G1CMIN Measure and Integration 2003-4

Prof. J.K. Langley

May 13, 2004

1 Introduction

Books: W. Rudin, *Real and Complex Analysis* ; H.L. Royden, *Real Analysis* (QA331).

Lecturer: Prof. J.K. Langley (jkl@maths, room C6c, 95 14964).
Lectures: Wed 11 C29; Th 1 C4; Th 5 C29.
There will be NO lectures in the week commencing Feb 9th.

Office hours: none specified but my door is open most of the time and you're free to consult me.

Assessment: 2.5 hour written examination. 5 questions, best 4 count.

PLEASE NOTE: G1CMIN is a "traditional" pure mathematics theory module along the familiar "definition, theorem, proof" structure; in particular there isn't much scope for calculations and the module is more like G12RAN or G13MTS than G12CAN.

Contents: review of real analysis, Lebesgue measure, Lebesgue integration.

Outline: there are two main themes to the module. The idea of measure is concerned with the size of sets. The first distinction we meet between sets is usually between finite and infinite sets. We then refine the idea of an infinite set to distinguish between the countable and the uncountable (we'll review this concept). When we discuss the Lebesgue measure of a subset of $\mathbb{R}$, it will give us an indication of how much of the line is filled up by our set. Thus we will be able to distinguish "big" uncountable sets from smaller ones.

The second main theme involves integration. G12RAN introduces the Riemann integral, which has advantages in that it is relatively easy to define, and displays well the link between integration and differentiation. Its main drawbacks are: (i) the class of Riemann integrable functions is too small; (ii) it has technical problems, particularly with regard to whether

$$\lim_{n\to\infty}\left(\int f_n\right) = \int\left(\lim_{n\to\infty} f_n\right). \tag{1}$$

Also, Riemann's integral is difficult to generalize to other settings.

Lebesgue measure gives a means for comparing the size of sets and leads to Lebesgue integration, which is used widely in pure maths, probability, mathematical physics, PDEs etc. This module will go as far as the main theorems concerning when (1) holds. More advanced topics will only be covered if time permits.

Aims and Learning Outcomes:

Aims: to teach the elements of measure theory and Lebesgue integration.

Learning Outcomes: a successful student will:

1. be able to state, and apply in the investigation of examples, the principal theorems as treated;
2. be able to prove simple propositions concerning sets, measure spaces, Lebesgue measure and integrable functions.

Coursework: is not part of the assessment for G1CMIN, but the questions should be helpful practice for the exam. There will be a *short* assignment each week, for handing in the week after.

For anyone taking this module for G13ES1 (Supplementary Maths), the assessment will consist of a **separate** coursework assignment handed out by the end of Week 6 of the semester. It will be due for handing in on the last day of the Spring term and will be based on material from the first half of the module. You need only attend enough lectures to cover this material.

Web notes: you can find printed notes at

www.maths.nott.ac.uk/personal/jkl/min03.pdf

The lectures will, however, cover all of the material (except for some proofs given in previous modules such as G1ALIM) so you may prefer to use the Web notes either not at all or just as backup. These notes may contain errors, omissions or obscure parts: these will be amended as and when I find them.

2 Sets

2.1 Countability

This is an important idea when deciding how "big" infinite sets are compared to each other. We shall see that $\mathbb{R}$ is a "bigger" set than $\mathbb{Q}$. We say that a set A is **countable** if either A is empty or there is a sequence $(a_n), n = 1, 2, 3, \ldots$, which "uses up" A, by which we mean that each $a_n \in A$ and each member of A appears at least once in the sequence. This is the same as saying that there is a surjective (onto) function $f : \mathbb{N} \to A$ (via $a_n = f(n)$).

FACT 1: Any finite set is countable.
Indeed, if $A = \{x_1, \ldots, x_N\}$, just put $a_n = x_n$ if $n \leq N$, and $a_n = x_N$ if $n > N$.

FACT 2: If $B \subseteq A$ and A is countable, then B is countable.
If B is finite, this is obvious. If B is not finite, then nor is A, so take a sequence which uses up A, and delete all entries in the sequence which don't belong to B. We then get an infinite sequence which uses up B.

FACT 3: Suppose that A is an infinite, countable set. Then there is a sequence $(b_n), n = 1, 2, \ldots$, of members of A in which each member of of A appears *exactly once*.

To see this, suppose that $(a_n), n = 1, 2, \ldots$ uses up A. Go through the list, deleting any entry which has previously occurred. So if $a_n = a_j$ for some $j < n$, we delete a_n. The resulting subsequence includes each member of A exactly once. We have thus arranged A into a sequence - first element, second element etc. - hence the name "countable" .

FACT 4: Suppose that $A_1, A_2, A_3, \ldots$ are countably many countable sets. Then the union $U = \bigcup_{n=1}^{\infty} A_n$, which is the set of all x which each belong to at least one A_n, is countable.

Proof: delete any A_j which are empty, and re-label the rest. Now suppose that the j'th set A_j is used up by the sequence $(a_{j,n}), n = 1, 2, \ldots$. Write out these sequences as follows:

$$\begin{array}{llll} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \ldots\ldots\ldots\ldots\ldots \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \ldots\ldots\ldots\ldots\ldots \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \ldots\ldots\ldots\ldots\ldots \end{array}$$

etc. Now the following sequence uses up all of U. We take

$$a_{1,1} \quad a_{1,2} \quad a_{2,1} \quad a_{1,3} \quad a_{2,2} \quad a_{3,1} \quad a_{1,4} \quad a_{2,3} \quad \ldots$$

FACT 5: The set of positive rational numbers is countable. The reason is that this set is the union of the sets $A_m = \{p/m : p \in \mathbb{N}\}$, each of which is countable. Similarly, the set of negative rational numbers is countable, and so is $\mathbb{Q}$ (the union of these two sets and $\{0\}$).

FACT 6: If A and B are countable sets, then so is the Cartesian product $A \times B$, which is the set of all ordered pairs (a, b), with $a \in A$ and $B \in B$.

Here "ordered" means that $(a, b) \neq (b, a)$ unless $a = b$.

Fact 6 is obvious if A or B is empty. Otherwise, if (a_n) uses up A and (b_n) uses up B, then $A \times B$ is the union of the sets $C_n = \{(a_n, b_m) : m = 1, 2, 3, \ldots\}$, each of which is countable.

FACT 7: the interval $(0, 1)$ is not countable, and therefore nor are $\mathbb{R}, \mathbb{C}$.

Proof: We prove the following stronger assertion. Consider the collection T of all real numbers $x = 0 \cdot d_1 d_2 d_3 d_4 \ldots\ldots$ in which each digit d_j is either 4 or 5. Then T is uncountable.

Suppose that the sequence $(a_n), n = 1, 2, \ldots$, uses up T. Write out each a_j as a decimal expansion

$$a_1 = 0 \cdot b_{1,1} b_{1,2} b_{1,3} \ldots\ldots$$

$$a_2 = 0 \cdot b_{2,1} b_{2,2} b_{2,3} \ldots\ldots$$

$$a_3 = 0 \cdot b_{3,1} b_{3,2} b_{3,3} \ldots\ldots$$

etc. Here each digit $b_{j,k}$ is 4 or 5. We make a new number $x = 0 \cdot c_1 c_2 c_3 c_4 \ldots$ as follows.

We look at $b_{n,n}$. If $b_{n,n} = 4$, we put $c_n = 5$, while if $b_{n,n} = 5$, we put $c_n = 4$. Now x cannot belong to the list above, for if we had $x = a_m$, then we'd have $c_m = b_{m,m}$, which isn't true.

Example

Let S be the collection of all sequences $a_1, a_2, \ldots$ with each entry a positive integer. Then S is uncountable: take the subset of S for which each a_j is 4 or 5 and use the above proof.

2.2 The real numbers $\mathbb{R}$

The key idea about $\mathbb{R}$ which we need is the existence of least upper bounds.

Let E be a non-empty subset of $\mathbb{R}$. We say that a real number M is an upper bound for E if $x \leq M$ for all x in E, and E is called bounded above. Among all upper bounds for E there is one which is the least, called the sup or l.u.b. of E.

We adopt the convention that if E is a subset of $\mathbb{R}$ which is not bounded above, then the sup of E is $+\infty$.

For example $\sup((0, 1) \cap \mathbb{Q}) = 1$ and $\sup \mathbb{N} = \infty$.

The greatest lower bound of E (denoted glb or inf) is defined similarly: it is the greatest real number which is less than or equal to every member of E. If E is not bounded below the glb is $-\infty$.

2.3 Real sequences

Let $(x_n), n = 1, 2, \ldots$ be a sequence of real numbers.

We say $\lim_{n\to\infty} x_n = L \in \mathbb{R}$ if to each positive real number ε corresponds an integer n_0 such that $|x_n - L| < \varepsilon$ for all $n \geq n_0$.

We say $\lim_{n\to\infty} x_n = +\infty$ if to each positive real number M corresponds an integer n_0 such that $x_n > M$ for all $n \geq n_0$.

We say $\lim_{n\to\infty} x_n = -\infty$ if $\lim_{n\to\infty}(-x_n) = +\infty$. Thus every negative real M has n_0 with $x_n < M$ for all $n \geq n_0$.

2.4 The monotone sequence theorem

If the real sequence (x_n) is non-decreasing (i.e. $x_n \leq x_{n+1}$) for $n \geq N$ then x_n tends to a limit (finite or $+\infty$).

We recall the proof. Let s be the supremum of the set $\{x_n : n \geq N\} = A$. Suppose first that A is not bounded above, so that $\sup A = +\infty$ by our convention. This means that if we are given some positive number M, then no matter how large M might be, we can find some member of the set A, say x_{n_1}, such that $x_{n_1} > M$. But then, because the sequence is non-decreasing, we have $x_n > M$ for all $n \geq n_1$, and this is precisely what we need in order to be able to say that $\lim_{n\to\infty} x_n = +\infty$.

Now suppose that A is bounded above, and let s be the sup. Suppose we are given some positive ε. Then we need to show that $|x_n - s| < \varepsilon$ for all sufficiently large n. But we know that $x_n \leq s$ for all n, so we just have to show that $x_n > s - \varepsilon$ for all large enough n.

This we do as follows. The number $s - \varepsilon$ is less than s and so is not an upper bound for A, and so there must be some n_2 such that $x_{n_2} > s - \varepsilon$. But then $x_n > s - \varepsilon$ for all $n \geq n_2$, and the proof is complete.

2.5 Lemma

Every real sequence $(x_n), n = 1, 2, \ldots,$ has a monotone subsequence.

Proof: For each n consider the set $E_n = \{x_m : m \geq n\}$. We look at two cases.

Suppose first that for every n the set E_n has a maximum element i.e. there is some $m \geq n$ such that $x_p \le x_m$ for all $p \ge n$.

To form our subsequence choose $n_1 \ge 1$ such that x_{n_1} is the maximum element of E_1. Then choose $n_2 \ge n_1 + 1$ such that x_{n_2} is the maximum element of E_{n_1+1}. Since $E_{n_1+1} \subseteq E_1$ we get $x_{n_2} \le x_{n_1}$. Now take the maximum element of E_{n_2+1}: this will be x_{n_3} for some $n_3 > n_2$ and we have $x_{n_3} \le x_{n_2}$. Repeating this we get a non-increasing subsequence (x_{n_k}).

Suppose now that some E_N has no maximum element. Let $n_1 = N$ and choose $n_2 > n_1$ such that $x_{n_2} > x_{n_1}$. Since E_N has no maximum element, there is an element greater than all of $x_{n_1}, x_{n_1+1}, \ldots, x_{n_2}$, and this is x_{n_3} for some $n_3 > n_2$. Carrying on in this way, we get a strictly increasing subsequence.

2.6 Corollary (Bolzano-Weierstrass theorem)

Every bounded real sequence has a bounded monotone subsequence and hence a convergent subsequence.

2.7 Nested intervals

Let $I_k = [a_k, b_k]$ be closed intervals in $\mathbb{R}$ such that $I_{k+1} \subseteq I_k$. Then $a_k \le a_{k+1} \le b_{k+1} \le b_1$ and $a_1 \le a_{k+1} \le b_{k+1} \le b_k$ so a_k converges, to A say, and b_k converges, to B say. We have $a_k \le A \le B \le b_k$ for all k, so $[A, B]$ is contained in all of the I_k. Thus the intersection of the I_k is non-empty.

The example $I_n = (0, 1/n)$ shows that open intervals do not in general have this property.

2.8 Open sets

A subset U of $\mathbb{R}$ is called open if the following is true. To each x in U corresponds $\delta_x > 0$ such that $(x - \delta_x, x + \delta_x) \subseteq U$. It is easy to check that if W_t is open for each t in some set T then the set $\bigcup_{t \in T} W_t$, which is the set of all y each belonging to at least one W_t, is open. Also the intersection of finitely many open sets is open.

2.9 Open intervals

By an open interval in $\mathbb{R}$ we mean any of the following: $(a, b), (a, +\infty), (-\infty, b), \mathbb{R}$. All are open sets.

A subset E of $\mathbb{R}$ is called closed if $\mathbb{R} \setminus E$ is open. Obviously a closed interval $[a, b]$ (where $-\infty < a \le b < \infty$) is closed, since the complement is $(-\infty, a) \cup (b, \infty)$, a union of two open sets.

Since every (non-empty) open interval contains a rational number and an irrational number, neither $\mathbb{Q}$ nor $\mathbb{R} \setminus \mathbb{Q}$ is open.

2.10 Lemma

Let $x \in \mathbb{R}$. For each t in some non-empty set T let W_t be an open interval containing x. Then $V = \bigcup_{t \in T} W_t$ is an open interval containing x.

Proof.
Let A be the inf of V and B the sup. We claim first that A and B are not in V. If B is in V then B is in some W_t. Since W_t is an open subset of $\mathbb{R}$ there is a $b > B$ in W_t, and b is in V, which is a contradiction. Thus A and B are not in V and clearly V is a subset of (A, B).

