Math 312 - Analysis 1

Lectures by Marianna Csörnyei
Notes by Zev Chonoles

The University of Chicago, Fall 2012

Introduction

Math 312 is one of the nine courses offered for first-year mathematics graduate students at the University of Chicago. It is the first of three courses in the year-long analysis sequence.

These notes were live-TeXed, though I edited for typos and added diagrams requiring the TikZ package separately. I used the editor TeXstudio.

I am responsible for all faults in this document, mathematical or otherwise; any merits of the material here should be credited to the lecturer, not to me.

Please email any corrections or suggestions to chonoles@math.uchicago.edu.

Acknowledgments

Thank you to all of my fellow students who sent me corrections, and who lent me their own notes from days I was absent. My notes are much improved due to your help.

Lecture 1 (2012-10-02)

The course will cover a mixture of real analysis and probability. Homeworks will be due on the Thursday of the following week. Homeworks will be 10% of the grade and the midterm and final together will be the other 90%. The exams will be in class.

The abstract setting for measure theory is as follows. We have a set X, with power set $P(X)$. A subset $\mathcal{A} \subseteq P(X)$ is called a σ-algebra when

1. $\varnothing \in \mathcal{A}$,
2. for any $A \in \mathcal{A}$, its complement $(X \setminus A) \in \mathcal{A}$, and
3. for any $A_1, A_2, \ldots, \in \mathcal{A}$, their union $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Some immediate consequences are:

- We also have $X \in \mathcal{A}$.
- $\mathcal{A}$ is also closed under countable intersections; for any $A_1, A_2, \ldots \in \mathcal{A}$, $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$. This is because
$$X \setminus \bigcap_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (X \setminus A_n).$$
- If $A, B \in \mathcal{A}$, then $A \setminus B \in \mathcal{A}$ because
$$A \setminus B = A \cap (X \setminus B).$$

The most extreme cases of σ-algebras are $\{\varnothing, X\}$ and $P(X)$.

A more interesting example is when X is a topological space and $\mathcal{A}$ is the σ-algebra generated by all open sets of X. This is called the Borel σ-algebra on X.

The σ-algebra generated by a collection of sets is the smallest σ-algebra that contains all of those sets. This can be constructed by considering all of the σ-algebras containing those sets, and taking their intersection. The intersection of σ-algebras can easily be seen to be a σ-algebra.

Let's consider a fundamental topological space, $\mathbb{R}$. What can we say about Borel sets in $\mathbb{R}$?

- Open sets are Borel.
- Closed sets, being complements of open sets, are Borel.
- Countable intersections of open sets (called G_δ sets) are Borel.
- Countable unions of closed sets (called F_σ sets) are Borel.

Some examples of F_σ sets that are neither open nor closed are $[0, 1)$ and $\mathbb{Q}$. Their complements are necessarily G_δ sets, and will also be neither open nor closed.

Homework. Is $\mathbb{Q}$ a G_δ set?

Continuing our list of Borel sets in $\mathbb{R}$,

- Countable unions of G_δ sets (called $G_{\delta\sigma}$ sets) are Borel.
- Countable intersections of F_σ sets (called $F_{\sigma\delta}$ sets) are Borel.

- Countable intersections of $G_{\delta\sigma}$ sets (called $G_{\delta\sigma\delta}$ sets) are Borel.
- Countable unions of $F_{\sigma\delta}$ sets (called $F_{\sigma\delta\sigma}$ sets) are Borel.
- $\cdots$

It is a theorem that each of these new classes is strictly bigger than the previous one; there is no finite step when we get no new sets. We do not even get all Borel sets when we look at all sequences of δ's and σ's. We must go all the way to ω_1 (the first uncountable ordinal) to get all Borel sets. Most of the time, we do not think beyond the first few steps here, but in descriptive set theory this is studied in more detail.

A pair $(X, \mathcal{A})$ of a set X together with a σ-algebra $\mathcal{A}$ on X is called a measure space (more precisely, a measurable space). The elements of $\mathcal{A}$ are called $\mathcal{A}$-measurable sets, or just measurable sets if the σ-algebra is understood.

Homework. What is the σ-algebra generated by the half-open intervals $[a, b)$? How is it related to the Borel σ-algebra - smaller, bigger, not comparable?

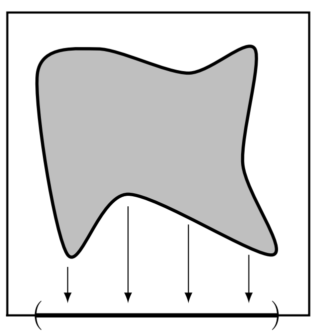
Definition. Given two measurable spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$, a map $f : X \to Y$ is called measurable if $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$.

Since it is difficult to understand what the Borel sets of $\mathbb{R}$ are, this would seem to be a difficult condition to check. But in fact, when $(Y, \mathcal{B}) = (\mathbb{R}, \text{Borel})$, a map $f : X \to \mathbb{R}$ is $\mathcal{A}$-measurable if

$$\{x \in X \mid f(x) < c\} \in \mathcal{A} \quad \text{and} \quad \{x \in X \mid f(x) > c\} \in \mathcal{A}$$

for all $c \in \mathbb{R}$. This is because the open half-lines already generate the entire Borel σ-algebra. More generally, if $\mathcal{C}$ is a collection of subsets of Y that generate the σ-algebra $\mathcal{B}$, then a map $f : X \to Y$ is measurable if $f^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$.

What about the image of Borel sets, instead of preimages? Is the image of a Borel set Borel?



The projection of an open set is open, and the projection of a countable union is the countable union of the projections. However, as you can see in the above image, the projection of the complement need not be the complement of the projection. Lebesgue famously made the mistake of assuming the projection of any Borel set is Borel, but in fact there is a G_δ set in $[0, 1]^2$ whose projection is not Borel.

Given a measurable space $(X, \mathcal{A})$, and Borel measurable functions $f_1, f_2 : X \to \mathbb{R}$, then

$$f_1 + f_2, \quad f_1 - f_2, \quad f_1 \cdot f_2, \quad f_1/f_2$$

are all Borel measurable (the last, of course, under the assumption that $f_2(x) \neq 0$ for all x).

Proof. The function $f_1 + f_2$ can be obtained as the composition

$$X \xrightarrow{F=(f_1,f_2)} \mathbb{R} \times \mathbb{R} \xrightarrow{+} \mathbb{R}$$

The composition of measurable functions is measurable, so it suffices to show that F and $+$ are measurable. The function $+$ is measurable (in fact, it is continuous). Because sets of the form $G_1 \times G_2$, where $G_1, G_2 \subseteq \mathbb{R}$ are open, generate the Borel σ-algebra on $\mathbb{R} \times \mathbb{R}$, it is enough to show that $F^{-1}(G_1 \times G_2) \in \mathcal{A}$ for any open $G_1, G_2 \subseteq \mathbb{R}$. But this is clear, because

$$F^{-1}(G_1 \times G_2) = f_1^{-1}(G_1) \cap f^{-1}(G_2). \qquad \square$$

Definition. The extended real line is $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. The open balls around $+\infty$ are sets of the form $(a, \infty]$ and the open balls around $-\infty$ are sets of the form $[-\infty, a)$.

Homework. If $f_1, f_2, \ldots : X \to \mathbb{R}$ are Borel measurable, prove that $g = \sup(f_n) : X \to \overline{\mathbb{R}}$ is Borel measurable.

Lecture 2 (2012-10-04)

Definition. Given a set X and a σ-algbera $\mathcal{A}$ on X, a function $\mu : \mathcal{A} \to [0, \infty]$ is a measure if

1. $\mu(\varnothing) = 0$
2. (σ-additivity) For disjoint sets $A_1, A_2, \ldots$,

$$\sum_{n=1}^{\infty} \mu(A_n) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right)$$

Examples.

- The counting measure on X: for any σ-algebra $\mathcal{A} \subseteq P(X)$, we define $\mu(A) = |A|$.
- The Dirac measure: given an $x \in X$,

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

- An atomic measure is one of the form $\mu = \sum c_j \delta_{x_j}$, so that

$$\mu(A) = \sum_{x_j \in A} c_j$$

  Note that any measure can be broken into an atomic part and a non-atomic part (i.e. a measure for which no points have positive measure). This is somewhat of a hint for the following homework:

Homework. Is there a σ-algebra $\mathcal{A}$ which is countably infinite?

Some properties of measures include:

- σ-additive $\implies$ additive (just take $A_j = \varnothing$ for large j)
- Monotonicity: if $A \subseteq B$, then $\mu(A) \leq \mu(B)$. Note that

$$\mu(B) = \mu(A) + \mu(B \setminus A)$$

  but you should not write this as

$$\mu(B \setminus A) = \mu(B) - \mu(A)$$

  because we could have $\mu(B) = \mu(A) = \infty$.

- Given $A_1 \subseteq A_2 \subseteq \cdots$,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

  Proof. The sets A_i are not disjoint, but consider instead the sets $A_1, A_2 \setminus A_1, A_3 \setminus A_2, \ldots$ which are disjoint.

Then

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup \bigcup_{n=1}^{\infty} (A_{n+1} \setminus A_n)$$

so

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu(A_1) + \mu(A_2 \setminus A_1) + \mu(A_3 \setminus A_2) + \cdots \\
&= \lim_{n\to\infty} (\mu(A_1) + \mu(A_2 \setminus A_1) + \cdots + \mu(A_n \setminus A_{n-1})) \\
&= \lim_{n\to\infty} \mu(A_n)
\end{aligned}$$

□

- What about when we have $A_1 \supseteq A_2 \supseteq \cdots$? It is not true in general that

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

For example, consider the counting measure on $\mathbb{N}$, and $A_1 = \mathbb{N}$, $A_2 = \{2, 3, \ldots, \}$, $A_3 = \{3, 4, \ldots\}$, etc. Then their intersection is $\varnothing$ which has $\mu(\varnothing) = 0$, even though $\mu(A_n) = \infty$ for all n. Howver, if there is at least one n where $\mu(A_n)$ is finite, then the statement is true.

Proof. Without loss of generality, we can assume that $\mu(A_1)$ is finite. Then

$$\underbrace{\mu(A_1 \setminus A_2) + \mu(A_2 \setminus A_3) + \cdots}_{\lim_{n\to\infty} \mu(A_1 \setminus A_n)} + \mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \mu(A_1).$$

Because $\mu(A_1)$ is finite, we *can* write $\mu(A_1 \setminus A_n) = \mu(A_1) - \mu(A_n)$, so

$$\lim_{n\to\infty} \mu(A_1 \setminus A_n) = \lim_{n\to\infty} (\mu(A_1) - \mu(A_n))$$

and therefore

$$\mu(A_1) - \lim_{n\to\infty} \mu(A_n) + \mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \mu(A_1).$$

□

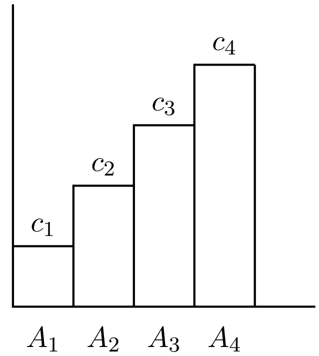
Let X be a compact metric space and $\mathcal{A}$ the Borel σ-algebra on X. Is a finite Borel measure μ determined by the measure of the balls?

Homework $(*)$**.** The answer is no; find an example. Federer wasn't able to find an example, after thinking about it for a day, but the construction is simple once you see it.

We say that $f : X \to \mathbb{R}$ is simple if it is measurable and takes only finitely many values. Equivalently, there are some disjoint sets $A_1, \ldots, A_n$ and $c_j \in \mathbb{R}$ such that

$$f = \sum_{j=1}^{n} c_j \chi_{A_j}.$$

You can visualize this as



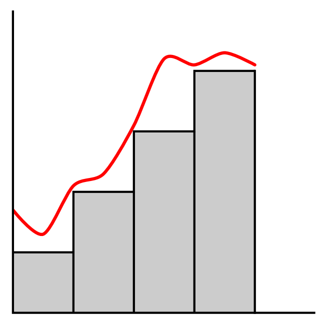
We define for a simple function $f \geq 0$

$$\int_A f\,d\mu = \sum_{j=1}^{n} c_j \mu(A \cap A_j).$$

If $g \geq 0$ is an arbitrary measurable function, then we define

$$\int_A g\,d\mu = \sup_{\substack{f \text{ simple} \\ f \leq g}} \int_A f\,d\mu.$$

Thus, we are approximating our function f from below by simple functions.



We need to restrict to nonnegative functions, even though $\pm\infty$ are not being allowed as values of our functions in this definition, because (for example) we cannot integrate $g(x) = \frac{1}{x}$ on $\mathbb{R}$ with this definition; there is no simple function $f \leq g$. This illustrates the difference between being finite everywhere and being bounded.

For an arbitrary measurable function g, we set

$$g^+ = \max(0, g), \quad g^- = -\min(0, g)$$

so that $g = g^+ - g^-$ and then define

$$\int_A g\,d\mu = \int_A g^-\,d\mu - \int_A g^-\,d\mu.$$

From now on, when I write a function, it will be assumed to be measurable, even if I don't say so.

Given a measure μ and an $f \geq 0$, we can make a new measure ν defined by

$$\nu(A) = \int_A f\,d\mu.$$

Theorem (Monotone Convergence Theorem)**.** *Given a sequence of functions*

$$0 \le f_1 \le f_2 \le f_3 \le \cdots$$

then for $f = \lim_{n\to\infty} f_n$*,*

$$\int f\,d\mu = \lim_{n\to\infty} \int f_n\,d\mu.$$

Proof. It is easy to see that

$$\int f_1 \le \int f_2 \le \int f_3 \le \cdots$$

so there is some $\alpha = \lim_{n\to\infty} \int f_n$. We need to show that $\alpha \le \int f$ and $\alpha \ge \int f$.

The former is trivial because $f_n \le f$ for all n. For the latter, consider the definition of the integral. We want to show that for any simple $g \le f$, we have $\int g \le \alpha$. Unfortunately, it is not true that for any such g, there is an n such that $g \le f_m$ for all $m \ge n$. It isn't even true pointwise, since we could have $g(x) = f(x)$ and $f_n(x) < f(x)$ for all n. We will need a different argument.

Given a simple function $g \le f$, let its values be $c_1, \ldots, c_N$. For any $\epsilon > 0$ that is smaller than all the c_i, define

$$g_\epsilon = \begin{cases} 0 & \text{if } g = 0, \\ g - \epsilon & \text{if } g > 0. \end{cases}$$

As $\epsilon \to 0$, we have that $\int g_\epsilon \to \int g$, so in fact it is enough to show that $\int g_\epsilon \le \alpha$ for all ϵ.

Define

$$B_n = \{x \in X \mid f_n(x) \ge g_\epsilon(x)\}.$$

We have that $\bigcup B_n = X$, and that $B_1 \subseteq B_2 \subseteq \cdots$. For any given $\epsilon > 0$, define a measure ν by

$$\nu(A) = \int_A g_\epsilon.$$

Then

$$\nu(X) = \nu\left(\bigcup B_n\right) = \lim_{n\to\infty} \nu(B_n)$$

and

$$\nu(B_n) = \int_{B_n} g_\epsilon \le \int_{B_n} f_n \le \int_X f_n.$$

Thus,

$$\int g_\epsilon = \nu(X) = \lim_{n\to\infty} \nu(B_n) \le \lim_{n\to\infty} \int_X f_n = \alpha.$$ □

Corollary (Beppo-Levi)**.** *Given a sequence of functions* $f_n \ge 0$*, then*

$$\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n.$$

Proof.

$$\sum_{n=1}^{\infty} \int f_n = \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n = \lim_{N\to\infty} \int \sum_{n=1}^{N} f_n \overset{\text{MCT}}{=} \int \lim_{N\to\infty} \sum_{n=1}^{N} f_n = \int \sum_{n=1}^{\infty} f_n.$$ □

Corollary (Fatou's Lemma)**.** *Given a sequence of functions* $f_n \geq 0$*,*

$$\int \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int f_n.$$

Homework. Find an example where this inequality is strict.

Proof. By definition,

$$\liminf_{n\to\infty} f_n = \lim_{n\to\infty} \underbrace{\inf\{f_n, f_{n+1}, f_{n+2}, \dots\}}_{g_n}.$$

We have that $0 \leq g_n \leq g_{n+1} \leq \cdots$ and therefore

$$\int \liminf_{n\to\infty} f_n = \int \lim_{n\to\infty} g_n = \lim_{n\to\infty} \int g_n \leq \liminf_{n\to\infty} \int f_n$$

because $g_n \leq f_n$. □

Corollary (Lebesgue Theorem)**.** *Given a sequence of functions* f_n *such that* $|f_n| \leq g$*, where* g *is a function such that* $\int g < \infty$*,*

$$\int \lim_{n\to\infty} f_n = \lim_{n\to\infty} \int f_n.$$

Homework. Find a sequence of functions f_n converging pointwise to a function f which do not satisfy the conclusion of this theorem.

Proof. Let $f = \lim_{n\to\infty} f_n$. In fact, a stronger statement is true: as $n \to \infty$,

$$\left| \int f_n - \int f \right| \leq \int |f_n - f| \to 0.$$

We have that $|f_n - f| \leq 2g$. Let $h_n = 2g - |f_n - f|$, so that $h_n \geq 0$. Apply Fatou:

$$\int \underbrace{\liminf_{n\to\infty} h_n}_{2g} \leq \liminf_{n\to\infty} \int h_n.$$

Therefore

$$\int 2g \leq \liminf_{n\to\infty} \left(\int 2g - |f_n - f| \right) = \int 2g + \liminf_{n\to\infty} \left(- \int |f_n - f| \right).$$

Therefore

$$\limsup_{n\to\infty} \left(\int |f_n - f| \right) \leq 0$$

which implies that in fact

$$\int |f_n - f| \to 0.$$

□

Lecture 3 (2012-10-09)

Last time, we talked about measure spaces $(X, \mathcal{A}, \mu)$.

Definition. We say that a measure space is complete if, for any $A \in \mathcal{A}$ such that $\mu(A) = 0$, we have $B \in \mathcal{A}$ for every subset $B \subseteq A$. By monotonicity of measures, they will also be null.

Claim. The σ-algebra generated by $\mathcal{A}$ and all subsets of null sets is the σ-algebra of all sets S such that there are $A_1, A_2 \in \mathcal{A}$ with $A_1 \subseteq S \subseteq A_2$ and $\mu(A_2 \setminus A_1) = 0$. We can then define $\mu(S) = \mu(A_1) = \mu(A_2)$.

Definition. An outer measure is a function $\mu : \mathcal{A} \to [0, \infty]$, where $\mathcal{A}$ is an arbitrary collection of sets, such that

- $\mu(\varnothing) = 0$,
- σ-**sub**-additivity: $\mu(\bigcup A_n) \leq \sum \mu(A_n)$ for any $A_1, A_2, \ldots \in \mathcal{A}$.

We can create an outer measure as follows: for an arbitrary collection of sets $\mathcal{A}$, and an arbitrary function $\alpha : \mathcal{A} \to [0, \infty]$, we can define for any $A \in P(X)$

$$\phi_\alpha(A) \overset{\text{def}}{=} \inf\{\textstyle\sum \mu(A_n) \mid A \subseteq \bigcup A_n,\ A_n \in \mathcal{A}\}.$$

If there is no cover of A by some $A_n \in \mathcal{A}$, then $\phi_\alpha(A) = \infty$.

