Spaces of measurable functions

We saw in Exercise REF that the set of strongly measurable functions and the set of Borel measurable functions are both vector spaces. We also saw that the $L^p_{\mathbb{X}}(\mu)$-norms define semi-normed spaces of strongly measurable functions (Exercise REF) and that the set of Bochner integrable functions is exactly the semi-normed space $\mathbb{L}^1_{\mathbb{X}}(\mu)$ (Theorem REF). In this chapter, we explore $\mathbb{L}^p_{\mathbb{X}}$-spaces more deeply and introduce another important notion of convergence for measurable functions.

1 $\mathbb{L}^{|}_{\mathbb{X}}$-spaces

Let us review Section REF. Given a measure space (Ω, Σ, μ) and a Banach space $\mathbb{X}$, the set of all strongly measurable functions mapping from Ω to $\mathbb{X}$ (with respect to $\bar{\Sigma}$ and the Borel σ-algebra of $\mathbb{X}$) comprises a vector space $\mathbb{M}_{\mathbb{X}}(\Omega, \bar{\Sigma})$. For $p \in [1, \infty)$, we defined the $L^p_{\mathbb{X}}(\Omega, \bar{\Sigma}, \mu)$-norm on these functions by

$$\|f\|_p := (\mathcal{M}\|f\|^p)^{1/p}.$$

(SIDENOTE: Based on the limiting behavior of this norm, the $p = \infty$ case is defined to be the essential supremum of $\|f\|$.) $L^p_{\mathbb{X}}(\Omega, \bar{\Sigma}, \mu)$ is the Banach space (Theorem REF) of equivalence classes of strongly measurable functions with finite $L^p_{\mathbb{X}}$-norm, and $\mathbb{L}^p_{\mathbb{X}}(\Omega, \bar{\Sigma}, \mu)$ is the semi-normed space of the functions themselves. When the subscript is omitted, the codomain is implied to be $\mathbb{R}$.

As we point out below, it turns out that almost all of the examples of normed spaces that we've seen so far are special cases of $L^p_{\mathbb{X}}$ spaces for some measure space. Thus establishing general properties of $L^p_{\mathbb{X}}$ spaces gets us a lot of mileage.

1.1 Examples

IS THIS A GOOD place to put the discussion of sequences as functions?

Revisit examples from metric space section? IT TURNS out that almost all important normed spaces seem to be special cases of L^p - DON'T BOTHER proving anything redundant about those spaces! e.g. in the sections introducing them, don't even bother showing that they satisfy norm properties - all we need is for $(\mathbb{R}, |\cdot|)$ to be a normed space - the rest follows from establishing that L^p spaces are normed spaces! (except some general results, e.g. supremum norm is norm, are useful beyond just the L^p cases - in fact, that is (essentially) the L^∞ case so I can just refer to it in part, when claiming that L^p spaces are normed spaces.

ALSO some non-examples maybe?

The space $C[a, b]$ of continuous functions on the real interval $[a, b]$ with norm

$$\|x\| := \max_{t \in [a,b]} |x(t)|$$

(SIDENOTE: REFERENCE the relevant theorem about continuous mappings from compact domains to remind the reader why the max exists. OR DID I just do that in the metric spaces section?)

I think $C[a, b]$ turns out to be a linear subspace of $L^\infty(\mu)$ where μ is Lebesgue measure on $[a, b]$ and zero elsewhere. (SIDENOTE: Any measure with positive density on $[a, b]$ and zero elsewhere results in this same L^∞ space.)

Are all supremum metrics also norms?

DISTINGUISH the isomorphism between vector spaces (and thus normed spaces) from the isomorphism between Hilbert spaces. All Hilbert spaces of the same cardinality have an isomorphic mapping that preserves all inner product values (and thus norms). On the other hand, all vector spaces with the same cardinality are isomorphic (right?) but different norms on those vector spaces don't necessarily have an *isometric* isomorphism. (right?)

1.2 Properties

Holder's inequality - does it still hold in this generality? YES i think so - two possible routes to prove it - Jensen's inequality method or Young's inequality method. START with the general statement for the product of a sequence of functions - then point out that the case of two factors and conjugate exponents is the most common statement - compare to Cauchy-Schwarz when using $p = q = 2$.

