1 Measurable function cont.

1.1 Convergence of measurable functions

Definition 1.1 (Almost everywhere / Almost surely)**.** We use the short hand *almost everywhere* (or *almost surely* in a probability space) to discuss properties that are true everywhere except a measure zero set.

Definition 1.2 (Convergence almost everywhere)**.** Let $(E, \mathcal{E}, \mu)$ be a measureable space. A sequence of measureable functions, $(f_n)_{n \geq 1} : E \to F$, *converges almost everywhere* to f if

$$\mu\left(\{x \in E \ : \ f_n(x) \nrightarrow f(x)\}\right) = 0$$

Definition 1.3 (Convergence in measure)**.** Let $(E, \mathcal{E}, \mu)$ be a measureable space. A sequence of real valued measureable functions, $(f_n)_{n \geq 1} : E \to F$, *converges in measure* to f if for every $\epsilon > 0$

$$\mu\left(\{x \ : \ |f(x) - f_n(x)| > \epsilon\}\right) \to 0, \quad \text{as } n \to \infty.$$

Example 1.4. The sequence of functions $f_n(x) = x^n$ converges to 0 Lebesgue almost everywhere on $[0, 1]$, and in measure, but it doesn't converge pointwise as it doesn't converge at $x = 1$.

Example 1.5. The- sequence of functions $f_n(x) = 1_{[n,n+1]}(x)$ converges to 0 Lebesgue almost everywhere (in fact everywhere) but not in measure.

Example 1.6. Consider the sequence of functions $f_1 = 1_{[0,1/2)}, f_2 = 1_{[1/2,1)}, f_3 = 1_{[0,1/4)}, f_4 = 1_{[1/4,1/2)}, f_5 = 1_{[1/2,3/4)}, f_6 = 1_{[3/4,1)}, f_7 = 1_{[0,1/8)}, f_8 = 1_{[1/8,1/4)} \dots$ then f_n converges to 0 in measure, but $f_n(x)$ does not converge for any x.

We can prove a quasi-equivalence between these two notions of measure. Before we do this we need to introduce a very useful lemma, the Borel-Cantelli Lemma. We introduce it here as it is used to prove the following theorem but it is a useful tool to have whilst doing measure theory, particularly probability theory. First let us also introduce some more notation

Definition 1.7. Let $(A_n)_n$ be a sequence of measurable sets then we have the following names

$$\limsup_n A_n = \bigcap_n \bigcup_{m \geq n} A_m = \{A_n \text{ infinitely often }\},$$

and

$$\liminf_n A_n = \bigcup_n \bigcap_{m \geq n} A_m = \{A_n \text{ eventually }\}.$$

The last names are more comon when the A_n are events in a probability space.

Lemma 1.8 (First Borel-Cantelli Lemma)**.** *Let $(E, \mathcal{E}, \mu)$ be a measure space. Then if $\sum_n \mu(A_n) < \infty$ it follows that $\mu(\limsup_n A_n) = 0)$.*

Proof. For any n we have

$$\mu(\limsup_n A_n) \leq \mu\left(\bigcup_{m \geq n} A_m\right) \leq \sum_{m \geq n} \mu(A_m).$$

Then the right hand side goes to zero as $n \to \infty$, so $\mu(\limsup_n A_n) = 0$. $\qquad \square$

Lemma 1.9 (Second Borel-Cantelli Lemma)**.** *Let $(E, \mathcal{E}, \mu)$ be a probability space ($\mu(E) = 1$). Then suppose that $\mu(A_i \cap A_j) = \mu(A_i)\mu(A_j)$ (the events are pairwise independent) for every i, j and that $\sum_n \mu(A_n) = \infty$ then it will follow that $\mu(\limsup_n A_n) = 1$.*

Proof. First we note that $\mu(A_i^c \cap A_j^c) = \mu((A_i \cup A_j)^c) = 1 - \mu(A_i \cup A_j) = 1 - \mu(A_i) - \mu(A_j) + \mu(A_i)\mu(A_j) = (1 - \mu(A_i))(1 - \mu(A_j))$. We use the inequality $1 - a \le e^{-a}$. Let $a_n = \mu(A_n)$ then

$$\mu\left(\bigcap_{m=n}^{N} A_m^c\right) = \Pi_{m=n}^{N}(1 - a_m) \le \exp\left(-\sum_{m=n}^{N} a_m\right) \to 0, \quad \text{as } N \to 0.$$

Therefore, $\mu\left(\bigcap_{m \ge n} A_m^c\right) = 0$ for every n. So $\mu(\limsup_n A_n) = 1 - \mu(\bigcup_n \bigcap_{m \ge n} A_m^c) = 1$. $\qquad\square$

Theorem 1.10. *Let $(E, \mathcal{E}, \mu)$ be a measure space and $(f_n)_n$ be a sequence of measurable functions. Then we have the following:*

