

$$\int (f+g) \le \liminf \int (f_n + g).$$

Since g is integrable,

$$\int f \le \liminf \int f_n.$$

Similarly, $g - f_n \ge 0$, so

$$\int (g - f) \le \liminf \int (g - f_n),$$

and hence

$$-\int f \le \liminf \int (-f_n) = -\limsup \int f_n.$$

Therefore

$$\int f \ge \limsup \int f_n,$$

which with the above proves the theorem. $\square$

Example. Suppose $f_n = n\chi_{(0,1/n)}$. Then $f_n \geq 0$, $f_n \to 0$ for each x, but $\int f_n = 1$ does not converge to $\int 0 = 0$. The trouble here is that the f_n do not increase for each x, nor is there a function g that dominates all the f_n simultaneously.

If in the monotone convergence theorem or dominated convergence theorem we have only $f_n(x) \to f(x)$ almost everywhere, the conclusion still holds. For if $A = \{x : f_n(x) \to f(x)\}$, then $f\chi_A \to f\chi_A$ for each x. And since A^c has measure 0, we see from Proposition 5.2(d) that $\int f\chi_A = \int f$, and similarly with f replaced by f_n.

Later on we will need the following two propositions.

Proposition 5.8. *Suppose f is measurable and for every measurable set A we have $\int_A f\,d\mu = 0$. Then $f = 0$ almost everywhere.*

Proof. Let $A = \{x : f(x) > \varepsilon\}$. Then

$$0 = \int_A f \geq \int_A \varepsilon = \varepsilon\mu(A)$$

since $f\chi_A \geq \varepsilon\chi_A$. Hence $\mu(A) = 0$. We use this argument for $\varepsilon = 1/n$ and $n = 1, 2, \ldots$, so $\mu\{x : f(x) > 0\} = 0$. Similarly $\mu\{x : f(x) < 0\} = 0$. $\square$

Proposition 5.9. *Suppose f is measurable and nonnegative and $\int f\,d\mu = 0$. Then $f = 0$ almost everywhere.*

Proof. If f is not almost everywhere equal to 0, there exists an n such that $\mu(A_n) > 0$ where $A_n = \{x : f(x) > 1/n\}$. But then since f is nonnegative,

$$\int f \geq \int_{A_n} f \geq \frac{1}{n}\mu(A_n),$$

a contradiction. $\square$

6. Product measures. If $A_1 \subset A_2 \subset \cdots$ and $A = \cup_{i=1}^\infty A_i$, we write $A_i \uparrow A$. If $A_1 \supset A_2 \supset \cdots$ and $A = \cap_{i=1}^\infty A_i$, we write $A_i \downarrow A$.

Definition. $\mathcal{M}$ is a *monotone class* is $\mathcal{M}$ is a collection of subsets of X such that
(a) if $A_i \uparrow A$ and each $A_i \in \mathcal{M}$, then $A \in \mathcal{M}$;
(b) if $A_i \downarrow A$ and each $A_i \in \mathcal{M}$, then $A \in \mathcal{M}$.

The intersection of monotone classes is a monotone class, and the intersection of all monotone classes containing a given collection of sets is the smallest monotone class containing that collection.

The next theorem, the monotone class lemma, is rather technical, but very useful.

Theorem 6.1. *Suppose $\mathcal{A}_0$ is a algebra, $\mathcal{A}$ is the smallest σ-algebra containing $\mathcal{A}_0$, and $\mathcal{M}$ is the smallest monotone class containing $\mathcal{A}_0$. Then $\mathcal{M} = \mathcal{A}$.*

Proof. A σ-algebra is clearly a monotone class, so $\mathcal{M} \subset \mathcal{A}$. We must show $\mathcal{A} \subset \mathcal{M}$.

Let $\mathcal{N}_1 = \{A \in \mathcal{M} : A^c \in \mathcal{M}\}$. Note $\mathcal{N}_1$ is contained in $\mathcal{M}$, contains $\mathcal{A}_0$, and is a monotone class. So $\mathcal{N}_1 = \mathcal{M}$, and therefore $\mathcal{M}$ is closed under the operation of taking complements.

Let $\mathcal{N}_2 = \{A \in \mathcal{M} : A \cap B \in \mathcal{M} \text{ for all } B \in \mathcal{A}_0\}$. $\mathcal{N}_2$ is contained in $\mathcal{M}$; $\mathcal{N}_2$ contains $\mathcal{A}_0$ because $\mathcal{A}_0$ is an algebra; $\mathcal{N}_2$ is a monotone class because $(\cup_{i=1}^\infty A_i) \cap B = \cup_{i=1}^\infty (A_i \cap B)$, and similarly for intersections. Therefore $\mathcal{N}_2 = \mathcal{M}$; in other words, if $B \in \mathcal{A}_0$ and $A \in \mathcal{M}$, then $A \cap B \in \mathcal{M}$.

Let $\mathcal{N}_3 = \{A \in \mathcal{M} : A \cap B \in \mathcal{M} \text{ for all } B \in \mathcal{M}\}$. As in the preceding paragraph, $\mathcal{N}_3$ is a monotone class contained in $\mathcal{M}$. By the last sentence of the preceding paragraph, $\mathcal{N}_3$ contains $\mathcal{A}_0$. Hence $\mathcal{N}_3 = \mathcal{M}$.