Dense subspaces

REMIND reader that the simple functions are dense

When Ω is $\mathbb{R}^d$ and $\mathbb{X}$ is $\mathbb{R}$, there are some other important dense subsets: - continuous functions with compact domains - the step functions

In considering the *separability* of $L^p_{\mathbb{X}}$-spaces, we will restrict our attention to real-valued functions on σ-finite measure spaces. In that case, the answer is straight-forward.

Theorem 1.1. *Let $(\Omega, \mathcal{A}, \mu)$ be a σ-finite measure space, and let $p \in [1, \infty)$. $L^p(\Omega, \mathcal{A}, \mu)$ is separable iff the Fréchet-Nikodym metric space is separable.*

An interesting consequence of this theorem is that if any one of these L^p spaces are separable, then they are all separable.

1.1 Prove Theorem 1.1.

1.2 Are the l^p-spaces separable?

1.3 Dual spaces

Theorem 1.2 (Cengiz, 1992, Corollary 1). *Let $p \in (1, \infty]$, and let q be its conjugate exponent. Suppose $(\Omega, \mathcal{A}, \mu)$ is a measure space and $\mathbb{X}$ is a Banach space with a separable dual. Given any $g \in L^q_{\mathbb{X}'}(\Omega, \mathcal{A}, \mu)$, the mapping*

$$f \mapsto \mu\, g(f)$$

is a continuous linear functional on $L^p_{\mathbb{X}}(\Omega, \mathcal{A}, \mu)$.

EXERCISE: Show that Theorem 1.2 also works for $p = 1$ if we add the condition that $(\Omega, \mathcal{A}, \mu)$ is localizable.

Theorem 1.3 (Cengiz, 1992, Corollary 1). *Let $p \in [1, \infty)$, and let q be its conjugate exponent. Suppose $(\Omega, \mathcal{A}, \mu)$ is a measure space and $\mathbb{X}$ is a Banach space with a separable dual. Then every continuous linear functional on $L^p_{\mathbb{X}}(\Omega, \mathcal{A}, \mu)$ must have the form*

$$f \mapsto \mu\, g(f)$$

for some $g \in L^q_{\mathbb{X}'}(\Omega, \mathcal{A}, \mu)$.

Theorems 1.2 and 1.3 together (along with Exercise REF) allow us the characterize the dual spaces under very general conditions.

Corollary 1.4. *Let $p \in (1, \infty)$, and let q be its conjugate exponent. Suppose $(\Omega, \mathcal{A}, \mu)$ is a measure space and $\mathbb{X}$ is a Banach space with a separable dual. The dual space of $L^p_{\mathbb{X}}(\Omega, \mathcal{A}, \mu)$ is isomorphic to $L^q_{\mathbb{X}'}(\Omega, \mathcal{A}, \mu)$. The conclusion holds for $p = 1$ as well if μ is localizable.*

We note that under the conditions of Corollary 1.4, $L^p_{\mathbb{X}}(\Omega, \mathcal{A}, \mu)$ is reflexive, i.e. the dual space of the dual space is the original space.

What about when $p = \infty$? If the codomain $\mathbb{X}$ is the reals, and μ is σ-finite, its dual space is isometric to the Banach space of finite charges $\mathrm{ba}(\Omega, \mathbb{A}, \mu)$.

2 Convergence in image measure

THE finite signed measures comprise the dual space of the bounded continuous functions (on a compact domain). The weak topology on the bounded continuous functions is the coarsest topology for which all of the finite signed measures are continuous as linear functionals. The weak-* topology on $\mathbb{M}$ is the coarsest topology for which the point-evaluation functionals are all continuous.

If $\mathbb{V}$ is reflexive, then the point-evaluation functionals make up the entirety of the dual space of $\mathbb{X}'$; in that case, all continuous linear functionals are (still) continuous with respect to the weak-* topology.

NEED TO INTRODUCE weak-* convergence of measures here - don't go into much detail, that should happen in the next chapter (including Levy-Prokhorov metric) - just define it here - then any facts about convergence in distribution that rely on more detail can be discussed (probably as exercises/solutions) in the next chapter.

Even if the sequence a functions aren't converging to the same function, their image measures may still be approaching some limiting distribution. In fact, the sequence of measurable functions don't even have to be defined on the same measure space for us to ask about the limiting behavior of their image measures; they just need to share a common codomain.

When the underlying space is a probability space, this is called convergence in distribution.

NEED TO cover the concept of CDF

2.1 Cumulative distribution functions

Define cdf. SPECIFIC to random variables. (Multivariate cdf? MAKE this a sidenote.)

