Exercise Class 6

Measurable Functions

With the notion of measurable sets at hand, we now turn our attention to the concept at heart of integration theory: measurable functions.

1 Definition

The notion of measurable function has some basic properties in common with another importannt class of functions, namely continuous ones. There are some analogies between the concepts of *topological spaces*, *open sets*, and *continuous functions* on one hand, and *measurable spaces*, *measurable sets*, and *measurable functions* on the other. These similarities become most apparent in an abstract setting. With this motivation, we introduce the following definition.

Definition 1. Let $(X, \mathcal{A}_X)$ and $(Y, \mathcal{A}_Y)$ be two measurable spaces. A map $\varphi : X \to Y$ is called $\mathcal{A}_X$-$\mathcal{A}_Y$-*measurable* if

$$\varphi^{-1}(A) = \{x \in X : \varphi(x) \in A\} \in \mathcal{A}_X \text{ for all } A \in \mathcal{A}_Y.$$

With this definition, one can make the following analogies with topology.

Topology	**Measure Theory**
A *topological space* (X, τ) is a set X endowed with a topology τ	A *measurable space* $(X, \mathcal{A})$ is a set X endowed with a σ-algebra $\mathcal{A}$
A set $O \subseteq X$ is *open* if $O \in \tau$	A set $A \subseteq X$ is *measurable* if $A \in \mathcal{A}$ (†)
A function $f : (X, \tau_X) \to (Y, \tau_Y)$ is *continuous* if $f^{-1}(O) \in \tau_X$ for every $O \in \tau_Y$	A function $f : (X, \mathcal{A}_X) \to (Y, \mathcal{A}_Y)$ is *measurable* if $f^{-1}(A) \in \mathcal{A}_X$ for every $A \in \mathcal{A}_Y$

Remark. (†) We slightly misuse the notion of measurable set. The above is not quite in line with Carathéodory's criterion for measurability. However, one often considers σ-algebras as domains of measures as the measures behave more nicely on σ-algebras — they satisfy σ-additivity. The σ-algebras thus contain the sets whose measure we can and want to define and thus we call those sets measurable; we do this without intending to imply that it is not possible to assign a measure to other sets. But if we choose $\mathcal{A}$ to be the σ-algebra of measurable sets, then the statement is trivially true.

In the lecture, we will (most likely) encounter a slightly less general version of measurability. We will mostly consider real-valued functions $f : X \to \overline{\mathbb{R}}$, i.e., functions that take not only finite values but also the values $+\infty$ and $-\infty$. We now assume that $\overline{\mathbb{R}}$ is equipped with the Borel σ-algebra $\mathcal{B}(\overline{\mathbb{R}})$, i.e., $(Y, \mathcal{A}_Y) = (\mathbb{R}, \mathcal{B}(\overline{\mathbb{R}}))$. Then the measurability of a function f means that the preimage of an arbitrary Borel set $B \subseteq \overline{\mathbb{R}}$ is measurable.

If a function of this sort is defined on a measure space (X, Σ, μ), where Σ is the collection of μ-measurable sets, then the function is also called μ-*measurable*. We state this properly in the following definition.

Definition 2. A function $f : X \to \overline{\mathbb{R}}$ is called μ-measurable, if

i) $f^{-1}(\{+\infty\})$ and $f^{-1}(\{-\infty\})$ are μ-measurable,

ii) $f^{-1}(B)$ is μ-measurable for every Borel set $B \subset \mathbb{R}$.

The following remark demonstrates that the definition of measurability of real-valued functions can be formulated in a much simpler form. For brevity, we denote the set $f^{-1}((-\infty, a)) = \{x \in \mathbb{R} \mid f(x) < a\}$ by $\{f < a\}$ — similarly for $\{f \leq a\}, \{f = a\}$, etc.

Remark. Condition ii) is equivalent to the following:

iii) $f^{-1}(U)$ is μ-measurable for every open set $U \subset \mathbb{R}$.
(Analogously for closed sets.)

iv) $\{f < a\}$ is μ-measurable for every $a \in \mathbb{R}$.
(Analogously for $\{f \leq a\}, \{f > a\}$, $\{f \geq a\}$, and plenty of other sets of this form.)