We claim that (A, B) is a subset of V. Let $y \in (A, B)$. Obviously if $y = x$ then y is in V. Suppose that $x < y < B$. Then (since B is the sup of V) there is some w with $x < y < w$ such that w lies in some W_s. Since x and w lie in the open interval W_s, so does y, and y is in V. The same proof works if $A < y < x$.

2.11 Theorem

Let V be a non-empty open subset of $\mathbb{R}$. Then V is the union of countably many pairwise disjoint open intervals.

Proof: let $x \in V$, and let C_x be the union of all open intervals W such that $x \in W \subseteq V$. Then C_x is well-defined (since V is open there is at least one such W) and open, and by Lemma 2.10 this C_x is an open interval.

We claim that if y is in C_x then $C_x = C_y$. To see this, note that C_x and C_y are both open intervals containing y and contained in V, and so is $C_x \cup C_y$, by Lemma 2.10. Since $x \in C_x \cup C_y$ we get $C_x \cup C_y \subseteq C_x$, as C_x is the union of all open intervals containing x and contained in V. Since $y \in C_x \cup C_y$, the same argument gives $C_x \cup C_y \subseteq C_y$.

It follows that if C_x and C_y have non-empty intersection, with t belonging to both, then both equal C_t and so $C_x = C_y$.

Now let d_n be a sequence using up all the rational numbers in V, and put $D_n = C_{d_n}$. The set of D_n is countable. Since each C_x contains a rational number, every C_x, for x in V, is one of the D_n. So V is the union of the D_n. Now we go along our sequence D_n, and delete any D_n which is equal to one which has occurred previously.

2.12 Lemma (Special case of the Heine-Borel theorem)

Let $I = [a, b]$ be a closed interval, and suppose that we have open subsets $V_1, V_2, V_3, \ldots$ of $\mathbb{R}$ such that I is contained in the union of the V_j. Then I is contained in the union of finitely many *of the V_j.*

We remark that this is related to the concept of compactness (G13MTS), but we do not need this concept here.

To prove the lemma, suppose the conclusion is false. Then we can find $x_1 \in I \setminus V_1, x_2 \in I \setminus (V_1 \cup V_2)$, and, in general, $x_n \in I \setminus (V_1 \cup \ldots \cup V_n)$. Since $a \leq x_n \leq b$, the sequence (x_n) is bounded, and so has a convergent subsequence x_{n_k}, with limit $\alpha \in [a, b]$. But then α is is some V_N, and there is some $u > 0$ with $(\alpha - u, \alpha + u) \subseteq V_N$, since V_N is open. Since $x_{n_k} \to \alpha$, we see that for large k we have $x_{n_k} \in (\alpha - u, \alpha + u) \subseteq V_N$. But $n_k \to \infty$ so that for large k we have $n_k > N$ and $x_{n_k} \notin V_N \subseteq V_1 \cup \ldots \cup V_{n_k}$. This contradiction proves the lemma.

Again, this is a property of closed intervals $[a, b]$ not shared by other intervals. For example $(0, 1]$ is contained in the union of the open intervals $(1/n, 2), n \in \mathbb{N}$, and $[0, \infty)$ is contained in the union of the open intervals $(-1, n), n \in \mathbb{N}$. In neither case will finitely many of those intervals suffice to cover the set.

3 Continuous functions

3.1 Basic facts

Let E be a subset of $\mathbb{R}$ and let $f : E \to \mathbb{R}$ be a function. We say f is continuous on E if the following is true. To each x_0 in E and each real $\varepsilon > 0$ corresponds a real $\delta > 0$ such that $|f(x) - f(x_0)| < \varepsilon$ for all x in E with $|x - x_0| < \delta$.

Fact 1: if t_n is a sequence in E converging to x_0 then $f(t_n)$ converges to $f(x_0)$. To see this, if $\varepsilon > 0$ take $\delta > 0$ as above. We have some n_0 such that $|t_n - x_0| < \delta$ for all $n \geq n_0$, giving $|f(t_n) - f(x_0)| < \varepsilon$ for all $n \geq n_0$.

Fact 2: if E is a closed interval $[a, b]$ and $f : E \to \mathbb{R}$ is continuous then f has a maximum and minimum on $[a, b]$ and in particular is bounded.

To see this, let M be the supremum of the set $f(E) = \{f(x) : x \in E\}$. Take a strictly increasing sequence (y_n) (thus $y_n < y_{n+1}$) with limit M. No y_n is an upper bound for $f(E)$, so we can find s_n in $f(E)$ with $y_n < s_n \leq M$. Thus s_n tends to M. So there exist t_n in E such that $f(t_n) \to M$. By the Bolzano-Weierstrass theorem we can assume WLOG that the sequence (t_n), being bounded, converges, to x_0 say, and x_0 is in the interval $[a, b]$, since $a \leq t_n \leq b$. Thus $f(x_0) = M$. Hence M is in $f(E)$ (and M is the max of $f(E)$). In particular M is finite.

3.2 Pointwise convergence

Let E be a subset of $\mathbb{R}$ and let $f_n, n \in \mathbb{N}$ and f be functions from E to $\mathbb{R}$. We say that f_n converges pointwise to f on E if for each x in E,

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

Thus for each $\varepsilon > 0$ and for each x in E, there is an integer $N(x)$ such that $|f_n(x) - f(x)| < \varepsilon$ for all $n \geq N(x)$.

If the f_n are continuous, does it follow that f is continuous? The answer is no.

Example
Let $g_n(x)$ be defined on $[0, 1]$ for $n \in \mathbb{N}$ by $g_n(x) = 1 - nx$ for $0 \leq x \leq 1/n$ and $g_n(x) = 0$ for $x > 1/n$. Then g_n is continuous on $[0, 1]$. Set $g(x)$ to be 1 for $x = 0$ and 0 otherwise. Then g is not continuous, but $g_n \to g$ pointwise on [0, 1].

Notice here also that

$$\lim_{n \to +\infty} (\lim_{x \to 0+} g_n(x)) = 1 \neq \lim_{x \to 0+} (\lim_{n \to +\infty} g_n(x)) = 0.$$

A second example displaying the same phenomenon comes from $h_n(x) = e^{-nx}$ on $[0, 1]$. Then again $h_n \to g$ pointwise on $[0, 1]$.

So we need a stronger condition which will force the limit function to be continuous. The idea is to make $N(x)$ independent of x.

3.3 Uniform convergence

If f_n and f are functions from E to $\mathbb{R}$ we say that f_n converges uniformly to f on E if the following is true. To each real $\varepsilon > 0$ corresponds an integer N such that $|f_n(x) - f(x)| < \varepsilon$ for all $n \geq N$ and for all $x \in E$.

The example g_n above does not converge uniformly to g on $[0, 1]$. To see this, take $\varepsilon = 1/4$ and $x = 1/2n$. Then $g(x) = 0$ but $g_n(x) = 1/2$. No matter how large we take N, we can choose $n \geq N$ with $|g_n(1/2n) - g(1/2n)| > \varepsilon$.

3.4 Theorem

If the real-valued functions f_n are continuous on E and converge uniformly to $f : E \to \mathbb{R}$ then f is continuous on E.

Proof: take x_0 in E and $\varepsilon > 0$. Take N so large that $|f_N(x) - f(x)| < \varepsilon/3$ for all x in E. Take $\delta > 0$ so that $|f_N(x) - f_N(x_0)| < \varepsilon/3$ for all x in E with $|x - x_0| < \delta$. For such x we get

$$|f(x) - f(x_0)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(x_0)| + |f_N(x_0) - f(x_0)| < 3\varepsilon/3.$$

4 Riemann Integration

The Riemann integral will be defined for continuous and some other functions. It is relatively easy to define and use, and displays the interplay between integration and differentiation well, but it has certain disadvantages.

4.1 Basic definitions for the Riemann integral

Let f be a bounded real-valued function on the closed interval $[a, b] = I$. Henceforth $a < b$ unless otherwise explicitly stated. Assume that $|f(x)| \leq M$ for all x in I.

A PARTITION P of I is a finite set $x_0, \ldots, x_n$ such that $a = x_0 < x_1 < \ldots < x_n = b$. The points x_j are called the vertices of P. We say that partition Q of I is a refinement of partition P of I if every vertex of P is a vertex of Q (i.e. P is a subset of Q). For P as above, we define

$$M_k(f) = \sup\{f(x) : x_{k-1} \leq x \leq x_k\} \leq M, \quad m_k(f) = \inf\{f(x) : x_{k-1} \leq x \leq x_k\} \geq -M.$$

Next, we define the UPPER SUM $U(P, f)$ and LOWER SUM $L(P, f)$ by

$$U(P, f) = \sum_{k=1}^{n} M_k(f)(x_k - x_{k-1}), \quad L(P, f) = \sum_{k=1}^{n} m_k(f)(x_k - x_{k-1}).$$

Notice that $L(P, f) \leq U(P, f)$. The reason we require f to be bounded is so that all the m_k and M_k are finite and the sums exist. Notice also that $-M \leq m_k \leq M_k \leq M$ for each k, and so

$$-M(b - a) \leq L(P, f) \leq U(P, f) \leq M(b - a).$$

Suppose that f is positive on I and that the area A under the curve exists. It is not hard to see that $L(P, f) \leq A \leq U(P, f)$ for every partition P of I.

Further, if you draw for yourself a simple curve, it is not hard to convince yourself that refining P tends to increase $L(P, f)$ and decrease $U(P, f)$. We prove this last statement as a lemma.

4.2 Lemma

Let f be a bounded real-valued function on $I = [a, b]$.

(i) If P, Q are partitions of I and Q is a refinement of P, then

$$L(P, f) \leq L(Q, f), \quad U(P, f) \geq U(Q, f).$$

(ii) If P_1 and P_2 are any partitions of I, then $L(P_1, f) \leq U(P_2, f)$. Thus any lower sum is $\leq$ any upper sum.

Proof:
(i) We first prove this for the case where Q is P plus one extra point. The general case then follows by adding points one at a time. So suppose that Q is the same as P, except that it has one extra vertex c, where $x_{k-1} < c < x_k$. Then $U(Q, f) - U(P, f)$ is

$$(\sup\{f(x) : x_{k-1} \leq x \leq c\})(c - x_{k-1}) + (\sup\{f(x) : c \leq x \leq x_k\})(x_k - c)$$

$$-(\sup\{f(x) : x_{k-1} \leq x \leq x_k\})(x_k - x_{k-1}).$$

This is using the fact that all other terms cancel. But the first two sups above are less than or equal to the third. Since $c-x_{k-1}, x_k-c, x_k-x_{k-1}$ are all positive, we get $U(Q,f)-U(P,f) \leq 0$. The proof for the lower sums uses the same idea, or can be proved by noting that $L(P,f) = -U(P,-f)$.

(ii) Here we just set P to be the partition obtained by taking all the vertices of P_1 and all those of P_2. We arrange these vertices in order, and P is a refinement of P_1 and of P_2. Now we can write

$$L(P_1, f) \leq L(P, f) \leq U(P, f) \leq U(P_2, f).$$

4.3 Definition of the Riemann integral

Let f be bounded, real-valued on $I = [a, b]$ as before, with $|f(x)| \leq M$ there. We define the UPPER INTEGRAL of f from a to b as

$$\overline{\int_a^b} f(x)dx = \inf_P \{U(P, f)\}$$

where the supremum is taken over *all* partitions P of I. This exists and is finite, because all the upper sums are bounded below by $-M(b-a)$. Similarly we define the LOWER INTEGRAL

$$\underline{\int_a^b} f(x)dx = \sup_P \{L(P, f)\}$$

taking the sup over all partitions P of I. Again this exists and is finite, because all the lower sums are bounded above by $M(b-a)$.

Now we define f to be Riemann integrable on I if the upper integral equals the lower integral, in which case we denote the common value by $\int_a^b f(x)dx$.

Notice that the lower integral is always $\leq$ the upper integral, because of Lemma 4.2, part (ii). Also, if f is Riemann integrable and positive on I and the area A under the curve exists, then the fact that $L(P, f) \leq A \leq U(P, f)$ for every partition P of I implies that the lower integral is $\leq A$ and the upper integral is $\geq A$, which means that A equals $\int_a^b f(x)dx$. As usual in integration, it does not matter whether you write $f(x)dx$ or $f(t)dt$ etc.

4.4 Example

Define f on $I = [0, 1]$ by $f(x) = 1$ if x is rational and $f(x) = 0$ otherwise. Let $P = \{x_0, \ldots, x_n\}$ be any partition of I. Then clearly $M_k(f) = 1$ for each k, since in each sub-interval $[x_{k-1}, x_k]$ there is a rational number. Thus $U(P, f) = \sum_{k=1}^n (x_k - x_{k-1}) = 1$ and so the upper integral is 1. Similarly, we have $m_k(f) = 0$ for each k, all lower sums are 0, and the lower integral is 0.

Before proving that continuous functions are Riemann integrable, we first deal with the rather easier case of monotone functions (non-decreasing or non-increasing).

4.5 Theorem

Suppose that f is a monotone function on $I = [a, b]$. Then f is Riemann integrable on I.

Proof: we only deal with the case where f is non-decreasing (i.e. $f(x) \leq f(y)$ for $x \leq y$). The non-increasing case is similar. Now if $f(b) = f(a)$ then f is constant on I and so the result follows trivially (all upper and lower sums are the same).