We now want to create a measure from this outer measure:

$$\underset{\text{(arbitrary)}}{\alpha,\ \mathcal{A}} \quad \longrightarrow \quad \underset{\text{(outer measure)}}{\phi_\alpha,\ P(X)} \quad \longrightarrow \quad \underset{\text{(measure)}}{\mu,\ \mathcal{M}_\phi}$$

Definition. Given an outer measure ϕ, we say that A is ϕ-measurable if, for any set S,

$$\phi(S) = \phi(S \cap A) + \phi(S \setminus A).$$

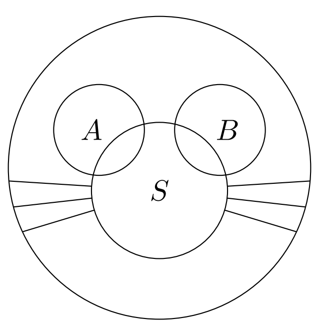
We will see that the collection of ϕ-measurable sets will form a complete measure space. First, note that we always have

$$\phi(S) \leq \phi(S \cap B) + \phi(S \setminus B).$$

If $\phi(A) = 0$, and $B \subseteq A$, then $S \cap B \subseteq B \subseteq A$ implies that $\phi(S \cap B) = 0$, and $S \setminus B \subseteq S$ implies that $\phi(S \setminus B) \leq \phi(S)$, so that

$$\phi(S) \leq 0 + \phi(S \setminus B)$$

hence $\phi(S) = \phi(S \setminus B)$, hence $\phi(B) = 0$.



Now we want to check that

$$\phi(S) = \phi(S \cap (A \cup B)) + \phi(S \setminus (A \cup B)).$$

We have that

$$\phi(S) = \phi(S \cap A) + \phi(S \setminus A).$$

Therefore

$$\phi(S \setminus A) = \phi(\underbrace{(S \setminus A) \cap B}_{S \cap B}) + \phi(\underbrace{(S \setminus A) \setminus B}_{S \setminus (A \cup B)})$$

and

$$\phi(S \cap A) + \phi(S \cap B) = \phi(S \cap (A \cup B))$$
$$\phi(S \cap (A \cup B)) = \phi(\underbrace{S \cap (A \cup B) \cap A}_{S \cap A}) + \phi(\underbrace{(S \cap (A \cup B)) \setminus A}_{S \cap B}).$$

The rest of the argument you should study on your own.

If our initial arbitrary collection of sets $\mathcal{A}$ is such that $\varnothing \in \mathcal{A}$, and for any $A, B \in \mathcal{A}$, we have $A \cup B, A \cap B, A \setminus B \in \mathcal{A}$, and we further assume that α is an outer measure on $\mathcal{A}$ such that for any $A, B \in \mathcal{A}$ we have

$$\alpha(A) = \alpha(A \cap B) + \alpha(A \setminus B)$$

then the resulting σ-algebra $\mathcal{M}$ will contain $\mathcal{A}$, i.e. $\mathcal{A} \subseteq \mathcal{M}$.

Here is a special case. Consider "bricks", i.e. sets of the form

$$\prod_{i=1}^{n} [a_i, b_i) \subseteq \mathbb{R}^n$$

and let $\mathcal{A}$ be the set of finite unions of such sets. Let α be the volume function. Then ϕ_α is the outer Lebesgue measure, $\mathcal{M}$ consists of the Lebesgue measurable sets, and μ is the Lebesgue measure. We can do the same construction on the bricks for any additive function α; the resulting measure will always contain at least the Borel sets.

Definition. We say that μ is a Borel measure on a σ-algebra $\mathcal{M}$ if

- The σ-algebra $\mathcal{M} \supseteq \{\text{Borel sets}\}$.
- Some people require that μ is σ-finite, i.e. there exist $A_1, A_2, \ldots$ with $\mu(A_n) < \infty$ such that $\mu(X \setminus \bigcup A_n) = 0$.
- Some people require that $\mu(K) < \infty$ for any compact K (in $\mathbb{R}^n$, this just says that bounded sets have finite measure).

Definition. We say that a measure μ is regular when for any $A \in \mathcal{A}$,

$$\mu(A) = \inf\{\textstyle\sum \mu(G_n) \mid G_n \text{ are open, } A \subseteq \bigcup G_n\} = \inf\{\mu(G) \mid G \text{ open}, A \subseteq G\}$$

For any $\epsilon > 0$, we can choose G such that $\mu(A) \leq \mu(G) \leq \mu(A) + \epsilon$. Letting $\epsilon \to 0$, we can see that for any A there is a G_δ set containing A of the same measure. By the same argument for the complement, we have an $F_\sigma \subseteq A$ of the same measure.

Now we'll start on a new topic.

Let λ be the Lebesgue measure on $\mathbb{R}^n$, and $\mathcal{B}(\mathbb{R}^n)$ the Borel sets in $\mathbb{R}^n$.

Definition. For any $X \subseteq \mathbb{R}^n$, a differential basis for X is a collection $\mathcal{D} \subseteq X \times \mathcal{B}(\mathbb{R}^n)$ such that

- For any $(x, A) \in \mathcal{D}$, we have $\lambda(A) > 0$.
- For any $x \in X$ and $r > 0$, there is some $(x, A) \in \mathcal{D}$ such that $A \subseteq B(x, r)$, the ball of radius r around x.

Given an $x \in \mathbb{R}^n$ and $A \subseteq \mathbb{R}^n$, let $r(A)$ be the smallest radius such that $A \subseteq B(x, r(A))$.

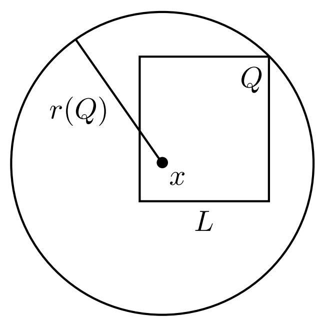
We say that a differential basis $\mathcal{D}$ is regular when it satisfies the following property: for any $x \in X$, there exist a $\delta > 0$ and $r_0 > 0$, depending on x, such that for any $(x, A) \in \mathcal{D}$ with $r(A) < r_0$,

$$\lambda(A) \geq \delta \cdot \lambda(B(x, r(A))).$$

In other words, $\mathcal{D}$ is regular when, for any $x \in X$, there is a positive ratio δ for which all sufficiently small $(x, A) \in \mathcal{D}$ take up at least δ of their bounding ball.

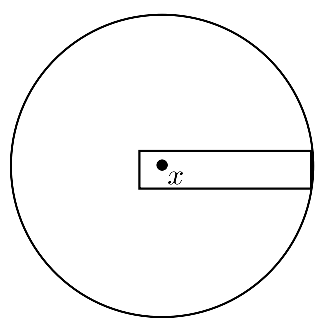
Examples.

- The symmetrical basis: all balls with center x. This is regular.
- The standard basis: all cubes Q that contain x.



  We know that $\lambda(Q) = \ell^n$, and that $r(Q) \leq \sqrt{n}\ell$, so $\lambda(B) \leq c_n \ell^n$ for some constant c_n depending only on n, which we can take as our δ. This is regular.

- The interval basis (also called the strong basis): all bricks that contain x. This is **not** regular.



Definition. Given a differential basis $\mathcal{D}$, we define

$$\overline{D}\mu(x) = \limsup_{r \to 0} \left\{ \frac{\mu(A)}{\lambda(A)} \;\middle|\; (x, A) \in \mathcal{D}, A \subseteq B(x, r) \right\}$$

to be the upper derivative of μ at x. The lower derivative $\underline{D}\mu(x)$ is the same except with lim inf, and the derivative is defined to be their common value if they are equal.

Example. Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function that is monotone increasing, and define $\mu([a, b)) = f(b) - f(a)$. This defines a Borel measure μ on $\mathbb{R}$. We'll choose $\mathcal{D}$ = symmetrical basis, so

$$\lim_{r \to 0} \frac{\mu(B(x, r))}{\lambda(B(x, r))} = \lim_{r \to 0} \frac{f(x + r) - f(x - r)}{2r}.$$

This is the symmetric derivative. Note that f did not have to be differentiable at x for this to exist. In contrast, if we chose the standard basis or interval basis (which are the same in one dimension), then we'd get the limit

$$\lim_{\substack{y,z \to x \\ x \in [y,z]}} \frac{f(y) - f(x)}{y - z}$$

which exists if and only if f is differentiable.

Definition. We define the maximal operator of a measure μ to be

$$M(x) = M_\mu(x) = \sup_{(x,A) \in \mathcal{D}} \frac{\mu(A)}{\lambda(A)}.$$

Note that if $\mu = \mu_1 + \mu_2$, then $M_\mu \leq M_{\mu_1} + M_{\mu_2}$. We only get an inequality because the set which maximizes $\mu_1 + \mu_2$ need not maximize μ_1 and μ_2 separately.

Lecture 4 (2012-10-11)

Today we'll be working in $\mathbb{R}^n$, and all our measures will be Borel measures that are σ-finite.

Let μ_1, μ_2 be measures.

Definition. We say that μ_1 is absolutely continuous with respect to μ_2, and we write $\mu_1 \ll \mu_2$, when $\mu_2(A) = 0$ implies $\mu_1(A) = 0$.

Definition. We say that μ_1 is singular with respect to μ_2, and we write $\mu_1 \perp \mu_2$, if there exist disjoint $A_1, A_2 \subset \mathbb{R}^n$ such that $\mu_1(\mathbb{R}^n \setminus A_1) = 0$ and $\mu(\mathbb{R}^n \setminus A_2) = 0$.

Theorem (Radon-Nykodim)**.** *If $\mu_1 \ll \mu_2$, then there is an f such that $\mu_1(A) = \int_A f\, d\mu_2$ for all A. We say that f is the Radon-Nykodim derivative of μ_1 with respect to μ_2, and we write that $f = \frac{d\mu_1}{d\mu_2}$.*

Theorem. *For any μ_1, μ_2, we can decompose $\mu_1 = \alpha + \beta$ such that $\alpha \ll \mu_2$ and $\beta \perp \mu_2$.*

Homework. Show that $\frac{d\mu_1}{d\mu_3} = \frac{d\mu_1}{d\mu_2} \cdot \frac{d\mu_2}{d\mu_3}$ for any $\mu_1 \ll \mu_2 \ll \mu_3$.

Last class, we defined the maximal operator M_μ of a measure μ, which is a function on $\mathbb{R}^n$, and noted that $M_{\mu_1+\mu_2} \le M_{\mu_1} + M_{\mu_2}$.

Theorem. *For any finite measure μ, we have*

$$\lambda\big(\{x \in \mathbb{R}^n \mid M_\mu(x) > t\}\big) \le 3^n \cdot \frac{\mu(\mathbb{R}^n)}{t}$$

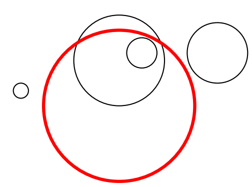
for any $t \ge 0$.

Lemma. *If $B_1, \dots, B_n$ are finitely many balls, there exists a subset $B_{i_1}, \dots, B_{i_k}$ of pairwise disjoint balls such that*

$$\bigcup_j (3B_{i_j}) \supseteq \bigcup_i B_i$$

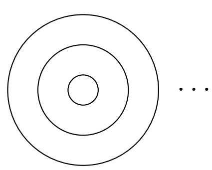
where $3B$ means the ball with the same center as B and three times the radius.

Proof of lemma. We use a greedy algorithm. At each step, choose the largest ball that is disjoint from all earlier chosen ones. This will obviously terminate because there are only finitely many balls.



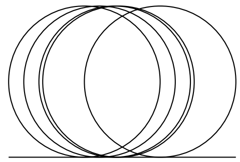
Any ball B_a not chosen will intersect some ball B_b that was chosen, and B_a must have radius less than or equal to B_b (otherwise B_b would not have been chosen), so that expanding B_b by a factor of 3 will cover all of B_a. $\square$

Note that this statement is false if we allow infinitely many balls; for example we could have nested balls around the same center whose radii go to infinity.



Homework. Prove that the statement of the lemma is true even with infinitely many balls, as long as the radii are bounded, allowing the chosen subcollection of balls to also be infinite, and replacing 3 with some arbitrary constant.

Is an arbitrary union of closed unit balls necessarily Borel? No. An easy construction is to choose a non-Borel subset of $\mathbb{R}$, and place balls which touch the line at exactly those points.



The union of the balls isn't Borel; otherwise, its intersection with the Borel set $\mathbb{R}$ would be Borel.

Homework. Is an arbitrary union of closed unit balls necessarily Lebesgue measurable?

Proof of theorem. We want to show, for any finite measure μ and $t > 0$, that

$$\lambda\big(\{x \in \mathbb{R}^n \mid M_\mu(x) > t\}\big) \leq 3^n \cdot \frac{\mu(\mathbb{R}^n)}{t}.$$

It is enough to show that

$$\lambda(K) \leq 3^n \cdot \frac{\mu(\mathbb{R}^n)}{t}$$

for all compact $K \subseteq \{x \mid M_\mu(x) > t\}$, because Lebesgue measure is inner regular, i.e. for any Lebesgue measurable set S,

$$\lambda(S) = \sup_{\substack{\text{compact} \\ K \subseteq S}} \lambda(K).$$

Now note that, by definition, $M_\mu(x) > t$ if and only if there is a ball B_x around x such that $\mu(B_x) > t \cdot \lambda(B_x)$.

Choose a cover of K by finitely many such balls $B_{x_1}, \dots, B_{x_k}$ (we can do this because K is compact), and by the lemma, we can choose $x_{i_1}, \dots, x_{i_s}$ such that $B_{x_{i_1}}, \dots, B_{x_{i_s}}$ are disjoint and

$$\bigcup_{j=1}^{s} (3B_{x_{i_j}}) \supseteq \bigcup_{i=1}^{k} B_{x_i}.$$

Then

$$\lambda(K) \leq \lambda\left(\bigcup_{i=1}^{k} B_{x_i}\right) \leq \lambda\left(\bigcup_{j=1}^{s} (3B_{x_{i_j}})\right) = 3^n \sum_{j=1}^{s} \lambda(B_{x_{i_j}})$$

$$\leq 3^n \sum_{j=1}^{s} \frac{\mu(B_{x_{i_j}})}{t} = 3^n \cdot \frac{\mu\left(\bigcup_{j=1}^{s} B_{x_{i_j}}\right)}{t} \leq 3^n \cdot \frac{\mu(\mathbb{R}^n)}{t}. \qquad \square$$

Corollary. *For almost every $x \in \mathbb{R}^n$, the upper derivative of μ at x is finite, i.e. $\overline{D}\mu(x) < \infty$.*

Proof. Our theorem implies that in symmetric basis, we have $M_\mu(x) < \infty$ Lebesgue-a.e. Now note that, by definition, $\overline{D}\mu \leq M_\mu$. $\square$

Note that, for a regular basis $\mathcal{D}$,

$$\frac{\mu(A)}{\lambda(A)} \leq \frac{\mu(B)}{\lambda(A)} = \underbrace{\frac{\mu(B)}{\lambda(B)}}_{<\infty} \cdot \underbrace{\frac{\lambda(B)}{\lambda(A)}}_{<\frac{1}{\delta}}$$

where $(x, A) \in \mathcal{D}$, with $A \subseteq B(x, r(A)) = B$, and $\frac{1}{\delta}$ is from the regularity of $\mathcal{D}$.

Theorem. *If μ is singular, then $D\mu(x) = 0$ a.e.*

Proof. By hypothesis, there exists a Lebesgue null set N such that $\mu(\mathbb{R}^n \setminus N) = 0$. We need to show that $D\mu(x) = 0$ for a.e. $x \in \mathbb{R}^n \setminus N$.

If N is closed, there is nothing to prove, because any $x \in \mathbb{R}^n \setminus N$ can be separated from N by a sufficiently small open ball.

Now choose a compact $K \subseteq N$ such that $\mu(N \setminus K) < \epsilon^2$.

Let $\mu = \mu_1 + \mu_2$, where $\mu_1(A) = \mu(A \cap K)$ and $\mu_2(A) = \mu(A \cap K^c)$. We have that $\overline{D}\mu \leq \overline{D}\mu_1 + \overline{D}\mu_2$, and that $\overline{D}\mu_1$ is 0 a.e.

Note that

$$\lambda\big(\{x \in \mathbb{R}^n \mid \overline{D}\mu_2(x) > t\big) \leq \lambda\big(\{x \in \mathbb{R}^n \mid M_{\mu_2}(x) > t\}\big) \leq \; 3^n \cdot \frac{\mu_2(\mathbb{R}^n)}{t} \leq 3^n \cdot \frac{\epsilon^2}{t}.$$

Letting $t = \epsilon$, and using that $\overline{D}\mu_1$ is 0 a.e.,

$$\lambda\big(\{x \in \mathbb{R}^n \mid \overline{D}\mu(x) > t\big) \leq \lambda\big(\{x \in \mathbb{R}^n \mid M_{\mu_2}(x) > t\}\big) \leq \; 3^n t.$$

Now let $t \to 0$. This shows that $\lambda(\{x \in \mathbb{R}^n \mid \overline{D}\mu(x) > 0\}) = 0$. $\square$

What happens when $\mu \ll \lambda$? By Radon-Nykodim, we have that $\mu = \int f \, d\lambda$ for some f, and it is a theorem (which we are about to prove) that

$$f = \frac{d\mu}{d\lambda} = D\mu \text{ a.e.}$$

Now let f be a function such that $\int |f| \, d\lambda < \infty$, i.e. $f \in L^1$.

Definition. We say that $x \in \mathbb{R}^n$ is a Lebesgue point of f if

$$\lim_{r \to 0} \frac{1}{\lambda(B(x,r))} \int_{B(x,r)} |f(t) - f(x)| \, dt = 0.$$

Theorem. *For any $f \in L^1$, almost every $x \in \mathbb{R}^n$ is a Lebesgue point of f.*

Proof. Define the measure

$$\mu_x(A) = \int_A |f(t) - f(x)| \, dt.$$

We need to show that $D\mu_x(x) = 0$ for almost all x. We will assume the following lemma for the moment:

Lemma. *For $f \in L^1$ and any $\epsilon > 0$, there is a continuous, compactly supported g such that $\int_{\mathbb{R}^n} |f - g| < \epsilon$.*

Now fix an $\epsilon > 0$, and choose g from the lemma such that $\int |f-g| < \epsilon^2$. Let $h = f - g$, so that $f = g + h$. We have

$$|f(t) - f(x)| \leq |g(t) - g(x)| + |h(t) - h(x)|$$

It is easy to see that

$$\lim_{r \to 0} \frac{1}{\lambda(B(x,r))} \int_{B(x,r)} |g(t) - g(x)| \, dt = 0$$

for all x (this essentially follows from g being continuous), so now we just need to look at h. Define

$$\nu_x(A) = \int_A |h(t) - h(x)| \, dt.$$

Our work so far shows that that $\overline{D}\mu_x \leq \overline{D}\nu_x$. Using the triangle inequality to produce a bound, and dividing by $\lambda(A)$,

$$\frac{\nu_x(A)}{\lambda(A)} \leq \frac{\int_A |h(t)| \, d\lambda}{\lambda(A)} + |h(x)| \cdot \frac{\lambda(A)}{\lambda(A)}.$$

Thus,

$$\begin{aligned} \{x \in \mathbb{R}^n \mid \overline{D}\mu_x > 2\epsilon\}] &\subseteq \{x \in \mathbb{R}^n \mid \overline{D}\nu_x > 2\epsilon\} \\ &\subseteq \left\{ x \in \mathbb{R}^n \,\middle|\, \begin{array}{c}\text{there exist arbitrarily} \\ \text{small } B \ni x \text{ such that}\end{array} \frac{\int_B |h(t)| \, dt}{\lambda(B)} > \epsilon \right\} \cup \{x \in \mathbb{R}^n \mid |h(x)| > \epsilon\}. \end{aligned}$$

Call the first set S_1 and the second set S_2. Note that the first set is precisely where the maximal operator of h is greater than ϵ. Thus, from the lemma,

$$\lambda(S_1) \leq 3^n \cdot \frac{\epsilon^2}{\epsilon} = 3^n \epsilon$$

and because $\int |f-g| = \int |h| < \epsilon^2$, we have

$$\lambda(S_2) \leq \frac{\int |h|}{\epsilon} \leq \epsilon.$$

Because $\epsilon > 0$ is arbitrary, we can conclude that

$$\lambda(\{x \in \mathbb{R}^n \mid \overline{D}\mu_x > 0\}) = 0.$$

Thus, the set of non-Lebesgue points of f is null. Now we prove our other claim. Given a measure $\mu \ll \lambda$, we can let $f = \frac{d\mu}{d\lambda}$, and then we see that as we take smaller and smaller balls $B \ni x$,

$$\left| \frac{\mu(B)}{\lambda(B)} - f(x) \right| = \left| \left(\frac{1}{\lambda(B)} \int_B f(t) \, d\lambda \right) - f(x) \right|$$

$$= \left| \frac{1}{\lambda(B)} \int_B (f(t) - f(x)) \right| \leq \frac{1}{\lambda(B)} \int_B |f(t) - f(x)| \to 0.$$

This demonstrates that we have $\frac{d\mu}{d\lambda} = \overline{D}\mu$ a.e. $\square$

Proof of lemma.

