Overview of Measure Theory

January 14, 2013

In this section, we will give a brief overview of measure theory, which leads to a general notion of an integral called the Lebesgue integral. Integrals, as we saw before, are important in probability theory since the notion of expectation or average value is an integral.

The ideas presented in this theory are fairly general, but their utility will not be immediately visible. However, a general note is that we are trying to quantify how "small" or how "large" a set is. Recall that if a bounded function has only finitely many discontinuities, it is still Riemann-integrable. The Dirichlet function was an example of a function with uncountably many discontinuities, and it failed to be Riemann-integrable. Is there a notion of "smallness" such that if a function is bounded except for a small set, then we can compute its integral? For example, we could say a set is "small" on the real line if it has at most countably infinitely many elements. Unfortunately, this is not enough to deal with the integral of the Dirichlet function. So we need a notion adequate to deal with the integral of functions such as the Dirichlet function.

We will deal with such a notion of integral in this section. It allows us to prove a fairly general theorem called the "Dominated Convergence Theorem", which does not hold for the Riemann integral, and is useful for some of the results in our course.

The general outline is the following - first, we deal with a notion of the class of *measurable sets*, which will restrict the sets whose size we can speak of. (We cannot speak of the size of any arbitrary set, for reasons beyond the scope of this course.) Second, we will deal with the notion of *measure*, which will be the generalized notion of size. Note that *probability* is just a finite measure, with the measure of the whole space being 1. Third, we will use the notion of measure to define what are *measurable functions*, which are the generalizations of continuous functions in the theory. Finally, we study the *Lebesgue integral* of these measurable functions.

1 σ-algebras and measurable sets

We motivate our discussion using probability. Suppose we consider an example like the throw of a die, and we can place bets on the events which occur. We can bet on individual outcomes like 3 or 4, but we could also bet on more sophisticated events like the occurrence of an even number, or the occurrence of a composite number. These sophisticated events are often unions or intersections or complements of simpler events. A theory of probability has to assign meaningful probabilities to these sophisticated events in a logically consistent manner - for example, a subset should have at most the probability of its superset.

We will first define which class of sets can we assign probabilities for. This will be defined using a class of sets, called a σ-algebra, and any of its member will be called an *event*.

Let Ω be the sample space of individual outcomes.

Definition 1.1. A set $\mathcal{S}$ of subsets of Ω is called an *algebra* if

1. $\emptyset, \Omega \in \mathcal{S}$.
2. $A, B \in \mathcal{S} \quad \Rightarrow \quad A \cap B, A \cup B, A \setminus B \in \mathcal{S}$.

Example 1.2. Suppose $\Omega = \{a, b\}$. Then $\{\emptyset, \Omega\}$ is an algebra, which is the smallest algebra possible. The only other algebra on Ω is $\{\emptyset, \{a\}, \{b\}, \{a, b\}\}$, the powerset of Ω.

Closure under finite unions and intersections is insufficient when we deal with several important probabilistic results. One example of such a result is called the law of large numbers which can be informally stated as follows. Suppose we consider the experiment of tossing a fair coin infinitely many times. Note that a single realization of this experiment involves infinitely many tosses. We would like to say that almost always, this experiment will yield heads half the time and tails the other half, whenever the number of coin tosses is large enough. Such a statement involves the probability of a countable intersection of events - namely, the probability that the first toss is a head, the probability that the second toss is a head, and so on.

So we possibly need closure under countable unions and intersections of sets.

Definition 1.3. An algebra $\mathcal{A}$ is called a σ-algebra if for any countably infinite collection of sets in $\mathcal{A}$, their union is also in $\mathcal{A}$. The pair $(\Omega, \mathcal{A})$ is called a *measurable space*.

This level of generality captures most of the interesting theorems in real analysis and probability theory, so we do not require any further closure properties (like closure under uncountable unions and uncountable intersections).

