24 Measure and Integration

This section supplies a minimal introduction to measure and integration theory. It is not an overview or a summary of measure theory. We will not explicitly construct the Lebesgue measure on the real line and the most important results for applications will be stated without proofs. We believe that by introducing a few fundamental notions of the theory we will make it possible to understand and use the concepts and theorems in measure and integration theory, as we will need them for our subsequent study of L^p spaces. For a more complete but concise introduction to this subject we recommend E.H. Lieb and M. Loss, *Analysis*, Graduate Studies in Mathematics volume 14, American Mathematical Society, 1997. A standard reference for the basics of measure theory is P.R. Halmos, *Measure Theory*, Graduate Texts in Mathematics volume 18, Springer Verlag, 1974.

24.1 Measures

In order to be able to define integrals of real-valued functions defined on a domain Ω, we need the notion of measure. Indeed, intuitively, we would like the integral of a characteristic function $\chi_A : \Omega \to \mathbb{R}$, for some subset $A \subset \Omega$, to be equal to the size of A, given by the *measure*, $\mu(A)$. A measure on Ω is a function defined on subsets of Ω, and for us its values will always be non-negative.

It turns out that it is not a good idea to try to define the size (measure) of arbitrary subsets of say $\Omega = \mathbb{R}$. Some sets are just too wild to be *measurable*. So, μ will be defined on a subset of all subsets of Ω. Because of the interpretation of $\mu(A)$ as the size of A, μ will be defined on a *sigma-algebra* Σ, which is defined as follows.

Definition 24.1 (sigma-algebra) *A set of subsets Σ of a set Ω is a sigma-algebra if it satisfies the following properties:*
i) $\emptyset \in \Sigma$,
ii) $A \in \Sigma \Rightarrow A^c := \Omega \setminus A \in \Sigma$,
iii) For any countable family $\{A_i\}$ of elements in Σ, $\bigcup_i A_i \in \Sigma$.

In other words, a sigma-algebra is a set of subsets of Ω that contains the empty set and that is closed under taking complements and countable unions. It follows that a sigma-algebra also contains Ω itself, and that it is closed under countable intersections as well.

Just as in the case of defining a topology (as a set of open sets satisfying the appropriate axioms) the most common way of defining a sigma-algebra is to specify a set of subsets, say $\mathcal{A}$, and consider the sigma-algebra $\Sigma(\mathcal{A})$ generated by it, i.e., $\Sigma(\mathcal{A})$ is the smallest sigma-algebra of subsets of Ω that contains $\mathcal{A}$. One can show that there is a unique such sigma-algebra.

Examples:

1. As any sigma-algebra should contain $\emptyset$, the smallest possible sigma-algebra of subsets of Ω is $\Sigma(\emptyset) = \{\emptyset, \Omega\}$.

2. The largest sigma-algebra of subsets of Ω is just the collection of *all* subsets.

3. The previous two examples are not particularly interesting. For most Ω, e.g., $\Omega = \mathbb{R}$, the first one is too small to be usefull and the second one is too large to be practical. For any topological space $(X, \mathcal{T})$, such as $\mathbb{R}^n$ with its standard topology, there is a natural sigma-algebra, called the *Borel sigma-algebra*, often denoted by $\mathcal{B}(X)$, defined as the smallest sigma-algebra of subsets of X that contains all the open subsets of X :
$$\mathcal{B}(X) = \Sigma(\mathcal{T})$$
E.g.,
$$\mathcal{B}(\mathbb{R}) = \Sigma(\{(a,b) \mid a < b\})$$
The elements of a Borel sigma-algebra are called *Borel sets.* Of course $\{(a,b) \mid a < b\}$ is only a subset of all open subsets of $\mathbb{R}$, but it is easy to see that this set is enough to generate all of the Borel sets.

Exercise 24.1 *Prove that*

$$\mathcal{B}(\mathbb{R}) = \Sigma(\{(a,b] \mid a < b\}) = \Sigma(\{[a,b) \mid a < b\}) = \Sigma(\{[a,b] \mid a < b\}) = \Sigma(\{(a,+\infty) \mid a \in \mathbb{R}\})$$

Definition 24.2 (measure) *Let Σ be a sigma-algebra of subsets of Ω. A measure on Ω is a function*
$$\mu : \Sigma \to [0, +\infty]$$
with the property of countable additivity, which, by definition, means that for any countable family $\{A_i\}$ of mutually disjoint elements of Σ, i.e., $A_i \cap A_j = \emptyset$, whenever $i \neq j$, we have
$$\mu(\bigcup_i A_i) = \sum_i \mu(A_i) \tag{24.1}$$

Note that (24.1) makes sense because Σ is closed under taking countable unions. This property is sometimes relaxed to just finite disjoint unions, but we will always assume countable additivity of any measure we consider. The elements of Σ are called *measurable sets*, for obvious reasons.