We thus have that $\mathcal{M}$ is a monotone class closed under the operations of taking complements and taking intersections. This shows $\mathcal{M}$ is a σ-algebra, and so $\mathcal{A} \subset \mathcal{M}$. □

Suppose $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are two measure spaces, i.e., $\mathcal{A}$ and $\mathcal{B}$ are σ-algebras on X and Y, resp., and μ and ν are measures on $\mathcal{A}$ and $\mathcal{B}$, resp. A *rectangle* is a set of the form $A \times B$, where $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Define a set function $\mu \times \nu$ on rectangles by

$$\mu \times \nu(A \times B) = \mu(A)\nu(B).$$

Lemma 6.2. *Suppose $A \times B = \cup_{i=1}^\infty A_i \times B_i$, where $A, A_i \in \mathcal{A}$ and $B, B_i \in \mathcal{B}$. Then*

$$\mu \times \nu(A \times B) = \sum_{i=1}^\infty \mu \times \nu(A_i \times B_i).$$

Proof. We have

$$\chi_{A \times B}(x, y) = \sum_{i=1}^\infty \chi_{A_i \times B_i}(x, y),$$

and so

$$\chi_A(x)\chi_B(y) = \sum_{i=1}^\infty \chi_{A_i}(x)\chi_{B_i}(y).$$

Holding x fixed and integrating over y with respect to ν, we have, using (5.4),

$$\chi_A(x)\nu(B) = \sum_{i=1}^\infty \chi_{A_i}(x)\nu(B_i).$$

Now use (5.4) again and integrate over x with respect to μ to obtain the result. □

Let $\mathcal{C}_0 = \{\text{finite unions of rectangles}\}$. It is clear that $\mathcal{C}_0$ is an algebra. By Lemma 6.2 and linearity, we see that $\mu \times \nu$ is a measure on $\mathcal{C}_0$. Let $\mathcal{A} \times \mathcal{B}$ be the smallest σ-algebra containing $\mathcal{C}_0$; this is called the *product σ-algebra*. By the Carathéodory extension theorem, $\mu \times \nu$ can be extended to a measure on $\mathcal{A} \times \mathcal{B}$.

We will need the following observation. Suppose a measure μ is σ-finite. So there exist E_i which have finite μ measure and whose union is X. If we let $F_n = \cup_{i=1}^n E_i$, then $F_i \uparrow X$ and $\mu(F_n)$ is finite for each n.

If μ and ν are both σ-finite, say with $F_i \uparrow X$ and $G_i \uparrow Y$, then $\mu \times \nu$ will be σ-finite, using the sets $F_i \times G_i$.

The main result of this section is Fubini's theorem, which allows one to interchange the order of integration.

Theorem 6.3. *Suppose $f : X \times Y \to \mathbb{R}$ is measurable with respect to $\mathcal{A} \times \mathcal{B}$. If f is nonnegative or $\int |f(x, y)|\, d(\mu \times \nu)(x, y) < \infty$, then*
(a) the function $g(x) = \int f(x, y)\nu(dy)$ is measurable with respect to $\mathcal{A}$;
(b) the function $h(y) = \int f(x, y)\mu(dx)$ is measurable with respect to $\mathcal{B}$;

(c) we have

$$\int f(x,y)\,d(\mu\times\nu)(x,y) = \int\Big(\int f(x,y)\,d\mu(x)\Big)\,d\nu(y)$$
$$= \int\Big(\int f(x,y)\,d\nu(y)\Big)\,\mu(dx).$$

Proof. First suppose μ and ν are finite measures. If f is the characteristic function of a rectangle, then (a)–(c) are obvious. By linearity, (a)–(c) hold if f is the characteristic function of a set in $\mathcal{C}_0$, the set of finite unions of rectangles.

Let $\mathcal{M}$ be the collection of sets C such that (a)–(c) hold for χ_C. If $C_i \uparrow C$ and $C_i \in \mathcal{M}$, then (c) holds for χ_C by monotone convergence. If $C_i \downarrow C$, then (c) holds for χ_C by dominated convergence. (a) and (b) are easy. So $\mathcal{M}$ is a monotone class containing $\mathcal{A}_0$, so $\mathcal{M} = \mathcal{A}\times\mathcal{B}$.

If μ and ν are σ-finite, applying monotone convergence to $C\cap(F_n\times G_n)$ for suitable F_n and G_n and monotone convergence, we see that (a)–(c) holds for the characteristic functions of sets in $\mathcal{A}\times\mathcal{B}$ in this case as well.

By linearity, (a)–(c) hold for nonnegative simple functions. By monotone convergence, (a)–(c) hold for nonnegative functions. In the case $\int|f|<\infty$, writing $f = f^+ - f^-$ and using linearity proves (a)–(c) for this case, too. □

7. The Radon-Nikodym theorem. Suppose f is nonnegative, measurable, and integrable with respect to μ. If we define ν by

$$\nu(A) = \int_A f\,d\mu,$$

then ν is a measure. The only part that needs thought is the countable additivity, and this follows from (5.4) applied to the functions $f\chi_{A_i}$. Moreover, $\nu(A)$ is zero whenever $\mu(A)$ is.

Definition. A measure ν is called *absolutely continuous* with respect to a measure μ if $\nu(A) = 0$ whenever $\mu(A) = 0$.

Definition. A function $\mu : \mathcal{A}\to(-\infty,\infty]$ is called a *signed measure* if $\mu(\emptyset) = 0$ and $\mu(\cup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i)$ whenever the A_i are disjoint and all the A_i are in $\mathcal{A}$.

Definition. Let μ be a signed measure. A set $A\in\mathcal{A}$ is called a *positive set* for μ if $\mu(B)\geq 0$ whenever $B\subset A$ and $A\in\mathcal{A}$. We define a *negative set* similarly.

Proposition 7.1. *Let μ be a signed measure and let $M>0$ such that $\mu(A)\geq -M$ for all $A\in\mathcal{A}$. If $\mu(F)<0$, then there exists a subset E of F that is a negative set with $\mu(E)<0$.*

Proof. Suppose $\mu(F)<0$. Let $F_1 = F$ and let $a_1 = \sup\{\mu(A) : A\subset F_1\}$. Since $\mu(F_1 - A) = \mu(F_1) - \mu(A)$ if $A\subset F_1$, we see that a_1 is finite. Let B_1 be a subset of F_1 such that $\mu(B_1)\geq a_1/2$. Let $F_2 = F_1 - B_1$, let $a_2 = \sup\{\mu(A) : A\subset F_2\}$, and choose B_2 a subset of F_2 such that $\mu(B_2)\geq a_2/2$. Let $F_3 = F_2 - B_2$ and continue.

One possibility is that this procedure stops after finitely many steps. This happens only if for some i every subset of F_i has nonpositive mass. In this case $E = F_i$ is the desired negative set.