The elements of a σ-algebra are called *events*, in the context of probability theory. We will now proceed to assign sizes to each of the sets in the σ-algebra. The σ-algebra is quite a large collection of subsets, and since we want to assign a measure to as large a collection of subsets as possible, it is natural to enquire if every subset of Ω can be assigned a measure. The answer is yes if Ω is a finite, or a countably infinite set. But when we have an uncountably infinite set, it is impossible to assign measures to every subset. [1]

Definition 1.4. Let $(\Omega, \mathcal{F})$ be a measurable space. A function $\mu : \mathcal{F} \to [0, \infty]$ is called a *measure* if it satisfies the following conditions.

1. $\mu(\emptyset) = 0$.
2. $\forall A \in \mathcal{A} \qquad \mu(A) \geq 0$.
3. (Countable Additivity). If $A_1, A_2, \ldots$ is a countable collection of pairwise disjoint sets from $\mathcal{A}$, then

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

A *probability measure* on a space $(\Omega, \mathcal{A})$ is a measure such that $\mu(\Omega) = 1$.

The triple $(\Omega, \mathcal{A}, \mu)$ will be called a measure space or a probability space, as appropriate.

Example 1.5. Let $\Omega = \{a, b\}$ and $\mathcal{F}$ be the powerset of Ω. Then $\mu : \mathcal{F} \to [0, 1]$ defined as

$$\mu(\emptyset) = 0, \quad \mu(\{a\}) = \frac{1}{4}, \quad \mu(\{b\}) = \frac{3}{4}, \quad \mu(\{a, b\}) = 1,$$

is a probability measure. Note that if we specify $\mu(\{a\})$ and $\mu(\{b\})$, it will uniquely specify the probability measure. (End of example 1.5)

We will see a famous example where we will specify the probability on the σ-algebra by specifying it on a special kind of class of subsets of Ω. [2]

[1] The existence of nonmeasurable sets is a result beyond the scope of this course.

[2] We cannot pick any class of subsets for this purpose. It suffices, for example, for the class to be a *semialgebra*. We omit such details.

Example 1.6. In this example, we explain the Borel probability measure on the unit interval. Let $\mathcal{F}$ be the smallest σ-algebra containing all sets of the form $[a, b)$ where a and b are reals, together with the empty set. We define $\mu([a, b)) = |b - a|$. Then this defines a unique probability measure on $([0, 1], \mathcal{F})$. This is called the Borel probability measure. Note that $\mathcal{F}$ is a fairly large collection of sets, which contains not only intervals, but rather strange sets like the Cantor Set. These arise due to the fact that any set defined by arbitrary countable intersections, unions and complements of the basic intervals are present inside $\mathcal{F}$.

However, not every set is inside $\mathcal{F}$. An example is the Vitali Set, which has no measure in the space. (End of example 1.6)

We establish a property of any finite measure with respect to a countable intersection of a nested sequence of sets. This property will be used in future.

Lemma 1.7. *Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space. If $F_1, F_2, \dots$ is a sequence of sets from $\mathcal{F}$ with $E_i \supseteq E_{i+1}$ for all positive integral i, then*

$$\mu\left(\cap_{i=1}^{\infty} F_i\right) = \lim_{n\to\infty} \mu(F_i).$$

Proof. First, note that $\cap_{i=1}^{\infty} F_i \in \mathcal{F}$. Let us call this set F. Then,

$$F_1 = \cap_{i=1}^{\infty} \left(F_i - F_{i+1}\right) \cap F,$$

which is a disjoint union of sets. By countable addititivity of the measure,

$$\mu(F_1) = \sum_{i=1}^{\infty} (\mu(F_i - F_{i+1}) + \mu(F) = \sum_{i=1}^{\infty} (\mu(F_i) - \mu(F_{i+1})) + \mu(F) = \mu(F_1) - [\lim_{n\to\infty} \mu(F_i)] + \mu(F).$$

Hence the conclusion follows. $\square$

Exercise

1. Let $\Omega = \{a, b, c\}$. List all the algebras on Ω.
2. Show that for a probability measure, $\mu(A^c) = 1 - \mu(A)$. Show that if $A \subseteq B$ and $A, B \in \mathcal{A}$, then $\mu(A) \leq \mu(B)$.
3. Show that if $A_1, A_2 \in \mathcal{F}$, then $\mu(A_1 \Delta A_2) = 0$ implies that $\mu(A_1) = \mu(A_2)$.