- *Suppose that $\mu(E) < \infty$ and that $f_n \to 0$ almost everywhere, then $f_n \to 0$ in measure.*

- *If $f_n \to 0$ in measure then there exists some subsequence $(n_k)_k$ such that $f_{n_k} \to 0$ almost everywhere.*

Proof. Suppose that $f_n \to 0$ almost everywhere. Then

$$\mu(\{x : |f_n(x)| \le \epsilon\}) \ge \mu\left(\bigcap_{m \ge n}\{x : |f_m(x)| \le \epsilon\}\right) \uparrow \mu(|f_n| \le \epsilon \text{ eventually}) \ge \mu(\{x : f_n(x) \to 0\}) = \mu(E),$$

therefore,
$$\mu(\{x : |f_n(x)| > \epsilon\}) = \mu(E) - \mu(\{x : |f_n(x)| \le \epsilon\}) \to 0.$$

Now suppose that $f_n \to 0$ in measure. We can find a subsequence n_k such that

$$\mu(\{x : |f_{n_k}(x)| > 1/k\}) \le 2^{-k}.$$

Therefore

$$\sum_k \mu(\{x : |f_{n_k}(x)| > 1/k\}) < \infty.$$

Therefore by the first Borel-Canteli lemma we have that

$$\mu\left(\{x : |f_{n_k}(x)| > 1/k \text{ infinitely often}\}\right) = 0.$$

Therefore $f_{n_k} \to 0$ almost everywhere. $\qquad\square$

1.2   Egoroff's Theorem and Lusin's Theorem

Theorem 1.11 (Egoroff's Theorem). *Let $(E, \mathcal{E}, \mu)$ be a finite measure space and $(f_n)_n$ be a sequence of real valued measurable functions on E. If μ is finite and if f_n converges μ-almost everywhere to f then for each positive ϵ there is a set A with $\mu(A^c) < \epsilon$, such that f_n converges uniformly on A to f.*

Proof. For each n let $g_n(x) = \sup_{j \ge n} |f_j(x) - f(x)|$. Then g_n is a positive function which is finite almost everywhere. The sequence $(g_n)_n$ converges to 0 almost everywhere and so in measure. Therefore, for each positive integer k we can find n_k such that

$$\mu\left(\{x : g_{n_k} > 1/k\}\right) < \epsilon 2^{-k}.$$

Define sets $A_k = \{x : g_{n_k}(x) \le 1/k\}$ and let $A = \bigcap_k A_k$. The set A has

$$\mu(A^c) = \mu\left(\bigcup_k A_k^c\right) \le \sum_k \mu(A_k^c) \le \sum_k \epsilon 2^{-k} = \epsilon.$$

We want to show that f_n converges uniformly to f on A. For each δ there exists a k such athat $1/k < \delta$, then as $A \subseteq A_k$, we have that for every $n \geq n_k$,

$$|f_n - f| \leq g_{n_k} \leq 1/k < \delta,$$

uniformly on all $x \in A_k$ and hence in A. $\qquad\square$

This proof motivates the following definition of convergence of functions.

Definition 1.12 (Almost uniform convergence). We say a sequence of functions $(f_n)_{n \geq 1}$ converges *almost uniformly* on a measure space $(E, \mathcal{E}, \mu)$ if for every $\epsilon > 0$ there exists a set A with $\mu(A^c) < \epsilon$ with $f_n \to f$ uniformly on A.

We can use Egoroff's theorem to prove a result called Lusin's theorem. First let us recall the definition of regularity

Definition 1.13. Let E be a topological space and μ be a measure on $(E, \mathcal{B}(E))$ then say μ is *regular* if for every $A \in \mathcal{B}(E)$ we have

- $\mu(A) = \inf\{\mu(U) \ : \ A \subseteq U, U \text{ is open}\}$,

- $\mu(A) = \sup\{\mu(K) \ : \ K \subseteq A, K \text{ is compact}\}$.

Theorem 1.14 (Lusin's Theorem). *Suppose that f is a measurable function and $A \subseteq \mathbb{R}^d$ is a Borel set and $\lambda(A) < \infty$ then for any $\epsilon > 0$ there is a compact subset K of A with $\lambda(A \setminus K) < \epsilon$ such that the restriction of f to K is continuous.*

Remark 1.15. This theorem can be generalised to locally compact Hausdorff spaces, see Cohn's book.

