Math 501, Fall '03

Outline of the Lebesgue Approach to Integration

The Riemann integral is too restrictive to handle interchange of limits and integration gracefully and is not adequate to the task of building complete metric spaces based on integral norms. The Lebesgue theory proposes an alternate definition that resolves these difficulties, yet coincides with Riemann integral on all Riemann integrable functions. The approach embodied in Lebesgue's theory generalizes easily and usefully beyond the context of integration on $\mathbb{R}^n$, and we intend to present the theory abstractly enough to see its general structure and to pursue several different applications.

The purpose of this lecture is to outline the basic definitions and strategy of Lebesgue's approach to integration. To motivate the discussion and to bring out the central ideas, we will first formulate a list of the minimum properties we want of an integral operator and its class of integrands. We then deduce some basic consequences of these requirements, which in turn reveal will reveal the basic structure of the theory we want. Although the initial list of properties is based on trying to extend the Riemann integral of a function on $\mathbb{R}^n$, the ideas we abstract from the discussion lead to a general theory of measure and integration.

This lecture will be informal. We will make precise statements, but we will not prove everything; that will be left to future lectures.

Minimal requirements for a theory of integration generalizing the Riemann integral

We consider integration over $\mathbb{R}^n$. Given a real-valued function f, f^+ and f^- will denote its positive and negative parts, respectively. Let $V = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be a rectangle in $\mathbb{R}^n$. The notation $(R)\int_V f\,dx$ will denote the Riemann integral of f over the rectange V. Given a subset G of $\mathbb{R}^n$, χ_G denotes the indicator function of G.

We want to have an integration theory that is well-behaved with respect to limit operations on integrands, so we will write down the minimal properties that we wish our integration theory to possess. There are two components to consider: a set $\mathbb{B}$ of real-valued functions on $\mathbb{R}^n$ we wish to integrate, and an integral operation $f \to \int f$.

Of $\mathbb{B}$ we want at the minimum:

F1 $\mathbb{B}$ contains all continuous functions.

F2 If $f(x) = \lim_n f_n(x)$ for all x and $f_n \in \mathbb{B}$ for every n, then $f \in \mathbb{B}$. ($\mathbb{B}$ is closed under point-wise limits.)

F3 If $f, g \in \mathbb{B}$ and $\alpha, \beta \in \mathbb{R}$, then $\alpha f + \beta g \in \mathbb{B}$. ($\mathbb{B}$ is a vector space.)

F4 If $f \in \mathbb{B}$, then so are $f^+ = \max\{f, 0\}$ and $f^- = -\min\{f, 0\}$.

F5 If $f \in \mathbb{B}$ and V is a rectangle, then $f\chi_V \in \mathbb{B}$.

The reason for the requirements F1]—[F3] is clear: our integral should be able to handle at least the continuous functions, but, unlike the Riemann integral, it should also handle limits of integrable functions; and of course, it should handle, sums of integrable functions. Requirements [F4] and [F5] are made for slightly more technical reasons that will become apparent in a moment.

The integration operation will assign to every non-negative f, a value $\displaystyle\int f$ in $\mathbb{R} \cup \{\infty\}$. Thus integrals with infinite values are allowed; this is very convenient.

For an arbitrary function f in $\mathbb{B}$, let

$$\int f \triangleq \int f^+ - \int f^-$$

if the right-hand side is the indeterminate form $\infty - \infty$. If both integrals $\int f^+$ and $\int f^-$ are infinite, $\int f$ is left undefined. Thus, the integral takes values in the extended reals, $\overline{\mathbb{R}} \triangleq \mathbb{R} \cup \{\infty, -\infty\}$. Of the integral, we require:

I1 If f is continuous and $V \subset \mathbb{R}^n$ is a rectangle, $\displaystyle\int f\chi_V = (R)\int_V f\,dx$.

($\displaystyle\int$ coincides with the Riemann integral on continuous functions on bounded sets.)

I2 (**Monotone convergence property**) If $f_n \in \mathbb{B}$ for every n and if $0 \leq f_1 \leq f_2 \cdots$, then

$$\int \lim_n f_n = \lim_n \int f_n.$$

I3 (**Linearity**) If $f, g \in \mathbb{B}$ and $\alpha, \beta \in \mathbb{R}$, then

$$\int [\alpha f + \beta g] = \alpha \int f + \beta \int g,$$

as long as the right-hand side is not the indeterminate form $\infty - \infty$.