Assume henceforth that $f(b) > f(a)$. Let $\varepsilon > 0$. We choose a partition $P = \{x_0, \ldots, x_n\}$ such that for each k we have $x_k - x_{k-1} < \varepsilon/(f(b) - f(a))$. Now, since f is non-decreasing we have $M_k(f) = f(x_k)$ and $m_k(f) = f(x_{k-1})$. Thus

$$U(P, f) - L(P, f) = \sum_{k=1}^{n} (M_k(f) - m_k(f))(x_k - x_{k-1}) = \sum_{k=1}^{n} (f(x_k) - f(x_{k-1}))(x_k - x_{k-1}) <$$

$$< \sum_{k=1}^{n} (f(x_k) - f(x_{k-1}))\varepsilon/(f(b) - f(a)) = (f(x_n) - f(x_0))\varepsilon/(f(b) - f(a)) = \varepsilon.$$

Therefore $U(P, f) \leq L(P, f) + \varepsilon$. So the upper integral of f (which is the inf of the upper sums) is at most $L(P, f) + \varepsilon$. But $L(P, f)$ is at most the lower integral (sup of the lower sums). Thus the upper and lower integrals differ by at most ε and, since ε is arbitrary, must be equal.

To handle the case of continuous functions, we need the following.

4.6 Uniform continuity

Let f be a real-valued function on the closed interval $I = [a, b]$. We say that f is uniformly continuous on I if the following is true. To each $\varepsilon > 0$ corresponds a $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ for all x and y in I such that $|x - y| < \delta$.

4.7 Theorem

If f is continuous on $[a, b]$ then f is uniformly continuous on $[a, b]$.

Proof: suppose that $\varepsilon > 0$ and that NO positive δ exists with the property in the statement. Then $1/n$, for $n \in \mathbb{N}$, is not such a δ. Thus there are points x_n and y_n in I with $|x_n - y_n| < 1/n$, but with $|f(x_n) - f(y_n)| \geq \varepsilon$.

Now (x_n) is a sequence in the closed interval I, and so is a bounded sequence, and therefore we can find a convergent subsequence (x_{k_n}), with limit B, say. Since $a \leq x_{k_n} \leq b$ for each n, we have $B \in I$. Now $|x_{k_n} - y_{k_n}| \to 0$ as $n \to \infty$, and so (y_{k_n}) also converges to B. Since f is continuous on I, we have $f(x_{k_n}) \to f(B)$ as $n \to \infty$ and $f(y_{k_n}) \to f(B)$ as $n \to \infty$, which contradicts the fact that $|f(x_{k_n}) - f(y_{k_n})|$ is always $\geq \varepsilon$. This contradiction proves the theorem.

Remark: the name UNIFORM continuity arises because the δ does not depend on the particular choice of x or y. The theorem is NOT true for open intervals, as the example $h(x) = 1/x$, $I = (0, 1)$ shows. To see this, just note that $h(1/n) - h(1/(n-1)) = 1$ for all $n \in \mathbb{N}$, but $|1/n - 1/(n-1)| = 1/n(n-1)$, which we can make as small as we like.

4.8 Theorem

Let f be continuous, real-valued, on $I = [a, b]$ $(a < b)$. Then f is Riemann integrable on I.

Proof: let $\varepsilon > 0$ be given. We choose a $\delta > 0$ such that for all x and y in I with $|x - y| < \delta$ we have $|f(x) - f(y)| < \varepsilon/(b-a)$. We choose a partition $P = \{x_0, \ldots, x_n\}$ of I such that, for each k, we have $x_k - x_{k-1} < \delta$. Now take a sub-interval $J = [x_{k-1}, x_k]$. We know that there exist c and d in J such that for all x in J we have $f(c) \leq f(x) \leq f(d)$. This means that $M_k(f) = f(d)$ and $m_k(f) = f(c)$. But $|c - d| < \delta$ and so $M_k(f) - m_k(f) = f(d) - f(c) < \varepsilon/(b-a)$. This holds for each k. Thus

$$U(P, f) - L(P, f) = \sum_{k=1}^{n} (M_k(f) - m_k(f))(x_k - x_{k-1}) < (\varepsilon/(b-a)) \sum_{k=1}^{n} (x_k - x_{k-1}) = \varepsilon.$$

The same argument as used for non-decreasing functions now applies.

4.9 Theorem

Let f and g be Riemann integrable functions on I. Let c, d be real numbers. Then $cf + dg$ is Riemann integrable on I and $\int_a^b cf(x) + dg(x)dx = c\int_a^b f(x)dx + d\int_a^b g(x)dx$.

Proof:
First consider cf, when $c > 0$. Obviously $L(P, cf) = cL(P, f)$ and $U(P, cf) = cU(P, f)$ and so $\int_a^b cf(x)dx = c\int_a^b f(x)dx$. It is also easy to see that $U(P, -f) = -L(P, f)$ and $L(P, -f) = -U(P, f)$, so $\int_a^b -f(x)dx = -\int_a^b f(x)dx$. This leaves only $f + g$ to consider. Take $\varepsilon > 0$ and partitions P_1, P_2 of I such that $L(P_1, f)$ and $U(P_2, f)$ are both within ε of $\int_a^b f(x)dx$. Note that we also have $L(P_1, f) \leq \int_a^b f(x)dx \leq U(P_2, f)$. We can assume that $P_1 = P_2$, because otherwise we can replace both by $P_1 \cup P_2$, and refinements can only push the lower and upper sums closer to the integral. Similarly, take Q such that $L(Q, g)$ and $U(Q, g)$ are both within ε of $\int_a^b g(x)dx$.

We can assume that $P = Q$, because otherwise we can replace both by $P \cup Q$. Now

$$L(P, f) + L(P, g) \leq L(P, f + g) \leq U(P, f + g) \leq U(P, f) + U(P, g).$$

Thus $L(P, f+g)$ and $U(P, f+g)$ both lie within 2ε of $\int_a^b f(x)dx + \int_a^b g(x)dx$. Therefore so do the upper and lower integrals of $f + g$ and, since ε is arbitrary, the result follows.

4.10 The fundamental theorem of the calculus

Suppose that F, f are real-valued functions on $[a, b]$, that F is continuous and f is Riemann integrable on $[a, b]$, and that $F'(x) = f(x)$ for all x in (a, b). Then $\int_a^b f(x)dx = F(b) - F(a)$.

Proof.
The proof is based on the mean value theorem. Let $P = \{x_0, \ldots, x_n\}$ be any partition of $[a, b]$. Then by the mean value theorem there exist points t_k satisfying $x_{k-1} < t_k < x_k$ such that

$$F(b) - F(a) = \sum_{k=1}^{n} (F(x_k) - F(x_{k-1})) = \sum_{k=1}^{n} f(t_k)(x_k - x_{k-1}).$$

But this means that $L(P, f) \leq F(b) - F(a) \leq U(P, f)$. Hence the lower integral of f is at most $F(b) - F(a)$, and the upper integral of f is at least $F(b) - F(a)$. But the lower and upper integrals of f are, by assumption, the same.

4.11 Examples

(i) The limit of a sequence of Riemann integrable functions need not be Riemann integrable. Let E be the countable set $\mathbb{Q} \cap [0, 1]$, and let (r_n) be a sequence in E, in which each element of E appears exactly once. Define g_n as follows. Set $g_N(x) = 1$ if x is one of the N points $r_1, \ldots, r_N$, and $g_N(x) = 0$ otherwise. Obviously all lower sums for g_N are 0. Take the partition $P = \{x_0, \ldots, x_n\}$ with $x_k = k/n, 0 \leq k \leq n$, and $n > 2N$ an integer. Then at most $2N$ intervals $[x_{k-1}, x_k]$ contain a point where $g_N(x) \neq 0$, so $U(P, g_N) \leq 2N/n$. So $\int_0^1 g_N(x)dx = 0$. But as $N \to \infty$ we see that g_N converges pointwise to the function of Example 4.4, which is not Riemann integrable.

(ii) Define h_n on $[0, 1]$, for integer $n > 2$, by $h_n(x) = n^2 x$ if $0 \leq x \leq 1/n$ and $h_n(x) = n^2(2/n - x)$ if $1/n \leq x \leq 2/n$ and $h_n(x) = 0$ if $x \geq 2/n$. Then h_n is continuous on $[0, 1]$ with Riemann integral 1. But $h_n \to 0$ pointwise on [0, 1], and 0 has Riemann integral 0.

5 Series

5.1 The extended real numbers

We define $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, \infty\}$. Later we will define *some* products involving ∞, but for now we just define

$$\infty + \infty = \infty, x + \infty = \infty$$

for all $x \in \mathbb{R}$. Note that $\infty + (-\infty)$ is not defined.

We can extend $<$ to $\mathbb{R}^*$ in the obvious way by saying that $-\infty < x < \infty$ for every x in $\mathbb{R}$.

Note that if we say a subset A of $\mathbb{R}$ is bounded above this will continue to mean that there is some $M \in \mathbb{R}$ such that $x \leq M$ for all $x \in A$: of course all x are also $\leq \infty$.

If A is any non-empty subset of $\mathbb{R}^*$, then A has a least upper bound (sup) and an inf. Note that $\sup A$ is ∞ if A has no upper bound in $\mathbb{R}$. In particular this is true if $\infty \in A$.

We can define limits of sequences in $\mathbb{R}^*$ exactly as in $\mathbb{R}$. In particular $x_n \to L \in \mathbb{R}$ means that given positive real ε we have $|x_n - L| < \varepsilon$ (and so $x_n \in \mathbb{R}$) for all $n \geq n_0(\varepsilon)$.

The monotone sequence theorem remains true: if x_n is a non-decreasing sequence in $\mathbb{R}^*$ then x_n tends to $\sup\{x_n\}$.

5.2 Series

We consider here only series with terms a_k in $[0, \infty]$. Given such a_k for $k \geq p \in \mathbb{N}$, the partial sums

$$s_n = \sum_{k=p}^{n} a_k$$

form a non-decreasing sequence in $\mathbb{R}^*$, and we set

$$S = \sum_{k=p}^{\infty} a_k = \lim_{n \to \infty} s_n.$$

This will be ∞ if any a_k is ∞, but of course

$$\sum_{k=1}^{\infty} 1/k = \infty.$$

Note that $s_n \leq S$ for each n.

5.3 Re-arrangements

Suppose that $a_k \in [0, \infty]$ for each $k \in \mathbb{N}$. Suppose that $\phi : \mathbb{N} \to \mathbb{N}$ is a bijection, and set $b_k = a_{\phi(k)}$ for each k. Then

$$S_1 = \sum_{k=1}^{\infty} b_k = a_{\phi(1)} + a_{\phi(2)} + \dots\dots\dots$$

is called a RE-ARRANGEMENT of

$$S_2 = \sum_{k=1}^{\infty} a_k,$$

and the sums S_1, S_2 are equal.

To see this, just note that if $n \in \mathbb{N}$, then we can find m such that $\phi(j) \leq m$ for $1 \leq j \leq n$, and so

$$\sum_{k=1}^{n} b_k = \sum_{k=1}^{n} a_{\phi(k)} \leq \sum_{p=1}^{m} a_p \leq S_2.$$

Letting $n \to \infty$ we get $S_1 \leq S_2$. Now just reverse the roles of the two series.

Surprisingly, this fails if we allow negative terms. The alternating series

$$S = \ln 2 = 1/1 - 1/2 + 1/3 - 1/4 + \ldots\ldots\ldots\ldots$$

re-arranges (one odd, followed by two evens) to

$$1/1 - 1/2 - 1/4 + 1/3 - 1/6 - 1/8 + 1/5 - 1/10 - 1/12 + \ldots\ldots.$$

which has sum $S/2$.

5.4 Double series

Suppose that m and p are (finite) integers and $m \leq n \leq \infty$ and $p \leq q \leq \infty$ and $a_{j,k} \in [0, \infty]$ for all *integers* j, k with $m \leq j \leq n, p \leq k \leq q$. N.B. j, k do not take the value ∞. Then

$$\sum_{j=m}^{n} \left(\sum_{k=p}^{q} a_{j,k} \right) = S_1$$

and

$$\sum_{k=p}^{q} \left(\sum_{j=m}^{n} a_{j,k} \right) = S_2$$

are equal.

Proof: this is clearly true if n and q are both finite. Also if $N \leq n$ is finite and $q = \infty$,

$$\sum_{j=m}^{N} \left(\sum_{k=p}^{\infty} a_{j,k} \right) = \sum_{j=m}^{N} \left(\lim_{M \to \infty} \sum_{k=p}^{M} a_{j,k} \right) =$$

$$= \lim_{M \to \infty} \sum_{j=m}^{N} \left(\sum_{k=p}^{M} a_{j,k} \right) = \lim_{M \to \infty} \sum_{k=p}^{M} \left(\sum_{j=m}^{N} a_{j,k} \right) = \sum_{k=p}^{\infty} \left(\sum_{j=m}^{N} a_{j,k} \right).$$

Setting $N = n$, this proves the result when n is finite. Now if n and q are both ∞, take any finite $N \geq m$ to get

$$\sum_{j=m}^{N} \left(\sum_{k=p}^{\infty} a_{j,k} \right) = \sum_{k=p}^{\infty} \left(\sum_{j=m}^{N} a_{j,k} \right) \leq \sum_{k=p}^{\infty} \left(\sum_{j=m}^{\infty} a_{j,k} \right).$$

Letting $N \to \infty$ we get $S_1 \leq S_2$, and the reverse inequality is proved the same way.

Note that if we take any sum of finitely many $a_{j,k}$ this will be at most the sum of finitely many rows, and so at most the double sum.

Again, examples show that this property of independence of order of summation fails if we allow terms of mixed sign.

5.5 Generalized series

If T is a countably infinite set, and $a_t \in [0, \infty]$ for every t in T, we can define

$$\sum_{t \in T} a_t$$

as follows. Let t_n be any sequence using up T, with every t in T appearing exactly once, set $b_n = a_{t_n}$ and put

$$\sum_{t \in T} a_t = \sum_{n=1}^{\infty} b_n.$$

It does not matter *which* particular sequence we take, because if we use a different sequence the series we get will be a re-arrangement of $\sum b_n$ and so have the same sum.