2 Measurable Functions

Once we have defined a measure space $(\Omega, \mathcal{F}, \mu)$, the next step is to consider what functions we can study in the new theory. We would obviously want to study continuous functions, but the class should also be much more general than that. The definition that follows is important from the viewpoint of probability theory as well. The realization that the notion of measurable functions captures the notion of random variables in probability theory is one of the most important components of Kolmogorov's axiomatization of probability.

Definition 2.1. Let $(X, \mathcal{F}, \mu)$ and $(X, \mathcal{G}, \nu)$ be measure spaces. Then a function $f : X \to Y$ is called *measurable* if for every $G \in \mathcal{G}$, we have that $f^{-1}(G)$ is a member of $\mathcal{F}$.

Recall that $\mathcal{F}$ is the set of events of X and $\mathcal{G}$ is the set of events of Y. Thus measurable functions are those functions such that every event $G \in \mathcal{G}$ has an inverse image $f^{-1}G$ which is in $\mathcal{F}$.

(We may attempt to make this definition a bit clearer: the σ-algebra over a set X is the collection of "events". In some sense, this forms the collection of knowable information about X, since the events are those sets which we can assign probabilities to. A larger σ-algebra over X contains more knowable events than a smaller σ-algebra over it. Thus we do not allow f to "refine" the event structure in the domain, though we allow it to "preserve" or to "coarsen" the event structure in the range.)

A few examples of measurable and non-measurable functions will make this clear.

Example 2.2. (Indicator functions of measurable sets) Let $(X, \mathcal{F})$ be a measurable space. Let the range be $\{0, 1\}$ with the powerset of $\{0, 1\}$ as the σ-algebra. Let A be some set in $\mathcal{F}$.

Then the function $I_A : X \to \{0, 1\}$ defined by $I_A(x) = 1$ if $x \in A$ and 0 otherwise, is a measurable function with respect to the given spaces. It is easy to see that $f^{-1}(\{1\}) = A$, $f^{-1}(\{0\}) = A^c$, $f^{-1}(\{0, 1\}) = X$, $f^{-1}(\emptyset) = \emptyset$. In each case, the inverse image of a measurable set in the range is measurable in the domain.

I_A is usually called the indicator function of A. (End of example)

Example 2.3. (Discrete, finite case, measurable function) Let $X = \{0, 1, 2, 3\}$, and $\mathcal{F}$ be the powerset of X. Let $Y = \{a, b, c\}$, and $\mathcal{G}$ be its powerset.

Consider the function $f : X \to Y$ defined by $f(0) = a, f(1) = b, f(2) = c, f(3) = a$. This function is measurable with respect to the given measurable spaces: Consider an arbitrary subset G of Y. Then,

$$f^{-1}G \quad = \quad \{f^{-1}(y) \mid y \in G\}$$

is a subset of X. Since $\mathcal{F}$ contains all the subsets of X, the above set is clearly contained in $\mathcal{F}$. (End of Example)

Example 2.4. (Discrete, countably infinite case, measurable function) Let us consider $\mathbb{Q}$ as the domain, with $\mathcal{F}$ being the powerset of $\mathbb{Q}$. Let the range be the same as the domain.

Consider $f : \mathbb{Q} \to \mathbb{Q}$ defined by $f(q) = 1/q$, when q is nonzero and $f(0) = 0$. Then the inverse image under f, of any set of rationals is also a set of rationals. Hence f is measurable with respect to the given measurable spaces. (End of example)

Example 2.5. (Continuous case, measurable) Consider a function mapping points in $[0, 1] \to [0, 1]$. Let the σ-algbra on both the domain and the range be the Borel space on $[0, 1]$.

If f is continuous, then f is measurable with respect to the space. To see this, it is enough to show that the inverse image of any set of the form $[a, \infty)$ is a Borel-measurable set.