It is easy to see that there is a smallest set of functions from $\mathbb{R}^n$ to $\mathbb{R}$ satisfying [F1] and [F2]. Just take the intersection of all sets of functions satisfying [F1] and [F2], and verify that this intersection still satisfies [F1] and [F2]. Functions in this smallest set are called the *Baire functions* on $\mathbb{R}^n$, as they were first defined by René Baire(1899). The set of Baire functions can be constructed by transfinite recursion, omitted here. The set of Baire functions on $\mathbb{R}^n$ is in fact a proper subset of the set of all real-valued functions on $\mathbb{R}^n$. Also, it automatically satisfies [F3]—[F5]. The set of continuous functions satisfies [F3] and the Baire functions inherit the vector space property by transfinite reduction, and [F4] and [F5] can also be proved by transfinite induction. But we need not worry about that here. Just think of the Baire functions as the smallest set of functions closed under [F1]—[F5]. Since we want to be conservative and ask for the minimum that our requirements entail, *we will henceforth take* $\mathbb{B}$ *to be precisely the set of Baire functions.* The Baire functions are also closed under the operation of multiplication; that is, if f and g are Baire functions, then so is fg. This can also be proved by transfinite induction. For the moment we ask you to admit it as a fact (or you can add it as a required property [F6].) When we get around to full proofs we will avoid all arguments by transfinite induction.

Basic consequences of [F1]-[F5] and [I1]-[I3]: Part I

We don't know yet whether there exists a theory satisfying the requirements [F1]-[F5], [I1]-I2]. But this does not prevent us exploring their hypothetical consequences! In particular, what do they imply about measuring volumes of sets? To frame the question mathematically, define the class of sets:

$$\mathcal{B} \triangleq \{A \subset \mathbb{R}^n \mid \chi_A \in \mathbb{B}\},$$

and, for each set A in $\mathcal{B}$, define its volume

$$|A| \triangleq m(A) \triangleq \int \chi_A.$$

Here is what can first be said about $\mathcal{B}$ and m, followed by incomplete proofs. It is assumed throughout that $\int$ satisfies [I1]-[I3].

Lemma 1 *$\mathcal{B}$ is the smallest class of subsets of $\mathbb{R}^n$ such that $\mathcal{B}$ contains all open sets and*

B1 $\mathbb{R} \in \mathcal{B}$;

B2 $A \in \mathcal{B}$ *implies* $A^c \in \mathcal{B}$;

B3 *If* $A_1, A_2, \ldots$ *are all in* $\mathcal{B}$, *then* $\bigcup_1^\infty A_n \in \mathcal{B}$.

B4 *If* $A_1, A_2, \ldots$ *are all in* $\mathcal{B}$, *then* $\bigcap_1^\infty A_n \in \mathcal{B}$.

Lemma 2 *a) If V is the rectangle* $[a_1, b_1) \times \cdots \times [a_n, b_n)$ *in* $\mathbb{R}^n$,

$$m(V) = \int \chi_V = \prod_1^n (b_i - a_i) \tag{1}$$

b) Let $A_1, A_2, \ldots$ belong to $\mathcal{B}$ and assume they are pairwise disjoint—$A_i \cap A_j = \emptyset$ for all $i \neq j$. Then

$$m\left(\bigcup_1^\infty A_i\right) = \sum_1^\infty m(A_i). \tag{2}$$

Remarks: In equation (1) of Lemma 2 we could as well have let the intervals in the product defining V be open or closed, or any combination of open, closed, and half open intervals. We chose the form appearing in (1) to fix a standard choice that will be convenient later on. Part b) of Lemma 2 is reassuring—the volume of a countable disjoint union is just the sum of the volumes of the component sets.

Proof of Lemma 1(mostly!) Let U be any open set in $\mathbb{R}^n$. Let $\text{dist}(x, U^c) \triangleq \inf\{|x-y| \mid y \in U^c\}$ denote the function which to every x assigns its distance to U^c. This is a continuous function as is $\max\{1, n\text{dist}(x, U^c)\}$, for every positive integer n. Now for every x

$$\chi_U(x) = \lim_n \max\{1, n\text{dist}(x, U^c)\}$$

exhibits χ_U as a pointwise limit of continuous functions, so χ_U is a Baire function and $U \in \mathcal{B}$.

To see that [B1] is true, simply note that $\chi_{\mathbb{R}} \equiv 1$ is a continuous function and so $\chi_{\mathbb{R}}$ is in $\mathbb{B}$ by [F1].