1. If r is large enough then $\int_{B(0,r)^c} |f| < \epsilon$.
2. If M is large enough then $\int_{\{x : |f(x)| > M\}} |f| < \epsilon$.

Let $\bigcup A_m = A$, so that $A_m = A \cap \{x \in \mathbb{R}^n \mid m\epsilon \leq f < (m+1)\epsilon\}$. Then

$$A = B(0,r) \cap \{x \in \mathbb{R}^n \mid |f| \leq M\}$$

Choose compact $K_m \subseteq A_m$ such that

$$\int_{A_m \setminus K_m} |f| \leq \epsilon'$$

and also such that $\lambda(A_m \setminus K_m) \leq \epsilon'$. We can do this because λ is inner regular, so that for each A_m there is a sequence $K_{m,i} \subseteq A_m$ of compact sets such that $\lambda(A_m \setminus K_{m,i}) < \frac{1}{i}$, and WLOG we can assume $K_{m,i} \subseteq K_{m,i+1}$, so letting $d\nu = f\, d\lambda$,

$$\lim_{n\to\infty} \nu(K_n) = \nu(A_n).$$

Define $g = m\epsilon$ on K_n. Extend it to a continuous function $g : B(0,r) \to [-M, M]$, with $g = 0$ outside $B(0,r)$. $\square$

Lecture 5 (2012-10-16)

What we've proved so far: the maximal operator theorem and the covering lemma. These involved balls and the symmetric basis.

We proved that $\overline{D}\mu(x) < \infty$ a.e., that it equals 0 a.e. if the measure μ is singular, and that it equals the Radon-Nykodim derivative of μ in every regular basis.

Maximal operator theorem for cubes: up to a constant, it is the same as for balls. If c_n is the ratio between the volume of a unit cube and a unit ball in dimension n,

$$\frac{1}{\lambda(Q)} \int_Q |f| \, d\mu \leq \frac{1}{\lambda(Q)} \int_B |f| = \frac{\lambda(B)}{\lambda(Q)} \frac{1}{\lambda(B)} \int_B |f| \leq c_n \frac{1}{\lambda(B)} \int_B |f|$$

Federer's *Geometric Measure Theory* is a good reference book. If you just want to know what's true, read this book; but it uses its own notation, so any time you read a theorem, you'll have to refer back to the previous one, and then the one before that, etc. You can also read Stein's *Harmonic Analysis* which covers a lot more than we'll get to in this course.

Let $A \subseteq \mathbb{R}^n$ be a Lebesgue measurable set, and define $f(x) = \chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$

Let $\mu = \int f \, d\lambda$, so that $\mu(B) = \lambda(A \cap B)$. For $x \in \mathbb{R}^n$, define

$$\underline{d}(x, A), \quad \overline{d}(x, A), \quad d(x, A)$$

to be the lower derivative, upper derivative, and derivative of μ at x with respect to the symmetric basis. These are, respectively, the lim inf, lim sup, and lim as $r \to 0$ of

$$\frac{\lambda(A \cap B(x, r))}{\lambda(B(x, r))}.$$

Almost every point in $\mathbb{R}^n$ is a Lebesgue point of f; this implies that $d(x, A) = 1$ for almost all $x \in A$, and $d(x, A) = 0$ for almost all $x \notin A$.

Theorem. *For any set A (not necessarily measurable),*

1. *$d(x, A) = 1$ for almost all $x \in A$.*
2. *$d(x, A) = 0$ for almost all $x \notin A$ $\iff$ A is measurable.*

Definition. Given any set A, we say that $H \supseteq A$ is a measurable hull of A if H is measurable and if, for any measurable $B \supseteq A$, we have $\lambda(H \setminus B) = 0$.

Any set has a hull; for any A, we can define

$$\lambda(A) = \inf_{\substack{G \supseteq A \\ G \text{ open}}} \lambda(G)$$

and then choose $G_1, G_2, \ldots$ such that $G_n \supseteq A$ and $\lambda(G_n) \to \lambda(A)$. Take $H = \bigcap G_n$.

Remark. If H_1 and H_2 are measurable hulls of A, then $\lambda(H_1 \triangle H_2) = 0$.

Remark. For any measurable B, $H \cap B$ is a measurable hull of $A \cap B$.

Proof of 1. For any set A, let H be a measurable hull. Then for all x,

$$\underline{d}(x,A) = \underline{d}(x,H), \quad \overline{d}(x,A) = \overline{d}(x,H), \quad d(x,A) = d(x,H)$$

so $d(x,A) = d(x,H) = 1$ for almost all $x \in H$, hence for almost all $x \in A$. $\square$

Proof of 2. The $\Longleftarrow$ implication is OK, and for $\Longrightarrow$, suppose that $d(x,A) = 0$ for a.e. almost all $x \in A$. We know that $d(x,H) = 1$ for almost all $x \in H$, so we must have that at almost every point in $H \setminus A$, the density is 1, and at almost every point in $H \setminus A$, the density is 0. Thus, $\lambda(H \setminus A) = 0$, so $A = H \setminus (H \setminus A)$, and H is measurable and $H \setminus A$ is measurable because it is a null set, and hence A is measurable. $\square$

Definition. The Denjoy topology, or the density topology, on $\mathbb{R}^n$, is defined by letting a set A be open if it is measurable and $d(x,A) = 1$ for all $x \in A$. We'll say that A is d-open because both Denjoy and density start with d.

Why is it a topology?

Finite intersections: It is trivial that if A_1 and A_2 are d-open, then $A_1 \cap A_2$ is d-open.

Arbitrary unions: If the A_α are d-open, then certainly the density at every $x \in \bigcup A_\alpha$ is 1, but why must $\bigcup A_\alpha$ be measurable? This need not be a countable union. It will suffice to check that $\mathbb{R}^n \setminus \bigcup A_\alpha$ is measurable. Thus, it will be enough to show that for **almost all** $x \in \bigcup A_\alpha$, then $d(x, \mathbb{R}^n \setminus \bigcup A_\alpha) = 0$. We will in fact check this for every point in the union. If $x \in \bigcup A_\alpha$, then $x \in A_\alpha$ for some α, so that $\mathbb{R}^n \setminus \bigcup A_\alpha \subseteq \mathbb{R}^n \setminus A_\alpha$. We know that $\mathbb{R}^n \setminus A_\alpha$ has density 0 at any $x \in A_\alpha$ (**because** of our assumption that $d(x, A_\alpha) = 1$ for all $x \in A_\alpha$), and therefore any smaller set must have density 0 at $x \in A_\alpha$, i.e.

$$d(x, \mathbb{R}^n \setminus \bigcup A_\alpha) \le d(x, \mathbb{R}^n \setminus A_\alpha) = 0.$$

We needed the assumption that $d(x, A_\alpha) = 1$ at every $x \in A_\alpha$ because, without it, we wouldn't have been able to get our hands on "almost all of $\bigcup A_\alpha$"; if we'd been taking arbitrary measurable A_α, we'd only have $d(x, A_\alpha) = 1$ for almost all $x \in A_\alpha$, and this wouldn't tell us anything about $d(x, A_\alpha)$ at almost all $x \in \bigcup A_\alpha$ because the arbitrary union can make things much bigger.

We have $\mathbb{R} = \varnothing \cup \mathbb{R}$; are there any other examples of d-open sets whose complements are also d-open? No:

Theorem. *The d-topology is connected; in other words, if $\mathbb{R}$ is the disjoint union of two d-open sets, then one of them is empty.*

Proof. Suppose that $d(x,A) = 1$ for all $x \in A$ and that $d(x,A) = 0$ for all $x \notin A$. Let's consider the function

$$f(x) = \int_0^x \chi_A = \begin{cases} \lambda((0,x) \cap A) & \text{if } x > 0, \\ -\lambda((0,x) \cap A) & \text{if } x < 0. \end{cases}$$

This is differentiable at any point of density for A, and in fact has derivative 1 at any point of density for A, essentially by the definition of the derivative; because we are assuming that any $x \in \mathbb{R}$ is either a point of density of A or of $\mathbb{R} \setminus A$, then f is differentiable everywhere, and

$$f'(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A, \end{cases}$$

but every derivative has the Darboux property, namely, that if $g(x_1) = a$ and $g(x_2) = b$, then for any $c \in [a,b]$, there is some $x \in [x_1, x_2]$ such that $g(x) = c$. Thus, either $A = \mathbb{R}$ or $A = \varnothing$. $\square$

We say that $\delta > 0$ is good if, for every measurable $A \subseteq \mathbb{R}$ such that $\lambda(A) > 0$, $\lambda(\mathbb{R} \setminus A) > 0$, there is an $x \in \mathbb{R}$ such that

$$\delta \leq \underline{d}(x, A) \leq \overline{d}(x, A) \leq 1 - \delta$$

Homework ($*$)**.** Prove that $\delta = \frac{1}{4}$ is good.

In fact, the optimal δ is the unique real root of $8\delta^3 + 8\delta^2 - \delta - 1 = 0$. It's approximately $\delta = 0.2684\ldots$

Corollary. *Every set of positive measure in $\mathbb{R}^2$ contains the vertices of a regular triangle.*

Proof. Choose a density point x in the set, and choose a ball around x such that

$$\lambda(A \cap B(x,r)) > \frac{1}{2}\lambda(B(x,r)).$$

Then if A^* denotes A rotated by $60°$, we also have $\lambda(A^* \cap B(x,r)) > \frac{1}{2}\lambda(B(x,r))$, so there is some $y \in A \cap A^* \cap B(x,r)$. $\square$

More generally, the same is true for the vertices of a square, or any finite configuration of points. Is it true that for any sequence $x_n \to 0$ in $\mathbb{R}$, that every measurable $A \subseteq \mathbb{R}$ contains a copy of it? Unknown in general; if $x_n \to 0$ really fast we know it's true.

There is a μ such that $\overline{D}\mu = \infty$ a.e. when we differentiate with respect to the *strong* differential basis (i.e. bricks). We can even choose a measure of the form $\mu = \int f$. However, we can't use a characteristic function $f = \chi_A$, because almost every $x \in A$ is a strong-density point of A.

Lecture 6 (2012-10-18)

Theorem (Steinhaus Theorem)**.** *Let $A \subseteq \mathbb{R}^n$ be measurable with $\lambda(A) > 0$. Let*

$$A - A = \{x - y \in \mathbb{R}^n \mid x, y \in A\}.$$

Then $A - A$ contains a ball around 0.

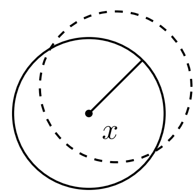
Proof. Choose $x \in A$ to be a density point, and take a small ball $B(x, r)$ around x which A almost entirely fills, say

$$\frac{\lambda(A \cap B(x,r))}{\lambda(B(x,r))} > 0.9.$$

Choose a $r_0 > 0$ such that

$$\frac{\lambda(B(x,r) \cap B(x+z,r))}{\lambda(B(x,r))} > 0.99$$

for all $|z| < r_0$, i.e. a distance r_0 such that shifting the ball $B(x, r)$ a distance $< r_0$ in any direction still mostly intersects the original ball.



Then comparing how much A can still intersect $B(x + z, r)$, we see that we must have that $A \cap (A + z) \neq \varnothing$ for all $|z| < r_0$, so that $A - A \supseteq B(0, r_0)$. $\square$

Homework. Given measurable $A, B \subseteq \mathbb{R}^n$ with $\lambda(A), \lambda(B) > 0$, prove that

$$A + B \stackrel{\text{def}}{=} \{x + y \mid x \in A, y \in B\}$$

contains a ball, i.e. its interior is non-empty.

The d-topology

Here are some facts:

- Open = every point is a density point
- Open in Euclidean topology $\implies$ open in d-topology
- Closed in Euclidean topology $\implies$ closed in d-topology
- Lebesgue null $\implies$ closed in d-topology
- Lebesgue measurable $\iff$ G_δ in d-topology, hence also
  $\iff$ F_σ in d-topology, hence also
  $\iff$ Borel in d-topology

Homework. What are the compact sets in the d-topology?

Homework. Show that the d-topology in $\mathbb{R}^2$ is not the same as the product topology from two copies of $\mathbb{R}$ with the d-topology.

Definition. For any finite p, we say that $f \in L^p(\mu)$ if $\int |f|^p \, d\mu < \infty$, and define its p-norm to be

$$\|f\|_p = \|f\|_{L^p(\mu)} = \left(\int |f|^p \, d\mu \right)^{1/p}.$$

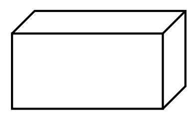
For $p = \infty$, we say that $f \in L^\infty(\mu)$ if there exists a K such that $|f| \leq K$ a.e., and we define its ∞-norm to be

$$\|f\|_\infty = \inf_{|f| \leq K \text{ a.e.}} K.$$

If we define a measure ν by $\nu = \int f \, d\lambda$ where $f \in L^1(\lambda)$ but $f \notin L^p(\lambda)$ for any $p > 1$, then if we try to differentiate ν with respect to a non-regular basis there can be problems.

The strong basis consists of all intervals, i.e. bricks,

$$I_1 \times \cdots \times I_n \subseteq \mathbb{R}^n$$



Given an $f \in L^p$ with $1 \leq p$, we define its maximal function (with respect to the symmetric basis) to be

$$Mf(x) = \sup_{B \ni x} \frac{1}{\lambda(B)} \int_B f \, d\lambda,$$

i.e. B ranges over all balls containing x. Changing between balls with center x and just balls containing x only changes this up to a constant factor.

Theorem. *For all $1 < p \leq \infty$, there is a constant $c = c(p, n)$ such that*

$$\|Mf\|_p \leq c\|f\|_p.$$

Note that this isn't quite true for $p = 1$; however, what is true is that

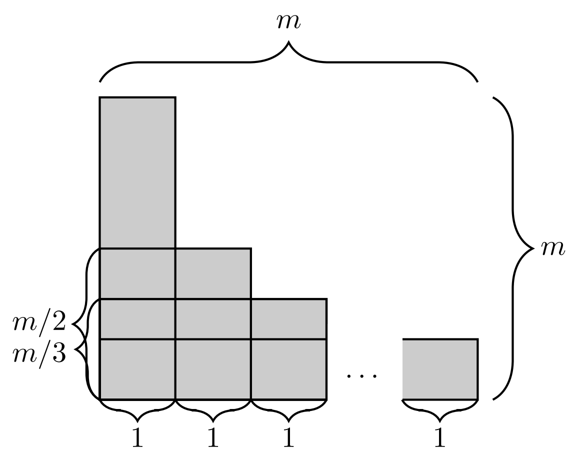
$$\int\limits_{Mf > 1} Mf \leq c \int |f|(1 + \log^+ |f|)$$

where $\log^+ = \max\{\log, 0\}$, or equivalently,

$$\int\limits_{Mf > 1} Mf \leq c \int |f| + c \int\limits_{|f| \geq 1} |f| \log |f|$$

Let's construct a "bad" function for $p = 1$; we'll also switch to the strong (a.k.a. interval) basis. Fix $m \in \mathbb{N}$, and define the set

$$S = \bigcup_{i=1}^{m} [0, i] \times [0, \tfrac{m}{i}).$$

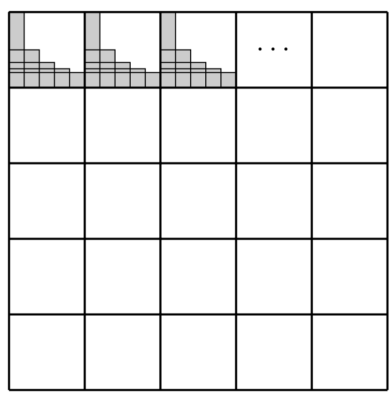


Note that

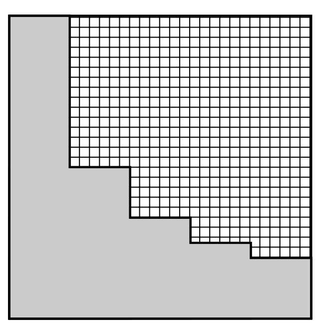
$$\lambda(S) = m\left(1 + \tfrac{1}{2} + \cdots + \tfrac{1}{m}\right) \sim m\log(m).$$

Let's say that $\lambda(S) = mL_m$, so that $L_m \sim \log(m)$.

Now we will "fill" the unit square with disjoint similar copes of S. We do this in countably infinitely many steps. In the first step, divide the unit square into $m \times m$ squares, and put a (scaled) copy of S in each one.



The measure of each scaled copy of S is $\frac{L_m}{m}$, so that the proportion of the unit square that is covered is exactly $\frac{L_m}{m}$. Thus, $\left(1 - \frac{L_m}{m}\right)$ is missing. In the second step, divide up the remaining area into small squares and do the same thing to each of them as what we just did to the unit square.



(one of the m^2 squares in the above picture)

After the nth step, there will be $\left(1 - \frac{L_m}{m}\right)^n$ missing from the square, which $\to 0$ as $n \to \infty$ because $L_m \sim \log(m)$. Thus after doing this infinitely many times, there is a only null set U in the unit square which we have not covered. Thus, we have

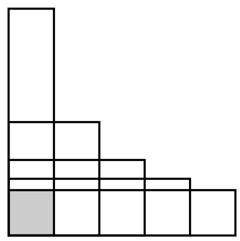
$$[0,1]^2 = \text{disjoint union of } S_1, S_2, \ldots \text{ together with the null set } U,$$

where the S_i's are scaled copies (with assorted scaling factors) of the original set S.

Let $G_n \supseteq U$ be open sets with $\lambda(G_n) < \frac{1}{2^n}$. Define, for an (as yet) unchosen constant k,

$$g_n = k \cdot \left(\chi_{G_n} + \chi_{\substack{\text{lower left corners}\\ \text{of all the } S_i}} \right),$$

where the "lower left corner" of an S_i means the following region:

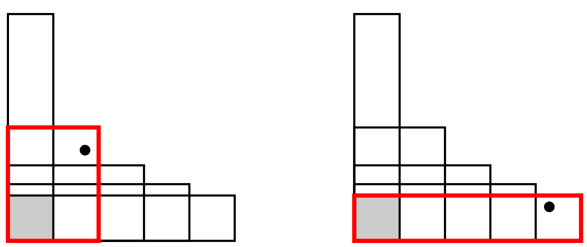


Remark. Obviously, g_n depends on k; but note that the function g_n also implicitly depends on our initial choice of m. This is important because we will later want to choose functions g_n constructed in the above manner, but where each of them uses a different value of m.

In S (the original), the lower left corner has measure 1 and the whole set S has measure mL_m, so the ratio of the measures of the lower left corners and the union of the S_i's is $\frac{1}{mL_m}$. Thus,

$$\int g_n \leq \frac{k}{2^n} + \frac{k}{mL_m}.$$

Key observation: every $x \notin U$ (i.e. every x in some S_i) is contained in a rectangle on which the average value of g_n is $\geq \frac{k}{m}$. For example,



Letting $M_S g_n$ denote the maximal function of g_n with respect the strong basis, this observation implies that

$$M_S g_n(x) = \sup_{R \ni x} \frac{1}{\lambda(R)} \int_R g_n \, d\lambda \geq \frac{k}{m}$$

for all $x \notin U$ (the supremum ranges over all rectangles R containing x).