(*) $[a, \infty)$ is a closed set in the standard topology over the real line. Since f is continuous, its inverse image is a closed set in the real line. Every closed set in the standard topology is Borel-measurable, proving that f is Borel-measurable.

(End of example 2.5)

Example 2.6. (A non-measurable function with respect to the trivial σ-algebra) Let $\mathbb{R}$ be the domain, and $\mathcal{F} = \{\emptyset, \mathbb{R}\}$. Let the range be $[0, 1]$, and $\mathcal{F}$ be the Borel σ-algebra on $[0, 1]$, (see Example 1.6). Then the function $f : \mathbb{R} \to [0, 1]$ defined by $f(x) = 0$ for nonpositive x and $f(x) = 1$ for positive x, is not measurable. For, $f^{-1}(\{1\}) = (0, \infty)$, but the latter set is not a member of the σ-algebra on the domain.

Thinking about this example gives a glimpse into the definition of a measurable function. Such a function cannot give "more information" than what is present in the domain. (End of example 2.6)

Exercise

1. Let us consider $\mathbb{Q}$ as the domain, with $\mathcal{F}$ being the powerset of $\mathbb{Q}$, and μ an arbitrary probability measure. Let the range be $\mathbb{Q} \cup \{\infty\}$, the σ-algebra be the powerset of $\mathbb{Q}$ (the same as over the domain), and the probability measure over the range be μ itself. Is $f : \mathbb{Q} \to \mathbb{Q} \cup \{\infty\}$ defined by $f(q) = 1/q$ measurable with respect to the given spaces? If yes, say why. If no, give an example of a measurable set whose inverse is not.

2. † A set of numbers $S \subseteq [0,1]$ is said to have measure 0 if for any $\epsilon > 0$, there is a set of open intervals covering S such that the Borel measure of the entire collection of open intervals is less than ϵ.

   Show that the set of rationals in $[0,1]$ has measure 0.

   Hint: Rational numbers in $[0,1]$ form a countable set. Take an enumeration of rationals. Can you enclose each rational in an open interval such that the collection of open intervals enclosing the rationals has length less than ϵ?.

3. † Show that the Dirichlet function is measurable with respect to the Borel σ-algebra on the unit interval.

4. Show that every continuous function $f : [0,1] \to [-\infty, \infty]$ is measurable with respect to the Lebesgue measure. *Hint: See Example 2.5.*

3 Lebesgue Integration

In this section, we restrict ourselves to the Borel σ-algebra on the unit interval, and the uniform probability measure. We will consider the problem of integrating *measurable* functions. (If we consider the Borel measure on $[0,1]$, every continuous function is measurable.)

Lebesgue's insight is that instead of partitioning the domain of a function(as in the Riemann integral), we can partition the range of a function. An analogy, given by Lebesgue is that of a shopkeeper who wants to count how much (s)he has earned during the day. The Riemann integral is to count how much each customer brought in, and then to compute the sum over all customers. Instead, (s)he could also do the following: count the number of Rs. 1000 notes, Rs. 500 notes and so on. The second approach leads to a radical generalization of the notion of integral.

A slight elaboration of the idea before we get into the definition might be useful. Suppose $f : [0,1] \to (-\infty, \infty)$ is a measurable function. We know that each subinterval $[a,b]$ of the range space is Borel-measurable. Hence,

$$f^{-1}([a,b]) = \{x \in [0,1] \mid f(x) \in [a,b]\}$$

is also Borel-measurable in the *domain*. This follows from the fact that the inverse of a measurable set under a measurable function is always measurable. Now, to compute a *lower* approximation of the integral, we can partition the *range* into a sequence of disjoint intervals $[a_1, a_2]$, $[a_2, a_3]$, $\ldots$, and compute

$$\sum_{i=1}^{\infty} f_-^i \times \mu(f^{-1}[a_i, a_{i+1}]),$$

where f_-^i is the infimum of all values of f in $[a_i, a_{i+1}]$, for $i \in \mathbb{Z}^+$.