To see that [B2] is true, note that if $A \in \mathcal{B}$, then $\chi_A \in \mathbb{B}$. By [F3], $\chi_{A^c} = 1 - \chi_A$ is also in $\mathbb{B}$ and hence $A^c \in \mathcal{B}$. The same reasoning shows

$$\text{if } A, B \in \mathcal{B} \text{ and } A \subset B, \text{ then } A - B \in \mathcal{B}. \tag{3}$$

Recalling that Baire functions are closed under multiplication (here is the only place we use this unproved fact), it follows that

$$A, B \in \mathcal{B} \quad \text{imply} \quad A \cap B \in \mathcal{B}, \tag{4}$$

because $\chi_{A\cap B} = \chi_A \chi_B$. Now if A and B are two sets in $\mathcal{B}$, it follows from (3) and (4) that $A - B = A - (A \cap B) \in \mathcal{B}$. Since $\chi_{A\cup B} = \chi_B + \chi_{A-B}$, [F2] implies, $A \cup B$ is in $\mathcal{B}$ also. By iteration $\mathcal{B}$ will be closed under finite unions.

Now consider a countable union $A = \cup_1^\infty A_i$ of sets in $\mathcal{B}$, since each finite union $\cup_1^n A_i$ is in $\mathcal{B}$ and since $\chi_A = \lim_n \chi_{\cup_1^n A_i}$, it follows again from [F2] that $A \in \mathcal{B}$. This proves [B3]. To prove [B4] use [B2], [B3], and the general identity $\cap_\lambda C_\lambda = (\cup_\lambda C_\lambda^c)^c$.

We have shown that $\mathcal{B}$ satisfies [B1]-[B4], but have yet to show it is the smallest such class containing open sets. We defer this, because it is more naturally proved later.

Proof of Lemma 2. This is a straightforward consequence of properties [I1] and [I2]. Let $A_1, A_2, \ldots$ be a sequence of pairwise disjoint sets in $\mathcal{B}$, and let $A \triangleq \cup_1^\infty A_i$. Let $f_n \triangleq \chi_{\cup_1^n A_i}$. By the disjointness of the sets $f_n = \sum_1^n \chi_{A_i}$, and so from linearity of the integration operation, property [I3],

$$\int f_n = \sum_1^n \int \chi_{A_i} = \sum_1^n m(A_i).$$

The functions f_n are non-negative, increase with increasing n, and $\lim_n f_n(x) = \chi_A$. Therefore, the monotone convergence property [I2] requires

$$\begin{aligned} m(A) &= \int \chi_{\cup_1^\infty A_i} = \lim_n \int f_n \\ &= \lim_n \sum_1^n m(A_i) = \sum_1^\infty m(A_i). \end{aligned}$$

Measures and σ-algebras

At this point we are ready to frame some abstract definitions to reformulate Lemmas 1 and 2.. Let $\mathcal{S}$ be an arbitrary but fixed nonempty set.

Definition 1 *A collection Σ of subsets of $\mathcal{S}$ is called a σ-algebra if*

S1 $S \in \Sigma$*;*

S2 $A \in \Sigma$ *implies* $A^c \in \Sigma$*;*

S3 *If* $A_n \in \Sigma$ *for every* n*,* $\bigcup_1^\infty A_i \in \Sigma$.

S4 *If* $A_n \in \Sigma$ *for every* n*,* $\bigcap_1^\infty A_i \in \Sigma$.

In words, a σ-algebra is a subset of a power set which is closed under complements, countable unions and countable intersections. Note that properties [S1], [S2], [S3] imply [S4], and properties [S1], [S2], [S4] imply [S3], so including both [S3] and [S4] in the definition is redundant; however, both are stated to emphasize that a σ-algebra is closed under both countable unions and countable intersections. Notice [S1]-[S4] simply generalize to an abstract setting the properties of [B1]-[B4] of $\mathcal{B}$ stated in Lemma 1. *Lemma 1 says that $\mathcal{B}$ is the smallest σ-algebra containing the open sets;* this will be explained further below.

Definition 2 *Let (S, Σ) be a non-empty set together with a σ-algebra of its subsets. A function μ which assigns to each set $A \in \Sigma$ a number in $[0, \infty]$ is called a* measure *if*

M1 $\mu(\emptyset) = 0$*;*

M2 *(Countable additivity) If $A_1, A_2, \ldots$ is a sequence of pairwise disjoint sets in Σ, then*

$$\mu\left(\bigcup_1^\infty A_i\right) = \sum_1^\infty \mu(A_i) \tag{5}$$

Observe that [M2] is just a generalization of the property (2) of Lemma 2, which was a consequence in turn of the requirement of monotone convergence for integration. Notice also, that if $A_1, \ldots, A_n$ is a finite sequence of pair-wise disjoint sets, then application of [M2] with $A_{n+1} = \emptyset, A_{n+2} = \emptyset$, etc., leads to finite additivity:

$$\mu\left(\bigcup_1^n A_i\right) = \sum_1^n \mu(A_i).$$

Moreover, if there exists at least one set A with $\mu(A) < \infty$, then application of finite additivity to $\mu(A) = \mu(A) + \mu(\emptyset)$, yields the property [M1], $\mu(\emptyset) = 0$.