If $\mu(\Omega) = 1$, μ is called a *probability measure*, and the triple (Ω, Σ, μ) is referred to as a *probability space.* In general, a *measure space* is a triple (Ω, Σ, μ), where Σ is a sigma-algebra of subsets of Ω, and μ is a measure on Ω.

Examples:

1. Consider an arbitrary set Ω, e.g., $\Omega = \mathbb{N}$, with the sigma-algebra consisting of all subsets of Ω. The *counting measure* on Ω is defined by

$$\mu(A) = \text{ the number of elements of } A$$

with the convention that if A is an infinite set, $\mu(A) = +\infty$. Check that this is a measure.

2. Let $\Omega = \{1, 2, \ldots, n\}$ be a finite set, and take again Σ to be the set of all subsets of Ω. All the probability measures on Ω are then determined by n non-negative numbers, $p_i \geq 0, i = 1, \ldots, n$, satisfying $p_1 + \cdots + p_n = 1$, and the measure is determined by

$$\mu(\{i\}) = p_i$$

In other contexts *signed measures*, which take arbitrary values in the reals, or *complex-valued measures*, are useful to consider. Here, we will always assume the measure is non-negative. Note that some sets A, may have infinite measure, i.e, $\mu(A) = +\infty$.

Exercise 24.2 *Let Σ be a sigma-algebra of subsets of Ω, and suppose μ is a measure on Ω. Prove the following properties:*
a) *If $A, B \in \Sigma$, then $A \setminus B \in \Sigma$.*
b) *If $A, B \in \Sigma$, and $A \subset B$, then $\mu(A) \leq \mu(B)$.*
c) *If $A, B \in \Sigma$, then $\mu(A \cup B) \leq \mu(A) + \mu(B)$.*

Another property that we will always assume is σ-finiteness: μ is *σ-finite* if there exists an at most countable family $\{A_i\}$ of measurable subsets of Ω such that

$$\bigcup_i A_i = \Omega, \qquad \text{and } \mu(A_i) < \infty$$

Exercise 24.3 *Is the counting measure on $[0, 1]$ sigma-finite? Prove or disprove.*

Theorem 24.3 *Suppose $\Sigma = \Sigma(A)$ is the sigma-algebra generated by the elements of $\mathcal{A}$, and suppose μ_1 and μ_2 are two measures such that*

$$\mu_1(A) = \mu_2(A), \qquad \textit{for all } A \in \mathcal{A}$$

and suppose there is an at most countable family of sets A_i in $\mathcal{A}$, such that $\bigcup_i A_i = \Omega$, and $\mu_1(A_i) < \infty$, then $\mu_1 = \mu_2$.

The above theorem gives a useful sufficient condition for a measure to be uniquely determined on all of $\Sigma(\mathcal{A})$ by its values on $\mathcal{A}$.

Examples:

1. Let $f : \mathbb{R} \to \mathbb{R}$ be a non-negative Riemann-integrable function. There is then a unique measure μ on the Borel sigma-algebra of $\mathbb{R}$ satisfying

$$\mu((a, b)) = \int_a^b f(x)\, dx$$

2. By choosing $f \equiv 1$ in the previous example we obtain a unique measure on $\mathbb{R}$ with the Borel sigma-algebra such that

$$\mu((a, b)) = |b - a|$$

Here, we have not explained why such a measure should exists at all, but the theorem above implies there can be at most one. One can prove the existence of this measure in Example 2) above by construction but it is not easy. It is called the *Lebesgue measure* on $\mathbb{R}$, or more precisely, it is the restriction of the Lebesgue measure to the Borel sigma-algebra. This is because it is convenient to have the property that any subset B of a measurable set A with $\mu(A) = 0$, is measurable and, then, of course, $\mu(B) = 0$. The sigma-algebra which contains all Borel sets, and all subsets of sets of measure zero is the standard sigma-algebra of subsets of $\mathbb{R}$ on which the Lebesgue measure is defined. If you are interested in the details, check out an analysis textbook, e.g., W. Rudin, *Real and Complex Analysis*, McGraw Hill, 3rd edition, 1987, or, G.B. Folland, *Real Analysis*, John Wiley & Sons, 1984.

24.2 Integration

Assuming that you are familiar with the *Riemann integral*, we will now define a general notion of integration of real- and complex-valued functions, called the *Lebesgue integral*, which includes the Lebesgue integral of functions on $\mathbb{R}$ as a special case.