For all small $\epsilon > 0$ and large K, we choose an m such that $L_m > \frac{2K}{\epsilon}$, and then set $k = mK$, so that

$$\frac{k}{mL_m} = \frac{K}{L_m} < \frac{\epsilon}{2}.$$

Finally, note that we can choose n such that $\frac{k}{2^n} < \frac{\epsilon}{2}$. Thus, for all $\epsilon > 0$ and K, we can choose m, k, n such that

$$\int g_n = \frac{k}{2^n} + \frac{k}{mL_m} < \epsilon \quad \text{and} \quad M_S g_n(x) \geq K \text{ almost everywhere.}$$

To summarize: for all ϵ and K, there is a $g = g_{\epsilon,K}$ such that $\int g < \epsilon$ and $M_S g > K$.

Now define

$$g = \sum_{n=1}^{\infty} g_{1/2^n,2^n}$$

so that

$$\int g = \sum_{n=1}^{\infty} \int g_{1/2^n,2^n} \leq \sum_{n=1}^{\infty} \frac{1}{2^n} = 1 \implies g \in L^1,$$

even though $M_S g \geq M_S g_{1/2^n,2^n} > 2^n$ for all n, i.e. $M_S g = \infty$.

This gets us an L^1 function whose maximal operator is infinite, but how can we modify this construction to a make a function whose derivative is also infinite?

In the construction of each function $g_{1/2^n,2^n}$ as n goes to ∞, let the maximum size of the staircases used go to 0 (i.e., given the m chosen in the construction of $g_{1/2^n,2^n}$, instead of initially subdividing the unit square into $m \times m$, divide it up even more finely). Then letting $g = \sum g_{1/2^n,2^n}$ again, we can get $\overline{D}_S g$ infinite almost everywhere.

This is necessary because the maximal operator is a supremum over all rectangles containing a given point, while the derivative is defined in terms of shrinking sequences of rectangles containing a given point, so the increasing "badness" of the functions $g_{1/2^n,2^n}$ needs to be visible at smaller and smaller scales in order for the derivative to notice.

Going back to our theorem, here are two lemmas.

Lemma. *For any f as in the theorem,*

$$\mu(\{x \mid Mf(x) > t\}) \leq \frac{c}{t} \int_{\{|f| > \frac{t}{2}\}} |f| \, d\mu.$$

Lemma. *For any $g \geq 0$,*

$$\int g \, d\mu = \int_0^\infty \mu(\{x \mid g(x) > t\}) \, dt.$$

Proof. Using Fubini's theorem, which we didn't cover in class,

$$\int_0^\infty \mu(\{x \mid g(x) > t\}) \, dt = \int_0^\infty \int \chi_{\{x|g(x)>t\}}(x) \, d\mu \, dt$$

$$= \int \underbrace{\int_0^\infty \chi_{\{x|g(x)>t\}}(x) \, dt}_{\int_0^{g(x)} 1 \, dt} \, d\mu = \int g(x) \, d\mu \qquad \square.$$

This second lemma makes intuitive sense because it is one way of capturing the idea that the integral is the area under a curve.

Lecture 7 (2012-10-23)

There will be no class on Thursday; there will instead be office hours in case anyone has questions before the exam, which is next week (October 30).

Let $\mathcal{M}$ denote the collection of all measurable functions on a set X, and by $\mathcal{M}^+$, the collection of all non-negative measurable functions $f : X \to [0, \infty]$.

Consider an operator $M : \mathcal{M} \to \mathcal{M}^+$, such as the maximal operator $M(f) = \sup_{B \ni x} \frac{1}{|B|} \int_B |f|$. Suppose that $\|Mf\|_\infty \leq \|f\|_\infty$ for all $f \in \mathcal{M}$ and $M(f+g) \leq Mf + Mg$, and also suppose that, for some measure μ on X, there is some $c > 0$ such that

$$\mu(\{x : Mf > t\}) < \frac{c\|f\|_1}{t}.$$

(This last condition is called the "weak 1-1 inequality".)

Claim. Then there are constants c_p, c' such that $\|Mf\|_p \leq c_p \|f\|_p$ for all $p > 1$ and every $f \in \mathcal{M}$, and

$$\int_{\{Mf>1\}} Mf(x)\, d\mu \leq c' \int_X |f|(1 + \log^+ |f|)\, d\mu.$$

Lemma. *We have that*

$$\mu(\{x \mid Mf(x) > t\}) \leq \frac{2c}{t} \int_{|f| > \frac{t}{2}} |f|\, d\mu.$$

Proof. "Cut" f at $\pm\frac{t}{2}$, and call this function f_1; in other words take

$$f_1 = \max\{\min\{f, \tfrac{t}{2}\}, -\tfrac{t}{2}\}.$$

Let $f_2 = f - f_1$. Then $|f_1| \leq \frac{t}{2}$ everywhere, so that by our assumptions about M, we have $\|Mf_1\|_\infty \leq \frac{t}{2}$, and hence $Mf_1 \leq \frac{t}{2}$ a.e. We also have

$$Mf \leq Mf_1 + Mf_2.$$

Let $A = \{x \mid Mf(x) > t\}$. Then for almost all $x \in A$,

$$t < Mf(x) \leq \frac{t}{2} + Mf_2(x)$$

so for almost all $x \in A$, $\frac{t}{2} \leq Mf_2(x)$. Now note that

$$\mu(A) \leq \mu(\{x \mid Mf_2(x) > \tfrac{t}{2}\}) \overset{\text{weak 1-1}}{\leq} \frac{c\|f_2\|_1}{t/2} = \frac{2c}{t}\int_X |f_2| = \frac{2c}{t} \int_{|f|>\frac{t}{2}} |f_2| \leq \frac{2c}{t} \int_{|f|>\frac{t}{2}} |f|. \qquad \square$$

Lemma. *For any $g \in \mathcal{M}^+$,*

$$\int_X g\, d\mu = \int_0^\infty \mu(\{x \mid g(x) > t\})\, dt.$$

Proof. This is clear from our intuition about integrals. $\square$

Proof of our claim. We have

$$\int_X |Mf|^p \, d\mu = \int_0^\infty \mu(\{x \mid Mf(x) > t^{1/p}\}) \, dt.$$

Making the change of variables $y = t^{1/p}$, this is equal to

$$\int_0^\infty py^{p-1} \mu(\{x \mid Mf(x) > y\}) \, dy.$$

From our first lemma, we have an inequality

$$\int_0^\infty py^{p-1} \mu(\{x \mid Mf(x) > y\}) \, dy \le \int_0^\infty py^{p-1} \frac{2c}{y} \int\limits_{|f| > \frac{y}{2}} |f| \, d\mu \, dy$$

and then letting $c_p = 2cp$,

$$\int_0^\infty py^{p-1} \mu(\{x \mid Mf(x) > y\}) \, dy \le \int_0^\infty py^{p-1} \frac{2c}{y} \int\limits_{|f| > \frac{y}{2}} |f| \, d\mu \, dy = \iint\limits_{0 < y < 2|f|} c_p y^{p-2} |f| \, d\mu \, dy$$

$$= c_p \int\limits_{|f| > 0} |f| \left(\int_0^{2|f|} y^{p-2} \, dy \right) d\mu \le c_p \int |f|^p$$

where, in the final expression above, the value of c_p has changed by a constant factor. □

Recall that a strong Lebesgue point just means a Lebesgue point with respect to the strong basis.

Theorem. *If $f \in L^p(\mathbb{R}^2)$ for $p > 1$, then almost every point is a strong Lebesgue point of f.*

Lemma. *For a measurable function $f : [0,1]^2 \to \mathbb{R}^+$ such that $f \in L^p$ for $p > 1$, let φ be the measure defined by $\varphi(A) = \int_A f \, d\mu$. Then*

$$\int_{[0,1]^2} \overline{D}_S \varphi \, d\lambda \le c_p \|f\|_p.$$

Proof of Lemma. Fix some y. Take the maximal function in the x coordinate:

$$m(x,y) = \sup_{a \le x \le b} \frac{1}{b-a} \int_a^b f(u,y) \, du$$

where $\sup_{a \le x \le b}$ means the supremum over all intervals $[a,b] \subseteq [0,1]$ containing x. We will see that $m \in L^p$. Define

$$E = \left\{ (x,y) \;\middle|\; \lim_{\substack{c \le y \le d \\ (d-c) \to 0}} \frac{1}{d-c} \int_c^d m(x,v) \, dv = m(x,y) \right\}.$$

If $x \in [a,b]$, then

$$\frac{1}{(b-a)(d-c)} \int\limits_{[a,b] \times [c,d]} f = \frac{1}{d-c} \int_c^d \left(\frac{1}{b-a} \int_a^b f(u,v) \, du \right) dv \le \frac{1}{d-c} \int_c^d m(x,v) \, dv.$$

If $(x,y) \in E$, then

$$\overline{D}_S \varphi(x,y) \leq m(x,y).$$

Therefore, if we show that almost every point in $[0,1]^2$ is in E, and additionally show that $\int_{[0,1]^2} m(x,y)$ is bounded by $c_p\|f\|_p$, we will have proved the lemma.

We know that for one-dimensional functions in L^1, almost every point is a Lebesgue point; thus, if we show that m is $L^1([0,1]^2)$, we will have that almost every one-dimensional "slice" of m is in $L^1([0,1])$, thereby implying that for almost every slice $\{x\} \times [0,1]$, almost every point of the slice is in E; this then implies that almost every point of $[0,1]^2$ is in E (apply Fubini's theorem to the characteristic function of E).

Now we see that will be enough to show that $\int_{[0,1]^2} m \leq c_p\|f_p\|$. This follows from the claim we proved earlier today:

$$\int\limits_{[0,1]^2} m \leq \left(\int\limits_{[0,1]^2} m^p\right)^{1/p} = \left(\int_0^1 \int_0^1 m^p(x,y)\,dx\,dy\right)^{1/p} \leq \left(\int_0^1 c_p \int_0^1 f(x,y)^p\,dx\,dy\right)^{1/p} = c_p\|f\|_p.$$

□

Proof of Theorem. WLOG, we can assume we are working in the unit square, because being a Lebesgue point is a local property; we are taking smaller and smaller balls around a point, so we can forget about what the function is doing far away.

Let $f \in L^p$. Choose a continuous g such that $\|f-g\|_p < \epsilon^2$. Let $h = f-g$, and define $\varphi(S) = \int_S |h|\,d\lambda$. Define two sets

$$A = \{x : |h(x)| > \epsilon\}, \quad B = \{x : \overline{D}_S\varphi(x) > \epsilon\}.$$

We will show that these sets are small.

Note that

$$\int\limits_{[0,1]^2} h \leq \|h\|_p < \epsilon^2$$

so that $\lambda(A) \leq \epsilon$. The lemma then implies that $\lambda(B) \leq c_p\epsilon$.

If $x \notin A \cup B$, then (letting R be a sufficiently small rectangle)

$$\int_R |f(t)-f(x)|\,dt \leq \underbrace{\int_R |f(t)-g(t)|\,dt}_{\leq \epsilon|R| \text{ since } x \notin B} + \underbrace{\int_R |g(t)-g(x)|\,dt}_{\substack{\leq \epsilon|R| \text{ since } g \text{ is continuous} \\ \text{and } R \text{ was chosen small enough}}} + \underbrace{\int_R |g(x)-f(x)|\,dx}_{\substack{=|g(x)-f(x)|\cdot|R| \\ \leq \epsilon|R| \text{ since } x \notin A}}.$$

Thus, for any given $\epsilon > 0$, the measure of the set of points x where

$$\limsup_{R \to x} \frac{1}{|R|}\int_R |f(t)-f(x)|\,dt > c\epsilon$$

is less than $c\epsilon$. □

We've shown that in any regular basis, we can differentiate any L^1 function (this fact, applied to characteristic functions, implies that almost every point of a set is a density point).

We've shown that in the strong basis, we can differentiate any L^p function for $p > 1$, but not necessarily L^1 functions (this fact, applied to characteristic functions, implies that almost every point of a set is a strong density point).

If, instead of axis-parallel rectangles, we take the basis consisting of all rectangles including rotated ones, then **NOTHING IS TRUE**.

Homework $(*)$**.** There exists a compact $K \subseteq \mathbb{R}^2$ of positive measure such that, for each $x \in K$, there exists a line segment that meets K at no other point. (Such a set is called a "hedgehog".)

A hedgehog set demonstrates that the rotated rectangle basis is bad; for any $x \in K$, we can choose some other $y \in K$ arbitrarily close to it such that, taking the line from x to y, we can find a line segment and then a very thin rectangle around that line segment where most of the rectangle is disjoint from the set K, making the density go to 0.

Lecture 8 (2012-10-25)

No class - office hours to ask questions before midterm.

Lecture 9 (2012-10-30)

Midterm.

Lecture 10 (2012-11-01)

Today we'll go back over some more basic material that not everyone has seen yet.

Recall the definition of L^p space:

Definition. Given a measure space (X, μ), and a function f on X such that $\int |f|^p \, d\mu < \infty$, we say that $f \in L^p(\mu)$. When there exists a K such that $|f(x)| \leq K$ almost everywhere, we say that $f \in L^\infty(\mu)$.

Definition. A normed space is a vector space V with a function $\|\cdot\| : V \to \mathbb{R}$ such that

- $\|f\| \geq 0$ for all f, with equality if and only if $f = 0$
- $\|cf\| = |c| \cdot \|f\|$
- $\|f_1 + f_2\| \leq \|f_1\| + \|f_2\|$

We say that V is complete if $\rho(f, g) = \|f - g\|$ is a complete metric. A complete normed space is called a Banach space.

As we will see later, the norms

$$\|f\|_p = \left(\int |f|^p \, d\mu \right)^{1/p}, \qquad \|f\|_\infty = \inf\{K : |f(x)| \leq K \text{ a.e.}\}$$

make L^p and L^∞, respectively, into complete normed spaces, but only after we identify functions f and g if $f = g$ a.e. (otherwise we can have $\|f\| = 0$ even when $f \neq 0$).

Theorem (Hölder's Inequality)**.** *For any $f \in L^p$ and $g \in L^q$ where $\frac{1}{p} + \frac{1}{q} = 1$, and either $1 < p, q < \infty$ or $p = 1, q = \infty$, we have $fg \in L^1$ and*

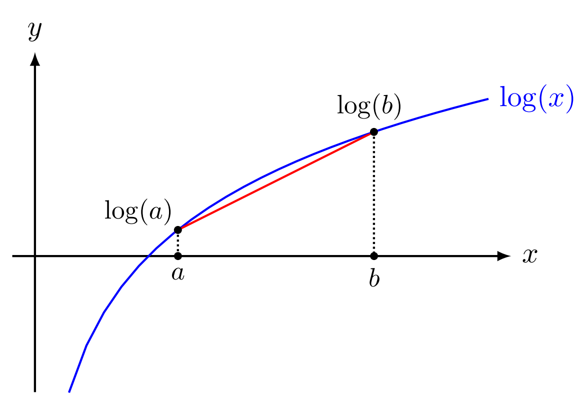
$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

Lemma. *For any $a, b \geq 0$ and $0 < \lambda < 1$, we have $a^\lambda b^{1-\lambda} \leq \lambda a + (1 - \lambda) b$.*

Proof of lemma. Taking logarithms of both sides,

$$\lambda \log(a) + (1 - \lambda) \log(b) \leq \log(\lambda a + (1 - \lambda) b).$$

But log is concave,



so this is true. $\square$

Proof of Hölder. Let's do the case when $\|f\|_p = 1$ and $\|g\|_q = 1$ first. Let $\lambda = \frac{1}{p}$ and $1 - \lambda = \frac{1}{q}$, and let $a = |f(x)|^p$ and $b = |g(x)|^q$. The lemma implies that

$$|f(x)| \cdot |g(x)| \leq \frac{1}{p}|f(x)|^p + \frac{1}{q}|g(x)|^q,$$

and therefore

$$\int |f(x)| \cdot |g(x)| \leq \frac{1}{p} \underbrace{\int |f(x)|^p}_{=1} + \frac{1}{q} \underbrace{\int |g(x)|^q}_{=1} = \frac{1}{p} + \frac{1}{q} = 1.$$

In the general case, we can just let $F = \frac{f}{\|f\|_p}$ and $G = \frac{g}{\|g\|_q}$, so that we can apply the special case to see that

$$\int |FG| \leq 1$$

and hence

$$\int \frac{|fg|}{\|f\|_p \|g\|_q} \leq 1 \implies \int |fg| \leq \|f\|_p \|g\|_q.$$

Finally, if $p = 1$ and $q = \infty$, we have that

$$\int |fg| \leq \int |f| \cdot K = K \int |f|$$

when $|g| \leq K$ almost everywhere. $\square$

Theorem (Minkowski Inequality). *For any $1 \leq p \leq \infty$, we have*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p$$

for any $f, g \in L^p$.

Proof. If $p = 1$ or $p = \infty$, this is trivial. Now suppose $1 < p < \infty$. First, let's show that $f + g \in L^p$: by the convexity of the logarithm again, we can see that

$$\left| \frac{f + g}{2} \right|^p \leq \frac{|f|^p + |g|^p}{2}$$

and therefore $|f + g|^p \leq 2^{p-1}(|f|^p + |g|^p)$. This shows that $f + g \in L^p$.

Now let q be the solution to $\frac{1}{p} + \frac{1}{q} = 1$ and let $F \in L^p$. Then we claim $|F|^{p-1} \in L^q$. This is because $pq - q = p$ implies

$$\left(|F|^{p-1}\right)^q = |F|^p$$

and moreover

$$\||F|^{p-1}\|_q^q = \|F\|_p^p.$$

Now for the final step. Note that

$$\|f + g\|_p^p = \int |f + g|^p = \int \underbrace{|f + g|^{p-1}}_{\in L^q} |f + g| \leq \int |f + g|^{p-1}|f| + \int |f + g|^{p-1}|g|$$

$$\overset{\text{Hölder}}{\leq} \left\||f + g|^{p-1}\right\|_q \cdot \|f\|_p + \left\||f + g|^{p-1}\right\|_q \cdot \|g\|_p = \|f + g\|_p^{p/q} \cdot \|f\|_p + \|f + g\|_p^{p/q} \|g\|_p.$$

Therefore

$$\|f+g\|_p^p \leq \|f+g\|_p^{p/q}(\|f\|_p + \|g\|_p),$$

but since $p - p/q = p(1-\frac{1}{q}) = 1$, this implies

$$\|f+g|_p \leq \|f\|_p + \|g\|_p. \qquad \square$$

Given a normed linear space B, and a sequence $b_1, b_2, \ldots \in B$, let $s_n = \sum_{j=1}^n b_j \in B$. Then we say that the series $\sum_{j=1}^\infty b_j$ converges if there is an $s \in B$ such that $\|s_n - s\| \to 0$. We say that it is absolutely convergent if $\sum_{j=1}^\infty \|b_j\| < \infty$.

Theorem (Riesz-Fisher). *For any $1 \leq p \leq \infty$, L^p is complete.*

Lemma. *For any normed linear space B, B is complete if and only if every absolutely convergent sequence converges.*

Proof of lemma. Suppose B is complete. For any absolutely convergent sequence $b_1, b_2, \ldots, \in B$,

$$\|s_n - s_m\| = \left\| \sum_{j=n+1}^m b_j \right\| \leq \sum_{j=n+1}^m \|b_j\| < \epsilon$$

if n is large enough. Thus the sequence of s_n's is Cauchy, and because B is complete the limit exists.

Now conversely, suppose that $b_1, b_2, \ldots \in B$ is a Cauchy sequence, and assume that every absolutely convergent sequence in B converges. Then for any k, there exists an N_k such that for all $n, m \geq N_k$, we have $\|b_m - b_n\| < \frac{1}{2^k}$. Thus,

$$b_{N_1} + (b_{N_2} - b_{N_1}) + (b_{N_3} - b_{N_2}) + \cdots$$

is absolutely convergent because

$$\|b_{N_1}\| + \|b_{N_2} - b_{N_1}\| + \|b_{N_3} - b_{N_2}\| + \cdots \leq \|b_{N_1}\| + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots < \infty$$

The partial sums of this series are just $b_{N_1}, b_{N_2}, \ldots$, so that by the assumption that every absolutely convergent series converges, there must be some $b \in B$ such that $b_{N_k} \to b$. But since $b_1, b_2, \ldots$ is a Cauchy sequence, we must also have that $b_n \to b$. $\square$

Homework. Prove the $p = \infty$ case of the Riesz-Fisher theorem.