1. Let F be a cdf of X. Show that F is right-continuous. That is, show that $x_n \downarrow x$ implies $F(x_n) \to F(x)$. What if the sequence is instead increasing to x?

Example cdfs - continuous, discrete.

Relationship between density and cdf.

THE INVERSE CDF TRANSFORM - THIS SHOULD BE AN EXERCISE rather than a section!

THE TEXT below was copied and pasted from one of my notecards - it needs to be revised until suitable for the book. e.g. MAKE some parts into exercises. AND MAKE it fit into the flow of this document.

HIGHLIGHT HOW central this is! One interesting implication: A uniform random variable is a good enough "source of randomness" to generate a draw from any distrn you want! (In fact, infinitely many draws.)

Let $G(t)$ be a cdf. This simply means that G is right continuous, monotone increasing, and has limits 0 and 1 as t goes to negative and positive infinity. (Any such function gives a probability distribution.) Define the inverse cdf transform by

$$G^{-1}(p) := \inf\{t : G(t) \geq p\}$$

(YOU CAN almost replace inf by min because of the right-continuity, I think. But you might want to worry about $U = 0$.) If $U \sim U[0, 1]$, then $G^{-1}(U)$ has G as its distribution function. This is easy to see, starting with the definition of the distribution function of the random variable $G^{-1}(U)$:

$$\begin{aligned} \mathbb{P}\{G^{-1}(U) \leq t\} &= \mathbb{P}\{U \leq G(t)\} \\ &= G(t) \end{aligned}$$

Sketch a plot of a cdf to see that $G^{-1}(U) \leq t$ is equivalent to $U \leq G(t)$. The only places that might cause you concern are the flat regions and the discontinuity points.

A closely-related fact is that the cdf transform of any *continuous* random variable is standard uniformly distributed. That is, let X have cdf F. Then $F(X) \sim U[0, 1]$.

$$\begin{aligned} \mathbb{P}\{F(X) \leq t\} &= \mathbb{P}\{X \leq F^{-1}(t)\} \\ &= F(F^{-1}(t)) \\ &= t \end{aligned}$$

(Because X is continuous, we know that F has a true inverse on the support of X.) With the inverse cdf transformation, we saw that we could get discreteness (if needed) from a draw from a continuous random variable. But in the cdf transformation, it's not possible to get something continuous from a random variable that has any point masses.

One can compose these two types of transformations in order to turn a draw from one distribution into a draw from another, via a standard uniform intermediary. If X is a continuous RV with cdf F and Y has cdf G, then $G^{-1}(F(X))$ has the same distribution as Y.

ANOTHER EXERCISE:

POINT out that the "average value of a function" formula from calculus is exactly an expectation using the uniform distribution. ALONG THE SAME LINES, the mean value theorem is just the observation that the average is between the inf and sup (strictly between if they aren't achieved - make sure to point this out in the section below!), and that by continuity every value between them is achieved. IS THERE A MORE GENERAL WAY OF STATING THE MEAN VALUE THEOREM?

If f is continuous on $[a, b]$ and differentiable on (a, b), then there exists a $c \in (a, b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

THE RIGHT HAND side is exactly the average value of f' $\mathbb{E}f'(X)$ for $X \sim \mathrm{Unif}[a, b]$. We know it's between the infimum and supremum of f' on $[a, b]$. IF f' is continuous, then it takes all of those values, so there exists a satisfactory $c \in (a, b)$. - need to be more careful about open and closed intervals here

THIS IS TRUE FOR ANY distribution of X - the uniform is only one particular case - but useful because we get to relate it to the antiderivative via the Fundamental Theorem of Calc.

SO the ordinary derivative is the "Lebesgue$[a, b]$ density" of the function somehow... THINK ABOUT THIS CAREFULLY

IS THERE a useful more general statement that allows for densities when X has a different distribution?

A useful re-expression of MVT:

$$f(b) = f(a) + (b - a)f'(c)$$

Let (X_n) be a sequence of random variables with cdfs (F_n), and let X have cdf F. We say that (X_n) converges in distribution to X (written $X_n \xrightarrow{\mathrm{d}} X$) if $F_n \to F$ at every point where F is continuous. (SIDENOTE: Sometimes people say "convergence in law" instead; this comes from "probability law" which is an old-fashioned term for probability distribution.) (If $X \sim P$, then we may also represent this convergence by $X_n \xrightarrow{\mathrm{d}} P$.) You'll see a generalization of this convergence in section REF. (WEAK CONVERGENCE for metric spaces - cover this in the section where weak [and weak-$*$] convergence are defined. INCLUDE generalizations of some of the observations made below.)