The other possibility is if this procedure continues indefinitely. In this case, let $E = \cap_{i=1}^\infty F_i$. Note $E = F - (\cup_{i=1}^\infty B_i)$, and the B_i are disjoint. So

$$\mu(E) = \mu(F) - \sum_{i=1}^\infty \mu(B_i),$$

and $\mu(E) \le \mu(F) < 0$. Also

$$\sum_{i=1}^{\infty} \mu(B_i) = \mu(F) - \mu(E) \le M.$$

This implies the series converges, so $\mu(B_i) \to 0$. Since $\mu(B_i) \ge a_i/2$, then $a_i \to 0$. Suppose E is not a negative set. Then there exists $A \subset E$ with $\mu(A) > 0$. Choose n such that $a_n < \mu(A)$. But A is a subset of F_n, so $a_n \le \mu(A)$, a contradiction. Therefore E is a negative set. □

Proposition 7.2. *Let μ be a signed measure and $M > 0$ such that $\mu(A) \ge -M$ for all $A \in \mathcal{A}$. There exist sets E and F that are disjoint whose union is X and such that E is a negative set and F is a positive set.*

Proof. Let $L = \inf\{\mu(A) : A \text{ is a negative set}\}$. Choose negative sets A_n such that $\mu(A_n) \to L$. Let $E = \cup_{n=1}^{\infty} A_n$. Let $B_n = A_n - (B_1 \cup \cdots \cup B_{n-1})$ for each n. Since A_n is a negative set, so is each B_n. Also, the B_n are disjoint. If $C \subset E$, then

$$\mu(C) = \lim_{n\to\infty} \mu(C \cap (\cup_{i=1}^{n} B_i)) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(C \cap B_i) \le 0.$$

So E is a negative set.

Since E is negative,

$$\mu(E) = \mu(A_n) + \mu(E - A_n) \le \mu(A_n).$$

Letting $n \to \infty$, we obtain $\mu(E) = L$.

Let $F = E^c$. If F were not a positive set, there would exist $B \subset F$ with $\mu(B) < 0$. By Proposition 7.1 there exists a negative set C contained in B with $\mu(C) < 0$. But then $E \cup C$ would be a negative set with $\mu(E \cup C) < \mu(E) = L$, a contradiction. □

We now are ready for the Radon-Nikodym theorem.

Theorem 7.3. *Suppose μ is a σ-finite measure and ν is a finite measure such that ν is absolutely continuous with respect to μ. There exists a μ-integrable nonnegative function f such that $\nu(A) = \int_A f\, d\mu$ for all $A \in \mathcal{A}$. Moreover, if g is another such function, then $f = g$ almost everywhere.*

Proof. Let us first prove the uniqueness assertion. For every set A we have

$$\int_A (f-g)\, d\mu = \nu(A) - \nu(A) = 0.$$

By Proposition 5.8 we have $f - g = 0$ a.e.

Since μ is σ-finite, there exist $F_i \uparrow X$ such that $\mu(F_i) < \infty$ for each i. Let μ_i be the restriction of μ to F_i, that is, $\mu_i(A) = \mu(A \cap F_i)$. Define ν_i, the restriction of ν to F_i, similarly. If f_i is a function such that $\nu_i(A) = \int_A f_i\, d\mu_i$ for all A, the argument of the first paragraph shows that $f_i = f_j$ on F_i if $i \le j$. If we define f by $f(x) = f_i(x)$ if $x \in F_i$, we see that f will be the desired function. So it suffices to restrict attention to the case where μ is finite.

Let

$$\mathcal{F} = \Big\{ g : 0 \le g, \int_A g\, d\mu \le \nu(A) \text{ for all } A \in \mathcal{A} \Big\}.$$

$\mathcal{F}$ is not empty because $0 \in \mathcal{F}$. Let $L = \sup\{\int g\, d\mu : g \in \mathcal{F}\}$, and let g_n be a sequence in $\mathcal{F}$ such that $\int g_n\, d\mu \to L$. Let $h_n = \max(g_1, \ldots, g_n)$.

If g_1 and g_2 are in $\mathcal{F}$, then $h_2 = \max(g_1, g_2)$ is also in $\mathcal{F}$. To see this,

$$\begin{aligned}\int_A h_2\,d\mu &= \int_{A\cap\{x:g_1(x)\geq g_2(x)\}} h_2\,d\mu + \int_{A\cap\{x:g_1(x)<g_2(x)\}} h_2\,d\mu\\ &= \int_{A\cap\{x:g_1(x)\geq g_2(x)\}} g_1\,d\mu + \int_{A\cap\{x:g_1(x)<g_2(x)\}} g_2\,d\mu\\ &\leq \nu(A\cap\{x: g_1(x)\geq g_2(x)\}) + \nu(A\cap\{x: g_1(x)<g_2(x)\}) = \nu(A).\end{aligned}$$

By an induction argument, h_n is in $\mathcal{F}$.

The h_n increase, say to f. By the monotone convergence theorem, $\int f\,d\mu = L$ and

$$\int_A f\,d\mu \leq \nu(A) \tag{7.1}$$

for all A.

Let A be a set where there is strict inequality in (7.1); let ε be chosen sufficiently small so that if π is defined by

$$\pi(B) = \nu(B) - \int_B f\,d\mu - \varepsilon\mu(B),$$

then $\pi(A) > 0$. π is a signed measure; let F be the positive set as constructed in Proposition 7.2. In particular, $\pi(F) > 0$. So for every B

$$\int_{B\cap F} f\,d\mu + \varepsilon\mu(B\cap F) \leq \nu(B\cap F).$$

We then have, using (7.1), that

$$\begin{aligned}\int_B (f+\varepsilon\chi_F)\,d\mu &= \int_B f\,d\mu + \varepsilon\mu(B\cap F)\\ &= \int_{B\cap F^c} f\,d\mu + \int_{B\cap F} f\,d\mu + \varepsilon\mu(B\cap F)\\ &\leq \nu(B\cap F^c) + \nu(B\cap F) = \nu(B).\end{aligned}$$

This says that $f+\varepsilon\chi_F \in \mathcal{F}$. However,

$$L \geq \int (f+\varepsilon\chi_F)\,d\mu = \int f\,d\mu + \varepsilon\mu(F) = L + \varepsilon\mu(F),$$

which implies $\mu(F) = 0$. But then $\nu(F) = 0$, and hence $\pi(F) = 0$, contradicting the fact that F is a positive set for F with $\pi(F) > 0$. $\square$

8. Differentiation of real-valued functions.

Let $E \subset \mathbb{R}$ be a measurable set and let $\mathcal{O}$ be a collection of intervals. We say $\mathcal{O}$ is a *Vitali cover* of E if for each $x \in E$ and each $\varepsilon > 0$ there exists an interval $G \in \mathcal{O}$ containing x whose length is less than ε. m will denote Lebesgue measure.