Proof of Riesz-Fisher. Let $1 \leq p < \infty$. Then for any sequence $f_1, f_2, \ldots \in L^p$, we want to show

$$\underbrace{\sum \|f_n\|_p}_{\sigma} < \infty \implies \sum f_n \text{ converges in } L^p \text{ to some function } f.$$

Let $g_n = \sum_{j=1}^n |f_j|$, which is in L^p. In particular, by Minkowski's inequality,

$$\|g_n\|_p \leq \sum_{j=1}^n \|f_j\|_p \leq \sigma,$$

and therefore $\int |g_n|^p \leq \sigma^p$. The sequence of functions g_n is monotone increasing, so $g_n^p \nearrow g^p$, and therefore

$$\int g^p = \lim \int g_n^p \leq \sigma^p \implies g^p < \infty \text{ a.e.} \implies g < \infty \text{ a.e.}$$

Thus, $\sum |f_n(x)|$ converges for almost all x, so

$$f(x) \stackrel{\text{def}}{=} \sum f_n(x)$$

converges for almost all x. But we still need to show that the partial sums $s_n = \sum_{j=1}^{n} f_j(x)$ satisfy $\|f - s_n\|_p \to 0$. To see this, observe that

$$\int |f - s_n|^p \leq \int (|f| + |s_n|)^p \leq \int (2g)^p,$$

and because $|f - s_n|^p \in L_1$ and $|f - s_n|^p \to 0$ a.e., we have that $|f - s_n|^p \to 0$ in L_1, which is the case if and only if $|f - s_n| \to 0$ in L^p. $\square$

For the rest of this class, assume we have a finite measure space, so that $\mu(X) < \infty$.

Proposition. *For any $f \in L^p$ and $\nu < p$, we also have $f \in L^\nu$. More generally,*

$$L^\infty \subseteq L^p \subseteq L^\nu \subseteq \cdots \subseteq L^1.$$

Proof. Intuitively, this is because raising to any power higher than 1 is no longer concave, it is convex, making Minkowski's inequality fail.

Letting $X_1 = \{x \in X \mid f(x) \leq 1\}$ and $X_2 = \{x \in X \mid f(x) > 1\}$, we have that

$$\int_X |f|^p = \underbrace{\int_{X_1} |f|^p}_{\leq \mu(X) < \infty} + \int_{X_2} |f|^p \geq \int_{X_2} |f|^\nu$$

because $|f|^p \geq |f|^\nu$ on X_2, so that

$$\int_X |f|^\nu = \underbrace{\int_{X_1} |f|^\nu}_{\leq \mu(X) < \infty} + \int_{X_2} |f|^\nu$$

is finite, and therefore $f \in L^\nu$. $\square$

Proposition. *If $f \in L^\infty$, then $f \in L^p$ for any p. Moreover, $\|f\|_p \to \|f\|_\infty$ as $p \to \infty$.*

Proof. For any $t < \|f\|_\infty$, we have that $\mu(A) > 0$ where $A = \{x \mid f(x) > t\}$. Then

$$\int_X |f|^p \geq \int_A |f|^p \geq \int_A t^p = t^p \mu(A).$$

Thus $\|f\|_p \geq t\mu(A)^{1/p}$. As $p \to \infty$, we have that

$$\liminf_{p \to \infty} \|f\|_p \geq t \cdot 1,$$

and therefore $\liminf \|f\|_p \geq \|f\|_\infty$. In the other direction, $|f| \leq \|f\|_\infty$ a.e., so that

$$\int |f|^p \leq \int \|f\|_\infty^p = \|f\|_\infty^p \cdot \mu(X),$$

and hence $\|f\|_p \leq \|f\|_\infty \cdot \mu(X)^{1/p}$. Therefore, $\limsup_{p \to \infty} \|f\|_p \leq \|f\|_\infty \cdot 1$. $\square$

Homework. If $\mu(X) < \infty$, is it true that if $f \in L^p$ for all $1 < p < \infty$, then $f \in L^\infty$?

Theorem. *For any $1 \le p \le \infty$, the simple functions are dense in L^p.*

Proof. The case of $p = \infty$ is easy - to define a simple function close to $f \in L^\infty$, break up the range of f in steps of size ϵ, and on the set $\{x \mid k\epsilon \le f(x) < (k+1)\epsilon\}$, set the simple function to be $k\epsilon$. This is a simple function because it will only take on finitely many values (since $\|f\|_\infty$ is finite), and it differs from f by at most ϵ everywhere.

Now let $1 \le p < \infty$. Note that it is enough to show the claim is true for $f \ge 0$. Fix an $\epsilon > 0$, and let $A = \{x \mid f(x) > \delta\}$ where δ is chosen such that

$$\int_{X/A} |f|^p < \frac{\epsilon}{4}.$$

Choose an n such that $A_n = \{x : f(x) \le n\}$ satisfies

$$\int_{X \setminus A_n} f^p < \frac{\epsilon}{4}.$$

Lastly, choose η such that

$$\frac{\epsilon}{4\mu(A_n)} = \eta^p.$$

Then, again breaking up the range of f, we define the set

$$M_\nu = \{x \in A_n \mid (\nu - 1)\eta \le f(x) \le \nu\eta\},$$

and now we can define the simple function

$$g(x) = \begin{cases} 0 & \text{if } x \notin A_n, \\ (\nu - 1) & \text{if } x \in M_\nu. \end{cases}$$

This satisfies

$$\int |f - g|^p = \int_{X \setminus A} |f|^p + \sum_\nu \int_{M_\nu} |f - g|^p + \int_{A/A_n} |f|^p < \left(\frac{\epsilon}{4}\right) + \underbrace{\eta^p \cdot \mu(A_n)}_{=\epsilon/4} + \left(\frac{\epsilon}{4}\right) < \epsilon. \qquad \square$$

Next time, we'll look more at the special properties of $\frac{1}{p} + \frac{1}{q} = 1$. In particular, we'll prove that the dual of L^p is L^q and the dual of L^1 is L^∞.

Lecture 11 (2012-11-06)

There will be no class on Thursday.

There were 3 questions on the exam, with a maximum score of 1 each. The grading scale for the midterm is

A	2
B	1.5
C	1
D	0.5

Some people were confused about the definition of a measurable function. Recall that, for measurable spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$, a function $f : X \to Y$ is $(\mathcal{A}, \mathcal{B})$-measurable if for all $B \in \mathcal{B}$, we have $f^{-1}(B) \in \mathcal{A}$. When Y is a topological space, then we just say $\mathcal{A}$-measurable, because (unless specified otherwise) Y will be given the Borel σ-algebra. This is equivalent to requiring that $f^{-1}(G) \in \mathcal{A}$ for all open $G \subseteq Y$.

Let's get back to what we talked about last class.

Definition. Let B be a normed linear space. A linear map $\Lambda : B \to \mathbb{R}$ (or $\Lambda : B \to \mathbb{C}$) is called a linear functional.

Definition. Let B_1, B_2 be normed linear spaces. A linear map $A : B_1 \to B_2$ is called a linear operator. We say that A is bounded if there is some K for which $\|Ax\| \leq K\|x\|$ for all $x \in B_1$. The best possible K is

$$\sup_{\|x\|=1} \|Ax\|.$$

We call this the norm of A. It turns out that the bounded linear operators themselves form a normed linear space under this norm; and in particular, we call

$$B^* = \{\text{bounded linear functionals } \Lambda \text{ on } B\}$$

the dual of B, which is a normed linear space with $\|\Lambda\| = \sup_{\|x\|=1} |\Lambda(x)|$.

Example. Let $B = \mathbb{R}^n$, with norm

$$\|x\| = \|x\|_2 = \left(\sum x_j^2\right)^{1/2}.$$

What is B^*? Let $e_1, \ldots, e_n$ denote the standard basis for $\mathbb{R}^n$, and given a linear functional Λ, let $\Lambda(e_i) = c_i$. Then

$$\Lambda(x) = \sum x_j c_j = \langle x, c \rangle \leq \|x\|_2 \cdot \|c\|_2,$$

so $\|\Lambda\| \leq \|c\|_2$, and because $|\Lambda(c)| = \|c\|_2^2$ we must have that $\|\Lambda\| = \|c\|_2$. The map identifying Λ with c is linear, and it preserves norms. Thus $B^* \cong B$.

Theorem. *For any $1 < p, q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, or $p = 1, q = \infty$, we have $(L^p)^* = L^q$.*

Remark. Note that this implies $(L^2)^* = L^2$. Also, $(L^\infty)^* \neq L^1$.

What do we really mean by this theorem? For any $\Lambda \in (L^p)^*$, there is some $g \in L^q$ such that $\Lambda(f) = \int fg \, d\mu$ for all $f \in L^p$.

Proof. Suppose that $g \in L^q$. Define $\Lambda(f) = \int fg \, d\mu$. We need to show that this is in fact a bounded linear operator.

It is obvious that Λ is linear, and it is bounded by Hölder's inequality:

$$|\Lambda(f)| = \left| \int fg \right| \leq \|f\|_p \|g\|_q.$$

This shows that $\|\Lambda\| \leq \|g\|_q$, and in fact we have equality, because choosing $f = g^{q-1}$, we have $\|f\|_p^p = \|g\|_q^q$, and therefore

$$\left| \int fg \right| = \int |g^q| = \|f\|_p \|g\|_q.$$

Now we want to prove that any element of $(L^p)^*$ can be obtained this way. We will prove it in the case that μ is finite, i.e. $\mu(X) < \infty$.

Homework. Prove this claim when μ is any σ-finite measure.

Given a linear functional $\Lambda \in (L^p)^*$, how can we construct the corresponding g? For any meausurable set A, we have that $\chi_A \in L^\infty \subseteq L^p$ (this inclusion holds because μ is finite). Now denote $\Lambda(\chi_A) = \varphi(A)$. We claim that φ is a measure (note that we don't know $\varphi(A)$ will be positive, so this could be a *signed* measure).

Given pairwise disjoint measurable $A_1, A_2, \ldots$ we have

$$A = \bigcup_{j=1}^{\infty} A_j = \left(\bigcup_{j=1}^{n} A_j \right) \cup \left(\underbrace{\bigcup_{j=n+1}^{\infty} A_j}_{B_n} \right).$$

Because Λ is linear,

$$\varphi(A) = \sum_{j=1}^{n} \varphi(A_j) + \varphi(B_n).$$

We need to show that $\varphi(B_n) \to 0$ as $n \to \infty$. Because Λ is bounded,

$$|\varphi(B_n)| = |\Lambda(\chi_{B_n})| \leq \|\Lambda\| \cdot \|\chi_{B_n}\|_{L^p} \underset{\substack{\uparrow \\ \text{not true} \\ \text{for } p=\infty}}{=} \|\Lambda\| \cdot \mu(B_n)^{1/p}.$$

Because

$$\mu(B_n) = \sum_{j=n+1}^{\infty} \mu(A_j) \to 0 \quad \text{as} \quad n \to \infty,$$

we are done. Note that this also demonstrates that $\varphi \ll \mu$. By the Radon-Nykodim theorem, there is some $g \in L^1$ such that

$$\Lambda(\chi_A) = \varphi(A) = \int g \chi_A \, d\mu.$$

So, we just need to show that $g \in L^q$. In the case that $p = 1$, we have

$$\left| \int_A g \, d\mu \right| = \Lambda(\chi_A) \leq \|\Lambda\| \cdot \mu(A),$$

and therefore

$$\left|\frac{1}{\mu(A)}\int_A g\,d\mu\right| \le \|\Lambda\|$$

for any meausurable A with $\mu(A) > 0$. This shows that $|g| \le \|\Lambda\|$ μ-a.e.

For $1 < p < \infty$, because

$$\Lambda(h) = \int hg\,d\mu \quad \text{for every characteristic function } h,$$

we therefore also have that

$$\Lambda(h) = \int hg\,d\mu \quad \text{for every simple function } h$$

because Λ and $\int$ are linear, and then

$$\Lambda(h) = \int hg\,d\mu \quad \text{for every } L^\infty \text{ function } h$$

because if $h \in L^\infty$ then for any $\epsilon > 0$ there is some simple h_1 such that $\|h - h_1\|_\infty < \epsilon$ (we've proved this before), and this means that

$$\left|\Lambda(h) - \int hg\right| = \left|\Lambda(h-h_1) - \int (h-h_1)g + \underbrace{\Lambda(h_1) - \int h_1 g}_{=0}\right| \le \underbrace{|\Lambda(h-h_1)|}_{\substack{\le \|\Lambda\|\cdot\|h-h_1\|_p \\ \le \|\Lambda\|\cdot\epsilon\mu(X)^{1/p}}} + \underbrace{\left|\int (h-h_1)g\right|}_{\le \epsilon\cdot\|g\|_1}.$$

Because $g \in L^1$, the sets $A_n = \{x : |g(x)| \le n\}$ have the property that $\mu(X \setminus A_n) \to 0$ as $n \to \infty$. Let

$$f = |g|^{q-1}\cdot \operatorname{sign}(g)\cdot \chi_{A_n}.$$

Because $f \in L^\infty$, we know that $\Lambda(f) = \int fg$. We have

$$\int_{A_n} |g|^q = \Lambda(f) \le \|\Lambda\| \cdot \left(\int_X |f|^p\right)^{1/p} = \|\Lambda\| \cdot \left(\int_{A_n} |g|^q\right)^{1/p}$$

and therefore

$$\left(\int_{A_n} |g|^q\right)^{1-1/p} = \left(\int_{A_n} |g|^q\right)^{1/q} \le \|\Lambda\|$$

for all n, which implies that $\|g\|_q \le \|\Lambda\|$. To show that there is actually equality, choose f such that $\|f\|_p^p = \|g\|_q^q$. □

Now we will prove Fubini's theorem. First, we need to discuss what it means to take the product of two measures.

Given two measure spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$, for any $A \in \mathcal{A}$ and $B \in \mathcal{B}$, we define $\varphi(A \times B) = \mu(A)\cdot\nu(B)$. We extend φ to a measure on the σ-algebra on $Z = X \times Y$ which is generated by sets of the form $A \times B$ for $A \in \mathcal{A}$, $B \in \mathcal{B}$.

Theorem (Fubini). *If φ is σ-finite and $\int_Z f\,d\varphi$ exists, then the function*

$$g(x) \stackrel{\text{def}}{=} \int_Y f(x,y)\,d\nu$$

exists for almost all $x \in X$, and

$$\int_X g(x)\,d\mu = \int_Z f\,d\varphi.$$

Homework. Let $X = Y = [0, 1]$, and define $f : X \times Y \to \mathbb{R}$ to be

$$f(x, y) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{if } x \neq y. \end{cases}$$

Let μ be Lebesgue measure on X, and ν be counting measure on Y. Calculate

$$\int_X \left(\int_Y f(x, y)\, d\nu \right) d\mu$$

and

$$\int_Y \left(\int_X f(x, y)\, d\mu \right) d\nu,$$

and explain why Fubini fails.

Homework. On the midterm, you showed that for measurable $A, B \subseteq [0, 1]$, the function

$$f(t) = \lambda(A \cap (B + t))$$

is continuous. Now, find $\int f(t)$.

Lecture 12 (2012-11-08)

No class.

Lecture 13 (2012-11-13)

Today we'll talk about calculus of several variables.

On the real line, we know

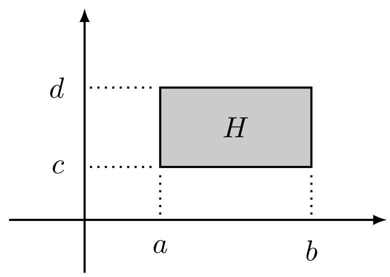
$$\int_a^b f(x)\,dx = F(b) - F(a), \qquad f = F'.$$

Note that $\{a, b\}$ is the boundary of the interval $[a, b]$.

Does this generalize to higher dimensions? That is, given some region H, can we say

$$\int_H (\text{something}) = \int_{\partial H} (\text{something else}) \quad ?$$

Let's start with the case when H is a rectangle $[a, b] \times [c, d]$,



Let F be a function, and F_x and F_y its partial derivatives. Then

$$\int F_y(x,y)\,dx\,dy = \int_a^b \left(\int_c^d F_y(x,y)\,dy \right) dx = \int_a^b (F(x,d) - F(x,c))\,dx = -\int_{\partial H} F(x,y)\,dx$$

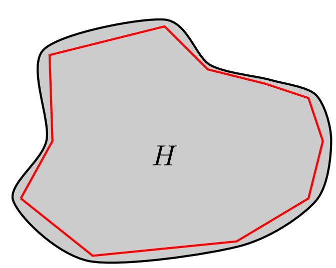
Similarly,

$$\int G_x(x,y)\,dx\,dy = \int_c^d \left(\int_a^b g_x(x,y)\,dx \right) dy = \int_c^d (G(b,y) - G(a,y))\,dy = \int_{\partial H} G(x,y)\,dy.$$

Putting these together, if F, G, F_y, and G_x are continuous, then

$$\int_H (G_x - F_y)\,dx\,dy = \int_{\partial H} F\,dx + G\,dy.$$

Thus, the statement is true for all rectangles H, and hence true for all triangles (divide a rectangle diagonally), hence true for all polygons H (triangulate the polygon). This then implies it is true for all closed rectifiable curves H, because we can approximate them with polygons:

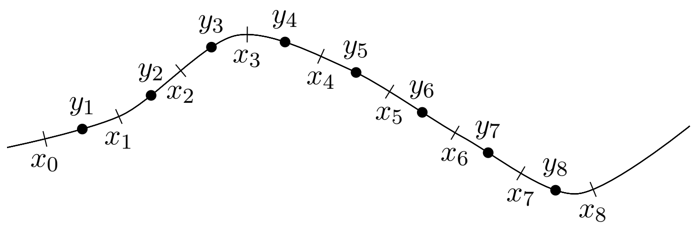


with

$$\int_P \to \int_H, \qquad \int_{\partial P} \to \int_{\partial H}.$$

Finally, this implies it is true for all H with rectifiable boundary (rectifiable means finite length).

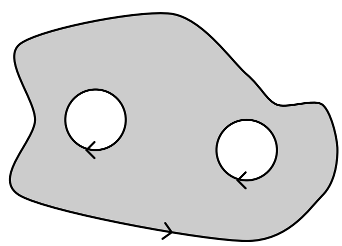
Remark. What exactly do we mean by $\int_\gamma f\,dg$ for a curve γ? We break up γ into smaller and smaller segments



and then define the integral to be the limit of the quantity

$$\sum f(y_j)(g(x_j) - g(x_{j-1})).$$

Given a domain $H \subseteq \mathbb{R}^d$ (which may have holes),

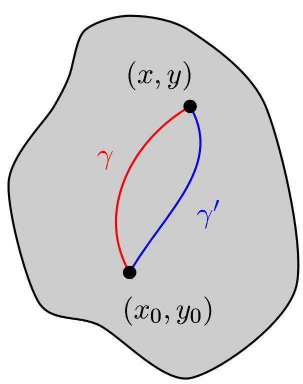


and a function $\varphi = (\varphi_1, \ldots, \varphi_d)$ on H, we say that a function $u : H \to \mathbb{R}$ is a primitive of φ if u is differentiable on H and $u' = \varphi$.

Does every function have a primitive? No. Suppose $u' = (f, g)$, and that u is twice differentiable; if $f = u_x$ and $g = u_y$, then

$$f_y = u_{xy} = u_{yx} = g_x.$$

Our result above implies that for any rectifiable curves γ and γ' with endpoints (x_0, y_0) and (x, y),



we will have

$$\int_\gamma f\,dx + g\,dy = u(x,y) - u(x_0, y_0) = \int_{\gamma'} f\,dx + g\,dy.$$

Therefore, a necessary condition is that $\int_\gamma = 0$ for any closed curve γ. When we assume everything is nice, it turns out this is also a sufficient condition.