Suppose Σ is a sigma-algebra of subsets of Ω, and $\mu : \Sigma \to [0, +\infty]$ is a measure. A function $f : \Omega \to \mathbb{R}$, is called *measurable* if all the sets $S_f(t)$, $t \in \mathbb{R}$, defined by

$$S_f(t) = \{x \in \Omega \mid f(x) > t\}$$

are measurable, i.e., $S_f(t) \in \Sigma$, for all $t \in \mathbb{R}$. The sets $S_f(t)$ are called the *level sets* of f.

Example:
Consider $f : \mathbb{R} \to \mathbb{R}$, defined by

$$f(x) = \frac{1}{1+x^2}$$

Then

$$\begin{array}{rcl} S_f(t) = & \emptyset & \text{if } t \geq 1 \\ S_f(t) = & (-\sqrt{\frac{1}{t}-1}, +\sqrt{\frac{1}{t}-1}) & \text{if } 0 < t < 1 \\ S_f(t) = & \mathbb{R} & \text{if } t \leq 0 \end{array}$$

Let, at first, $f : \Omega \to \mathbb{R}$, be a non-negative measurable function. Following Lieb & Loss, we will define $\int f\,\mu(dx)$ as the Riemann integral of the function $F(t)$ defined by

$$F(t) = \mu(S_f(t)) \tag{24.2}$$

F is well-defined because f is measurable. From the properties of Exercise 24.2, it is obvious that $F(t)$ is a non-negative, non-increasing monotone function of t. Therefore, its Riemann integral is well-defined (allowing for the value $+\infty$ for the integral). Hence, we can define the *Lebesgue integral* of f by

$$\int_\Omega f(x)\,\mu(dx) := \int_0^{+\infty} F(t)\,dt \tag{24.3}$$

Give this formula a minute thought and you will realize that this is a very natural definition of the integral. One can prove that if f is itself Riemann integrable, its Riemann integral and its Lebesgue integral as defined in (24.3) conincide.

In the case of $\Omega = \mathbb{R}$, and with μ the Lebesgue measure, we have defined the standard Lebesgue integral of any non-negative Lebesgue-measurable function on $\mathbb{R}$.

Example:
For $f(x) = 1/(1+x^2)$, $F(t) = 2\sqrt{\frac{1}{t}-1}$. So, in this case, $F = 2f^{-1}$, and it is then obvious that

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_0^{+\infty} F(t)\,dt$$

A non-negative function f is called *integrable* (or *summable*), if f is measurable and

$$\int f(x)\mu(dx) < \infty$$

A function $\Omega \to \mathbb{C}$ is defined to be measurable if the four non-negative functions $f_1, \ldots, f_4$ appearing in the decomposition

$$f = f_1 - f_2 + i f_3 - i f_4$$

are all measurable. If $f_1, \ldots, f_4$ are all integrable, f is called integrable and the integral of f is then defined by linear extension:

$$\int_\Omega f \mu(dx) := \int_\Omega f_1 \, \mu(dx) - \int_\Omega f_2 \, \mu(dx) + i \int_\Omega f_3 \, \mu(dx) - i \int_\Omega f_4 \, \mu(dx)$$

Note that, for measurable f, f is integrable iff $|f|$ is integrable.

24.3 Sets of measure zero, almost everywhere, and all that

We have introduced the Borel sigma-algebra $\mathcal{B}(\mathbb{R})$ of subsets of $\mathbb{R}$. It can be shown that there is a unique measure λ defined on $\mathcal{B}(\mathbb{R})$, with the following two properties:

1. For every measurable set A, and $a \in \mathbb{R}$ one has $\lambda(A + a) = \lambda(A)$ (translation invariance of λ);
2. $\lambda([0, 1]) = 1$

The standard Lebesgue measure on $\mathbb{R}$ is an extension of λ to a sigma-algebra Σ that contains $\mathcal{B}(\mathbb{R})$, but that also has the property that if $A \in \Sigma$, $\lambda(A) = 0$, and $B \subset A$, then $B \in \Sigma$ and $\lambda(B) = 0$. A measurable set A with the property $\lambda(A) = 0$, is called a *set of measure zero.* So, for the Lebesgue measure any subset of a set of measure zero is a set of measure zero. This is a convenient property to have. The Borel sigma-algebra is too small in the sense that it does not contain necessaritly all subsets of a set of measure zero, i.e., not all subsets of a set of measure zero are necessarily measurable. We will use the same symbol λ for the measure discussed above extended to Σ, and called it the *(standard) Lebesgue measure.*

Examples:

1. Any finite subset $A \subset \mathbb{R}$ has measure zero. Indeed, for any $a, b \in \mathbb{R}$, $\lambda(\{a\} = \lambda(\{b\})$, by the translation invariance of the measure. Then, for any n we can find $a_1 < a_2 < \cdots a_n \in [0, 1]$, for which we then have

$$n\lambda(\{a_1\}) = \sum_{i=1}^{n} \lambda(\{a_i\} = \lambda(\{a_1, \ldots, a_n\} \leq \lambda([0, 1]) = 1$$

where we have used the elementary properties of measures and the translation invariance of λ. As, n is arbitrary, it follows that $\lambda(\{a_1\}) = 0$, and hence $\lambda(\{a\} = 0$, for all $a \in \mathbb{R}$.

2. Any countable subset of $\mathbb{R}$ has measure zero. This is a consequence of the previous example and the countable additivity of measures. If $A = \{a_i\}_{i=1}^{\infty}$, we have
$$\lambda(A) = \lambda(\bigcup_{i=1}^{\infty}\{a_i\}) = \sum_{i=1}^{\infty} \lambda(\{a_i\}) = 0$$
E.g., the set $\mathbb{Q}$ of all rational numbers has measure zero. It follows that the irrational numbers between 0 and 1 have measure 1.

3. There are also sets of measure zero that are uncountably infinite. An example is the Cantor set.

The expression *almost everywhere*, abbreviated as *a.e.*, means "with possible exception of a set of measure zero". E.g for two functions f and g, $f = g$ a.e., means that there is a set of measure zero E, such that $f(x) = g(x)$, for all $x \in \Omega \setminus E$. If necessary the measure μ for which this exceptional set has measure zero may be indicated by writing μ-a.e. A sequence of measurable funtions (f_n) converges pointwise μ-a.e. to f, if there is a measurable set E with $\mu(E) = 0$, and such that
$$\lim_{n\to\infty} f_n(x) = f(x), \qquad \text{for all } x \in \Omega \setminus E$$
It can be show that the a.e. pointwise limit of a sequence of measurable functions is measurable. Strictly speaking, for this to make sense one has to (re)define f on E by hand in a suitable way, e.g., put $f(x) = 0$ for all $x \in E$.

From the definition of the integral it is easy to see that if f and g are integrable functions and $f = g$ μ-a.e., then $\int f(x)\,\mu(dx) = \int g(x)\,\mu(dx)$.

In probability theory almost everywhere is ususally called *almost surely.*

A related concept is the *essential supremum* of a set of real numbers $A \subset \mathbb{R}$: indexsupremum!essential
$$\text{ess sup}A := \inf\{C > 0 \mid |x| \leq C, \text{ for a.e. } x \in A\}$$

24.4 The Monotone and Dominated Convergence Theorems

By now, you know from experience that convergent sequences of functions are at the core of the majority of the arguments in analysis. Therefore, the following theorems about integrals of sequences of functions are extremely important. In this section the notions of measurability, integrability etc. will refer to an arbitrary measure space (Ω, Σ, μ), and we will implicitly assume that μ is sigma-finite. Functions are $\Omega \to \mathbb{C}$ unless otherwise mentioned.

The easiest to prove of the convergence theorems is the Monotone Convergence Theorem.

Theorem 24.4 (Monotone Convergence Theorem) *Let (f_n) be a sequence of non-negative functions, converging pointwise and in a monotone increasing way to a function f, i.e.,*

$$f_n(x) \leq f_{n+1}(x), \quad \text{for all } n, \text{ and } \lim_{n\to\infty} f_n(x) = f(x), \quad \text{for a.e. } x$$

Then

$$\lim_{n\to\infty} \int f_n(x)\,\mu(dx) = \int f(x)\,\mu(dx) \tag{24.4}$$

Equality also holds when the LHS is $+\infty$. In this case it means that f is measurable and non-negative but has an infinite integral in the sense of (24.3).

While it is obvious that the same result holds for a decreasing sequence of non-positive functions converging pointwise, (24.4) is generally false for arbitrary pointwise converging sequences or even decreasing sequences of non-negative functions.

Exercise 24.4 *Give an example of a sequence of non-negative integrable functions converging pointwise and monotonically to a function f, but such that the equality (24.4) does not hold.*

Exercise 24.5 *Prove Theorem 24.4 by reducing (24.4) to the case of Riemann integrable functions. Hint: first, show that the sequence of functions (F_n), where F_n is related to f_n as in (24.2), is monotone increasing. Use the countable additivity of the measure to show that $F_n \to F$, where F is also given by (24.2).*

For arbitrary pointwise converging sequences of non-negative integrable functions there is a inequality instead of the equality (24.4). This is a consequence of Fatou's Lemma.