Lemma 8.1. *Let E have finite measure and let $\mathcal{O}$ be a Vitali cover of E. Given $\varepsilon > 0$ there exists a finite subcollection of disjoint intervals $I_1, \ldots, I_n$ such that $m(E - \cup_{i=1}^n I_n) < \varepsilon$.*

Proof. We may replace each interval in $\mathcal{O}$ by a closed one, since the set of endpoints of a finite subcollection will have measure 0.

Let O be an open set of finite measure containing E. Since $\mathcal{O}$ is a Vitali cover, we may suppose without loss of generality that each set of $\mathcal{O}$ is contained in O. Let $a_1 = \sup\{m(I) : I \in \mathcal{O}\}$. Let I_1 be an element of $\mathcal{O}$ with $m(I_1) \geq a_1/2$. Let $a_2 = \sup\{m(I) : I \in \mathcal{O}, I \text{ disjoint from } I_1\}$,and choose $I_2 \in \mathcal{O}$ disjoint from I_1 such that $m(I_2) \geq a_2/2$. Continue in this way, choosing I_{n+1} disjoint from $I_1, \ldots, I_n$ and in $\mathcal{O}$ with length at least one half as large as any other such interval in $\mathcal{O}$ that is disjoint from $I_1, \ldots, I_n$.

If the process stops at some finite stage, we are done. If not, we generate a sequence of disjoint intervals $I_1, I_2, \ldots$ Since they are disjoint and all contained in O, then $\sum_{i=1}^{\infty} m(I_i) \leq m(O) < \infty$. So there exists N such that $\sum_{i=N+1}^{\infty} m(I_i) < \varepsilon/5$.

Let $R = E - \cup_{i=1}^{N} I_i$; we will show $m(R) < \varepsilon$. Let J_n be the interval with the same center as I_n but five times the length. Let $x \in R$. There exists an interval $I \in \mathcal{O}$ containing x with I disjoint from $I_1, \ldots, I_N$. Since $\sum m(I_n) < \infty$, then $\sum a_n \leq 2 \sum m(I_n) < \infty$, and $a_n \to 0$. So I must either be one of the I_n for some $n > N$ or at least intersect it, for otherwise we would have chosen I at some stage. Let n be the smallest integer such that I intersects I_n; note $n > N$. We have $m(I) \leq a_{n-1} \leq 2m(I_n)$. Since x is in I and I intersects I_n, the distance from x to the midpoint of I_n is at most $m(I) + m(I_n)/2 \leq (5/2)m(I_n)$. Therefore $x \in J_n$.

Then $R \subset \cup_{i=N+1}^{\infty} J_n$, so $m(R) \leq \sum_{i=N+1}^{\infty} m(J_n) = 5 \sum_{i=N+1}^{\infty} m(I_n) < \varepsilon$. $\square$

Given a function f, we define the *derivates* of f at x by

$$D^+ f(x) = \limsup_{h \to 0+} \frac{f(x+h) - f(x)}{h}, \qquad D^- f(x) = \limsup_{h \to 0-} \frac{f(x) - f(x-h)}{h}$$
$$D_+ f(x) = \liminf_{h \to 0+} \frac{f(x+h) - f(x)}{h}, \qquad D_- f(x) = \liminf_{h \to 0-} \frac{f(x) - f(x-h)}{h}.$$

If all the derivates are equal, we say that f is *differentiable* at x and define $f'(x)$ to be the common value.

Theorem 8.2. *Suppose f is nondecreasing on $[a, b]$. Then f is differentiable almost everywhere, f' is measurable, and $\int_a^b f'(x)\,dx \leq f(b) - f(a)$.*

Proof. We will show that the set where any two derivates are unequal has measure zero. We consider the set E where $D^+ f(x) > D_- f(X)$, the other sets being similar. Let $E_{u,v} = \{x : D^+ f(x) > u > v > D_- f(x)\}$. If we show $m(E_{u,v}) = 0$, then taking the union of all pairs of rationals with $u > v$ rational shows $m(E) = 0$.

Let $s = m(E_{u,v})$, let $\varepsilon > 0$, and choose an open set O such that $E_{u,v} \subset O$ and $m(O) < s + \varepsilon$. For each $x \in E_{u,v}$ there exists an arbitrarily small interval $[x - h, x]$ contained in O such that $f(x) - f(x - h) < vh$. Use Lemma 8.1 to choose $I_1, \ldots, I_n$ which are disjoint and whose interiors cover a subset of A of $E_{u,v}$ of measure greater than $s - \varepsilon$. Suppose $I_n = [x_n - h_n, x_n]$. Summing over these intervals,

$$\sum_{n=1}^{N} [f(x_n) - f(x_n - h_n)] < v \sum_{n=1}^{n} h_n < vm(O) < v(s + \varepsilon).$$

Each point $y \in A$ is the left endpoint of an arbitrarily small interval $(y, y + k)$ that is contained in some I_n and for which $f(y + k) - f(y) > u(k)$. Using Lemma 8.1 again, we pick out a finite collection $J_1, \ldots, J_M$ whose union contains a subset of A of measure larger than $s - 2\varepsilon$. Summing over these intervals yields

$$\sum_{i=1}^{M} [f(y_i + k_i) - f(y_i)] > u \sum k_i > u(s - 2\varepsilon).$$

Each interval J_i is contained in some interval I_n, and if we sum over those i for which $J_i \subset I_n$ we find

$$\sum [f(y_i + k_i) - f(y_i)] \leq f(x_n) - f(x_n - h_n),$$

since f is increasing. Thus

$$\sum_{n=1}^{N}[f(x_n) - f(x_n - h_n)] \geq \sum_{i=1}^{M}[f(y_i + k_i) - f(y_i)],$$

and so $v(s+\varepsilon) > u(s - 2\varepsilon)$. This is true for each ε, so $vs \geq us$. Since $u > v$, this implies $s = 0$.