What we have gained here is that $\mu(f^{-1}[a_i, a_{i+1}])$ is defined, and we will not encounter issues similar to what the Riemann integral encounters with the Dirichlet function. This is the key idea from which we start the definition of the Lebesgue integral.

A function $f : [a,b] \to [-\infty, \infty]$ is called bounded (almost everywhere) if there is a constant C such that except for a set of measure 0, $f(x) < C$. That is,

$$m\{x \mid f(x) > C\} = 0.$$

For example, the function defined by $f(x) = 0$ for $x \in (0,1)$ and $f(0) = f(1) = \infty$ is bounded almost everywhere, since the set where the function attains infinity, $\{0,1\}$ has Borel measure 0. On the other hand, $x \mapsto 1/x$ defined on $[0,1]$ is an unbounded function. [3] In this course, we will need to deal only with the Lebesgue integral of functions bounded (almost everywhere) defined on a probability measure.

[3] For any constant $C > 1$, the function will be greater than C in the interval $[0, \frac{1}{C})$, and this set has measure $1/C$. If $0 < C \leq 1$, then f is greater than C in $[0,1]$, *i. e.* with probability 1.

We will now give a definition of the Lebesgue integral, in steps. First, we define the integral of a step function. This definition coincides with that of the Riemann integral. Then we define a new concept - a simple function, and its integral. This can already handle integrals beyond Riemann integration. [4] Finally, we will define the concept of the integral of a bounded measurable function. This is more general than the concept of a Riemann integral of a bounded function.

3.1 Lebesgue integral of a step function

Definition 3.1. (Integral of a step function). Consider a function $f : [a, b] \to \mathbb{R}$ defined as follows. Let $a = x_0 < x_1 < \cdots < x_n = b$ be real numbers. Let

$$f(x) = c_i \qquad \text{if } x \in [x_i, x_{i+1}],$$

where $c_0, c_1, \ldots, c_{n-1}$ are real numbers. Then

$$\int_a^b f(x)dx = \sum_{i=0}^{n-1} c_i(x_{i+1} - x_i).$$

3.2 Lebesgue integral of a simple function

Now, we define the notion of a simple function. Let E be a set of reals. The function χ_E defined by

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{otherwise ,} \end{cases}$$

is called the *characteristic function of E.* [5] If E is measurable, then χ_E is measurable with respect to the Borel σ-algebra. (Exercise 2) Note that the Dirichlet function is the characteristic function of the set of irrationals in $[0, 1]$.

Definition 3.2. A *simple function* is a linear combination of characteristic functions of measurable sets.

Thus, if $E_1, E_2, \ldots, E_n$ are a finite collection of disjoint measurable sets, then

$$f(x) = \sum_{i=1}^{n} a_i \; \chi_{E_i}(x), \qquad a_1, \ldots, a_n \in \mathbb{R}$$

is a simple function.

Since we are dealing with a finite measure, we define the integral of a simple function as follows.

Definition 3.3. if $E_1, E_2, \ldots, E_n$ are a finite collection of disjoint measurable subsets of $[a, b]$, and

$$f(x) = \sum_{i=1}^{n} a_i \; \chi_{E_i}(x), \qquad a_1, \ldots, a_n \in \mathbb{R}$$

is a simple function, then

$$\int_a^b f(x)dx = \sum_{i=1}^{n} a_i \; m(E_i),$$

where $m(E)$ is the measure of the set E.

[4] At this stage, the power of Lebesgue and Riemann integrals are incomparable - for example, Riemann integrals are defined for continuous functions, which are not simple functions.

[5] This is also called the *indicator function of E*.

3.3 Lebesgue integral of a bounded measurable function

Let $f : [a, b] \to \mathbb{R}$ be a bounded function on a finite measure and E be a measurable subset of $[a, b]$. We could examine "lower approximations" to the integral and "upper approximations" to the integral of f, as in Riemann integration. Unlike Riemann integral, we approximate using f simple functions.