Proof. Exercise 6.2. □

Conditions ii) and iii) can actually be used for any topological space (Y, τ) and not just $\mathbb{R}$ equipped with the standard topology.

Condition iv) seems particularly useful to prove measurability of functions. Let us consider a few examples — the first of which will be useful later.

Example 3. 1. Consider the characteristic function of a set $A \subset X$ given by

$$\chi_A(x) = \begin{cases} 1 & \text{for } x \in A, \\ 0 & \text{for } x \notin A. \end{cases}$$

Observe that for each $a \in \mathbb{R}$,

$$\{\chi_A < a\} = \begin{cases} \varnothing & \text{for } a \leq 0, \\ A^c & \text{for } 0 < a \leq 1, \\ X & \text{for } a > 1. \end{cases}$$

It follows immediately that χ_A is a measurable function if and only if A is a measurable set.

2. Suppose now that f is a measurable function. Then $|f|$ is measurable as well, as

$$\{|f| < a\} = \begin{cases} \varnothing & \text{for } a \leq 0, \\ \{f < a\} \cap \{f > -a\} & \text{for } a > 0. \end{cases}$$

The sets on the r.h.s. are measurable by assumption.

3. Let g be another measurable function. We now prove that $f + g$ is measurable. Let $(q_k)_{k \in \mathbb{N}}$ be a sequence of rational numbers. We show that

$$\{f + g < a\} = \bigcup_{k=1}^{\infty} \left(\{f < q_k\} \cap \{g < a - q_k\}\right)$$

for all $a \in \mathbb{R}$.

Proof. Let $x \in \{f + g < a\}$. Then $f(x) < a - g(x)$. Hence there exists $q_k \in \mathbb{Q}$ such that $f(x) < q_k < a - g(x)$. Therefore $x \in \{f < q_k\} \cap \{g < a - q_k\}$ and thus x belongs to the r.h.s.

Now suppose that x belongs to the r.h.s. Then we have $x \in \{f < q_k\} \cap \{g < a - q_k\}$ for at least one k. That is, $f(x) < q_k$ and $g(x) < a - q_k$ and thus $f(x) + g(x) < a$. This concludes the proof. □

4. Lastly, we show that for f and g measurable, their maximum $\max(f, g)$ is measurable too. This follows directly by
$$\{\max(f, g) < a\} = \{f < a\} \cap \{g < a\}.$$

2 Necessity for Integration Theory

With the notion of measurable functions at hand, we can start our construction of integrals. The starting point is the notion of characteristic functions. The next step is to pass to the functions that are the building blocks of integration theory. For the Riemann integral it is the class of *step functions*, with each given as a finite sum
$$f = \sum_{k=1}^{N} a_k \chi_{R_k},$$
where each R_k is a rectangle and the a_k are constants.

For the Lebesgue integral, we use a slightly more general notion. We use so-called *simple functions*, which are given as
$$f = \sum_{k=1}^{N} a_k \chi_{E_k},$$
where each E_k is a measurable set of finite measure and the a_k are constants. To get rid of some technical complications, we assume that the E_k are disjoint and the representation of a simple function is unique. It is clear from the previous examples that simple functions are measurable. This now allows us to define the integral.

If f is a simple function with unique form $f(x) = \sum_{k=1}^{N} a_k \chi_{E_k}(x)$, then we define the *Lebesgue integral* of f by
$$\int_{\mathbb{R}^d} f(x) d\mu(x) = \sum_{k=1}^{N} a_k \mu(E_k),$$
where we denote by μ the Lebesgue measure.

If A is a measurable set of $\mathbb{R}^d$ with finite measure, then $f(x)\chi_A(x)$ is also a simple function and we define
$$\int_E f(x) dx = \int_{\mathbb{R}^d} f(x)\chi_E(x) d\mu(x).$$

Defining the integral for simple functions is fairly straightforward. The following theorem allows us to consider an integral for all measurable functions (although we need a few more assumptions to have well-definedness).

Theorem 4. *Suppose f is μ-measurable on $\mathbb{R}^d$. Then there exists a sequence of simple functions $(f_k)_{k=1}^{\infty}$ that satisfies*
$$\left|f_k(x)\right| \le \left|f_{k+1}(x)\right| \quad \text{and} \quad \lim_{k\to\infty} f_k(x) = f(x) \quad \forall\, x.$$
In particular, we have $\left|f_k(x)\right| \le \left|f(x)\right|$ for all x and k.