This shows that

$$g(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

is defined almost everywhere and that f is differentiable wherever g is finite. Define $f(x) = f(b)$ if $x \geq b$. Let $g_n(x) = n[f(x + 1/n) - f(x)]$. Then $g_n(x) \to g(x)$ for almost all x, and so g is measurable. Since f is increasing, $g_n \geq 0$. By Fatou's lemma

$$\begin{aligned}\int_a^b g \leq \liminf \int_a^b g_n &= \liminf n \int_a^b [f(x+1/n) - f(x)]dx \\ &= \liminf \Big[n\int_b^{b+1/n} f - n\int_a^{a+1/n} f\Big] = \liminf \Big[f(b) - n\int_a^{a+1/n} f \Big] \\ &\leq f(b) - f(a).\end{aligned}$$

This shows that g is integrable and hence finite almost everywhere. □

A function is of *bounded variation* if $\sup\{\sum_{i=1}^{k} |f(x_i) - f(x_{i-1})|\}$ is finite, where the supremum is over all partitions $a = x_0 < x_1 < \cdots < x_k = b$ of $[a, b]$.

Lemma 8.3. *If f is of bounded variation on $[a, b]$, then f can be written as the difference of two nondecreasing functions on $[a, b]$.*

Proof. Define

$$P(y) = \sup\Big\{\sum_{i=1}^{k}[f(x_i) - f(x_{i-1})]^+\Big\}, \qquad N(y) = \sup\Big\{\sum_{i=1}^{k}[f(x_i) - f(x_{i-1})]^-\Big\},$$

where the supremum is over all partitions $a = x_0 < x_1 < \cdots < x_k = y$ for $y \in [a, b]$. Since

$$\sum_{i=1}^{k}[f(x_i) - f(x_{i-1})]^+ = \sum_{i=1}^{k}[f(x_i) - f(x_{i-1})]^- + f(y) - f(a),$$

taking the supremum over all partitions of $[a, y]$ yields

$$P(y) = N(y) + f(y) - f(a).$$

Clearly P and N are nondecreasing in y, and the result follows by solving for $f(y)$. □

Define the *indefinite integral* of an integrable function f by

$$F(x) = \int_a^x f(t)\, dt.$$

Lemma 8.4. *If f is integrable, then F is continuous and of bounded variation.*

Proof. The continuity follows from the dominated convergence theorem The bounded variation follows from

$$\sum_{i=1}^{k} |F(x_i) - F(x_{i-1})| = \sum_{i=1}^{k}\Big|\int_{x_{i-1}}^{x_i} f(t)\, dt\Big| \leq \sum_{i=1}^{k}\int_{x_{i-1}}^{x_i} |f(t)|\, dt \leq \int_a^b |f(t)|\, dt$$

for all partitions. $\square$

Lemma 8.5. *If f is integrable and $F(x) = 0$ for all x, then $f = 0$ a.e.*

Proof. For any interval, $\int_c^d f = \int_a^d f - \int_a^c f = 0$. By dominated convergence and the fact that any open set is the countable union of disjoint open intervals, $\int_O f = 0$ for any open set O.

If E is any measurable set, take O_n open that such that χ_{O_n} decreases to χ_E a.e. By dominated convergence,

$$\int_E f = \int f\chi_E = \lim \int f\chi_{O_n} = \lim \int_{O_n} f = 0.$$

This with Proposition 5.8 implies f is zero a.e. $\square$

Proposition 8.6. *If f is bounded and measurable, then $F'(x) = f(x)$ for almost every x.*

Proof. By Lemma 8.4, F is of bounded variation, and so F' exists a.e. Let K be a bound for $|f|$. If

$$f_n(x) = \frac{F(x + 1/n) - F(x)}{1/n},$$

then

$$f_n(x) = n \int_x^{x+1/n} f(t)\,dt,$$

so $|f_n|$ is also bounded by K. Since $f_n \to F'$ a.e., then by dominated convergence,

$$\begin{aligned}\int_a^c F'(x)\,dx &= \lim \int_a^c f_n(x)\,dx = \lim \int_a^c [F(x + 1/n) - F(x)]\,dx \\ &= \lim n \int_c^{c+1/n} F(x)\,dx - n \int_a^{a+c} F(x)\,dx = F(c) - F(a) = \int_a^c f(x)\,dx,\end{aligned}$$

using the fact that F is continuous. So $\int_a^c [F'(x) - f(x)]\,dx = 0$ for all c, which implies $F' = f$ a.e. by Lemma 8.5. $\square$

Theorem 8.7. *If f is integrable, then $F' = f$ almost everywhere.*

Proof. Without loss of generality we may assume $f \geq 0$. Let $f_n(x) = f(x)$ if $f(x) \leq n$ and let $f_n(x) = n$ if $f(x) > n$. Then $f - f_n \geq 0$. If $G_n(x) = \int_a^x [f - f_n]$, then G_n is nondecreasing, and hence has a derivative almost everywhere. By Lemma 8.6, we know the derivative of $\int_a^x f_n$ is equal to f_n almost everywhere. Therefore

$$F'(x) = G_n'(x) + \Big[\int_a^x f_n\Big]' \geq f_n(x)$$

a.e. Since n is arbitrary, $F' \geq f$ a.e. So $\int_a^b F' \geq \int_a^b f = F(b) - F(a)$. On the other hand, by Theorem 8.2, $\int_a^b F'(x)\,dx \leq F(b) - F(a) = \int_a^b f$. We conclude that $\int_a^b [F' - f] = 0$; since $F' - f \geq 0$, this tells us that $F' = f$ a.e. $\square$

A function is *absolutely continuous* on $[a, b]$ if given ε there exists δ such that $\sum_{i=1}^k |f(x_i') - f(x_i)| < \varepsilon$ whenever $\{x_i, x_i')\}$ is a finite collection of nonoverlapping intervals with $\sum_{i=1}^k |x_i' - x_i| < \delta$.