For example,

$$\sum_{t \in \mathbb{Z}} 2^t = 2^0 + 2^1 + 2^{-1} + 2^2 + \ldots\ldots\ldots = \infty.$$

6 Measures

We first need:

6.1 σ-algebras

Let X be a set, and let Π be a collection of subsets of X. Then Π is called a σ-algebra if Π is non-empty and the following two conditions are satisfied:

(i) for every A in Π, the complement $X \backslash A$ is in Π;

(ii) if we have countably many A_j, say $A_1, A_2, \ldots$, all in Π, then their union $\bigcup_j A_j$ is in Π. In particular, the union of finitely many elements of Π is an element of Π.

By taking $A \cup (X \setminus A)$, we see that X is always in Π, and so is the empty set.

It follows from (i) and (ii) that the intersection of countably many elements of Π is an element of Π (see problem sheet).

The simplest example of a σ-algebra is the power set $P(X)$, the collection of all subsets of X.

Note that some books omit the requirement that Π be non-empty: however, an empty collection of subsets of X is not very interesting. Also it is easy to check (optional) that this definition is equivalent to what Dr. Feinstein calls a σ-field in his G1CMIN lecture notes and exam papers.

6.2 Lemma

Let X be a set. If for every t in some set T, the collection Π_t is a σ-algebra of subsets of X, then the intersection $\bigcap_t \Pi_t$, which is the collection of all subsets of X each belonging to all of the Π_t, is itself a σ-algebra of subsets of X.

The proof is trivial: since $X \in \Pi_t$ for every t, the intersection is non-empty, and conditions (i) and (ii) are obviously satisfied.

It follows that every non-empty collection H of subsets of X is a sub-collection of a 'minimal' σ-algebra of subsets of X. To do this, take the intersection of all σ-algebras of subsets of X each of which contain H. This is not a vacuous definition, as there is always at least one, namely $P(X)$. This is called the σ-algebra 'generated' by H.

The σ-algebra generated by the open sets is called the σ-algebra of Borel sets. We shall see that not every subset of $\mathbb{R}$ is a Borel set.

6.3 Measures

Let X be a non-empty set, and let Π be a (N.B. non-empty) σ-algebra of subsets of X. By a measure μ on Π we mean a function $\mu : \Pi \to [0, +\infty]$ which satisfies the conditions $\mu(\emptyset) = 0$ and

$$\mu(E) = \sum_j \mu(E_j)$$

whenever we have a countable family of pairwise disjoints sets E_j (all in Π) whose union is E. The elements of Π will be called μ-measurable (or just measurable) sets, and we will often talk about μ as a measure on X (with the existence of Π taken for granted).

6.4 Examples

Let X be a set and let $\mu(U)$ be the number of elements of each subset U of X. Then μ is a measure on $P(X)$, called the *counting measure*.

If we have a measure μ on a set X and $\mu(X) = 1$, then μ is called a probability measure. Measurable sets correspond to events. The countable additivity corresponds to the probabilities of pairwise mutually disjoint events. In particular, $\mu(X \backslash U) = 1 - \mu(U)$.

Let X be a set, let $x \in X$, and let $\mu(U)$ be 1 if $x \in U$, and 0 otherwise. Then μ is a measure on $P(X)$ (point mass at x).

Let X be an uncountable set and, for $U \subseteq X$, let $\mu(U)$ be 0 if U is countable and ∞ if U is uncountable.

6.5 Some properties of measures μ

(i) If $A \subseteq B$ then $\mu(A) \leq \mu(B)$.

To see this, write $C = B \backslash A = B \cap A^c$. Then $B = A \cup C$ and A, C are disjoint, and

$$\mu(B) = \mu(A) + \mu(C) \geq \mu(A).$$

(ii) $\mu(A \cup B) = \mu(A) + \mu(B \setminus A) \leq \mu(A) + \mu(B)$ (using (i)).

(iii) $\mu(\bigcup_{j=1}^{n} G_j) \leq \sum_{j=1}^{n} \mu(G_j)$. (By induction, using (ii)).

(iv) If $B_j \subseteq B_{j+1}$ and B is the union of the (countably many) B_j then $\mu(B_j) \to \mu(B)$ as $j \to \infty$.

To prove (iv), write $E_1 = B_1$ and $E_{j+1} = B_{j+1} \backslash B_j$. Then B_j is the union of $E_1, \ldots, E_j$ (by induction on j) and B is the union of all the E_j and the E_j are disjoint. (If $m < n$ we have $E_m \subseteq B_m \subseteq B_{n-1}$ and $E_n = B_n \setminus B_{n-1}$ and so $E_m \cap E_n = \emptyset$. Also if m is the smallest j for which $x \in B_j$ then $x \in E_m$). We now get

$$\mu(B_j) = \sum_{m=1}^{j} \mu(E_m) \to \sum_{m=1}^{\infty} \mu(E_m) = \mu(B).$$

(v) It follows from (iv) that if the F_j are all μ-measurable then $\mu(\bigcup_{j=1}^{\infty} F_j) = \lim_{n\to\infty} \mu(\bigcup_{j=1}^{n} F_j)$.

(vi) We then have, using (iii) and (v), subadditivity:

$$\mu(\bigcup_{j=1}^{\infty} G_j) = \lim_{n\to\infty} \mu(\bigcup_{j=1}^{n} G_j) \leq \lim_{n\to\infty} \sum_{j=1}^{n} \mu(G_j) = \sum_{j} \mu(G_j).$$

(vii) Note that some books omit the condition $\mu(\emptyset) = 0$. However $\mu(\emptyset) = \mu(\emptyset) + \mu(\emptyset)$ and so the only possible values for $\mu(\emptyset)$ are 0 and ∞. If $\mu(\emptyset) = \infty$ then $\mu(A) = \infty$ for every $A \in \Pi$.

7 Lebesgue outer measure

We are going to construct the Lebesgue measure λ on a σ-algebra of subsets of $\mathbb{R}$, which may be thought of as a generalization of the idea of the length of an interval.

Now the idea of length makes sense for an interval, so that the length of (a, b) is $b - a$ (and is ∞ if $b = \infty$ or $a = -\infty$). However, for other sets such as $\mathbb{Q}$ the idea of length isn't defined in any "obvious" way. So we start by constructing the Lebesgue outer measure, which is defined for every subset A of $\mathbb{R}$. The idea is to "minimize" the total length of open intervals which together cover A. We will then prove some properties, including the fact that for intervals the outer measure equals the length.

7.1 Definition

Let $A \subseteq \mathbb{R}$. It is always possible to choose countably many open intervals which together cover A i.e. whose union contains A: in fact we can do this with one interval $(-\infty, \infty)$.

So we define the outer measure $\lambda^*(A)$ as follows. Consider all countable collections C of open intervals (a_k, b_k) such that $A \subseteq \bigcup (a_k, b_k)$: these intervals have total length $L(C) = \sum_k (b_k - a_k)$.

We do this for all such countable collections C of open sets covering A, and take the inf of $L(C)$ over all possible C. In particular, since $A \subset (-\infty, +\infty)$, this is always defined. Formally,

$$\lambda^*(A) = \inf \left\{ \sum_k (b_k - a_k) : A \subseteq \bigcup_k (a_k, b_k) \right\},$$

in which we consider only covering of A by *countably many* open intervals.

Another way to express this is that $\lambda^*(A)$ is the infimum of those positive t such that there exist countably many open intervals (a_k, b_k) of total length $\sum_k (b_k - a_k) = t$ with $A \subseteq \bigcup_k (a_k, b_k)$.

Obviously, if $A \subseteq B \subseteq \mathbb{R}$ then any collection of open intervals covering B also covers A, and so $\lambda^*(A) \le \lambda^*(B)$.

Also $\{x\} \subseteq (x - 1/n, x + 1/n)$ for $n \in \mathbb{N}$, so $\lambda^*(\{x\}) \le 2/n$ and so $\lambda^*(\{x\}) = 0$. Thus we have $\lambda^*(\emptyset) = 0$.

7.2 Theorem

λ^* is countably sub-additive, which means the following. If we have countably many subsets $E_1, E_2, \ldots$ of $\mathbb{R}$, and $E = \bigcup_n E_n$ is their union, then $\lambda^*(E) \le \sum_n \lambda^*(E_n)$.

Proof. This is obvious if any $\lambda^*(E_n)$ is infinite. Now assume that all $\lambda^*(E_n)$ are finite. Let $\infty > \delta > 0$. For each n, we have $\lambda^*(E_n) + \delta 2^{-n} > \lambda^*(E_n)$ and so we can choose a countable family of open intervals $I_{j,n}$ of length $L_{j,n}$, such that

$$E_n \subseteq \bigcup_j I_{j,n}, \quad \sum_j L_{j,n} < \lambda^*(E_n) + \delta 2^{-n}.$$

Then E is contained in the union of all the $I_{j,n}$. There are countably many of these intervals $I_{j,n}$ and together they cover E.

Consider the sum of the lengths of all the $I_{j,n}$. A partial sum s for this series is the sum of finitely many $L_{j,n}$. So if N is large enough we get

$$s \le \sum_{n=1}^{N} \sum_j L_{j,n} \le \sum_n (\lambda^*(E_n) + \delta 2^{-n}) \le \sum_n \lambda^*(E_n) + \sum_{n=1}^{\infty} \delta 2^{-n} = \delta + \sum_n \lambda^*(E_n).$$

Since we're taking an arbitrary partial sum, the sum of all the $L_{j,n}$ is at most $\delta + \sum_n \lambda^*(E_n)$. Since δ is arbitrary we get $\lambda^*(E) \le \sum_n \lambda^*(E_n)$.

7.3 Examples

We have $\lambda^*(\mathbb{Q}) = 0$, while $\lambda^*([0,1] \backslash \mathbb{Q}) = 1$.

7.4 Lemma

Let $A \subseteq \mathbb{R}$, let $x \in \mathbb{R}$, and let $A + x = \{y + x : y \in A\}$. Then $\lambda^*(A) = \lambda^*(A + x)$.

To prove this, let $s > \lambda^*(A)$. Then s is not a lower bound for the set of $t \in (0, \infty]$ such that A can be covered by the union of countably many open intervals of total length t. So there exists t with $t < s$ such that A is contained in the union of the (a_k, b_k), where $\sum_k (b_k - a_k) = t$. But then $A + x$ is contained in the union of the countably many intervals $(a_k + x, b_k + x)$, and these have total length t. So $\lambda^*(A + x) \leq \lambda^*(A)$. The reverse inequality holds, because $A = (A + x) + (-x)$.

7.5 Theorem

If A is an interval then $\lambda^*(A)$ is the length of A.

Proof. Suppose that the interval A has finite end-points a, b. Then for $n \in \mathbb{N}$ we have $A \subseteq (a - 1/n, b + 1/n)$ so $\lambda^*(A) \leq b - a + 2/n$. Since n is arbitrary we get $\lambda^*(A) \leq b - a$. Thus for a finite interval A, we have $\lambda^*(A)$ not more than the length of A, and this is obviously also true for an interval of infinite length. So we need to show that $\lambda^*(A)$ is at least the length of A.

To prove this, we assume first that $A = [a, b]$, with a, b finite, $a < b$. Suppose that we have a countable family of open intervals (a_k, b_k) $(k \in T \subseteq \mathbb{N})$ which together cover A, and $\sum_k (b_k - a_k) < b - a$. By Lemma 2.12 we can assume that there are only finitely many of these intervals. Thus A is contained in the union of N open intervals (a_k, b_k). Let n be large and partition A into n closed intervals $I_1, \ldots, I_n$ of equal length $s = (b - a)/n$, with vertices in the set $\{x_p = a + ps : p \in \mathbb{Z}\}$. For each k, the total length of those I_j which meet (a_k, b_k) is at most $b_k - a_k + 2s$ (it is at most the difference between the least x_p which is $\geq b_k$ and the greatest x_q which is $\leq b_k$).

Since every I_j is contained in $[a, b]$ and so meets at least one (a_k, b_k), we see that the total length of the I_j is

$$(b - a) = ns \leq \sum_k ((b_k - a_k) + 2s) \leq 2Ns + \sum_k (b_k - a_k).$$

Since n can be chosen arbitrarily large, with s consequently arbitrarily small, we get

$$b - a = ns \leq \sum_k (b_k - a_k).$$

So $\lambda^*(A)$ is at least the length of A when A is a closed interval with finite end-points. The general case follows, because if A is an interval and t is positive but less than the length of A, we can choose a closed interval B contained in A, of length t, to get $\lambda^*(A) \geq \lambda^*(B) \geq t$.

This idea and proof are easily generalized to higher dimensions. In $\mathbb{R}^2$ we take the infimum of the sum $\sum$ (area of B_j), over all coverings of A by open rectangles B_j (open means no boundary points included). Consider e.g. the straight line A from $(0,0)$ to $(1,1)$. The part with $j/n \leq x, y \leq (j+1)/n$ lies in the open rectangle $(j-1)/n < x, y < (j+2)/n$, which has area $9n^{-2}$. So $\lambda^*(A) \leq 9n^{-1}$ and, since n can be arbitrarily large, A has two-dimensional outer measure 0.

8 Lebesgue measure

The outer measure λ^* of the last section turns out not to be a measure on the whole power set $P(\mathbb{R})$. However, we can find a σ-algebra of subsets of $\mathbb{R}$ on which it is a measure. The idea is the following. If λ^* *were* a measure on the power set of $\mathbb{R}$ then we'd have

$$\lambda^*(A) = \lambda^*(A \cap E) + \lambda^*(A \backslash E)$$

for every A, E, by disjointness. So we consider those E for which this is true for every A.

8.1 Definition

A subset E of $\mathbb{R}$ is said to be Lebesgue measurable if we have

$$\lambda^*(A) = \lambda^*(A \cap E) + \lambda^*(A \backslash E)$$

for every subset A of $\mathbb{R}$. Note that

$$\lambda^*(A) \leq \lambda^*(A \cap E) + \lambda^*(A \backslash E)$$

always holds, so that to show that some set E is Lebesgue measurable it suffices to prove that

$$\lambda^*(A) \geq \lambda^*(A \cap E) + \lambda^*(A \backslash E)$$

for every A.