Let ψ and ϕ denote simple functions. Consider

$$\inf_{\psi\,:\,f<\psi} \int_E \psi,$$

the least upper approximation to the area under f, using simple functions ψ. Also, consider

$$\sup_{\phi\,:\,\phi<f} \int_E \phi,$$

the greatest lower approximation to the area under f, using simple functions ϕ. We can ask, for which functions f do these coincide. This is the great result that leads us to Lebesgue integration.

Lemma 3.4. *Let f be bounded on a measurable set E of finite measure. Then*

$$\inf_{\psi\,:\,f<\psi} \int_E \psi \qquad = \qquad \sup_{\phi\,:\,\phi<f} \int_E \phi$$

for all simple functions ϕ and ψ if and only if f is measurable.

We omit the proof of this lemma, and use it to define the notion of the Lebesgue integral of a bounded measurable function.

Definition 3.5. If f is a bounded measurable function defined on a measurable set E of finite measure, the *Lebesgue integral of f over E* is

$$\int_E f dx = \inf_{\psi\,:\,f\le\psi} \int_E \psi(x)dx,$$

where the infimum is taken over all simple functions ψ lower bounded by f.

Exercise

1. † Show that the Dirichlet function is integrable with respect to the Borel-measure on $[0, 1]$, and compute its integral over the unit interval.
2. Prove that if E is Borel-measurable, then χ_E is measurable with respect to the Borel σ-algebra.

4 Convergence Theorems for the Lebesgue integral

Suppose we have a sequence of functions f_n tending to f as $n \to \infty$. Then we could ask, does the sequence $\int f_1 dx$, $\int f_2 dx$, $\ldots$ tend to $\int f dx$? Note that if f_ns are restricted to simple functions, then the answer is yes, since $\int f$ is defined to be the suprema of such approximations by simple functions. What if f_ns are bounded measurable functions? Does the integral of their limit function f exist?

> This kind of a question will be important in our study of the optimality of compressors - for example, $f_n(x)$ could be the compression ratio achieved on the first n bits of x. If the average compression ratio over n-long strings is finite for every n, then is the average compression ratio over infinite sequences well-defined?

We will consider a famous result called the Dominated Convergence Theorem, which we will use in the course. For this, temporarily, we will consider bounded nonnegative functions. The approach is to upper bound each f_n by some bounded measurable g_n in such a way that the limit function g of the sequence g_n is integrable. This will ensure that f is integrable. Let us consider an example to see what exactly this process is.

Consider $\sum_{n=1}^{\infty} \frac{1}{n^{1.5}}$. We want to know whether this series converges. Unfortunately, it is not easy to get a closed form for the sum, and we may not know whether such a form exists. So we approach the problem as follows.

$$\begin{aligned} 1 + \left(\frac{1}{2^{1.5}} + \frac{1}{3^{1.5}}\right) + \left(\frac{1}{4^{1.5}} + \frac{1}{5^{1.5}} + \frac{1}{6^{1.5}} + \frac{1}{7^{1.5}}\right) + \dots \\ &\leq 1 + \left(\frac{2}{2^{1.5}}\right) + \left(\frac{4}{4^{1.5}}\right) + \dots \\ &= 1 + \frac{1}{2^{0.5}} + \frac{1}{4^{0.5}} + \dots \\ &= \frac{1}{1 - (1/2)^{0.5}} < \infty. \end{aligned}$$

We have bounded the infinite sum by considering an upper bound to the sum whose closed form was easier to obtain and analyze. Thus, we have established that our original series converges.

More minutely, if we say that

$$f_n = \sum_{i=1}^{2^n - 1} \frac{1}{i^{1.5}},$$

then

$$g_n = \sum_{i=1}^{n} \frac{2^i}{2^{i \times 1.5}}$$

is an upper bound for each n. g_n converges to

$$g = \sum_{i=1}^{\infty} \frac{1}{2^{i \times 0.5}},$$

which is an convergent infinite geometric series. Hence the limit of f_n, *viz.* the series

$$\sum_{n=1}^{\infty} \frac{1}{n^{1.5}}$$

is convergent. (We do not know what the value is, but we know that it is finite.)