Lemma 8.8. *If $F(x) = \int_a^x f(t)\,dt$ for f integrable on $[a,b]$, then F is absolutely continuous.*

Proof. Let $\varepsilon > 0$. Choose a simple function s such that $\int_a^b |f - s| < \varepsilon/2$. Let K be a bound for $|s|$ and let $\delta = \varepsilon/2K$. If $\{(x_i, x_i')\}$ is a collection of nonoverlapping intervals, the sum of whose lengths is less than δ, then set $A = \cup_{i=1}^k (x_i, x_i')$ and note $\int_A |f - s| < \varepsilon/2$ and $\int_A s < K\delta = \varepsilon/2$. $\square$

Lemma 8.9. *If f is absolutely continuous, then it is of bounded variation.*

Proof. Let δ correspond to $\varepsilon = 1$ in the definition of absolute continuity. Given a partition, add points if necessary so that each subinterval has length at most δ. We can then group the subintervals into at most K collections, each of total length less than δ, where K is an integer larger than $(1 + b - a)/\delta$. So the total variation is then less than K. $\square$

Lemma 8.10. *If f is absolutely continuous on $[a,b]$ and $f'(x) = 0$ a.e., then f is constant.*

Proof. Let $c \in [a,b]$, let $E = \{x \in [a,c] : f'(x) = 0\}$, and let $\varepsilon > 0$. For each point $x \in E$ there exists arbitrarily small intervals $[x, x+h] \subset [a,c]$ such that $|f(x+h) - f(x)| < \varepsilon h$. By Lemma 8.1 we can find a finite collection of such intervals that cover all of E except for a set of measure less than δ, where δ is the δ in the definition of absolute continuity. If the intervals are $[x_i, y_i]$ with $x_i < y_i \le x_{i+1}$, then $\sum |f(x_{i+1}) - f(y_i)| < \varepsilon$ by the definition of absolute continuity, while $\sum |f(y_i) - f(x_i)| < \varepsilon \sum (y_i - x_i) \le \varepsilon(c - a)$. So adding these two inequalities together,

$$|f(c) - f(a)| = \Big|\sum [f(x_{i+1}) - f(y_i)] + \sum [f(y_i) - f(x_i)]\Big| \le \varepsilon + \varepsilon(c - a).$$

Since ε is arbitrary, then $f(c) = f(a)$, which implies that f is constant. $\square$

Theorem 8.11. *F is an indefinite integral if and only if it is absolutely continuous.*

Proof. One direction was Lemma 8.11. Suppose F is absolutely continuous on $[a,b]$. Then F is of bounded variation, $F = F_1 - F_2$ where F_1 and F_2 are nondecreasing, and F' exists a.e. Since $|F'(x)| \le F_1'(x) + F_2'(x)$, then $\int |F'(x)|\,dx \le F_1(b) + F_2(b) - F_1(a) - F_2(a)$, then F' is integrable. If $G(x) = \int_a^x F'(t)\,dt$, then G is absolutely continuous by Lemma 8.11, so $F - G$ is absolutely continuous. Then $(F - G)' = 0$ a.e., and therefore $F - G$ is constant. Thus $F(x) = \int_a^x F'(t)\,dt + F(a)$. $\square$

9. L^p spaces.

For $1 \le p < \infty$, define the L^p norm of f by

$$\|f\|_p = \Big(\int |f(x)|^p d\mu \Big)^{1/p}.$$

For $p = \infty$, define the L^∞ norm of f by

$$\|f\|_\infty = \inf\{M : \mu(\{x : |f(x)| \ge M\}) = 0\}.$$

For $1 \le p \le \infty$ the space L^p is the set $\{f : \|f\|_p < \infty\}$.

The L^∞ norm of a function f is the supremum of f provided we disregard sets of measure 0.

It is clear that $\|f\|_p = 0$ if and only if $f = 0$ a.e.

Proposition 9.1. (Hölder's inequality) *If* $1 < p, q < \infty$ *and* $p^{-1} + q^{-1} = 1$, *then*

$$\int f(x)g(x)d\mu \le \|f\|_p \|g\|_q.$$

This also holds if $p = \infty$ *and* $g = 1$.

Proof. If $M = \|f\|_\infty$, then $\int fg \le M \int |g|$ and the case $p = \infty$ and $q = 1$ follows. So let us assume $1 < p, q < \infty$. If $\|f\|_p = 0$, then $f = 0$ a.e and $\int fg = 0$, so the result is clear if $\|f\|_p = 0$ and similarly if $\|g\|_q = 0$. Let $F(x) = |f(x)|/\|f\|_p$ and $G(x) = |g(x)|/\|g\|_q$. Note $\|F\|_p = 1$ and $\|G\|_q = 1$, and it suffices to show that $\int FG \le 1$.

The second derivative of the function e^x is again e^x, which is positive, and so e^x is convex. Therefore if $0 \le \lambda \le 1$, we have

$$e^{\lambda a + (1-\lambda)b} \le \lambda e^a + (1-\lambda)e^b.$$

If $F(x), G(x) \ne 0$, let $a = p \log F(x)$, $b = q \log G(x)$, $\lambda = 1/p$, and $1 - \lambda = 1/q$. We then obtain

$$F(x)G(x) \le \frac{F(x)^p}{p} + \frac{G(x)^q}{q}.$$

Clearly this inequality also holds if $F(x) = 0$ or $G(x) = 0$. Integrating,

$$\int FG \le \frac{\|F\|_p^p}{p} + \frac{\|G\|_q^q}{q} = \frac{1}{p} + \frac{1}{q} = 1.$$

$\square$

One application of Hölder's inequality is to prove Minkowski's inequality, which is simply the triangle inequality for L^p.

Proposition 9.2. (Minkowski's inequality) *If* $1 \le p \le \infty$, *then*

$$\|f + g\|_p \le \|f\|_p + \|g\|_p.$$

Proof. Since $|(f+g)(x)| \le |f(x)| + |g(x)|$, integrating gives the case when $p = 1$. The case $p = \infty$ is also easy. So let us suppose $1 < p < \infty$. If $\|f\|_p$ or $\|g\|_p$ is infinite, the result is obvious, so we may assume both are finite. The inequality $(a+b)^p \le 2^p a^p + 2^p b^p$ with $a = |f(x)|$ and $b = |g(x)|$ yields, after an integration,

$$\int |(f+g)(x)|^p d\mu \le 2^p \int |f(x)|^p d\mu + 2^p \int |g(x)|^p d\mu.$$

So we have $\|f + g\|_p < \infty$. Clearly we may assume $\|f + g\|_p > 0$.