We will sometimes write E^c for $\mathbb{R} \setminus E$.

8.2 Some basic facts

(i) $\mathbb{R}$ and the empty set are Lebesgue measurable.

(ii) Any set E with $\lambda^*(E) = 0$ is Lebesgue measurable.

(iii) If E is Lebesgue measurable, so is $\mathbb{R} \backslash E$.

(iv) If E is Lebesgue measurable, so is $x + E$ for every $x \in \mathbb{R}$.

Property (iii) is obvious. To prove (ii), assume $\lambda^*(E) = 0$. Then $\lambda^*(A \cap E)$ will be 0 for all A, which gives

$$\lambda^*(A \cap E) + \lambda^*(A \setminus E) = \lambda^*(A \setminus E) \leq \lambda^*(A)$$

as required. Property (i) now follows from (ii) and (iii).

Now we prove (iv). Let $E \subseteq \mathbb{R}$ be Lebesgue measurable, and note that for $A \subseteq \mathbb{R}$ we have

$$A \cap (E + x) = \{y : y \in A, y - x \in E\} = \{z + x : z + x \in A, z \in E\} =$$

$$= x + \{z : z \in A - x, z \in E\} = x + ((A - x) \cap E)$$

and hence $\lambda^*(A \cap (E + x)) = \lambda^*((A - x) \cap E)$, using Lemma 7.4.

Next, note that $(E + x)^c = E^c + x$. This gives

$$A \setminus (E + x) = A \cap (E + x)^c = A \cap (E^c + x) = x + ((A - x) \cap E^c)$$

which gives $\lambda^*(A \setminus (E + x)) = \lambda^*((A - x) \cap E^c)$.

Since E is Lebesgue measurable we now get

$$\lambda^*(A) = \lambda^*(A - x) = \lambda^*((A - x) \cap E) + \lambda^*((A - x) \cap E^c) = \lambda^*(A \cap (E + x)) + \lambda^*(A \setminus (E + x)).$$

8.3 Theorem

The union or intersection of finitely many Lebesgue measurable sets is Lebesgue measurable.

We first prove that if E and F are Lebesgue measurable subsets of $\mathbb{R}$, then so is $G = E \cup F$. Write E^c for $\mathbb{R} \setminus E$. We have

$$\lambda^*(A) = \lambda^*(A \cap E) + \lambda^*(A \cap E^c) = \lambda^*(A \cap E \cap F) + \lambda^*(A \cap E \cap F^c) + \lambda^*(A \cap E^c \cap F) + \lambda^*(A \cap E^c \cap F^c) =$$

$$= \lambda^*(A \cap E \cap F) + \lambda^*(A \cap E \cap F^c) + \lambda^*(A \cap E^c \cap F) + \lambda^*(A \cap G^c) \geq \lambda^*(A \cap G) + \lambda^*(A \cap G^c),$$

since λ^* is sub-additive and

$$G = E \cup F = (E \cap F) \cup (E \cap F^c) \cup (E^c \cap F).$$

Now, if we are given $n \geq 3$ and Lebesgue measurable sets $F_1, \ldots F_n$, we write

$$\bigcup_{j=1}^{n} F_j = F_n \cup \bigcup_{j=1}^{n-1} F_j$$

and so the assertion about unions of finitely many sets follows by induction. Intersections work as well, because

$$\bigcap_{j=1}^{n} F_j$$

is the complement of

$$\bigcup_{j=1}^{n} F_j^c$$

and each F_j^c is Lebesgue measurable.

8.4 Theorem

(i) Suppose that $F_1, F_2, \ldots$ are countably many pairwise disjoint Lebesgue measurable sets with union F. Then F is Lebesgue measurable, and $\lambda^*(F) = \sum_j \lambda^*(F_j)$.

(ii) The union of countably many Lebesgue measurable sets is Lebesgue measurable.

Proof. (i) We may assume that there are infinitely many F_j, by making them all empty from some j on, if necessary. Let $G_n = \bigcup_{j=1}^n F_j$. The G_n are Lebesgue measurable.

We claim that for every $n \in \mathbb{N}$, and for every $A \subseteq \mathbb{R}$,

$$\lambda^*(A \cap G_n) = \sum_{j=1}^{n} \lambda^*(A \cap F_j).$$

This is clearly true for $n = 1$. Assuming it true for n and using the Lebesgue measurability of F_{n+1}, we get

$$\lambda^*(A\cap G_{n+1}) = \lambda^*(A\cap G_{n+1}\cap F_{n+1}) + \lambda^*(A\cap G_{n+1}\cap F_{n+1}^c) = \lambda^*(A\cap F_{n+1}) + \lambda^*(A\cap G_n) = \sum_{j=1}^{n+1} \lambda^*(A\cap F_j)$$

and the result follows.

Now, by the previous claim, if A is any subset of $\mathbb{R}$,

$$\lambda^*(A) = \lambda^*(A\cap G_n) + \lambda^*(A\cap G_n^c) \geq \lambda^*(A\cap G_n) + \lambda^*(A\cap F^c) = \sum_{j=1}^{n} \lambda^*(A\cap F_j) + \lambda^*(A\cap F^c).$$

Since n is arbitrary,

$$\lambda^*(A) \geq \sum_j \lambda^*(A\cap F_j) + \lambda^*(A\cap F^c). \tag{2}$$

Since λ^* is countably sub-additive, we get, from (2),

$$\lambda^*(A) \geq \lambda^*(A\cap F) + \lambda^*(A\cap F^c),$$

which proves that F is Lebesgue measurable. Now choosing $A = F$ in (2) and using sub-additivity again gives

$$\lambda^*(F) \geq \sum_j \lambda^*(F_j) \geq \lambda^*(F).$$

To prove (ii), just note that if we have $E_1, E_2, \ldots$ with union E, then setting $H_1 = E_1, H_{n+1} = E_{n+1}\backslash(\bigcup_{j=1}^n E_j)$, the H_j are Lebesgue measurable and pairwise disjoint, and their union is E. (If $m < n$ then $H_m \subseteq E_m$ and $H_n \cap E_m$ is empty).

8.5 Corollary

For Lebesgue measurable F we define $\lambda(F) = \lambda^*(F)$. The Lebesgue measurable sets form a σ-algebra Π of subsets of $\mathbb{R}$, and λ is a measure on Π.

8.6 Theorem

Let $a \in \mathbb{R}$. Then the open interval $(a, +\infty)$ is Lebesgue measurable.

Proof. Let A be any subset of $\mathbb{R}$, and let $A_1 = A \cap (a, +\infty), A_2 = A \cap (-\infty, a]$. We only need to show that $\lambda^*(A) \geq \lambda^*(A_1) + \lambda^*(A_2)$, which is obvious if $\lambda^*(A) = \infty$. Assume now that $\lambda^*(A)$ is finite. Take a real $\delta > 0$ and choose a countable collection of open intervals I_j which cover A, such that

$$\sum_j |I_j| < \lambda^*(A) + \delta,$$

using $|I|$ for the length. Let

$$P_j = I_j \cap (a, +\infty), \quad Q_j = I_j \cap (-\infty, a].$$

Then $A_1 \subseteq \bigcup_j P_j$ so

$$\lambda^*(A_1) \leq \sum_j \lambda^*(P_j) = \sum_j |P_j|$$

since P_j is an interval. Doing the same for A_2,

$$\lambda^*(A_1) + \lambda^*(A_2) \leq \sum_j |P_j| + \sum_j |Q_j| = \sum_j |I_j| < \lambda^*(A) + \delta.$$

Since δ is arbitrary the theorem is proved.

8.7 Corollary

All Borel sets (in particular, all open sets) are Lebesgue measurable.

8.8 Theorem

Let $E \subseteq \mathbb{R}$. The following statements are equivalent.

(i) E is Lebesgue measurable.
(ii) For every real $\varepsilon > 0$ there exists an open set U with $E \subseteq U$ and $\lambda^*(U \setminus E) < \varepsilon$.
(iii) There exists a set V, such that:
V is the intersection of countably many open sets;
V contains E;
$\lambda^*(V \setminus E) = 0$.

Thus Lebesgue measurable sets may in a certain sense be "well-approximated" by open sets.

Proof. (ii) implies (iii): for each $n \in \mathbb{N}$ take an open set V_n containing E, such that $\lambda^*(V_n \setminus E) < 1/n$. Now let V be the intersection of the V_n. Then $E \subseteq V$, and for each n we have $\lambda^*(V \setminus E) \leq \lambda^*(V_n \setminus E) < 1/n$.

(iii) implies (i): take V as in (iii), and set $W = \mathbb{R} \setminus V$. Then $W \subseteq F = \mathbb{R} \setminus E$, and $F \setminus W = V \setminus E$.

Since V is the intersection of open U_j, say, then $W = \mathbb{R} \setminus \bigcap U_j = \bigcup(\mathbb{R} \setminus U_j)$ is a union of closed sets, and so is Lebesgue measurable. Hence we can write F as the union of a closed set W and a set $F \setminus W$ which has outer measure 0. Thus F is Lebesgue measurable and so is E.

(i) implies (ii). First set $E_n = E \cap [-n, n], n \in \mathbb{N}$. Each E_n is Lebesgue measurable. For each $n \in \mathbb{N}$ choose, directly from the definition of outer measure, a countable union U_n of open intervals containing E_n and having sum of lengths less than $\lambda(E_n) + \varepsilon 2^{-n}$. Then $A_n = U_n \setminus E_n$ has $\lambda(E_n) + \lambda(A_n) = \lambda(U_n)$ and so $\lambda(A_n) = \lambda^*(A_n) < \varepsilon 2^{-n}$. Let U be the union of the $U_n, n \in \mathbb{N}$. Then U is open, and $U \setminus E = (\bigcup U_n) \setminus E = \bigcup(U_n \setminus E) \subseteq \bigcup(U_n \setminus E_n) = \bigcup A_n$ has, by the subadditivity of λ^*, outer measure at most $\sum_{n\in\mathbb{N}} \lambda^*(A_n) < \varepsilon(1/2 + 1/4 + \ldots) = \varepsilon$.

9 A set which is not Lebesgue measurable

9.1 The Axiom of Choice

The version of this axiom which we will use is the following:

Suppose that we have a set T, and that A_t is a non-empty set, for each $t \in T$, and that $A_t \cap A_s = \emptyset$, for $s, t \in T, s \neq t$. Then we can form a set $B = \{c_t : t \in T\}$ by choosing one c_t from each A_t.

9.2 Theorem

There exists a subset E of $[0, 1]$ such that E is not Lebesgue measurable.

Proof. We define a relation $\sim$ on $[0, 1]$ by $x \sim y$ iff $x - y$ is rational. Then $\sim$ is an equivalence relation. To see this, obviously $x \sim x$ (so $\sim$ is reflexive) and $x \sim y$ iff $y \sim x$ (symmetric) and $x \sim y$ and $y \sim z$ imply that $y - x$ and $z - y$ are rational, so that $z - x$ is rational (transitive).

For each $x \in I = [0, 1]$ we form the equivalence class

$$[x] = \{y \in I : y \sim x\}.$$

Then either $[x] \cap [y]$ is empty, or $[x] = [y]$. Thus I is the union of these disjoint equivalence classes. We have a set of pairwise disjoint equivalence classes, whose union is I.

(To see that we have a *set* T of these, use the mapping $\phi : I \to T$ given by $\phi(x) = [x]$, so that T is just the image $\phi(I)$.)

Using the Axiom of Choice, we form a set E which contains precisely one element of each equivalence class. So for each x in $[0, 1]$ there is a unique y in E such that $x - y$ is rational. Note that $-1 \leq x - y \leq 1$.

Now use the fact that $H = \mathbb{Q} \cap [-1, 1]$ is countable, and write this set as $\{r_1, r_2, r_3, \ldots\}$, with the r_j distinct rational numbers, using up H. Then every x in $[0, 1]$ belongs to one of the E_j defined by $E_j = E + r_j$. Also, if $j \neq k$ then $E_j \cap E_k = \emptyset$. For if u lies in both then $u - r_j$ and $u - r_k$ are both in E. But $(u - r_j) \neq (u - r_k)$ and $(u - r_j) \sim (u - r_k)$, and E contains just one element of each equivalence class.

Suppose that E is Lebesgue measurable. Then so is each E_j. Now, $[0,1]$ is contained in the countable union of the pairwise disjoint E_j, so

$$1 = \lambda([0,1]) = \lambda^*([0,1]) \le \sum_{j\in\mathbb{N}} \lambda^*(E_j) = \sum_{j\in\mathbb{N}} \lambda(E_j) = \sum_{j\in\mathbb{N}} \lambda(E).$$

So $\lambda(E) > 0$. But the E_j are disjoint and each is a subset of $[-1,2]$, so

$$\infty = \sum_{j\in\mathbb{N}} \lambda(E) = \sum_{j\in\mathbb{N}} \lambda(E_j) \le \lambda([-1,2]) = 3,$$

which is impossible.

Hence it is not always true that

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

so the Lebesgue outer measure described earlier is not additive on the whole power set of $\mathbb{R}$, and so is not a measure on the power set of $\mathbb{R}$.

We also see that not every subset of $\mathbb{R}$ is a Borel set.

Not all mathematicians accept the Axiom of Choice. It can be shown to imply that every set C has an ordering $<^*$ with the properties that, for all a, b, c in C,

(i) either $a <^* b$ or $b <^* a$ or $b = a$, and $a <^* b$ and $b <^* c$ implies $a <^* c$, as well as

(ii) every non-empty subset D of C has a least element i.e. there exists $d \in D$ such that $d <^* c$ for all c in D with $c \neq d$.