Theorem 24.5 (Fatou's Lemma) *Let (f_n) be a sequence of non-negative integrable functions. Then $f(x) := \liminf_{n\to\infty} f_n(x)$ defines a measurable function and*

$$\liminf_{n\to\infty} \int f_n(x)\,\mu(dx) \geq \int f(x)\,\mu(dx)$$

The most useful of the convergence theorems is the Dominated Convergence Theorem. It does not require the functions to be non-negative nor the convergence to be monotone.

Theorem 24.6 (Dominated Convergence Theorem) *Let (f_n) be a sequence of integrable functions, converging to f pointwise a.e.. Suppose there exists an integrable function g such that $|f_n(x)| \leq g(x)$, μ-a.e., then*

$$\lim_{n\to\infty} \int f_n(x)\,\mu(dx) = \int f(x)\,\mu(dx)$$

Proof: First, show that it is sufficient to prove the statement for non-negative functions f_n. Then, apply Fatou's Lemma 24.5 twice, once to the sequence (f_n), and once to the sequence $(g - f_n)$. ■

Exercise 24.6 *Fill in the details of the proof of the Dominated Convergence Theorem.*

24.5 Fubini's Theorem

The Lebesgue measure on $\mathbb{R}^n$ defined on the Borel sigma-algebra of $\mathbb{R}^n$ (or, again, an extension of it with lots of sets of measure zero), is the unique measure with the properties of translation and rotation invariance and such that the measure of the unit cube is 1. We denote it again by λ. It is the product of the Lebesgue measure on n copies of $\mathbb{R}$ in the sense that, if $A_i \in \mathcal{B}(\mathbb{R})$, then $A = A_1 \times \cdots \times A_n \in \mathcal{B}(\mathbb{R}^n)$ and

$$\lambda(A) = \lambda(A_1)\lambda(A_2)\cdots\lambda(A_n)$$

This is a particular case of a general construction of a measure $\mu = \mu_1 \times \mu_2$ on $\Omega = \Omega_1 \times \Omega_2$, given any two measure spaces $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$. Take for Σ the sigma-algebra of subsets of Ω generated by the subsets of the form $A_1 \times A_2$, where $A_1 \in \Sigma_1$ and $A_2 \in \Sigma_2$, and put

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

It can be shown that μ extends uniquely to a measure on all of Σ. It is called the *product measure* of μ_1 and μ_2.

Example:
Let μ_1 and μ_2 be two probability measures on Ω_1 and Ω_2 respectively. In the case of probability measures, the measurable subsets are called *events*. So, if A_1 is a measurable subset of Ω_1 and A_2 is a measurable subset of Ω_2, then $\mu_1(A_1)$ and $\mu_2(A_2)$ are the probabilities with which these events occur. One says that A_1 and A_2 are *independent* if the probability that A_1 and A_2 both happen is the product of their respective probabilities. If all events in Ω_1 are independent of all events in Ω_2, the measure on $\Omega = \Omega_1 \times \Omega_2$ describing their joint probabilities must then satisfy

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2),$$

i.e., μ is the product measure $\mu_1 \times \mu_2$.

We are now faced with an apparent ambiguity when we integrate functions of more then one variable, say $f : \mathbb{R}^2 \to \mathbb{C}$. Do we use the Lebesgue measure on $\mathbb{R}^2$ to integrate $f(x, y)$, or do we first fix one of the variables, say x and integrate over y using the Lebesgue measure on $\mathbb{R}$, which yields a function of x that we can then integrate by using the Lebesgue measure on $\mathbb{R}$ again? Or the other way around, first integrate over x, then over y? Fortunately there is a theorem that makes this question irrelevant, as it should be.

Theorem 24.7 (Fubini's Theorem) *Suppose μ_1 and μ_2 are two sigma-finite measures on Ω_1 and Ω_2 respectively. Let f be a measurable function defined on $\Omega_1 \times \Omega_2$. If f is integrable for $\mu = \mu_1 \times \mu_2$, i.e.,*

$$\int |f(x, y)| \, \mu(dx\, dy) < \infty$$

then

$$\begin{aligned} & \int f(x, y) \, \mu(dx\, dy) \\ = \quad & \int \left[\int f(x, y) \, \mu_1(dx) \right] \mu_2(dy) \\ = \quad & \int \left[\int f(x, y) \, \mu_2(dy) \right] \mu_1(dx) \end{aligned}$$

The equality also holds if f is a non-negative measurable function with an infinite integral.