Now write

$$|f + g|^p \le |f|\,|f + g|^{p-1} + |g|\,|f + g|^{p-1}$$

and apply Hölder's inequality with $q = (1 - \frac{1}{p})^{-1}$. We obtain

$$\int |f + g|^p \le \|f\|_p \Big(\int |f + g|^{(p-1)q} \Big)^{1/q} + \|g\|_p \Big(\int |f + g|^{(p-1)q} \Big)^{1/q}.$$

Since $p^{-1} + q^{-1} = 1$, then $(p-1)q = p$, so we have

$$\|f + g\|_p^p \le \Big(\|f\|_p + \|g\|_p \Big) \|f + g\|_p^{p/q}.$$

Dividing both sides by $\|f + g\|_p^{p/q}$ and using the fact that $p - (p/q) = 1$ gives us our result. $\square$

Minkowski's inequality says that L^p is a normed linear space, provided we identify functions that are equal a.e. The next proposition says that L^p is complete. This is often phrased as saying that L^p is a Banach space, i.e., a complete normed linear space.

Before proving this we need two easy preliminary results. The first is sometimes called Chebyshev's inequality.

Lemma 9.3. *If $1 \le p < \infty$,*

$$\mu(\{x : |f(x)| \ge a\}) \le \frac{\|f\|_p^p}{a^p}.$$

Proof. If $A = \{x : |f(x)| \ge a\}$, then

$$\mu(A) \le \int_A \frac{|f(x)|^p}{a^p} d\mu \le \frac{1}{a^p} \int |f|^p d\mu.$$

$\square$

The next lemma is sometimes called the Borel-Cantelli lemma.

Lemma 9.4. *If $\sum \mu(A_j) < \infty$, then*

$$\mu(\cap_{j=1}^\infty \cup_{m=j}^\infty A_m) = 0.$$

Proof.

$$\mu(\cap_{j=1}^\infty \cup_{m=j}^\infty A_m) = \lim_{j\to\infty} \mu(\cup_{m=j}^\infty A_m) \le \lim_{j\to\infty} \sum_{m=j}^\infty \mu(A_m) = 0.$$

$\square$

Proposition 9.5. *If $1 \le p \le \infty$, then L^p is complete.*

Proof. We do only the case $p < \infty$; the case $p = \infty$ is easy. Suppose f_n is a Cauchy sequence in L^p. Given $\varepsilon = 2^{-(j+1)}$, there exists n_j such that if $n, m \ge n_j$, then $\|f_n - f_m\|_p \le 2^{-(j+1)}$. Without loss of generality we may assume $n_j \ge n_{j-1}$ for each j.

Set $n_0 = 0$ and define $f_0 \equiv 0$. If $A_j = \{x : |f_{n_j}(x) - f_{n_{j-1}}(x)| > 2^{-j/2}$, then from Lemma 9.3, $\mu(A_j) \le 2^{-jp/2}$. By Lemma 9.4, $\mu(\cap_{j=1}^\infty \cup_{m=j}^\infty A_m) = 0$. So except for a set of measure 0, for each x there is a last j for which $x \in \cup_{m=j}^\infty A_m$, hence a last j for which $x \in A_j$. So for each x (except for the null set) there is a j_0 (depending on x) such that if $j \ge j_0$, then $|f_{n_j}(x) - f_{n_{j-1}}(x)| \le 2^{-j}$.

Set

$$g_j(x) = \sum_{m=1}^\infty |f_{n_m}(x) - f_{n_{m-1}}(x)|.$$

$g_j(x)$ increases for each x, and the limit is finite for almost every $x$by the preceding paragraph. Let us call the limit $g(x)$. We have

$$\|g_j\|_p \le \sum_{m=1}^j 2^{-j} + \|f_{n_1}\|_p \le 2 + \|f_{n_1}\|_p$$

by Minkowski's inequality, and so by Fatou's lemma, $\|g\|_p \le 2 + \|f_{n_1}\|_p < \infty$. We have

$$f_{n_j}(x) = \sum_{m=1}^j (f_{n_m}(x) - f_{n_{m-1}}(x)).$$

Suppose x is not in the null set where $g(x)$ is infinite. Since $|f_{n_j}(x) - f_{n_k}(x)| \le |g_{n_j}(x) - g_{n_k}(x)| \to 0$ as $j, k \to \infty$, then $f_{n_j}(x)$ is a Cauchy series (in $\mathbb{R}$), and hence converges, say to $f(x)$. We have $\|f - f_{n_j}\|_p = \lim_{m\to\infty} \|f_{n_m} - f_{n_j}\|_p$; this follows by dominated convergence with the function g defined above as the dominating function.

We have thus shown that $\|f - f_{n_j}\|_p \to 0$. Given $\varepsilon = 2^{-(j+1)}$, if $m \ge n_j$, then $\|f - f_m\|_p \le \|f - f_{n_j}\|_p + \|f_m - f_{n_j}\|_p$. This shows that f_m converges to f in L^p norm. $\square$

The following is very useful.

Proposition 9.6. *For $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$,*

$$\|f\|_p = \sup\Big\{ \int fg : \|g\|_q \le 1 \Big\}. \tag{9.1}$$

When $p = 1$ (9.1) holds if we take $q = \infty$, and if $p = \infty$ (9.1) holds if we take $q = 1$.

Proof. The right hand side of (9.1) is less than the left hand side by Hölder's inequality. So we need only show that the right hand side is greater than the left hand side.

First suppose $p = 1$. Take $g(x) = \operatorname{sgn} f(x)$, where $\operatorname{sgn} a$ is 1 if $a > 0$, is 0 if $a = 0$, and is -1 if $a < 0$. Then g is bounded by 1 and $fg = |f|$. This takes care of the case $p = 1$.

Next suppose $p = \infty$. Since μ is σ-finite, there exist sets F_n increasing up to X such that $\mu(F_n) < \infty$ for each n. If $M = \|f\|_\infty$, let a be any finite real less than M. By the definition of L^∞ norm, the measure of $A = \{x \in F_n : |f(x)| > a\}$ must be positive if n is sufficiently large. Let $g(x) = (\operatorname{sgn} f(x))\chi_A(x)/\mu(A)$. Then the L^1 norm of g is 1 and $\int fg = \int_A |f|/\mu(A) \ge a$. Since a is arbitrary, the supremum on the right hand side must be M.