Such an ordering is called a well-ordering. $\mathbb{N}$ is well-ordered by ordinary $<$, but $\mathbb{R}$ is not. Every countable set A has an obvious well-ordering. Write A as $\{a_j\}$ without repetition, and order by $a_1 <^* a_2 <^* a_3 <^* \ldots$.

10 Measurable functions

We have constructed Lebesgue measure for (some) subsets of $\mathbb{R}$, and we now return to a general measure space. So assume we have a non-empty set X, a (non-empty) σ-algebra Π of subsets of X, and a measure $\mu : \Pi \to [0,\infty]$. We will sometimes refer to the elements of Π as μ-measurable sets.

The aim is to construct the integral $\int_E f d\mu$ for measurable E, and to develop its properties, and in this chapter we determine which functions can be used. Since we occasionally need products of functions, we define *some* products involving ∞.

10.1 Products involving infinity

We now define

$$x.\infty = \infty$$

if $x > 0$ and

$$0.\infty = 0.$$

The purpose of the latter is to make certain integrals take their expected values.

10.2 Lemma

Let $E \subseteq X$ with E in Π, and let f be a function from E to R^*. The following four conditions are equivalent:

(i) For each real y, the set $A = \{x \in E : f(x) > y\}$ is in Π.

(ii) For each real y, the set $B = \{x \in E : f(x) \leq y\}$ is in Π.

(iii) For each real y, the set $\{x \in E : f(x) \geq y\}$ is in Π.

(iv) For each real y, the set $\{x \in E : f(x) < y\}$ is in Π.

Each of these conditions implies that:

(v) for each extended real number y, the set $\{x \in E : f(x) = y\}$ is in Π.

Proof. (i) and (ii) are equivalent, because $B = E \backslash A = E \cap A^c$. If A is in Π then so are A^c and B. Similarly, (iii) and (iv) are equivalent. Also, (i) and (iii) are equivalent, because

$$\{x \in E : f(x) > y\} = \bigcup_{n \in \mathbb{N}} \{x \in E : f(x) \geq y + 1/n\}$$

and

$$\{x \in E : f(x) \geq y\} = \bigcap_{n \in \mathbb{N}} \{x \in E : f(x) > y - 1/n\}.$$

When y is finite, (v) clearly follows, since the intersection of elements of Π is in Π. Finally

$$\{x \in E : f(x) = +\infty\} = \bigcap_{n \in \mathbb{N}} \{x \in E : f(x) > n\}$$

and

$$\{x \in E : f(x) = -\infty\} = \bigcap_{n \in \mathbb{N}} \{x \in E : f(x) < -n\}.$$

We define f to be μ-measurable (on E) if f satisfies any of (i) to (iv).

10.3 Lemma

If f is μ-measurable then so are $-f$ and f^2 and $|f|$ and cf, for any constant $c > 0$.

Proof. Take $y \in \mathbb{R}$. Then $\{x \in E : -f(x) > y\} = \{x \in E : f(x) < -y\} \in \Pi$. Also $\{x \in E : cf(x) > y\} = \{x \in E : f(x) > y/c\}$. For $y < 0$ we clearly have $|f(x)| > y$ and $f(x)^2 > y$ on all of E, while for $y \geq 0$,

$$\{x \in E : f(x)^2 > y\} = \{x \in E : f(x) > y^{1/2}\} \cup \{x \in E : f(x) < -y^{1/2}\}$$

and the idea for $|f|$ is the same.

Sums and products are only difficult insofar as $f + g$ is undefined where $f = +\infty$ and $g = -\infty$ and vice versa.

10.4 Theorem

If f, g are μ-measurable functions on $E \in \Pi$, taking values in $\mathbb{R}$, then $f+g$ and fg are μ-measurable.

Proof. Suppose that $f(u) + g(u) > y \in \mathbb{R}$. Then $f(u) > y - g(u)$ and there is a rational number r such that $f(u) > r > y - g(u)$ and so $g(u) > y - r$. The set $\{x \in E : f(x) > r, g(x) > y - r\}$ is the intersection of μ-measurable sets and so μ-measurable. Then

$$\{x \in E : f(x) + g(x) > y\} = \bigcup_{r \in \mathbb{Q}} \{x \in E : f(x) > r, g(x) > y - r\}.$$

Now fg is also μ-measurable, since

$$fg = \frac{1}{2}((f+g)^2 - f^2 - g^2).$$

Remark: we can modify the same proof for $f+g$ and fg if f, g are measurable on E taking values in $[0, \infty]$. Let $M = \{x \in E : f(x) = \infty\}$ and $N = \{x \in E : g(x) = \infty\}$ and $D = E \backslash (M \cup N)$. Then M, N and D are in Π, and f, g are μ-measurable on D (we always get the intersection of D with an element of Π). So for all $y \in \mathbb{R}$ we have

$$\{x \in E : f(x) + g(x) > y\} = \{x \in D : f(x) + g(x) > y\} \cup M \cup N,$$

which is in Π.

Similarly if $y < 0$ then $f(x)g(x) > y$ for all x in E, while if $y \geq 0$ then

$$\{x \in E : f(x)g(x) > y\} = \{x \in D : f(x)g(x) > y\} \cup V \cup W,$$

where V is the set on which $f = \infty$ and $g > 0$ (an intersection of measurable sets) and W is the set on which $g = \infty$ and $f > 0$.

10.5 Lemma

Let $f_1, f_2, \ldots$ be countably many μ-measurable extended real-valued functions on $E \in \Pi$. We can assume that f_n is defined for each $n \in \mathbb{N}$ by making all of them the same from some point on, if necessary. Then the functions g, h defined by

$$g(x) = \inf\{f_n(x) : n \in \mathbb{N}\}, \quad h(x) = \sup\{f_n(x) : n \in \mathbb{N}\}$$

are μ-measurable.

If $f_n \to f$ pointwise on E, then f is μ-measurable on E.

Proof. If $y \in \mathbb{R}$ then

$$\{x \in E : h(x) > y\} = \bigcup_{n=1}^{\infty} \{x \in E : f_n(x) > y\}$$

and this is in E. Similarly for g (take the set of x where $g(x) < y$).

Now suppose that $f_n \to f$. For each $n \in \mathbb{N}$ define g_n by

$$g_n(x) = \sup\{f_k(x) : k \geq n\}.$$

Then each g_n is μ-measurable, by the first part. We claim that for each $x \in E$ we have

$$f(x) = \lim_{n\to\infty} g_n(x).$$

If $y < f(x)$ then for large n we have $g_n(x) \geq f_n(x) > y$. Now suppose $y > f(x)$. Then for large n we have $f_k(x) < y$ for $k \geq n$ and so $g_n(x) \leq y$. So $g_n(x) \to f(x)$ as $n \to \infty$ and our claim is proved.

But we clearly have $g_{n+1}(x) \leq g_n(x)$ (sup of a subset), and so for each $x \in E$ we get

$$\lim_{n\to\infty} g_n(x) = \inf\{g_n(x) : n \in \mathbb{N}\}.$$

So f is an infimum of measurable functions and so measurable.

We saw in the chapter on Riemann integration that continuous and monotone functions are Riemann integrable. Here we show that they are measurable with respect to Lebesgue measure.

10.6 Theorem

Let $f : \mathbb{R} \to \mathbb{R}$ be continuous, and let $g : \mathbb{R} \to \mathbb{R}$ be non-decreasing. Then f and g are measurable with respect to Lebesgue measure.

Proof. Take y in $\mathbb{R}$. The set $\{x \in \mathbb{R} : f(x) > y\}$ is open, so Lebesgue measurable.

If $g(x) > y$ for all $x \in \mathbb{R}$ then obviously $\{x \in \mathbb{R} : g(x) > y\} = \mathbb{R}$, which is Lebesgue measurable. Now let a be the sup of x such that $g(x) \leq y$, assuming henceforth that there exists at least one such x. Then $g(x) > y$ for $x > a$ by definition. Next, if $x < a$ then (again by definition) there exists x' with $x < x' \leq a$ and $g(x') \leq y$, so that since g is non-decreasing we have $g(x) \leq y$ for $x < a$. So the set $\{x \in \mathbb{R} : g(x) > y\}$ is either $\emptyset$ or $\mathbb{R}$ or (a, ∞) or $[a, \infty)$ and all of these are Lebesgue measurable.

10.7 The characteristic function

Let (X, Π, μ) be a measure space. Let $A \subseteq X$, and define the characteristic function χ_A by $\chi_A(x) = 1$ if $x \in A$ and 0 if $x \notin A$. It is clear that this function is μ-measurable if and only if A is in Π.

10.8 Simple functions

Lt (X, Π, μ) be a measure space. A simple function is a function $s : X \to \mathbb{R}$ which takes only finitely many different values (all in $\mathbb{R}$ and so finite). We restrict simple functions to taking only finite values so that, for example, the sum of two simple functions is always defined (we avoid $\infty - \infty$). There are therefore finitely many disjoint sets $A_j, 1 \le j \le n$, whose union is X, and pairwise distinct real numbers α_j, such that we have

$$s(x) = \sum_{j=1}^{n} \alpha_j \chi_{A_j}(x).$$

Note that since $A_j = \{x \in X : s(x) = \alpha_j\}$, it follows that s is μ-measurable if and only if all the A_j are in Π.

10.9 Theorem

Let (X, Π, μ) be a measure space and let $f : X \to [0, \infty]$ be a non-negative extended real-valued function on X. Then f is μ-measurable if and only if there are μ-measurable simple functions s_n such that $0 \le s_1 \le s_2 \le s_3 \le \ldots \le f$ on X and $s_n \to f$ pointwise on X.

Proof. The 'if' part follows at one from Lemma 10.5. Now suppose that f is μ-measurable. For $n \in \mathbb{N}$ divide the interval $[0, 2^n]$ into 4^n closed intervals of length 2^{-n}, their end-points forming the set

$$T_n = \{0, 1/2^n, 2/2^n, 3/2^n, \ldots, 2^n - 1/2^n, 2^n\}.$$

For $x \in X$ and $n \in \mathbb{N}$ let $s_n(x)$ be the largest element of T_n which is $\le f(x)$. Clearly $0 \le s_n(x) \le s_{n+1}(x)$ since $T_n \subseteq T_{n+1}$. If $f(x) = \infty$ then we have $s_n(x) = 2^n \to f(x)$. If $f(x)$ is finite then for large n we have $f(x) < 2^n$ and $s_n(x) \le f(x) < s_n(x) + 1/2^n$, and so $s_n(x) \to f(x)$. Finally s_n is measurable because the sets $\{x : f(x) \ge 2^n\}$ and $\{x : j/2^n \le f(x) < (j+1)/2^n\}$ are all measurable.

11 The Lebesgue integral of a non-negative simple function

Throughout this section we assume that X is a set, and that μ is a (non-negative) measure defined on a σ-algebra of subsets of X. When we write *measurable* for a set or function we will always mean with respect to μ.

11.1 The integral of a simple function

To motivate this idea, consider the function s given by

$$s(x) = 0 \quad (x < 0), \quad s(x) = -1 \quad (0 \le x < 1), \quad s(x) = 2 \quad (1 \le x \le 2), \quad s(x) = 0 \quad (x > 2).$$

Then s is a Lebesgue measurable non-negative simple function. The area under the curve $y = s(x)$ is obviously $-1 + 2 = 1$. Let

$$\alpha_0 = 0, \quad A_0 = (-\infty, 0) \cup (2, \infty), \quad \alpha_1 = -1, \quad A_1 = [0, 1), \quad \alpha_2 = 2, \quad A_2 = [1, 2].$$

Then

$$s(x) = \sum_{j=0}^{2} \alpha_j \chi_{A_j}(x)$$

and the area under the curve is $\alpha_0\lambda(A_0) + \alpha_1\lambda(A_1) + \alpha_2\lambda(A_2)$.

In the general case suppose that $s : X \to [0, \infty)$ is a non-negative simple function which is μ-measurable. Note that s doesn't take the value ∞. Hence there are distinct real numbers $\alpha_j \geq 0$ and pairwise disjoint measurable sets A_j for $1 \leq j \leq n$ such that X is the union of the A_j and

$$s(x) = \sum_{j=1}^{n} \alpha_j \chi_{A_j}(x)$$

on X. We define

$$\int_X s d\mu = \sum_{j=1}^{n} \alpha_j \mu(A_j).$$

Note that it suffices to sum over those j such that $\alpha_j \neq 0$.

For example, let $X = \mathbb{R}$ and let $A = \mathbb{Q}$. Then χ_A is a simple function and $\int_{\mathbb{R}} \chi_A d\lambda = 1.\lambda(\mathbb{Q}) = 0$.

Similarly, $\int_{\mathbb{R}} 0 d\lambda = 0$. This is the reason why we define $0.\infty = 0$. We restrict attention (at least for now) to $s \geq 0$ because we need to avoid $\infty - \infty$.

Clearly

$$\int_X c s d\mu = c \int_X s d\mu$$

if c is a non-negative real constant.

Note also that if s is a non-negative μ-measurable simple function, then

$$\int_X s d\mu = 0$$

iff the set $\{x \in X : s(x) > 0\}$ has μ measure 0.

11.2 The integral over a subset

If s is a non-negative μ-measurable simple function and E is a measurable subset of X then $\chi_E s$ is a non-negative measurable simple function and we define

$$\int_E s d\mu = \int_X \chi_E s d\mu.$$

In fact we can write

$$s = \sum \alpha_j \chi_{A_j},$$

in which we just sum over those j with $\alpha_j \neq 0$, and

$$\chi_E s = \sum \alpha_j \chi_{A_j \cap E}.$$

Thus

$$\int_E s d\mu = \sum \alpha_j \mu(A_j \cap E) = \sum_{\alpha_j \neq 0} \alpha_j \mu(A_j \cap E) = \sum_{\alpha_j \neq 0} \alpha_j \mu(\{x \in E : s(x) = \alpha_j\}).$$

In effect we are just restricting s to E and taking the measure of the portion of each set which lies in E.