Proof. Suppose first that f only takes countably many values, $a_1, a_2, a_3, \ldots$ on A the let $A_k = \{x \in A \ : \ f(x) = a_k\}$, by measurablility of f we can see that $A_k = f^{-1}(\{a_k\})$ is measurable. We know that $A = \bigcup_n A_n$ so by continuity of measure $\lambda(\bigcup_{k=1}^n A_k) \uparrow \lambda(A)$. Since $\lambda(A) < \infty$ we have that for any $\epsilon > 0$ there exists n such that $\lambda(A \setminus \bigcup_{k=1}^n A_k) < \epsilon/2$. By the regularity of Lebesgue measure we can find compact subsets $K_1, \ldots, K_n$ such that $\lambda(A_n \setminus K_n) \leq \epsilon/2n$. Then let $K = \bigcup_{k=1}^n K_k$. This is a compact subset of A and

$$\lambda(A \setminus K) \leq \lambda(A \setminus \bigcup_{k=1}^n A_k) + \lambda(\bigcup_{k=1}^n A_k \setminus \bigcup_{k=1}^n K_k) < \epsilon/2 + \epsilon/2.$$

Now f restricted to K is continuous since the K_i are disjoint and f is constant on each K_i.

Now we have proved the special case where f takes countably many values we can use this to prove the theorem for general f. Let $f_n = 2^{-n}\lfloor 2^n f \rfloor$ then $2^{-n} \geq f(x) - f_n(x) \geq 0$ so $f_n(x) \to f(x)$. Now by Egoroff's theorem there exists K with $\lambda(A \setminus K) < \epsilon/2$ such that $f_n \to f$ uniformly on K. Now, f_n can only take countably many values, so by our special case of Lusin's theorem there exists a $K_n \subseteq K$, compact, such that $\lambda(K \setminus K_n) \leq \epsilon 2^{-n-1}$, and f_n is continuous on K_n. Now let $K_\infty = \bigcap_n K_n$, then K_∞ is compact and $\lambda(A \setminus K_\infty) = \lambda(A \setminus K) + \lambda(K \setminus K_\infty) = \lambda(A \setminus K) + \lambda(\bigcup_n (K \setminus K_n)) \leq \epsilon/2 + \sum_n \epsilon 2^{-n-1} = \epsilon$. Now we have that f_n converges uniformly to f on K_∞ and f_n is continuous on K_∞ for each n. As the uniform limit of continuous functions is continuous this shows that f is continuous on K. $\qquad\square$

2   Integration

We now get to the definition of the Lebesgue integral which is the second important object that we construct in this course. There are several different notations for the integral of a function f with respect to a measure μ. We have

$$\mu(f) = \int_E f \mathrm{d}\mu = \int_E f(x)\mu(\mathrm{d}x).$$

When you are integrating with respect to Lebesgue measure the most common notation is

$$\int_E f(x)\mathrm{d}x.$$

Before we start constructing the integral we'll briefly discuss the motivations for how to construct it. Firstly, you've already seen the Riemann integral. We can describe the strategy of Riemann integration - very loosly - as splitting the *domain* of the function into equal sized chunks, estimating the height of the function on each chunk then adding them together. Broadly what happens with Lebesgue integration is that we split the *range* of the function into equal sized chunks, estimate the size of the part of the domain which will end up in that chunk of range then sum everything up. We need the theory of measure in order to do this because the bit of the domain corresponding to chuncks of the range can be quite weird sets whose size it wouldn't be possible to measure. The first motivation for this is that whilst Riemann integration only works for functions from subsets of $\mathbb{R}^d$ to $\mathbb{R}$, Lebesgue integration allows the domain on the function to be quite weird, (as long as it is a measure space). As an example, this is helpful for taking expectations rigorously because expectations are integral of random variables and the domain of a random variable is a probability space which may not be explicit.

The second big motivation for introducing a new theory of integration is the issue of convergence. It is important in many practical applications of integration theory to know when $\lim_n \int f_n(x)\mathrm{d}x = \int \lim_n f_n(x)\mathrm{d}x$ or when $\int_{E_x} \int_{E_y} f(x,y)\mathrm{d}x\mathrm{d}y = \int_{E_y} \int_{E_x} f(x,y)\mathrm{d}y\mathrm{d}x$. Lebesgue integration allows us to rigorously find conditions on f under which these statements will be true. This is often not possible in a satisfactory way with the Riemann theory of integration. We will see some of these convergence theorems next week and then switching the order of integration towards the end of the course (currently planned for week 9). The most important motivation for developing good convergence theorems was the development of Fourier series. We want to know when it is possible to integrate a Fourier series term by term.

The strategy for constructing the integral is to begin by defining $\mu(f)$ when f belongs to a special class of measurable functions that we call *simple functions*. We then define the integral to progressively larger classes of functions.

Definition 2.1 (Simple functions). Let $(E, \mathcal{E}, \mu)$ be a measure space. The set of simple functions on this space taking values in $\mathbb{R}$ are functions of the form

$$f(x) = \sum_{k=1}^{n} a_k 1_{A_k}(x).$$

Here, the A_k are disjoint sets in $\mathcal{E}$, 1_A represents the indicator function of the set, and the a_k are non-negative real numbers. We note that this representation of f is not unique.

Definition 2.2 (The integral of a simple function). Still working in the setting above, let $f(x) = \sum_{k=1}^{n} a_k 1_{A_k}(x)$, then we can define

$$\mu(f) = \sum_{k=1}^{n} a_k \mu(A_k).$$