Theorem. *Given continuous f, g on H, there is a primitive u for $\varphi = (f, g)$ if and only if $\int_\gamma f\,dx + g\,dy = 0$ for any closed curve γ.*

Proof. We just did the $\implies$ direction.

To see $\impliedby$, fix some $(x_0, y_0) \in H$, and define

$$u(x, y) = \int_\gamma f\,dx + g\,dy$$

where γ is any curve that connects (x_0, y_0) to (x, y).

Because

$$\frac{u(x, y+h) - u(x, y)}{h} = \frac{\int_y^{y+h} g(t)\,dt}{h} \to g(y) \qquad \text{as } h \to 0$$

and similarly with x and f, we are done. □

Theorem. *If f, g are differentiable on H, and H is simply connected, then there exists a primitive u for (f, g) if and only if $g_x = f_y$.*

Proof. We've done the $\implies$ direction.

To see $\impliedby$, note that because H is simply connected, the region bounded by any curve γ in H is a domain A entirely contained inside H, and therefore

$$\int_\gamma f\,dx + g\,dy = \int_A \underbrace{g_x - f_y}_{=0}\,dxdy = 0.$$ □

Definition. If u is twice differentiable, we say that u is harmonic if

$$\Delta u = u_{xx} + u_{yy} = 0.$$

Δ is called the Laplace operator.

Examples. Clearly, any linear function is harmonic. For a second order polynomial $u = ax^2 + bxy + cy^2$, we have $\Delta u = 2a + 2c$, so that $x^2 - y^2$ and xy are a basis for the vector space of harmonic second order polynomials.

Homework. Find a basis for the vector space of harmonic polynomials in x and y of degree 6.

The key property of harmonic functions is that their value at a point is determined by their integral on a circle around that point, which is what we'll prove now.

We know that

$$\int_H G_x = \int_{\partial H} G, \qquad \int_H F_y = -\int_{\partial H} F.$$

Let's choose $G = u_x v$ and $F = u_y v$, for some function v. We get that

$$\int_H (u_{xx}v + u_x v_x)\,dxdy = \int_{\partial H} u_x v, \qquad \int_H (u_{yy}v + u_y v_y)\,dx\,dy = -\int_{\partial H} u_y v.$$

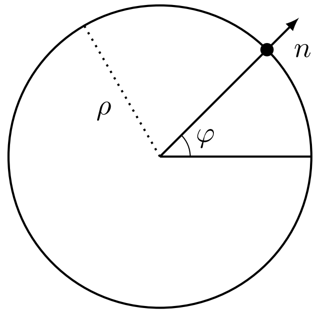
Thus
$$\int_H (\Delta u)v + \langle u', v'\rangle \, dxdy = \int_{\partial H} v(u_x \, dy - u_y \, dx) = \int_{\partial H} v \cdot \frac{\partial u}{\partial n} \, ds$$
where $u_x \, dy - u_y \, dx = \langle u', (dy, -dx)\rangle = \langle u', n'\rangle$ (?) and n is the normalized unit vector in the radial direction (see picture below). This is known as the first Green formula.

The second Green formula (or symmetric Green formula) says that

$$\int_H (\Delta u \cdot v - \Delta v \cdot u) = \int_{\partial H} v \cdot \frac{\partial u}{\partial n} - u \cdot \frac{\partial v}{\partial n}.$$

Choosing $v \equiv 1$, we have
$$\int_H \Delta u = \int_{\partial H} \frac{\partial u}{\partial n},$$
and therefore, if u is harmonic, then
$$\int_{\partial H} \frac{\partial u}{\partial n} \, ds = 0.$$

On a circle,



we have
$$\frac{\partial u}{\partial n} = \frac{\partial u}{\partial r}, \qquad ds = \rho \, d\varphi$$
so that the integrals $\int \frac{\partial f}{\partial n} \, ds$ and $\int \frac{\partial f}{\partial r} \, r d\varphi$ are equivalent for any f. Thus, if u is harmonic,

$$\int_{\partial H} \frac{\partial u}{\partial n} \, ds = 0$$

$$\int_0^{2\pi} \left.\frac{\partial u}{\partial r}\right|_{r=\rho} \rho \, d\varphi = 0$$

$$\rho \int_0^{2\pi} \left.\frac{\partial u}{\partial r}\right|_{r=\rho} d\varphi = 0$$

$$\frac{\partial}{\partial r} \int_0^{2\pi} u \, d\varphi = 0.$$

Therefore, u has the same integral on concentric circles.

Letting $I(r) = \int_0^{2\pi} u(r\cos(\varphi), r\sin(\varphi)) \, d\varphi$, this just means that $I'(r) = 0$, so that $I(r) =$ a constant. In fact, it is easy to see that we must have $I(r) = 2\pi u(0)$ for any r.

If u is harmonic and $v = \frac{1}{4}(x^2 + y^2)$, then $\Delta v = 1$. Letting $H = B(0, \rho)$,

$$\int_H u = \int_0^{2\pi} \left(u \cdot \frac{1}{2}\rho - \frac{1}{4}\rho^2 \frac{\partial u}{\partial r} \right) \rho \, d\varphi = \frac{\rho^2}{2} \int_0^{2\pi} u \, d\varphi.$$

Therfore

$$\frac{1}{\pi\rho^2} \int_{B(0,\rho)} u = \frac{1}{2\pi} \int_0^{2\pi} u \, d\varphi.$$

This says that the average of u on the disc is equal to its average on a circle, which is equal to $u(0)$.

Let's consider harmonic functions that depend only on one variable. For example, if $u(x, y) = u(x)$, then we have $u_{xx} + u_{yy} = u_{xx} = 0$, so $u = ax + b$.

Homework. If $u(x, y)$ depends only on r, what form must u have? What if u depends only on φ?

Let's consider the Laplacian in polar coordinates. We will write $u(r, \varphi) = u(r\cos(\varphi), r\sin(\varphi))$. What does it mean to say $\Delta u = 0$? We have

$$u_r = u_x \cos(\varphi) + u_y \sin(\varphi)$$

so

$$u_{rr} = (u_{xx}\cos(\varphi) + u_{xy}\sin(\varphi))\cos(\varphi) + (u_{yx}\cos(\varphi) + u_{yy}\sin(\varphi))\sin(\varphi).$$

We also have

$$u_\varphi = -u_x \cdot r\sin(\varphi) + u_y \cdot r\cos(\varphi)$$

so

$$u_{\varphi\varphi} = -r(-u_{xx}r\sin(\varphi) + u_{xy}r\cos(\varphi))\sin(\varphi) - u_x\cos(\varphi) + r\cos(\varphi)(-u_{yx}r\sin(\varphi) + u_{yy}r\cos(\varphi)) - u_y r\sin(\varphi)$$

Taken all together, we therefore have $\Delta u = 0$ implies

$$u_{\varphi\varphi} + r^2 u_{rr} + r u_r = 0.$$

If u depends only on r, then $r^2 \cdot u'' + ru' = 0$, so that $r \cdot f' + f = 0$ where $u' = f$, and therefore

$$\frac{f'}{f} = -\frac{1}{r}$$

$$(\log(f))' = -\log(r)'$$

so $u = c\log(r) + d$.

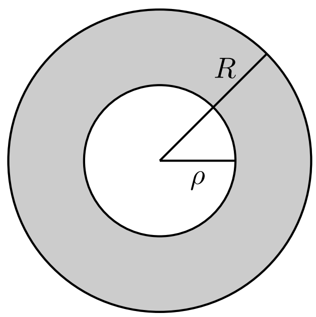
Lecture 14 (2012-11-15)

Homework. Does the function $\left(\frac{y}{x^2+y^2}, \frac{-x}{x^2+y^2}\right)$ have a primitive on the following domains? If yes, find one.

1. The upper half plane
2. The lower half plane
3. The right half plane
4. The left half plane
5. $\mathbb{R}^2 \setminus \{0\}$

Last time, we showed that $\log(r)$ is harmonic on $\mathbb{R}^2 \setminus \{0\}$.

Let u, v be harmonic, and assume that v is of the form $v = -\log(r) + w$.



Because u and v are harmonic (i.e. $\Delta u = 0$ and $\Delta v = 0$) and using the second Green's identity,

$$\begin{aligned}
0 &= \int_0^{2\pi} R\left[u\left(-\frac{1}{R} + \left.\frac{\partial u}{\partial r}\right|_{r=R}\right) - (-\log(R)+w)\left.\frac{\partial u}{\partial r}\right|_{r=R}\right] \\
&\quad - \int_0^{2\pi} \rho\left[u\left(-\frac{1}{\rho} + \left.\frac{\partial u}{\partial r}\right|_{r=\rho}\right) - (-\log(\rho)+w)\left.\frac{\partial u}{\partial r}\right|_{r=\rho}\right] \\
&= \int_0^{2\pi} \underbrace{\left(u\left.\frac{\partial w}{\partial r}\right|_{r=R} - w\left.\frac{\partial u}{\partial r}\right|_{r=R}\right)}_{=0} R - \int_0^{2\pi} \underbrace{\left(u\left.\frac{\partial w}{\partial r}\right|_{r=\rho} - w\left.\frac{\partial u}{\partial r}\right|_{r=\rho}\right)}_{=0} \rho + \int_0^{2\pi} \left.\frac{-u}{R}\cdot R\right|_{r=R} \\
&\quad + R\log(R) \underbrace{\int_0^{2\pi} \left.\frac{\partial u}{\partial r}\right|_{r=R}}_{=0} - \rho\log(\rho) \underbrace{\int_0^{2\pi} \left.\frac{\partial u}{\partial r}\right|_{r=\rho}}_{=0} + \int_0^{2\pi} \left.\frac{u}{\rho}\cdot\rho\right|_{r=\rho}
\end{aligned}$$

Thus, letting

$$I(r) = \int_0^{2\pi} u(r\cos(\varphi), r\sin(\varphi)),$$

we have that $0 = -I(r) + I(\rho)$, so that $I(r) = \text{constant} = 2\pi u(0,0)$.

Corollary (Maximum Principle)**.** *If u is harmonic on a domain H, then if there is any $(x_0, y_0) \in H$ such that $u(x_0, y_0) = \max_{(x,y)\in H} u(x,y)$, then u is constant. Moreover, the same is true for a local maximum.*

Proof. Suppose that $u(x_0, y_0)$ is maximal on $B((x_0, y_0), R)$. Then for any $0 < r < R$,

$$\frac{1}{2\pi}\int_0^{2\pi} u(x_0 + r\cos(\varphi), y_0 + r\sin(\varphi)) = u(x_0, y_0),$$

so

$$\frac{1}{2\pi}\int_0^{2\pi} \underbrace{(u(x_0, y_0) - u(x_0 + r\cos(\varphi), y_0 + r\sin(\varphi)))}_{\geq 0} = 0,$$

so because u is continuous, we must have that u is constant on $B((x_0, y_0), R)$.

For the case of a global maximum, suppose $S = \sup_H u$. If $S = \infty$, then nothing to prove, and if $S < \infty$, let $A = \{z \mid u(z) = S\}$. By what we've proved about local maxima, A must be open; but because u is continuous, A must be closed. Because H is a domain, it is connected, so this forces $A = \varnothing$ or $A = H$. $\square$

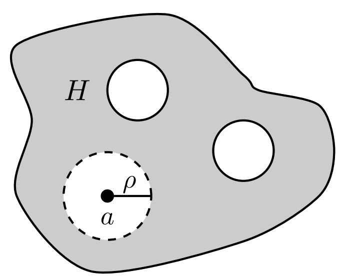
The minimum principle is also true, because the negative of a harmonic function is still harmonic.

Corollary (Uniqueness)**.** *If u and v are harmonic on H, and continuous on the closure $\overline{H}$, then $u|_{\partial H} = v|_{\partial H}$ implies that $u = v$ on H.*

Proof. If u and v are harmonic on H, then $u - v$ is also harmonic on H, and $u - v|_{\partial H} = 0$. If $u - v$ is not constant, then $\max_{\overline{H}}(u - v)$ can only be attained on ∂H, and same for the minimum. $\square$

Remark. We are assuming throughout that the boundaries of our domains H are rectifiable curves, so in particular, domains H are assumed to be precompact.

How can we find u from its values on the boundary of H?



For any $z = (x, y)$, let $r = |a - z|$. Take a small disk of radius ρ around a, and let $H_\rho = H \setminus \overline{B}(a, \rho)$.

Let u be harmonic on H, and let $v = -\log(r) + w$, where w is an arbitrary harmonic function on H. Then v is harmonic on H_ρ (it wouldn't have been defined if we hadn't removed the small disk around a) and because u is harmonic on H, it is also harmonic on H_ρ, so by the second Green's identity,

$$0 = \int_{\partial H_\rho} v\,\frac{\partial u}{\partial n} - u\,\frac{\partial v}{\partial n}.$$

But $\partial H_\rho = \partial H - \partial B(a, \rho)$, so we get that

$$\begin{aligned}\int_{\partial H} v\,\frac{\partial u}{\partial n} - u\,\frac{\partial v}{\partial n} &= \int_{\partial B(a,\rho)} v\,\frac{\partial u}{\partial n} - u\,\frac{\partial v}{\partial n} \\ &= \int_0^{2\pi} \left[(-\log(r) + w)\frac{\partial u}{\partial r}\bigg|_{r=\rho} - u\left(-\frac{1}{r} + \frac{\partial w}{\partial r}\right)\bigg|_{r=\rho}\right]\rho\end{aligned}$$

$$= -\log(\rho) \underbrace{\int_0^{2\pi} \frac{\partial u}{\partial r}\bigg|_{r=\rho}}_{=0} \rho + \underbrace{\int_0^{2\pi} \left(w\frac{\partial u}{\partial r} - u\frac{\partial w}{\partial r} \right)}_{=0,\text{ by second Green's}} \rho + \int_0^{2\pi} u$$

$$= 2\pi u(a)$$

and therefore

$$\int_{\partial H} \left((-\log(r) + w)\frac{\partial u}{\partial n} - \frac{\partial(-\log(r) + w)}{\partial n} u \right) = 2\pi u(a).$$

This is known as the third Green's identity.

The Dirichlet problem is, for a given function f, how to find a w with $\Delta w = 0$ and $w|_{\partial H} = f$. It turns out that if F is twice differentiable, and $F|_{\partial H} = f$, then we can take $w = \min_F \int_H |F'|^2$; that is, the minimum is attained at a function F which is harmonic (we leave this without proof).

Definition. The Green function $g_a(z) = -\log(r) + w_a$ is defined as follows. We want

1. continuous on $\overline{H}$ and $\equiv 0$ on ∂H.
2. $g_a(z) \geq 0$ (in fact $g_a > 0$ except on the boundary); this is because $-\log(r)$ is arbitrarily large close to $r = a$ (we threw out a small disk around a though), so the function is positive at some point, so 0 (the value on the boudnary) must be the minimum.
3. $g_a + \log|a - z|$ is harmonic on H.

It turns out that these properties specify $g_a(z)$ uniquely.

Theorem. *For any $a \neq b$, $g_a(b) = g_b(a)$.*

Proof. We take the domain H and throw out disks of radius ρ around a and b, to avoid the singularities of the logarithms there. Call the resulting domain $H_{\rho\rho}$, i.e.

$$H_{\rho\rho} = H \setminus (B(a,\rho) \cup B(b,\rho)).$$

Then because g_a and g_b are harmonic on $H_{\rho\rho}$, and because they are 0 on ∂H, we have

$$\int_{\partial H_{\rho\rho}} \left(g_a\frac{\partial g_b}{\partial n} - g_b\frac{\partial g_a}{\partial n} \right) = 0 \qquad \text{and} \int_{\partial H} \left(g_a\frac{\partial g_b}{\partial n} - g_b\frac{\partial g_a}{\partial n} \right) = 0$$

which, because $\partial H = \partial H_{\rho\rho} + \partial B(a,\rho) + \partial B(b,\rho)$, implies

$$\int_{\partial B(a,\rho)} \left(g_a\frac{\partial g_b}{\partial n} - g_b\frac{\partial g_a}{\partial n} \right) + \int_{\partial B(b,\rho)} \left(g_a\frac{\partial g_b}{\partial n} - g_b\frac{\partial g_a}{\partial n} \right) = 0.$$

Note that

$$\int_{\partial B(a,\rho)} \left(g_a\frac{\partial g_b}{\partial r}\bigg|_{r=\rho} - g_b\frac{\partial g_a}{\partial r}\bigg|_{r=\rho} \right)\rho$$

$$= \int_{\partial B(a,\rho)} \left[(-\log(r) + w_a)\frac{\partial g_b}{\partial r}\bigg|_{r=\rho} - g_b\frac{\partial(-\log(r) + w_a)}{\partial r}\bigg|_{r=\rho} \right]\rho,$$

and (because w_a is harmonic) the only non-zero term in this is

$$\int g_b \frac{1}{\rho}\rho = 2\pi g_b(a).$$

However, we have to be careful in applying the same calculation to the other integral, namely $\int_{\partial B(b,\rho)}$, because in the expression

$$\int_{\partial B(a,\rho)} \left(g_a \frac{\partial g_b}{\partial r}\bigg|_{r=\rho} - g_b \frac{\partial g_a}{\partial r}\bigg|_{r=\rho} \right) \rho$$

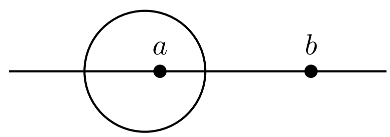
switching a and b gives a minus sign. Thus,

$$2\pi g_b(a) - 2\pi g_a(b) = 0,$$

and we are done. $\square$

Now we will compute the Green function for $B(0,1)$. We write $g_a(z) = -\log|z-a| + w$. We know that we want $g_a(z) = 0$ if $|z| = 1$, and choosing $a = 0$, we see that $g_0(z) = -\log(z)$ works. What happens for $a \neq 0$ though?

We apply Appollonius's theorem, which is that the set of points z such that $\frac{|z-a|}{|z-b|} =$ constant is a circle,



Homework. Prove Appollonius's theorem.

We choose a b such that $|a| \cdot |b| = 1$, and we see that the circle for which $\frac{|z-b|}{|z-a|} = \frac{1}{|a|}$ is precisely the circle $|z| = 1$.

We have

$$\log|z-b| - \log|z-a| = \log\left(\frac{1}{|a|}\right)$$

and

$$w = \log|z-b| + \log|a|,$$

so that

$$g_a(z) = \log\left(|a| \cdot \frac{|z-b|}{|z-a|}\right).$$

Let $z = (r\cos(\varphi), r\sin(\varphi))$, $a = (\rho\cos(\theta), \rho\sin(\theta))$, and $b = (\frac{1}{\rho}\cos(\theta), \frac{1}{\rho}\sin(\theta))$.

Then

$$\log|z-a| = \frac{1}{2}\log|z-a|^2 = \frac{1}{2}\log\left(r^2 + \rho^2 - 2r\rho\cos(\varphi-\theta)\right),$$

so that

$$\frac{\partial}{\partial r}\log|z-a| = \frac{1}{2} \cdot \frac{2r - 2\rho(\cos(\varphi-\theta))}{r^2+\rho^2-2r\rho\cos(\varphi-\theta)}$$

and

$$\frac{\partial}{\partial r}\log|z-b| = \frac{1}{2} \cdot \frac{2r - 2\frac{1}{\rho}(\cos(\varphi-\theta))}{r^2+\frac{1}{\rho^2}-2r\frac{1}{\rho}\cos(\varphi-\theta)}.$$

Therefore

$$-\frac{\partial}{\partial n}\log\left(\frac{|a||z-b|}{|z-a|}\right) = \frac{1-\rho^2}{1-2\rho\cos(\varphi-\theta)+\rho^2}.$$

Theorem. *If u is harmonic on $B(0,1)$, then*

$$u(a) = \frac{1}{2\pi}\int_0^{2\pi} u(\cos(\varphi), \sin(\varphi)) \cdot \frac{1-\rho^2}{1-2\rho\cos(\varphi-\theta)+\rho^2}$$

for every $\rho < 1$.