Let us now summarize the results so far using the language of these new definitions. *A theory of integration satisfying properites [F1]-[F3] and [I1]-[I3] entails at least the existence of a countably additive measure m on the smallest σ-algebra containing the open sets, such that $m(V)$ is the usual volume of V for all rectangles V (see(1) in Lemma 2).*

Before continuing, we give some other examples of our newly defined objects and make some simple observations.

Examples of σ-algebras. Let S be non-empty.

1. The power set $\mathcal{P}(S)$ of S is a σ-algebra.

2. $\{S, \emptyset\}$ is a σ-algebra.

3. Let $S = B_1 \cup B_2 \cup \cdots \cup B_m$ and suppose that the B_i's are pairwise disjoint. Consider the collection $\mathcal{C}$ of all unions $\cup_{i \in K} B_i$, where K ranges over all subsets of $\{1, 2, \ldots, m\}$. This is also a σ-algebra. Since $\mathcal{C}$ is finite, any union or intersection of its elements reduces to a finite union or intersection. The sets $B_1, \ldots, B_m$ are called the *atoms* of $\mathcal{C}$.

Examples of measures

1. (Trivial measures) On $(S, \mathcal{P}(S))$, let $\rho(A) = 0$ for every $A \in \mathcal{P}$. Let $\nu(A) = \infty$ for every $A \in \mathcal{P}(S)$. Both ρ and ν are measures.

2. (Counting measure on a finite set) Let S be a finite set. On $(S, \mathcal{P}(S)$, let $\mu(A)$ be the number of elements in A.

3. (Finite atomic measures) Take a general non-empty set S with an arbitrary σ-algebra Σ of its subsets. Let $x_1, \ldots, x_n$ be distinct points in S. For each $A \in \Sigma$, let $\mu(A)$ be the number of points in the set $\{x_1, \ldots, x_n\} \cap A$.

We have described $\mathcal{B}$ as the smallest σ-algebra of subsets of $\mathbb{R}^n$ containing the open sets, but we have yet to show that this description is really meaningful. This we do next. Only the following observation is needed: if $\{\Sigma_\lambda \mid \lambda \in \Lambda\}$ is a nonempty family of σ-algebras of a fixed set S, then the intersection $\bigcap_{\lambda\in\Lambda} \Sigma_\lambda$ is again a σ-algebra. Now let $\mathcal{R}$ be a non-empty collection of subsets of S. The family $\Lambda_{\mathcal{R}} \triangleq \{\mathcal{C} \mid \mathcal{R} \subset \mathcal{C},\ \mathcal{C} \text{ is a } \sigma\text{-algebra}\}$ is nonempty because it contains $\mathcal{P}(S)$. Define

$$\sigma(\mathcal{R}) \triangleq \bigcap_{\mathcal{C}\in\Lambda} \mathcal{C}.$$

This is the smallest σ-algebra containing $\mathcal{R}$ and is called the σ-algebra generated by $\mathcal{R}$. Specifying σ-algebras by generating sets is very convenient. Beyond simple examples such as those above, it is in general not possible to give explicit constructions of σ-algebras except by transfinite recursion.

Definition 3 *In $\mathbb{R}^n$, the σ-algebra generated by the collection of open sets is called the Borel σ-algebra, and its members are called Borel sets.*

Using these definitions, Lemma 1 says that the class of sets whose indicators are Baire functions is the σ-algebra of Borel sets.

More Consequences and Abstract Definitions

In this section we return to considering integrals of functions $f \in \mathbb{B}$. To do this, it is useful to give another, simple characterization of $\mathbb{B}$, which will be introduced in an abstract setting. The essential facts behind the theory are the following identities, which say the operations of union and intersection commute with forming inverse images of a function. Let X and Y be any sets and let f be a function from X into Y. Let $\{A_\lambda \mid \lambda \in \Lambda\}$ be a family of subsets of Y. Then

$$f^{-1}\left(\bigcup_{\lambda\in\Lambda} A_\lambda\right) = \bigcup_{\lambda\in\Lambda} f^{-1}(A_\lambda) \tag{6}$$

$$f^{-1}\left(\bigcap_{\lambda\in\Lambda} A_\lambda\right) = \bigcap_{\lambda\in\Lambda} f^{-1}(A_\lambda) \tag{7}$$

$$f^{-1}(U^c) = \left(f^{-1}(U)\right)^c \tag{8}$$

These identities are easy to derive from the definitions of inverse image, union, and intersection.