Now suppose $1 < p < \infty$. We may suppose $\|f\|_p > 0$. Let q_n be a sequence of nonnegative simple functions increasing to f^+, r_n a sequence of nonnegative simple functions increasing to f^-, and $s_n(x) = (q_n(x) - r_n(x))\chi_{F_n}(x)$. Then $s_n(x) \to f(x)$ for each x, $|s_n(x)| \le |f(x)|$ for each x, s_n is a simple function, and $\|s_n\|_p < \infty$ for each n. If $f \in L^p$, then $\|s_n\|_p \to \|f\|_p$ by dominated convergence. If $\int |f|^p = \infty$, then $\int |s_n|^p \to \infty$ by monotone convergence. For n sufficiently large, $\|s_n\|_p > 0$.

Let

$$g_n(x) = (\operatorname{sgn} f(x)) \frac{|s_n(x)|^{p-1}}{\|s_n\|_p^{p/q}}.$$

Since $(p-1)q = p$, then

$$\|g_n\|_q = \frac{(\int |s_n|^{(p-1)q})^{1/q})}{\|s_n\|_p^{p/q}} = \frac{\|s_n\|_p^{p/q}}{\|s_n\|_p^{p/q}} = 1.$$

On the other hand, since $|f| \ge |s_n|$,

$$\int f g_n = \frac{\int |f|\,|s_n|^{p-1}}{\|s_n\|_p^{p/q}} \ge \frac{\int |s_n|^p}{\|s_n\|_p^{p/q}} = \|s_n\|_p^{p-(p/q)}.$$

Since $p - (p/q) = 1$, then $\int f g_n \ge \|s_n\|_p$, which tends to $\|f\|_p$. □

The above proof also establishes

Corollary 9.7. *For $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$,*

$$\|f\|_p = \sup\{ \int fg : \|g\|_q \le 1, g \text{ simple}\}.$$

The space L^p is a normed linear space. We can thus talk about its dual, namely, the set of bounded linear functionals on L^p. The dual of a space Y is denoted Y^*. If H is a bounded linear functional on L^p, we define the norm of H to be $\|H\| = \sup\{H(f) : \|f\|_p \le 1\}$.

Theorem 9.8. *If $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$, then $(L^p)^* = L^q$.*

Proof. If $g \in L^q$, then setting $H(f) = \int fg$ for $f \in L^p$ yields a bounded linear functional; the boundedness follows from Hölder's inequality. Moreover, from Hölder's inequality and Proposition 9.6 we see that $\|H\| = \|g\|_q$.

Now suppose we are given a bounded linear functional H on L^p and we must show there exists $g \in L^q$ such that $H(f) = \int fg$. First suppose $\mu(X) < \infty$. Define $\nu(A) = H(\chi_A)$. If A and B are disjoint, then

$$\nu(A \cup B) = H(\chi_{A\cup B}) = H(\chi_A + \chi_B) = H(\chi_A) + H(\chi_B) = \nu(A) + \nu(B).$$

To show ν is countably additive, it suffices to show that if $A_n \uparrow A$, then $\nu(A_n) \to \nu(A)$. But if $A_n \uparrow A$, then $\chi_{A_n} \to \chi_A$ in L^p, and so $\nu(A_n) = H(\chi_{A_n}) \to H(\chi_A) = \nu(A)$; we use here the fact that $\mu(X) < \infty$. Therefore ν is a countably additive signed measure. Moreover, if $\mu(A) = 0$, then $\chi_A = 0$ a.e., hence $\nu(A) = H(\chi_A) = 0$. By writing $\nu = \nu^+ - \nu^-$ and using the Radon-Nikodym theorem for both the positive and negative parts, we see there exists an integrable g such that $\nu(A) = \int_A g$ for all sets A. If $s = \sum a_i \chi_{A_i}$ is a simple function, by linearity we have

$$H(s) = \sum a_i H(\chi_{A_i}) = \sum a_i \nu(A_i) = \sum a_i \int g\chi_{A_i} = \int gs.$$

By Corollary 9.7,

$$\|g\|_q = \sup\Big\{ \int gs : \|s\|_p \le 1, s \text{ simple}\Big\} \le \sup\{H(s) : \|s\|_p \le 1\} \le \|H\|.$$

If s_n are simple functions tending to f in L^p, then $H(s_n) \to H(f)$, while by Hölder's inequality $\int s_n g \to \int fg$. We thus have $H(f) = \int fg$ for all $f \in L^p$, and $\|g\|_p \le \|H\|$. By Hólder's inequality, $\|H\| \le \|g\|_p$.

In the case where μ is σ-finite, but not finite, let $F_n \uparrow X$ be such that $\mu(F_n) < \infty$ for each n. Define functionals H_n by $H_n(f) = H(f\chi_{F_n})$. Clearly each H_n is a bounded linear functional on L^p. Applying the above argument, we see there exist g_n such that $H_n(f) = \int fg_n$ and $\|g_n\|_q = \|H_n\| \le \|H\|$. It is easy to see that g_n is 0 if $x \notin F_n$. Moreover, by the uniqueness part of the Radon-Nikodym theorem, if $n > m$, then $g_n = g_m$ on F_m. Define g by setting $g(x) = g_n(x)$ if $x \in F_n$. Then g is well defined. By Fatou's lemma, g is in L^q with a norm bounded by $\|H\|$. Since $f\chi_{F_n} \to f$ in L^p by dominated convergence, then $H_n(f) = H(f\chi_{F_n}) \to H(f)$, since H is a bounded linear functional on L^p. On the other hand $H_n(f) = \int_{F_n} fg_n = \int_{F_n} fg \to \int fg$ by dominated convergence. So $H(f) = \int fg$. Again by Hölder's inequality $\|H\| \le \|g\|_p$. $\square$

References.

1. G.B. Folland, *Real analysis: modern techniques and their applications*, New York, Wiley, 1984.

2. H.L. Royden, *Real analysis*, New York, Macmillan, 1963.

3. W. Rudin, *Real and complex analysis*, New York, McGraw-Hill, 1966.