The function $P(\rho, \delta) = \frac{1-\rho^2}{1-2\rho\cos(\delta)+\rho^2}$ is very special, so it has a name: the Poisson kernel.

Corollary.

$$\frac{1}{2\pi}\int_0^{2\pi} P(\rho, \varphi-\theta) = 1.$$

Homework. Prove that $P(x, y) = P(\rho\cos(\delta), \rho\sin(\delta))$ is harmonic.

Lecture 15 (2012-11-20)

Recall that last time, we defined the Poisson kernel

$$P(\rho, \delta) = \frac{1-\rho^2}{1-2\rho\cos(\delta)+\rho^2}.$$

We know that P is harmonic.

Theorem. *If u is harmonic on $\{|z|<1\}$ and continuous on $\{|z|\le 1\}$, then*

$$u(a) = u(\rho\cos(\nu), \rho\sin(\nu)) = \frac{1}{2\pi}\int_0^{2\pi} u \cdot P(\rho, \varphi-\nu)\,d\varphi.$$

Theorem. *If $f = f(\cos(\varphi), \sin(\varphi))$ is continuous, then*

$$u(a) \overset{\text{def}}{=} \frac{1}{2\pi}\int_0^{2\pi} f \cdot P(\rho, \varphi-\nu)\,d\varphi$$

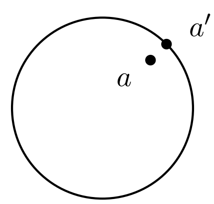
is harmonic on the open disc $\{|z|<1\}$. and

$$u^*(a) = \begin{cases} u(a) & \text{if } |a|<1, \\ f(a) & \text{if } |a|=1 \end{cases}$$

is continuous. Thus, the Poisson kernel preserves continuity.

Proof. Because P is harmonic, and finite sums of harmonic functions are harmonic, then in the limit, the integral defining $u(a)$ is harmonic.

Now, we want to show that $u^*(a)$ is continuous, i.e. that $u(a)$ is close to $f(a')$ if a is close to a':



This is because the Poisson kernel will weight values close to a' more than further points. Thus, we are taking the weighted average concentrated around a'. When a is close to a', we have that $\rho \approx 1$ and $\nu \approx \varphi$, so that the Poisson kernel is approximately

$$\frac{1-\rho^2}{2[1-\cos(\nu-\varphi)]]} \approx \frac{\text{small}}{\text{small}}.$$

□

Homework. Formally prove that $u^*(a)$ is continuous.

For the following material, you should read the proofs somewhere yourself.

Definition. Given $f \in L^1([0, 2\pi])$, the Fourier coefficients of f are defined to be

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-inx}, \qquad -\infty < n < +\infty.$$

Theorem.

1. *$f \in L^1$ implies that $c_n \to 0$ as $n \to \pm\infty$.*
2. *$f \in L^2$ implies that $\sum |c_n|^2 < \infty$. Moreover, $\sum |c_n|^2 = \|f\|_2$.*

Remark. In the measure space $(\mathbb{Z}, P(\mathbb{Z}), \#)$ (i.e., $\mathbb{Z}$ with the counting measure), then $c(n) \in L^2$ if and only if $\sum |c_n|^2 < \infty$. L^2 of this measure space is usually referred to as ℓ^2. The Fourier transform is a bijection between L^2 functions on $\mathbb{S}^1$ and L^2 functions on $\mathbb{Z}$.

Remark. We can define the partial sum functions

$$S_N = \sum_{n=-N}^{N} c_n e^{int}.$$

The sequence of S_N's is Cauchy in L^2, and therefore it converges in L^2.

Note that

$$\frac{1}{2\pi} \int_0^{2\pi} e^{inx} e^{-imx} \, dx = \begin{cases} 1 & \text{if } n = m, \\ 0 & \text{if } n \neq m. \end{cases}$$

In fact, defining an inner product on L^2 by $\langle f, g \rangle = \int f\overline{g}$, we have that the functions $\{\frac{e^{inx}}{\sqrt{2\pi}}\}$ form an orthonormal basis of L^2 (note that we have normalized them with a factor of $\frac{1}{\sqrt{2\pi}}$). We also see that

$$\langle f, f \rangle = \int |f|^2 = \|f\|_2.$$

Convolution

Integrating against the Poisson kernel is a special case of something called convolution.

Given $f : \mathbb{Z} \to \mathbb{C}$ such that $\sum_{n=-\infty}^{\infty} |f(n)| < \infty$, then

$$\widehat{f}(x) = \sum_{n=-\infty}^{\infty} f(n) e^{ixn},$$

($\widehat{\mathbb{Z}}$ is the unit circle, i.e. $\mathbb{R}/[0, 2\pi]$).

Now consider

$$\widehat{f}(x) \cdot \widehat{g}(x) = \bigg(\sum_{n=-\infty}^{\infty} f(n) e^{ixn} \bigg) \bigg(\sum_{m=-\infty}^{\infty} g(m) e^{imx} \bigg).$$

By Fubini's theorem, this is equal to

$$\sum_{n+m} f(n) g(m) e^{ix(n+m)} = \sum_k e^{ikx} \sum_{k=m+n} f(m) g(n) = (\widehat{f * g})(x),$$

where $f * g$ is the convolution of f and g.

Proposition. *If $f, g \in L^1(\mathbb{Z})$, then $(f * g) \in L^1(\mathbb{Z})$.*

Proof.

$$\begin{aligned}\sum_k |(f*g)(k)| &= \sum_k \Big| \sum_{k=m+n} f(n)g(m) \Big| \\ &\le \sum_k \sum_{k=m+n} |f(n)||g(m)| \\ &= \Big(\sum_n |f(n)| \Big) \Big(\sum_k |g(k-m)| \Big) = \|f\|_1 \|g\|_1 \qquad \square\end{aligned}$$

Now, instead of considering the functions $n \mapsto e^{inx}$, let's look at $t \mapsto e^{ixt}$, where $x, t \in \mathbb{R}^d$ and xt means the dot product of x and t.

Definition. For $f, g \in L^1(\mathbb{R}^d)$, define their convolution to be

$$(f*g)(x) = \int_{\mathbb{R}^d} f(y)g(x-y)\,dy.$$

Theorem. *For any $f, g \in L^1(\mathbb{R}^d)$,*

1. *$f * g$ exists at almost every x.*
2. *$f * g \in L^1$ and $\|f*g\|_1 \le \|f\|_1 \|g\|_1$.*
3. *$*$ is commutative and associative.*

Proof. Observe that

$$\begin{aligned}\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(y)||g(x-y)|\,dy\,dx &= \int_{\mathbb{R}^d} |f(y)| \Big(\int_{\mathbb{R}^d} |g(x-y)|\,dx \Big) dy \\ &= \int |f(y)| \cdot \|g\|_1\,dy = \|f\|_1 \|g\|_1,\end{aligned}$$

which proves claim 2. This also implies that

$$\int |f(y)||g(x-y)|\,dy < \infty$$

at almost every x, so indeed $f * g$ exists a.e., which is claim 1.

Claim 3 is clear; for example, swap f and g in the sum for commutativity. $\square$

Homework. If $f \in L^1$ and $g \in L^p$, does it follow that $f * g \in L^p$ and $\|f*g\|_p \le \|f\|_1 \|g\|_p$?

Homework. If $\frac{1}{p} + \frac{1}{q} = 1$ and $f \in L^p$, $g \in L^q$, show that $f * g$ is continuous and tends to 0 as $|x| \to \infty$.

Homework. Prove that there is no "unit" function, i.e. that there is no $f \in L^1$ such that $f * g = g$ for every $g \in L^1$.

Lemma. *Let $f \in L^1$ and $y \in \mathbb{R}^d$. Define $f_y(x) = f(x+y)$. Then*

$$\lim_{y \to 0} \|f_y - f\|_1 = 0.$$

Proof. If f is continuous and compactly supported, then this is trivial; if we integrate a small ϵ on a bounded set, we'll get something small.

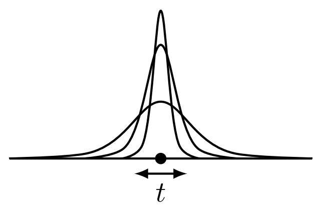
Now, take an arbitrary $f \in L^1$, and choose a g which is continuous and compactly supported such that $\|f-g\|_1 < \epsilon$. Then

$$f_y - f = (f_y - g_y) + (g_y - g) + (g - f),$$

Note that $f_y - g_y = (f-g)_y$, so that

$$\|f_y - f\|_1 \le \|f_y - g_y\|_1 + \|g_y - g\|_1 + \|g - f\|_1 < 3\epsilon. \qquad \square$$

For $t \in \mathbb{R}$, let f_t (**not** translation) be a non-negative function such that $\int f_t = 1$ and $f_t(x) = 0$ for $|x| > t$. These functions are "approximate units".



We can think of $f_t * g$ as a weighted average of g around 0.

Theorem. *For any $g \in L^1$, we have $\|f_t * g - g\|_1 \to 0$ as $t \to 0$.*

Proof. First, note that

$$\begin{aligned}
\|f_t * g - g\|_1 &= \int \left| \int f_t(y) g(x-y)\, dy - g(x) \right| dx \\
&= \int \left| \int f_t(y)(g(x-y) - g(x))\, dy \right| dx \\
&\le \int_{-t}^{t} \int_{\mathbb{R}^d} |g(x-y) - g(x)|\, dx\, f_t(y)\, dy
\end{aligned}$$

and, by the lemma, once t is sufficiently small we have that

$$\le \epsilon \int_{|y|<t} f_t(y)\, dy = \epsilon. \qquad \square$$

Example. We could take

$$f_t(y) = \frac{\chi_{B(0,t)}(y)}{\lambda(B(0,t))},$$

in which case

$$(f_t * g)(x) = \int g(x-y) f_t(y)\, dy \overset{\text{commutativity}}{=} \frac{1}{\lambda(B(0,t))} \int_{\mathbb{R}^d} g(y) \chi_{B(0,t)}(x-y)\, dy = \frac{1}{\lambda(B(0,t))} \int_{B(0,t)} g,$$

which goes to $g(x)$ a.e.

Example. Let $\frac{1}{p} + \frac{1}{q} = 1$. We could take $f \in L^p$, and map $g \in L^q$ to $\int f\overline{g}$. We know that

$$\|f\|_p = \|\Lambda_f\| = \sup_{\substack{g \in L^q \\ \|g\|_q = 1}} |\Lambda_f(g)|.$$

Theorem. *Let $\frac{1}{p} + \frac{1}{q} = 1$, with $p < \infty$. For any $f \in L^p$ and $g \in L^q$,*

$$(f * g)(x) = \int f(x-y)g(y)\,dy$$

exists for every x, and in fact is continuous. Moreover, $\|f * g\|_\infty \le \|f\|_p \|g\|_q$.

Theorem. *Let f_t be approximate units, and let $g \in L^p$ with $1 < p < \infty$. Then $\|f_t * g - g\|_p \to 0$ as $t \to 0$.*

Proof. It will suffice to prove that for any $h \in L^q$,

$$\int_{\mathbb{R}^d} (f_t * g - g)h \le (\text{small constant})\|h\|_q,$$

because L^p is the dual of L^q and this would show that the operator norm of the integrate-against-$(f_t * g - g)$ operator on L^q (which is equal to the L^p norm of $f_t * g - g$) is small. We know that

$$g(x) = \int f_t(y)g(x)\,dy,$$

so that

$$\int\int (g(x-y) - g(x))f_t(y)\,h(x)\,dy\,dx \le \int_{|y|<t} \|g_y - g\|_p \|h\|_q\,dy.$$

Therefore, it will suffice to show that

$$\int_{|y|<t} \|g_y - g\|_p \to 0$$

as $t \to 0$; but this is clear. $\square$

Lecture 16 (2012-11-27)

Today we'll be starting probability. We'll mainly discuss some terminology and notations.

Definition. We say that $(\Omega, \mathcal{F}, P)$ is a probability space when it is measure space (Ω being the underlying set, $\mathcal{F}$ the σ-algebra, and P the measure) such that $P(\Omega) = 1$. We say that Ω is the set of outcomes, $\mathcal{F}$ is the σ-algebra of events, and P is the probability measure.

Examples.

- We can make a probability space representing flipping a coin n times. We let $\Omega = \{H, T\}^n$, $\mathcal{F} =$ all subsets of Ω, and $P(\omega) = \frac{1}{2^n}$ for each $\omega \in \Omega$.
- Now suppose we are flipping a coin infinitely many times. Then Ω consists of all infinite sequences of H and T, and $\mathcal{F}$ is the σ-algebra generated by cylinder sets, e.g.

$$\{(x_1, x_2, \ldots) \in \Omega \mid x_1 = H, x_2 = T, \ldots, x_n = H\}.$$

  Finally, we define P to be the product measure $(\frac{1}{2}\delta_H + \frac{1}{2}\delta_T)^{\mathbb{N}}$, or in other words the measure such that $P(C) = \frac{1}{2^n}$ when C is a cylinder set in which the first n coordinates have been fixed.

Definition. A random variable X is a function $X : \Omega \to (-\infty, \infty)$ such that $X^{-1}(B) \in \mathcal{F}$ for every Borel $B \subseteq \mathbb{R}$. We define

$$\mu_X(B) \stackrel{\text{def}}{=} P(X \in B) = P(X^{-1}(B)),$$

which specifies a measure on $\mathbb{R}$. We say that this is the distribution of X. The distribution function of X is the function defined by $F(x) = \mu_X((-\infty, x]) = P(X \leq x)$.

Here are some basic properties of the distribution function.

- $\lim\limits_{x\to\infty} F(x) = 1$.
- $\lim\limits_{x\to-\infty} F(x) = 0$.
- F is monotone increasing.
- F is right continuous, but not necessarily left continuous:

$$\lim_{\epsilon\to 0^+} F(x+\epsilon) = F(x) \neq \lim_{\epsilon\to 0^-} F(x+\epsilon).$$

It turns out these properties actually characterize the functions F which are distribution functions. We can see this by defining

$$\mu_X((-\infty, x]) \stackrel{\text{def}}{=} F(x),$$

then extending μ_X to all Borel sets.

Definition. If there is a function $f : \mathbb{R} \to [0, \infty)$ such that $P(a \leq x \leq b) = \int_a^b f$, we say that f is the density of X. If it exists, we then have that $\int_{-\infty}^{\infty} = 1$, and if f is continuous at x then $f(x) = F'(x)$.

Definition. The characteristic function of an event $E \in \mathcal{F}$, i.e. the function

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

is called an indicator function when we are doing probability.

Examples.

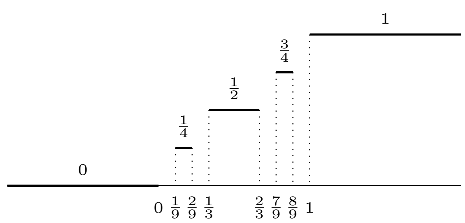
- Returning to flipping coins, we let $(\Omega, \mathcal{F}, P)$ represent flipping a coin infinitely many times. Define
$$X_n(\omega_1, \omega_2, \ldots) = \omega_n = \begin{cases} 1 & \text{if } n\text{th flip is head,} \\ 0 & \text{if } n\text{th flip is tail,} \end{cases}$$
and let $S_n = X_1 + \cdots + X_n$. Then S_n is the number of heads in the first n flips. Define $\mathcal{F}_n$ to be the σ-algebra of events that depend only on the first n flips. Then S_n is also a random variable on $(\Omega, \mathcal{F}_n, P)$, but S_{n+1} is not.

- Let μ be a probability measure on $(\mathbb{R}, \text{Borel})$. Consider the random variable $X : \mathbb{R} \to \mathbb{R}$ which is the identity function on $\mathbb{R}$. Then $\mu_X = \mu$.

- We say that X has normal distribution with mean μ and variance σ^2 if X has density
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma}.$$
In particular, the standard normal distribution $(\mu = 0, \sigma = 1)$ has density $f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$, so that
$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt.$$

- If X is a random variable and $g : \mathbb{R} \to \mathbb{R}$ is a Borel-measurable function, then $Y = g(X)$ is also a random variable.

- The Cantor function



is a continuous function from $\mathbb{R}$ to $[0, 1]$, and it is a distribution of some random variable (it meets all of the criteria we set above). It has no atoms (i.e., no points of positive measure), and no density function. Its distribution is also not absolutely continuous with respect to the Lebesgue measure, because the Cantor function maps the Cantor set (which has Lebesgue measure 0) onto $[0, 1]$ (which has positive Lebesgue measure).

Definition. The expected value $E(X)$ of a random variable X is defined to be
$$E(X) = \int X\, dP = \int x\, d\mu_X.$$
However, $E(X)$ may not exist, because the function could approach both positive and negative infinity.

It is easy to see that for any Borel-measurable g, we have
$$E(g(X)) = \int g(x)\, d\mu_X,$$

and that if X has density function f, we have

$$E(g(X)) = \int f(x)g(x)\,dx.$$

Definition. The variance $\mathrm{Var}(X)$ of a random variable X is defined to be

$$\begin{aligned}\mathrm{Var}(X) &= E((X - E(X))^2)\\ &= E(X^2 - 2X\cdot E(X) + E(X)^2)\\ &= E(X^2) - 2E(X)E(X) + E(X)^2\\ &= E(X^2) - E(X)^2.\end{aligned}$$

The variance is often denoted by σ^2, so that $\sigma = \sqrt{\mathrm{Var}(X)}$ (you can see that $\mathrm{Var}(X)$ is non-negative via Cauchy-Schwarz, Jensen's inequality, or simply by noting that $\mathrm{Var}(X)$ is the expected value of something non-negative).

Theorem (Markov's inequality)**.** *For any random variable X, we have*

$$P(|X| \geq c) \leq \frac{E(|X|)}{c}.$$

Proof. Define a new random variable X_c by $X_c = c\chi_{|X|\geq c}$. It is easy to see that $X_c \leq |X|$. Now note that

$$c\cdot P(|X| \geq c) = c\cdot P(X_c = c) = E(X_c) \leq E(|X|).$$ □

Theorem (Chebyshev's inequality)**.** *For any random variable X, we have*

$$P(|X - E(X)| \geq c) \leq \frac{\mathrm{Var}(X)}{c^2}.$$

Proof. Applying Markov, we immediately get that

$$P(|X - E(X)|^2 \geq c^2) \leq \frac{\mathrm{Var}(X)}{c^2}.$$ □

Homework. Show that for any Borel-measurable, non-decreasing $f : [0,\infty) \to [0,\infty)$ and any non-negative random variable X, we have, for all c,

$$P(X \geq c) \leq \frac{E(f(X))}{f(c)}.$$

Definition. We say that events $A, B \in \mathcal{F}$ are independent if $P(A\cap B) = P(A)\cdot P(B)$. More generally, the collection of events $\{A_\alpha\}_{\alpha\in I}$ is independent if

$$P(A_{\alpha_1} \cap\cdots\cap A_{\alpha_n}) = P(A_{\alpha_1})\cdots P(A_{\alpha_n})$$

for all finite subsets $\{\alpha_1,\ldots,\alpha_n\} \in I$. Note that this is **not** the same as the A_α's being pairwise independent.

Example. Suppose we are rolling a die twice. Let

$$\begin{aligned}A_1 &= \{\text{sum of the rolls is } 7\},\\ A_2 &= \{\text{first roll is } 1\},\end{aligned}$$

$$A_3 = \{\text{second roll is } 6\}.$$

It is easy to see that

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{6}.$$

We have

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{36},$$

so the events A_1, A_2, A_3 are pairwise independent, but

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{36} \neq \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6},$$

so the collection is not independent.