For example,

$$\int_{[0,1]} \chi_{\mathbb{Q}} d\lambda = 1.\lambda([0,1] \cap \mathbb{Q}) = 0.$$

This integral is very easily computed, whereas the Riemann integrable of this function over $[0,1]$ does not exist.

11.3 Lemma

Suppose that s, t are non-negative μ-measurable simple functions on X and $0 \leq s \leq t$ on $E \in \Pi$. Then $\int_E s d\mu \leq \int_E t d\mu$.

Proof. Suppose first that $E = X$ and

$$s = \sum_j \alpha_j \chi_{A_j}, \quad t = \sum_k \beta_k \chi_{B_k}.$$

Here χ_H means the characteristic function of H. The A_j are disjoint and their union is X, and the α_j are distinct, and the same thing is true for B_k, β_k. Let $s = c_{j,k}, t = d_{j,k}$ on $A_j \cap B_k$. So $c_{j,k} = \alpha_j, d_{j,k} = \beta_k$. Then

$$\int_X s d\mu = \sum_j \alpha_j \mu(A_j) = \sum_j \alpha_j \sum_k \mu(A_j \cap B_k) = \sum_j \sum_k c_{j,k} \mu(A_j \cap B_k) \leq$$

$$\leq \sum_j \sum_k d_{j,k} \mu(A_j \cap B_k) = \sum_k \sum_j d_{j,k} \mu(A_j \cap B_k) =$$

$$= \sum_k \beta_k \sum_j \mu(A_j \cap B_k) = \sum_k \beta_k \mu(B_k) = \int_X t d\mu.$$

In the general case we just note that $\chi_E s \leq \chi_E t$. Lemma 11.3 is proved.

11.4 Lemma

Let s be a non-negative μ-measurable simple function on X. For each μ-measurable $E \subseteq X$ (i.e. each $E \in \Pi$), set $\psi(E) = \int_E s d\mu = \int_X \chi_E s d\mu$. Then ψ is a measure.

Obviously $\psi(\emptyset) = \int_X 0 d\mu = 0$. Suppose that E is a countable union of pairwise disjoint sets $E_k \in \Pi$. Suppose $s = \sum_j \alpha_j \chi_{A_j}$. Then

$$\psi(E) = \sum_j \alpha_j \mu(A_j \cap E) = \sum_j \alpha_j \sum_k \mu(A_j \cap E_k) = \sum_j \sum_k \alpha_j \mu(A_j \cap E_k) =$$

$$= \sum_k \sum_j \alpha_j \mu(A_j \cap E_k) = \sum_k \int_{E_k} s d\mu = \sum_k \psi(E_k).$$

We can change the order of summation since all terms are non-negative.

11.5 Lemma

Let s, t be non-negative μ-measurable simple functions on X, and let E be a μ-measurable subset of X. Then

$$\int_E (s+t) d\mu = \int_E s d\mu + \int_E t d\mu.$$

Proof. We only need prove this when $E = X$, since $\chi_E s, \chi_E t$ are simple. Let (as before)

$$s = \sum_j \alpha_j \chi_{A_j}, \quad t = \sum_k \beta_k \chi_{B_k}, \quad E_{j,k} = A_j \cap B_k.$$

Then

$$\int_{E_{j,k}} (s+t) d\mu = (\alpha_j + \beta_k) \mu(E_{j,k}) = \int_{E_{j,k}} s d\mu + \int_{E_{j,k}} t d\mu.$$

But then, by the previous lemma, since $s + t$ is simple,

$$\int_X (s+t) d\mu = \sum_{j,k} \int_{E_{j,k}} (s+t) d\mu = \sum_{j,k} \int_{E_{j,k}} s d\mu + \sum_{j,k} \int_{E_{j,k}} t d\mu = \int_X s d\mu + \int_X t d\mu.$$

12 The Lebesgue integral of a general non-negative function

Note first that if s is a non-negative μ-measurable simple function on X then Lemma 11.3 gives

$$\int_X s d\mu = \sup \int_X t d\mu$$

in which the sup is taken over all non-negative μ-measurable simple t such that $0 \leq t \leq s$ on X.

Motivated partly by this fact, partly by lower sums for the Riemann integral and partly by Theorem 10.9, we do the following. If f is any non-negative μ-measurable function on X, taking values in $[0, \infty]$, we define

$$\int_X f d\mu = \sup \int_X s d\mu,$$

in which the sup is taken over all measurable simple functions s such that $0 \leq s \leq f$ on X. Note that if f is itself a simple function, then the sup is a max.

For example, we show that

$$\int_{\mathbb{R}} |x| d\lambda = \infty.$$

To see this, let $s(x)$ be 0 if $x < 1$, with $s(x) = 1$ if $x \geq 1$. Then s is a non-negative Lebesgue measurable simple function, with $0 \leq s(x) \leq |x|$ on $\mathbb{R}$. Hence

$$\int_{\mathbb{R}} |x| d\lambda \geq \int_{\mathbb{R}} s d\lambda = 1.\lambda([1, \infty)) = \infty.$$

(i) If c is a positive real number then $\int_X c f d\mu = c \int_X f d\mu$ for every non-negative measurable f. We just write

$$\int_X c f d\mu = \sup_{0 \leq s \leq cf} \int_X s d\mu = \sup_{0 \leq ct \leq cf} \int_X ct d\mu = \sup_{0 \leq t \leq f} c \int_X t d\mu = c \int_X f d\mu,$$

in which each s is μ-measurable and simple and we write $s = ct$, and use the obvious fact that $ct \leq cf$ iff $t \leq f$.

(ii) If $f = 0$ on X then f is simple and $\int_X f d\mu = 0$ (even if $\mu(X) = \infty$).

12.1 The integral over a subset

Let E be a μ-measurable subset of X. Let f be a non-negative μ-measurable extended real-valued function on X. Then $g = \chi_E.f$ is a μ-measurable function.

We define

$$\int_E f d\mu = \int_X \chi_E.f d\mu.$$

Note that this agrees with our earlier definition when f is simple.

If f is not defined on all of X, but only on some measurable F with $E \subseteq F \subseteq X$, we extend f to X by making it 0 off F. Then $\chi_E.f$ is the same regardless of which F we have.

We list some properties of the integral so defined.

(i) If $f \leq g$ on E apart from on a set of μ-measure 0, then $\int_E f d\mu \leq \int_E g d\mu$.

To see this, suppose first that $E = X$, and that we can write $X = F \cup G$, where G has measure 0, $F \cap G = \emptyset$, and $f \leq g$ on F. Then for non-negative μ-measurable simple s, taking values α_j on pairwise disjoint μ-measurable sets A_j,

$$\int_X s d\mu = \sum \alpha_j \mu(A_j) = \sum \alpha_j \mu(A_j \cap F) + \sum \alpha_j \mu(A_j \cap G) = \int_X \chi_F.s d\mu.$$

So if $0 \le s \le f$ on X then $\chi_F s \le \chi_F f \le \chi_F g \le g$ and so

$$\int_X s d\mu = \int_X \chi_F s d\mu \le \int_X g d\mu.$$

Taking the sup over all s with $0 \le s \le f$ we get

$$\int_X f d\mu \le \int_X g d\mu.$$

In the general case we just note that $\chi_E.f \le \chi_E.g$ on all of X apart from a set of measure 0.

Note that is is standard to say that a property holds "almost everywhere" (or a.e.) on E if it holds on $E \setminus G$, where G has μ-measure 0. This is a very useful concept but it is important to remember that where G has 0 measure will in general depend on which measure we are using.

(ii) If $f = g$ a.e. on E (i.e. apart from on a set of measure 0) then $\int_E f d\mu = \int_E g d\mu$.

This is easy, by (i), but is useful, and implies for example that we can change f to be, say 0, on a set of measure 0 without changing any integral.

(iii) If $\mu(E) = 0$ then $f = 0$ a.e. on E and so $\int_E f d\mu = \int_E 0 d\mu = 0$.

(iv) If $A \subseteq B$ and A, B are μ-measurable subsets of X then $\int_A f d\mu = \int_X \chi_A f d\mu \le \int_X \chi_B f d\mu = \int_B f d\mu$.

(v) Suppose that $\int_E f d\mu$ is finite. Then the set $F = \{x \in E : f(x) = +\infty\}$ has measure zero.

To see this, put $s_n(x) = n\chi_F(x), n \in \mathbb{N}$. Then $s_n \le \chi_E.f$ on X so $n\mu(F) = \int_X s_n d\mu \le \int_E f d\mu$.

We can express this conveniently by saying that f is finite a.e. on E.

(vi) Suppose that $\int_E f d\mu = 0$. Then the set $F = \{x \in E : f(x) > 0\}$ has measure 0.

If not, the set $F_n = \{x \in E : f(x) > 1/n\}$ has positive measure for some $n \in \mathbb{N}$, and we put $s = (1/n)\chi_{F_n}$. Then $0 \le s \le \chi_E.f$ on X and $0 < (1/n)\mu(F_n) = \int_X s d\mu \le \int_E f d\mu = 0$.

Combining (ii) and (vi) we see that for a μ-measurable non-negative function f on E we have $\int_E f d\mu = 0$ if and only if f vanishes a.e. on E.

12.2 Monotone convergence theorem

Let f_n be non-negative measurable functions on a measurable subset E of X, such that
(i) $0 \le f_1 \le f_2 \le \ldots$ a.e. on E, and
(ii) $f_n \to f$ pointwise a.e. on E
i.e. $0 \le f_1 \le f_2 \le \ldots \le f$ and $f_n \to f$ pointwise on $F = E \setminus G$, where $\mu(G) = 0$.

Then

$$\int_E f_n d\mu \to \int_E f d\mu, \quad n \to \infty.$$

Proof. Strictly speaking f is not defined *a priori* on all of E. However, f is measurable on F, by 10.5. If we change f_n and f to all be 0 on G then (i) is still satisfied and $f_n \to f$ pointwise on all of E, the function f is measurable on E, and we have not changed the values of any integrals. Assume for now that $E = X$. Now clearly, by the monotone sequence theorem,

$$\int_X f_{n-1} d\mu \leq \int_X f_n d\mu \to y \in [0, +\infty]$$

as $n \to \infty$. Since $f_n \leq f$, we have $y \leq \int_X f d\mu$.

The proof will be completed for the case $E = X$ if we can show that $y \geq \int_X f d\mu$. Suppose that s is μ-measurable and simple, with $0 \leq s \leq f$ on X, and set $\psi(U) = \int_U s d\mu$ for each $U \in \Pi$. This is a measure, as shown in the last chapter.

Let $0 < c < 1$. Let $E_n = \{x \in X : f_n(x) \geq cs(x)\}$. Then $E_n \subset E_{n+1}$ (obvious). We claim that every $x \in X$ is in the union of the E_n, which is obvious if $s(x) = 0$, because then x is in E_1. If $s(x) > 0$ then $f(x) > 0$ and $cs(x) < s(x) \leq f(x)$ so $x \in E_n$ for large n, since $f_n(x) \to f(x) \geq s(x)$. Now

$$\int_X f_n d\mu \geq \int_{E_n} f_n d\mu \geq c \int_{E_n} s d\mu = c\psi(E_n) \to c\psi(X),$$

since X is the union of the expanding sets E_n. Thus $y \geq c\psi(X) = c \int_X s d\mu$ so $y \geq \int_X s d\mu$ since c is arbitrary, and so since s is arbitrary we get $y \geq \int_X f d\mu$.

For the general case in which we have $E \in \Pi$ we just extend each f_n and f to be 0 on $X \setminus E$, and we get

$$\int_E f_n d\mu = \int_X \chi_E f_n d\mu \to \int_X \chi_E f d\mu = \int_E f d\mu.$$

This is by applying the result for X to $\chi_E.f_n$ and $\chi_E.f$.

12.3 Examples and remarks

(i) We will later see that this result is not true without the condition that $f_1 \geq 0$. Let $s_n(x) = 0$ for $x \leq n$ and $s_n(x) = -1$ for $x > n$. Then s_n is simple, $s_n \leq s_{n+1}$ and $s_n \to 0$ pointwise. But we'll see that $-\infty = \int_{\mathbb{R}} s_n d\lambda \not\to 0 = \int_{\mathbb{R}} 0 d\lambda$.

(ii) We also cannot drop the hypothesis that $f_1 \leq f_2$ etc. If we set $f_n(x) = n\chi_{[0,1/n]}(x)$ then $\int_{\mathbb{R}} f_n d\lambda = 1$ but f_n converges pointwise to the function which is ∞ at 0 and 0 everywhgere else, and this function has integral 0.

(iii) If f is a non-negative measurable function and we choose a choose a non-decreasing sequence of non-negative simple functions s_n tending to f pointwise (as in Theorem 10.9), then $\int_X s_n d\mu \to \int_X f d\mu$.

(iv) If $a_m \in [0, \infty]$ for each $m \in \mathbb{N}$ and we set $f(m) = a_m$ then $\int_{\mathbb{N}} f d\mu = \sum_{m=1}^{\infty} a_m$, in which μ is the counting measure on $\mathbb{N}$. To see this, let $f_n(m) = a_m$ for $1 \leq m \leq n$, with $f_n(m) = 0$ for $m > n$. Assume first that all a_m are finite. Then each f_n is a simple function and f_n, f satisfy the hypotheses of the MCT and so

$$\int_{\mathbb{N}} f d\mu = \lim_{n\to\infty} \int_{\mathbb{N}} f_n d\mu = \lim_{n\to\infty} \sum_{m=1}^{n} a_m = \sum_{m=1}^{\infty} a_m.$$

If any a_m is infinite then so is the integral of f (since f is infinite on a set of non-zero μ-measure), and so is the sum of the series (look at partial sums).

12.4 Theorem

Let f_n be non-negative measurable functions on a measurable subset E of X and let $f = \sum_{n=1}^{\infty} f_n$. Then $\int_E f d\mu = \sum_{n=1}^{\infty} \int_E f_n d\mu$.