Thus we can approximate measurable functions as monotone limits of simple functions. It is now not quite obvious how to define an integral for measurable functions, however, this shall give us a solid basis to understand what is going to happen in the lecture.

Intuition Let us now compare the building blocks of the Riemann and Lebesgue integral to get a grasp on the differences between the two integrals.

When we calculate a Riemann integral, we usually take an infinitesimal step in the domain (usually $\mathbb{R}$), this can be thought of as an infinitesimally small rectangle which is usually denoted by dx. We thus approximate f using step functions.

In Lebesgue integration, we do a different thing. In some sense, we take infinitesimal steps along the in codomain and then we ask: For which x is the function f equal to y? That is, we want to determine the size of the preimage of y under f. To be more precise;

$$df^{-1}(y) = \{x : y \leq f(x) \leq y + dy\}.$$

Integration now means that we multiply the size of this set (i.e., its measure) by the corresponding value $f(x)$ and then sum the results for all y. However, the set $f^{-1}(y)$ can only be measured if the function f is measurable. This goes to show the necessity of the notion of measurability and yields (at least some sort of) an intuitive reason for the definition using preimages. Using measurable sets instead of just rectangles gives us more freedom in our approximations since we are not restricted by the structure of the domain $\mathbb{R}$ anymore.

This sort of computation is very similar to the idea of taking expectations of a random variable X. One considers the possible values X can take and then multiplies said value with the size of $X^{-1}(x) = \{X = x\}$. Then one takes the sum over all possible values of X. As it turns out, this is not a coincidence. From an abstract point of view, taking expectations is the same thing as integrating.

3 A Link to (Abstract) Probability Theory

Measurable functions are fundamental objects in probability theory — they are just better known under a different name.

Definition 5. Let $f : X \to \mathbb{R}$ be a μ-measurable function. If μ is a probability measure, then f is also called a *random variable*.

To be consistent with the notation from probability theory we call our domain $(\Omega, \mathcal{F}, P)$ instead of (X, Σ, μ) and we denote the random variable by X instead of f.

The following theorem shows that a measurable function can transfer a measure on one space to a measure on another space.

Theorem 6. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a random variable. Then $\nu : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ defined by*

$$\nu(A) := P(X^{-1}(A)),\ A \in \mathcal{B},$$

is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Proof. Clearly, we have that $\nu(\varnothing) = 0$. Next, let (A_n) be a countable family of disjoint Borel sets. Then

$$\nu(\bigcup_n A_n) = P(X^{-1}(\bigcup_n A_n)) = P(\bigcup_n X^{-1}(A_n)) = \sum_n P(X^{-1}(A_n)) = \sum_n \nu(A_n).$$

□

Definition 7. The measure PX^{-1} is called *pushforward measure* or *image measure*

The name follows from the fact, that one obtains the measure by transferring ("pushing forward") a measure to another using a measurable function.

Remark. The same proof works for measurable maps between any two measurable spaces. Thus Theorem 5 holds in a much more general setting.

In fact, the pushforward measure PX^{-1} is the *probability distribution* or the *law* of the random variable X. The function $F : \mathbb{R} \to \mathbb{R}$ defined by

$$F(x) = P(\underbrace{X^{-1}(-\infty, x]}_{=\{X \leq x\}}), \quad x \in \mathbb{R}$$

is the *distribution function* or *cdf* of X.

The most useful property of the pushforward measure is the so-called *change-of-variables* formula. This is a crucial step in the generalization of probability theory to abstract probability spaces.

Theorem 8. *A measurable function $g : \mathbb{R} \to \mathbb{R}$ is integrable with respect to the pushforward measure PX^{-1} if and only if the composition $g \circ X$ is integrable with respect to the measure P. Then the integrals coincide, i.e.,*

$$\int_\Omega (g \circ X)(\omega) dP(\omega) = \int_\mathbb{R} g(x) dPX^{-1}(x).$$

Since we have yet to define the integral of an arbitrary measure properly, we omit the proof for the moment.