Definition. If $\mathcal{F}_\alpha$ for $\alpha \in I$ are σ-algebras, we say that the $\mathcal{F}_\alpha$'s are independent if

$$P(A_{\alpha_1} \cap \cdots \cap A_{\alpha_n}) = P(A_{\alpha_1}) \cdots P(A_{\alpha_n})$$

for all $A_{\alpha_1} \in \mathcal{F}_{\alpha_1}, \ldots, A_{\alpha_n} \in \mathcal{F}_{\alpha_n}$.

To any random variable X, there is a corresponding σ-algebra $\mathcal{F}_X = \{X^{-1}(B) \mid \text{Borel } B \subseteq \mathbb{R}\}$, which is the smallest σ-algebra on which X is measurable.

Lecture 17 (2012-11-29)

Independence of Random Variables

Let $X_1, \ldots, X_n$ be random variables.

Definition. Their joint distribution is the function

$$F(t_1, \ldots, t_n) = P(X_1 \leq t_1, \ldots, X_n \leq t_n).$$

Their joint density, if it exists, is the function f such that

$$\int_B f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = P((X_1, \ldots, X_n) \in B)$$

for all Borel $B \subseteq \mathbb{R}^n$. We define

$$\mu(B) \stackrel{\text{def}}{=} P((X_1, \ldots, X_n) \in B).$$

The following statements are equivalent:

- $X_1, \ldots, X_n$ are independent
- The σ-algebras $\mathcal{T}_1, \ldots, \mathcal{T}_n$ are independent
- $\mu = \mu_1 \times \cdots \times \mu_n$
- If the densities exist, $f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n)$
- $P(X_1 \in B_1, \ldots, X_n \in B_n) = P(X_1 \in B_1) \cdots P(X_n \in B_n)$.

For any X and Y, we know that $E(X + Y) = E(X) + E(Y)$ because the integral is additive. If X and Y are independent, it is also true that $E(XY) = E(X)E(Y)$. We can see this by looking at discrete events (simple functions): if $X = c_j$ with probability p_j (we call this event A_j) and $Y = d_m$ with probability q_m (we call this event B_m), then

$$E(X) = \sum p_j c_j, \qquad E(Y) = \sum q_m d_m,$$

and because $P(A_j \text{ and } B_m) = p_j q_m$, we have

$$E(XY) = \sum_{j,m} p_j q_m c_j d_m.$$

We can then pass to arbitrary measurable functions in the standard way (take monotone limits of non-negative simple functions to get arbitrary non-negative measurable functions, then consider differences of non-negative measurable functions).

Homework. We've shown that independent random variables X and Y are orthogonal, i.e. they satisfy $E(XY) = E(X)E(Y)$. Is the converse true?

Proposition. *If $X_1, \ldots, X_n$ are pairwise orthogonal, then*

$$\operatorname{Var}(X_1 + \cdots + X_n) = \operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n).$$

Proof. Because $\mathrm{Var}(X) = \mathrm{Var}(X + c)$ for a constant c, we can assume that $E(X_j) = 0$ for all j. Then

$$\begin{aligned}\mathrm{Var}(X_1 + \cdots + X_n) &= E\big[(X_1 + \cdots + X_n)^2\big] \\ &= E(\textstyle\sum X_j^2 + 2(\sum X_i X_j)) \\ &= \textstyle\sum E(X_i^2) + 2\sum E(X_i X_j) \\ &= \textstyle\sum \mathrm{Var}(X_i) + 2\underbrace{E(X_i)}_{=0}\underbrace{E(X_j)}_{=0} \\ &= \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)\end{aligned}$$

□

There are extreme cases where Var is not additive:

$$\mathrm{Var}(X - X) = 0, \qquad \mathrm{Var}(X + X) = 4\mathrm{Var}(X).$$

Definition. When we are doing probability, we refer to convergence in measure as convergence in probability. Thus, $X_n \to X$ in probability when for all $\epsilon > 0$, $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$.

Definition. When we are doing probability, we refer to convergence a.e. as almost sure convergence. Thus, $X_n \to X$ almost surely when there exists an event A with $P(A) = 1$ such that $X_n(\omega) \to X(\omega)$ for all $\omega \in A$.

Homework. Suppose that $X_1, X_2, \ldots$ have the property that $E(X_n) \to \mu$ and $\mathrm{Var}(X_n) \to 0$. Show that $X_n \to \mu$ in probability, but not necessarily almost surely.

Definition. We say $X_1, X_2, \ldots$ are independent indentically distributed (i.i.d.) random variables when they are independent and have the same distribution. Intuitively, this means they are different occurrences of the same variable, e.g. X_1 = flipping a coin, X_2 = flipping the coin again, etc. Denoting $E(X_j) = \mu$ and $\mathrm{Var}(X_j) = \sigma^2$ for all j, we have that

$$E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \mu, \qquad \mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}.$$

Theorem (Weak law of large numbers)**.** *If $X_1, X_2, \ldots$ are i.i.d., then*

$$\frac{X_1 + \cdots + X_n}{n} \to \mu$$

in probability.

Theorem (Strong law of large numbers)**.** *If $X_1, X_2, \ldots$ are i.i.d., then*

$$\frac{X_1 + \cdots + X_n}{n} \to \mu$$

almost surely.

Recall that for a sequence of sets $A_1, A_2, \ldots$, we define $\limsup_{j\to\infty} A_j = \bigcap_{m=1}^{\infty} \bigcup_{j=m}^{\infty} A_j$.

Theorem (Borel-Cantelli lemma)**.**

1. *If $\sum P(A_j) < \infty$, then $P(\limsup_{j\to\infty} A_j) = 0$.*
2. *If $\sum P(A_j) = \infty$ and the A_j are independent, then $P(\limsup_{j\to\infty} A_j) = 1$.*

Proof of 1. This is clear because the tail of a convergent sum goes to 0. ☐

Proof of 2. For any m,

$$P(A_m^c \cap A_{m+1}^c \cap \cdots) \overset{\text{independence}}{=} \prod_{n=m}^{\infty} (1 - P(A_n)) = 0$$

because $\sum_{n=m}^{\infty} P(A_n) = \infty$. Then $A_m \cup A_{m+1} \cup \cdots$ happens almost surely, i.e. $P(\bigcup_{n=m}^{\infty} A_n) = 1$, so that

$$P\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n\right) = 1.$$

☐

Theorem (Kolmogorov 0-1 Law)**.** *Let $X_1, X_2, \ldots$ be random variables. Define the σ-algebras*

$$\begin{aligned} \mathcal{F}_1 &= \sigma(X_1) & \qquad \mathcal{G}_1 &= \sigma(X_2, X_3, \ldots) \\ \mathcal{F}_2 &= \sigma(X_1, X_2) & \qquad \mathcal{G}_2 &= \sigma(X_3, X_4, \ldots) \\ &\cdots & &\cdots \end{aligned}$$

Then $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots$ and $\mathcal{G}_1 \supseteq \mathcal{G}_2 \supseteq \cdots$, but the limit $\mathcal{F}_0 = \bigcup_{n=1}^{\infty} \mathcal{F}_n$ is only an algebra, not necessarily a σ-algebra. However, the limit $\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{G}_n$ is a σ-algebra.

If A is measurable with respect to $\mathcal{T}$, then either $P(A) = 0$ or $P(A) = 1$.

Lemma. *Suppose that $\mathcal{F}^0$ is an algebra, and $\mathcal{F}$ is the σ-algebra generated by $\mathcal{F}^0$. Then for every $A \in \mathcal{F}$ and every $\epsilon > 0$, there is some $B \in \mathcal{F}^0$ such that $P(A \triangle B) < \epsilon$.*

Proof of lemma. It will suffice to show that the collection $\mathcal{G}$ of sets A which have this property is a σ-algebra. Trivially, we have that $\mathcal{F}^0 \subseteq \mathcal{G}$. If $A \in \mathcal{G}$, then $A^c \in \mathcal{G}$ because $P(A \triangle B) < \epsilon$ implies that $P(A^c \triangle B^c) < \epsilon$, and $B^c \in \mathcal{F}^0$ because $B \in \mathcal{F}^0$.

Now we need to show that if $A_1, A_2, \ldots \in \mathcal{G}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{G}$.

First approach (from class): Choose $B_j \in \mathcal{F}^0$ such that $P(A_j \triangle B_j) < \frac{\epsilon}{2^j}$. We know that

$$P\left(\bigcup_{j=1}^{N} A_j \triangle \bigcup_{j=1}^{\infty} B_j\right) < \epsilon$$

if N is large enough, and

$$P\left(\bigcup_{j=1}^{N} B_j \triangle \bigcup_{j=1}^{\infty} B_j\right) < \epsilon,$$

so

$$P\left(\bigcup_{j=1}^{N} A_j \triangle \bigcup_{j=1}^{N} B_j\right) < 2\epsilon,$$

and hence

$$P\left(\bigcup_{j=1}^{\infty} A_j \triangle \bigcup_{j=1}^{N} B_j\right) < 4\epsilon \qquad (?)$$

proving the lemma.

Second approach (not from class): Let $C = \bigcup_{n=1}^{\infty} A_n \in \mathcal{G}$. Let $\epsilon > 0$, and choose an m such that

$$P\left(\bigcup_{j=1}^{m} A_j\right) \geq P(C) - \frac{\epsilon}{2}.$$

For $j = 1, \ldots, m$, choose $B_j \in \mathcal{F}^0$ such that $P(A_j \vartriangle B_j) \leq \epsilon/2^{j+1}$. Let $B = \bigcup_{j=1}^{m} B_j$, and note that

$$C \vartriangle B \subseteq \left(\bigcup_{j=1}^{m} A_j \vartriangle B_j\right) \cup \left(C \setminus \bigcup_{j=1}^{m} A_j\right),$$

so that $P(C \vartriangle B) < \epsilon$, and hence $C = \bigcup_{j=1}^{m} A_j \in \mathcal{G}$. $\square$

Proof of Kolmogorov. We want to show that if $A \in \mathcal{T}$, then $P(A) = 0$ or $P(A) = 1$.

For any $\epsilon > 0$, choose a $B_\epsilon \in \bigcup_{j=1}^{\infty} \mathcal{F}_j$ (remember, this is only an algebra) such that $P(A \vartriangle B_\epsilon) < \epsilon$. Then $B_\epsilon \in \mathcal{F}_n$ for some n; recall that $A \in \mathcal{G}_n$ for every n, and $\mathcal{F}_n$ and $\mathcal{G}_n$ are independent. Therefore, $P(A \cap B_\epsilon) = P(A)P(B_\epsilon)$. Also note that

$$P(A \cap B_\epsilon) \geq P(A) - P(A \vartriangle B_\epsilon) \geq P(A) - \epsilon.$$

Thus, as $\epsilon \to 0$,

$$\begin{array}{ccc} P(A \cap B_\epsilon) & = & P(A)P(B_\epsilon) \\ \downarrow & & \downarrow \\ P(A) & = & P(A)P(A) \end{array}$$

hence $P(A) = P(A)^2$, hence $P(A) = 0$ or $P(A) = 1$. $\square$

Definition. We define the Fourier transform of a function g to be

$$\widehat{g}(z) = \int_{-\infty}^{\infty} e^{-ixz}\, g(x)\, dx.$$

The inverse Fourier transform is

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixz}\, \widehat{g}(x)\, dx.$$

Definition. We say that g is a Schwartz function when g is C^∞, and all of the derivatives $g^{(j)}(x)$ tend to 0 as $x \to \pm\infty$ faster than any polynomial.

It turns out that if g is Schwartz, then $\widehat{g}$ is Schwartz.

Lecture 18 (2012-12-04)

Today, we'll prove the Central Limit Theorem.

Let g be a Schwartz function, and let $\widehat{g}$ be its Fourier transform, i.e.

$$\widehat{g}(y) = \int_{-\infty}^{\infty} e^{-ixy}\, g(x)\, dx.$$

Recall that the inverse Fourier transform can be obtained as

$$g(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixy}\, \widehat{g}(x)\, dx.$$

The function $e^{-x^2/2}$ acts (almost) as an identity:

$$\widehat{e^{-x^2/2}} = \sqrt{2\pi}\, e^{-y^2/2}.$$

We will also need the fact that

$$\int f\widehat{g} = \int f(x) \int e^{-ixy}\, g(y)\, dy\, dx = \int g(y) \underbrace{\int e^{-ixy} f(x)\, dx}_{\widehat{f}}\, dy = \int g\widehat{f}.$$

Definition. The characteristic function φ of X is defined by $\varphi(t) = E(e^{iXt})$. In general, this is a complex-valued function.

Observe that $\varphi(0) = 1$, and that for any t, we have $|\varphi(t)| \leq 1$. Also, φ is continuous because the dominated convergence theorem implies that

$$\lim_{s \to t} \varphi(s) = \lim_{s \to t} E(e^{iXs}) = \varphi(t).$$

If X has density f, then

$$\varphi(t) = \int e^{ixt}\, f(x)\, dx = \widehat{f}(-t).$$

It turns out that if X and Y have the same characteristic function, then they have the same distribution. To prove this, we need some sort of "inversion" (when X and Y have densities, we could apply actual Fourier inversion, but we need something that works in general). If $X_1, \ldots, X_n$ are independent random variables, then

$$\varphi_{X_1+\cdots+X_n}(t) = E(e^{i(X_1+\cdots+X_n)t}) = E(e^{iX_1 t} \cdots e^{iX_n t}) \underset{\substack{\uparrow \\ \text{independence}}}{=} \varphi_{X_1}(t) \cdots \varphi_{X_n}(t).$$

If X has a normal distribution with mean 0 and variance 1, then

$$\varphi(t) = \int e^{ixt} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx = e^{-t^2/2}$$

because $e^{ixt}e^{-x^2/2} = e^{-(x-t)^2/2}e^{-t^2/2}$.

For a random variable X and any a, b,

$$\varphi_{aX+b} = e^{ibt}\varphi_X(at),$$

so the characteristic function of a normal distribution with mean μ and variance σ^2 is $e^{i\mu t}e^{-\sigma^2 t^2/2}$.

If $E(|X|) < \infty$, then $\varphi'(0) = iE(X)$. More generally, if the higher moments of X are finite, i.e. if we have $E(|X|^k) < \infty$ for some k, then $\varphi^{(j)}(0) = i^j E(X^j)$ for all $1 \leq j \leq k$.

Lemma. *Let $X_1, \ldots, X_n$ be i.i.d. random variables, with mean μ and variance σ^2. Let*

$$Y_n = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{\sigma^2 n}}.$$

Then $\varphi_{Y_n}(t) \to e^{-t^2/2}$ (the normal distribution) as $n \to \infty$.

Proof. We can write

$$\varphi(t) = 1 + \varphi'(0)t + \frac{1}{2}\varphi''(0)t^2 + \epsilon_t t^2,$$

where $\epsilon_t \to 0$ as $t \to 0$. We can assume WLOG that $\mu = 0$ and $\sigma = 1$, so we get

$$\varphi(t) = 1 - \frac{t^2}{2} + \epsilon_t t^2.$$

We have that

$$\varphi_{Y_n}(t) = \left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n,$$

so

$$\lim_{n\to\infty} \varphi_{Y_n}(t) = \lim_{n\to\infty} \left(1 - \frac{t^2}{2n} + \epsilon_{t/\sqrt{n}}\frac{t^2}{n}\right)^n = \left(1 - \frac{t^2}{2n} + \frac{\delta_n}{n}\right)^n,$$

where $\delta_n \to 0$ as $n \to \infty$ (because $\epsilon_{t/\sqrt{n}} \to 0$ as $n \to \infty$). Taking logarithms,

$$\lim_{n\to\infty} \log(\varphi_{Y_n}(t)) = \lim_{n\to\infty} n \underbrace{\log\left(1 - \frac{t^2}{2n} + \frac{\delta_n}{n}\right)}_{-\frac{t^2}{2n} + \frac{\rho_n}{n}}$$

where $\rho_n \to 0$ as $n \to \infty$, so

$$\lim_{n\to\infty} \log(\varphi_{Y_n}(t)) = -\frac{t^2}{2},$$

and hence $\varphi_{Y_n}(t) = e^{-t^2/2}$. $\square$

Theorem (Central limit theorem)**.** *Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2. Then*

$$\lim_{n\to\infty} P\left(a \le \frac{X_1 + \cdots + X_n - \mu n}{\sigma\sqrt{n}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx.$$

Proof. We need to prove that if μ_n is a sequence of distributions with $\varphi_n(t) \to e^{-t^2/2}$, then $\lim_{n\to\infty} \mu_n([a,b]) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx$.

Our approach will be to approximate $\chi_{[a,b]}$ with Schwartz functions. We can find a Schwartz function $g = g_\epsilon$ with $0 \le g \le 1$, $g(x) = 1$ on $[a,b]$, and $g(x) = 0$ on $(-\infty, a-\epsilon) \cup (b+\epsilon, \infty)$. We claim that

$$\lim_{n\to\infty} \int g(x)\, d\mu_n(x) = \frac{1}{\sqrt{2\pi}} \int g(x) e^{-x^2/2}.$$

Note that, if this is true, then as $\epsilon \to 0$, we get

$$\mu_n([a,b]) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx$$

as desired. Now, we need to prove this claim.

Observe that

$$\int_{-\infty}^{\infty} g(x)\,d\mu_n(x) = \int \frac{1}{2\pi} \int e^{-xy}\,\widehat{g}(y)\,dy\,d\mu_n(x)$$

where $\widehat{g}(y) = \int e^{-ixy}\,g(x)\,dx$, so

$$\int_{-\infty}^{\infty} g(x)\,d\mu_n(x) = \frac{1}{2\pi} \int \left(\int e^{ixy}\,d\mu_n(x) \right) \widehat{g}(y)\,dy = \frac{1}{2\pi} \int \varphi_n(y)\widehat{g}(y)\,dy.$$

Taking the limit as $n \to \infty$,

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g(x)\,d\mu_n(x) = \frac{1}{2\pi} \int e^{-y^2/2}\,\widehat{g}(y)\,dy.$$

Now use that $\int f\widehat{g} = \int \widehat{f} g$ and $\widehat{e^{-y^2/2}} = \sqrt{2\pi} e^{-y^2/2}$. □

Let X be a random variable with characteristic function φ and distribution μ. We want to express φ in terms of μ somehow. We claim that

$$\mu([a,b]) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-iya} - e^{-iyb}}{iy}\,\varphi(y)\,dy,$$

as long as a, b are points where F (the distribution function) is continuous, or equivalently, as long as a, b are not atoms for the measure μ. The integrand in the above expression can tend to both $-\infty$ and ∞, so the integral over all of $\mathbb{R}$ may not exist, but this limit will exist because there will be cancellations. The denominator of iy doesn't lead to infinite values, because

$$\left| \varphi(y) \cdot \frac{e^{-iya} - e^{-iyb}}{iy} \right| \leq \left| \frac{e^{-iya} - e^{-iyb}}{iy} \right| \leq b - a.$$

Proof. Note that

$$\begin{aligned}
\frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-iya} - e^{-iyb}}{iy}\,\varphi(y)\,dy &= \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-iya} - e^{-iyb}}{iy} \int_{-\infty}^{\infty} e^{ixy}\,d\mu(x)\,dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-T}^{T} \frac{e^{iy(x-a)} - e^{iy(x-b)}}{iy}\,dy\,d\mu(x)
\end{aligned}$$

For any $c > 0$,

$$\int_{-T}^{T} \frac{e^{icx}}{ix}\,dx = 2 \int_{0}^{cT} \frac{\sin(x)}{x}\,dx$$

(we can see this by splitting the exponential into sine and cosine). Also, recall that

$$\lim_{T\to\infty} \int_{0}^{T} \frac{\sin(x)}{x} = \frac{\pi}{2}.$$

Thus, if $x - a$ and $x - b$ have the same sign, then

$$\lim_{T\to\infty} \int_{-T}^{T} \frac{e^{iy(x-a)} - e^{iy(x-b)}}{iy}\,dy = 0,$$

and if $x - a > 0$ and $x - b < 0$, we get that it equals $4 \cdot \frac{\pi}{2} = 2\pi$. □