Proof. As usual we can assume $E = X$ because otherwise we can extend the f_n and f to X by making them 0 on $X \setminus E$. We first prove the theorem for finite sums. Let s_n be non-negative μ-measurable simple functions such that $s_n \leq s_{n+1} \leq f_1$ and $s_n \to f_1$ pointwise, and let $t_n \to f_2$ in the same fashion. Then $0 \leq s_n + t_n \leq s_{n+1} + t_{n+1}$ and so, by the MCT,

$$\int_X s_n d\mu + \int_X t_n d\mu = \int_X s_n + t_n d\mu \to \int_X f_1 + f_2 d\mu.$$

Here we've used the result, already proved, that the integral of the sum of two simple functions is the sum of the integrals. But $\int_X s_n d\mu \to \int_X f_1 d\mu$ and $\int_X t_n d\mu \to \int_X f_2 d\mu$. The theorem is thus proved for the sum of two functions and, by induction, for the sum of finitely many functions. Now by the MCT

$$\int_X f d\mu = \lim_{n\to\infty} \int_X \sum_{j=1}^{n} f_j d\mu = \lim_{n\to\infty} \sum_{j=1}^{n} \int_X f_j d\mu.$$

12.5 Theorem

Let f be a non-negative measurable function on X. Define $\psi(E) = \int_E f d\mu$. Then ψ is a measure and $\int_E g d\psi = \int_E g f d\mu$ for non-negative μ-measurable g.

Proof. Let E be a countable union of pairwise disjoint measurable E_j. Then

$$\psi(E) = \int_X \chi_E f d\mu = \int_X \sum_j \chi_{E_j} f d\mu = \sum_j \int_X \chi_{E_j} f d\mu = \sum_j \psi(E_j).$$

Thus ψ is a measure (obviously $\psi(\emptyset) = 0$). Now if $g = \chi_E$ for some E then

$$\int_X g f d\mu = \int_E f d\mu = \psi(E) = \int_X \chi_E d\psi = \int_X g d\psi.$$

Suppose now that g is non-negative, measurable and simple. Write

$$g = \sum_j \alpha_j \chi_{A_j}$$

as usual (a finite sum, with $\alpha_j \in [0, \infty), A_j \in \Pi$). Then Theorem 12.4 gives

$$\int_X g f d\mu = \sum_j \alpha_j \int_X \chi_{A_j} f d\mu = \sum_j \alpha_j \int_{A_j} f d\mu = \sum_j \alpha_j \psi(A_j) = \int_X g d\psi.$$

For a general non-negative measurable g, take simple s_n with limit g and $0 \leq s_1 \leq s_2 \leq \ldots \leq g$. Then

$$\int_X g d\psi = \lim \int_X s_n d\psi = \lim \int_X s_n f d\mu = \int_X g f d\mu$$

by the MCT.

13 The integral of a general measurable function

Suppose now that f is any measurable function on a measurable subset E of X, taking values in $\mathbb{R}^*$. Then $f^+ = \max\{f, 0\}$ and $f^- = \max\{-f, 0\}$ are measurable, and $f = f^+ - f^-$, and we can define

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$$

provided this is not $\infty - \infty$ i.e. at least one of $\int_E f^+ d\mu, \int_E f^- d\mu$ is finite.

Example: let $s(x) = 0$ for $x \leq a \in \mathbb{R}$, with $s(x) = -1$ for $x > a$. Then $s^+ = 0, s^- = \chi_{(a,\infty)}$ and

$$\int_{\mathbb{R}} s d\lambda = 0 - \lambda((a, \infty)) = -\infty.$$

13.1 Integrable functions

We say that a μ-measurable function $f : E \to [-\infty, \infty]$ is integrable if both $\int_E f^+ d\mu$ and $\int_E f^- d\mu$ are finite, in which case

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$$

definitely exists. Note that in this case f^+ and f^- are finite a.e., and so by changing f on a set of measure 0 we can assume that in fact f maps E into $\mathbb{R}$.

We list some easy properties of integrable f. We have

$$\int_E f d\mu \leq \int_E f^+ d\mu \leq \int_E |f| d\mu.$$

Since we also have, obviously,

$$-\int_E f d\mu = \int_E -f d\mu \leq \int_E |f| d\mu,$$

we get the standard inequality

$$\left| \int_E f d\mu \right| \leq \int_E |f| d\mu.$$

Finally, if f and g are integrable functions and $f \leq g$ then $f^+ \leq g^+$ and $g^- \leq f^-$ and so $\int_E f d\mu \leq \int_E g d\mu$.

13.2 Lemma

Suppose that f, g are integrable functions and that α is a real number. Then $f+g$ and αf are integrable and

$$\int_E \alpha f d\mu = \alpha \int_E f d\mu, \quad \int_E f + g d\mu = \int_E f d\mu + \int_E g d\mu.$$

Proof. Since f and g are integrable we can assume that both take values in $\mathbb{R}$, and so $f+g$ and αf are certainly measurable. It remains only to prove the assertion about $h = f + g$. We have

$$h^+ - h^- = h = f + g = f^+ - f^- + g^+ - g^-$$

and so $h^+ + f^- + g^- = h^- + f^+ + g^+$. Now Theorem 12.4 gives

$$\int_E h^+ d\mu + \int_E f^- d\mu + \int_E g^- d\mu = \int_E h^- + \int_E f^+ d\mu + \int_E g^+$$

and re-arranging gives the result.

13.3 The dominated convergence theorem

Suppose that f_n and g are measurable functions from a measurable subset E of X to $\mathbb{R}^*$, with $|f_n| \leq g$ a.e. on E, and $\int_E g d\mu < +\infty$. Suppose that $f_n \to f$ pointwise a.e. on E. Then f is integrable and $\int_E |f_n - f| d\mu \to 0$ and $\int_E f_n d\mu \to \int_E f d\mu$ as $n \to +\infty$.

Proof. As usual we can assume that $E = X$, by extending the functions to be 0 off E. We can also assume that $|f_n| \leq g$ and $f_n \to f$ pointwise on all of X, by changing the functions to all be 0 on the set of measure 0 where this may fail. Doing this does not change the value of any integrals.

We know that f is measurable. But $|f| \leq g$, so $f^+ \leq g, f^- \leq g$, and so f is integrable.

Since $\int_X g d\mu$ is finite, g is finite off a set F of measure 0, and again we can change f_n and g to be 0 on F without changing the value of any integrals.

For each $n \in \mathbb{N}$, define h_n by

$$h_n(x) = \sup\{|f_k(x) - f(x)| : k \geq n\}.$$

Then, since $f_n(x) \to f(x) \in \mathbb{R}$, we see that $h_n \to 0$ pointwise on X. We also have $h_{n+1} \leq h_n$. Moreover, $h_n \leq 2g$, since $|f_k(x) - f(x)| \leq 2g(x)$ for all $k \geq n$ and for all $x \in X$.

Put $g_n = 2g - h_n$. Then $0 \leq g_1 \leq g_2 \leq \ldots \leq 2g$, and $g_n \to 2g$ pointwise. Hence the MCT tells us that

$$\int_X g_n d\mu \to \int_X 2g d\mu.$$

This gives

$$\int_X h_n d\mu = \int_X 2g d\mu - \int_X g_n d\mu \to 0,$$

using the fact that g is integrable. Hence

$$\int_X |f_n - f| d\mu \leq \int_X h_n d\mu \to 0$$

as $n \to +\infty$. Thus

$$\int_X f_n - f d\mu \to 0,$$

and

$$\int_X f_n d\mu = \int_X f_n - f d\mu + \int_X f d\mu \to \int_X f d\mu.$$

13.4 Remark

The existence of g is necessary for the theorem to work. Let $f_n(x) = n^2$ for $0 < x \leq 1/n$, and let $f_n(x) = 0$ otherwise. Then $f_n \to 0$ pointwise, but $\int_{\mathbb{R}} f_n d\lambda = n \to \infty$.

13.5 Theorem

Let the bounded real-valued function f be Riemann integrable on $I = [a, b]$, with $|f| \leq M < \infty$ there. Then f is measurable with respect to Lebesgue measure on I and the Riemann integral $J_1 = \int_a^b f(x)dx$ equals the Lebesgue integral $J_2 = \int_I f d\lambda$.

Proof. Suppose first that $f \geq 0$ on $[a, b]$, say $0 \leq f \leq N$. Let P be a partition of I. Then $L(P, f)$ is equal to $\int_I s d\lambda$ for some simple function s with $0 \leq s \leq f$. Also $U(P, f)$ equals $\int_I S d\lambda$ for some simple S with $N \geq S \geq f$.

We first show that, *assuming* f is measurable, we have $J_1 = J_2$.

We have $L(P, f) = \int_I s d\lambda \leq \int_I f d\lambda = J_2$, and taking the supremum over all P we get $J_1 \leq J_2$. Similarly $J_2 \leq \int_I S d\lambda = U(P, f)$, and taking the inf over P we get $J_2 \leq J_1$.

We now prove that f is a Lebesgue measurable function. Since f is Riemann integrable, we can take partitions P_n with P_{n+1} a refinement of P_n and $L(P_n, f) \to J_1, U(P_n, f) \to J_1$ as $n \to \infty$. This gives us simple functions s_n, S_n with

$$0 \leq s_n \leq s_{n+1} \leq f \leq S_{n+1} \leq S_n \leq N$$

such that

$$\int_I S_n - s_n d\lambda = \int_I S_n d\lambda - \int_I s_n d\lambda = U(P_n, f) - L(P_n, f) \to 0$$

as $n \to \infty$. Let $S^*(x) = \inf_n S_n(x), s^*(x) = \sup_n s_n(x)$ on $I = [a, b]$. Then

$$s^* \leq f \leq S^*, \quad |S_n - s_n| \leq N,$$

and $S_n - s_n \to S^* - s^*$ pointwise on I. So the DCT tells us that

$$\int_I S^* - s^* d\lambda = 0, \quad S^* - s^* = 0 \quad a.e.$$

So we get $S^* = s^* = f$ a.e. on I, say on $F = I \setminus G$, where $\lambda(G) = 0$. Since S^* is measurable (an infimum of measurable functions) we deduce that f is measurable on F. We then have, for $y \in \mathbb{R}$,

$$\{x \in I : f(x) > y\} = \{x \in F : f(x) > y\} \cup \{x \in G : f(x) > y\},$$

which is the union of a Lebesgue measurable set and a set of measure 0, and so Lebesgue measurable.

It remains only to consider the case where $f : I \to [-M, M]$ is Riemann integrable. Here we just apply the above proof to $f + M$, and note that adding M to f just increases the Lebesgue and Riemann integrals by $M(b - a)$.

14 Pointwise and uniform convergence revisited

14.1 Example

Let $h_n(x) = e^{-nx}$ on $I = [0, 1]$, and let

$$h(0) = 1, \quad h(x) = 0 \quad (0 < x \leq 1).$$

Then $h_n \to h$ pointwise on I.

It is not true that $h_n \to h$ uniformly on I: take $\varepsilon = \frac{1}{4}$ and set

$$x_n = \frac{\ln 2}{n}.$$

Then

$$h_n(x_n) = \frac{1}{2}, \quad |h_n(x_n) - h(x_n)| = \frac{1}{2} > \varepsilon.$$

So there is no N such that $|h_n(x) - h(x)| < \varepsilon$ for all $n \geq N$ and all $x \in I$.

However, it is true that if we fix $\delta > 0$ then $h_n \to h$ uniformly on $[\delta, 1]$. In fact, on this set, we have

$$|h_n(x) - h(x)| = h_n(x) \leq e^{-n\delta} < \varepsilon$$

provided n is large enough.

14.2 Lemma

Let E be a subset of $\mathbb{R}$ with finite Lebesgue measure $m(E)$. Let $f_n : E \to \mathbb{R}^*$ be measurable functions converging pointwise on E to a function $f : E \to \mathbb{R}$. Let ε, δ be positive. Then there exist N and a subset A of E such that $\lambda(A) < \delta$ and $|f_n(x) - f(x)| < \varepsilon$ for all $n \geq N$ and all $x \in E \backslash A$.

Proof. We know that f is also measurable, by Lemma 10.5. Let E_N be the set of all $x \in E$ for which there exists $n \geq N$ with $|f_n(x) - f(x)| \geq \varepsilon$. In fact this set is

$$E \cap \bigcup_{n \geq N} \{x : |f_n(x) - f(x)| \geq \varepsilon\}$$

and so is Lebesgue measurable. Then $E_{N+1} \subseteq E_N$ and the intersection P of the E_N is empty. Since

$$\lambda(E) = \lambda(E \backslash P) = \lim \lambda(E \backslash E_N),$$

there exists N such that $A = E_N$ has $\lambda(A) < \delta$.

14.3 Egorov's theorem

Let E be a subset of $\mathbb{R}$ of finite Lebesgue measure $\lambda(E)$. Let $f_n : E \to \mathbb{R}^*$ converge pointwise on E to $f : E \to \mathbb{R}$. Let $\eta > 0$. Then there exists a subset A of E, with $\lambda(A) < \eta$, such that $f_n \to f$ uniformly on $E \backslash A$.

Proof. For each $n \in \mathbb{N}$ use Lemma 14.2 to choose $A_n \subseteq E$ and p_n such that $\lambda(A_n) < \eta/2^n$ and $|f_m(x) - f(x)| < 1/n$ for all $m \geq p_n$ and all $x \in E \backslash A_n$. Let $A = \bigcup A_n$. Then $E \backslash A \subseteq E \backslash A_n$, and so $m \geq p_n$ implies that $|f_m(x) - f(x)| < 1/n$ for all $x \in E \backslash A$.

Note that no such theorem holds for E of infinite measure. Let $f_n = \chi_{[n,\infty)}$, for $n \in \mathbb{N}$. Then $f_n \to 0$ pointwise, but not uniformly on the complement of any set of finite measure.