The following abstract definitions are vitally important.

Definition 4 *a) Given a nonempty set S and a σ-algebra Σ of its subsets, a function $f : S \to \mathbb{R}$ is said to be Σ-measurable if $f^{-1}(U) = \{x \mid f(x) \in U\} \in \Sigma$ for every Borel set U contained in $\mathbb{R}$. (When the identity of Σ is clear from the context, one simply says "measurable" instead of "σ-measurable.)*
b) A measurable simple function on (S, Σ) is a function of the form $\sum i = 1^n c_i \chi_{A_i}$, where each c_i is a real number and each A_i belongs to Σ.
c) A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be Borel measurable if $f^{-1}(U)$ is a Borel subset of $\mathbb{R}^n$ for every Borel set U of $\mathbb{R}$.

It is easy to check that a measurable simple function is measurable.

The following lemmas show why these definitions are useful. To avoid repetition S always denotes a general non-empty set and Σ a σ-algebra of its subsets.

Lemma 3 *a) If f is a measurable function, there exists a sequence $\{f_n\}$ of measurable simple functions such that $\lim_n f_n(x) = f(x)$ for every $x \in S$.*
b) The set of measurable functions on (S, Σ) is closed under pointwise limits.

Lemma 4 *The set $\mathbb{B}$ of Baire functions on $\mathbb{R}^n$ is equal to the set of Borel measurable functions on $\mathbb{R}^n$.*

We shall defer proofs to the formal development of the theory, except for the proof of Lemma 3 a), since that is nice and simple, and important to what we want to see next about the integral. It suffices to show this claim for $f \geq 0$ because to prove it for arbitrary, write f as a difference of its positive and negative parts $f = f^+ - f^-$, and express each part as a limit of simple measurable functions.

We will show that, in fact, every nonnegative, measurable f is an *increasing limit* of simple measurable functions. Indeed, for every positive integer n, let

$$f_n(x) \triangleq \sum_{k=0}^{n2^n} k2^{-n} \chi_{f^{-1}((k/2^n,(k+1)/2^n])}(x).$$

In other words, if $\frac{k}{2^n} < x \leq \frac{k+1}{2}$ and $0 \leq k \leq n2^n$, set $f(x) = \frac{k}{2^n}$; otherwise, let $f(x) = 0$. Any interval $(k/2^n, (k+1)/2^n]$ is a Borel set in $\mathbb{R}$, and so, since f is measurable by assumption, $f^{-1}((k/2^n, (k+1)/2^n])$ is in Σ. Thus f_n is a measurable simple function. Clearly, if $0 \leq x \leq n$, $0 < f(x) - f_n(x) < 2^{-n}$, and so f is a pointwise limit of the sequence $\{f_n\}$. The reader should check that $f_n \leq f_{n+1}$ for every n.

Return to our hypothetical integral $\int$ on $\mathbb{R}^n$. Observe first that the integral of any Borel measurable simple function can be expressed in terms of the volume measure m. Indeed, by the linearity property [I3],

$$\int \sum_1^n c_i \chi_{A_i} = \sum_1^n c_i \int \chi_{A_i} = \sum_1^n c_i m(A_i) \tag{9}$$

Now let f be any nonnegative Borel measurable function and let $\{f_n\}$ be the sequence just constructed converging to f. Then we have from the monotone convergence principle [I2] and (9) that

$$\int f = \lim_n \int f = \lim_n \sum_{k=0}^{n2^n} \frac{k}{2^n} m \left(\{x \mid \frac{k}{2^n} < f(x) \leq \frac{k}{2^n}\} \right). \tag{10}$$

This formula expresses $\int f$ in terms of the measure m.

This completes the picture we wanted to draw of the consequences of [F1]-[F5] and [I1]-[I3]. We worked from our hypothetical integral to a measure on Borel sets and then, from the monotone convergence property, to formula (10). The idea of Lebesgue's theory is to work backward from measure to integral. First, construct a measure m on the Borel sets, or even a σ-algebra containing the Borel sets, such that for rectangles V $m(V)$ gives the usual volume. Once that is accomplished, *define* the integral by a formula like (10). The theory will be successful, if one can prove properties [I1] and [I2] from the new definition of integral.