Any sequence of random variables that converges in probability also converges in distribution.

Theorem 2.1. *If $X_n \xrightarrow{\mathbb{P}} X$, then $X_n \xrightarrow{d} X$.*

Another connection between convergence in distribution and our previous notions of convergence is supplied by the following remarkable result.

Theorem 2.2 (Skorokhod's Representation Theorem). *If $X_n \xrightarrow{d} X$, then one can construct $Y_1, Y_2, \ldots$ and Y with matching distributions (i.e. each $Y_i \stackrel{d}{=} X_i$ and $Y \stackrel{d}{=} X$) such that $Y_n \xrightarrow{a.s.} Y$.*

(SIDENOTE: The notation $Y \stackrel{d}{=} X$ means that Y and X have the same distribution.)

2. Use inverse cdf transformation to prove Skorokhod's Representation Theorem.

One implication of this result is for interchanging limits and expectations. If the corresponding $Y_1, Y_2, \ldots$ are dominated [INSTEAD SAY "uniformly integrable" if I introduced that concept in the previous document] (or monotonically increasing and non-negative), then

$$\lim \mathbb{E}X_n = \lim \mathbb{E}Y_n = \mathbb{E} \lim Y_n = \mathbb{E}Y = \mathbb{E}X \tag{1}$$

We've seen the Continuous Mapping Theorem for almost sure convergence and convergence in measure (Theorem **??**). A third part of the theorem extends the same result to convergence in distribution. (SIDENOTE: As before, this result is stated a bit over-ambitiously. For now, just think of (X_n) as a sequence of random variables. A more general definition of convergence in distribution that works for random elements in general will be introduced in CITE SECTION.)

Theorem 2.3 (Continuous Mapping Theorem (for convergence in distribution)). *If g is a continuous function from one metric space to another and X_n be a sequence of random elements, then*

$$X_n \xrightarrow{d} X \qquad \textit{implies} \qquad g(X_n) \xrightarrow{d} g(X)$$

(SIDENOTE: As with the previous part of the Continuous Mapping Theorem, it is sufficient to require that the probability that X takes a value at which g is discontinuous has probability zero.)

3. Let $X_n \xrightarrow{d} X$ and let g be a bounded continuous function. Explain why

$$\lim \mathbb{E}g(X_n) = \mathbb{E}g(X)$$

Exercise CITE is related to a more general way of defining convergence in distribution; WE'LL COME BACK TO THIS IN cite the relevant functional analysis section.

There is also an analog to Theorem **??** for convergence in distribution, but it requires one of the sequences to be converging to a constant.

Theorem 2.4 (Slutsky's Theorem). *Let (X_n) and (Y_n) be sequences of random elements with metric space codomains $\mathbb{A}$ and $\mathbb{B}$. Let g be a continuous function from $\mathbb{A} \times \mathbb{B}$ to another metric space. Let c be a constant element of $\mathbb{B}$. Then*

$$X_n \xrightarrow{d} X, Y_n \xrightarrow{\mu} c \qquad \textit{implies} \qquad g(X_n, Y_n) \xrightarrow{\mu} g(X, c)$$

(SIDENOTE: As you've come to expect, g is allowed to be discontinuous with probability zero.)

Most statements of Slutsky's Theorem state the condition that Y_n conver in probability to c rather than in distribution. It turns out that convergence in distribution to a constant implies converge in probability, in a sense. The Y_n may be defined on different probability space, but each probability space has its own constant c function. The sequence of sets where $|Y_n - c| \geq \epsilon$ has probability going to zero.

If, for instance, the X_n and Y_n have the same vector space codomain, ther $X_n + Y_n$ converges [in distribution] to $X + c$, because addition is a continuous function. Similarly, if a continuous multiplication operation is defined, then $X_n Y_n$ converges to Xc. (SIDENOTE: But for instance, if Y_n is a sequence of scalars, $X_n / Y_n \xrightarrow{\mathrm{d}} X/c$ requires $c \neq 0$.)

DO PART of portmanteau theorem HERE?

One corollary of Theorem 2.2 is that bounded convergence carries over to convergence in distribution. That is, let $X_n \xrightarrow{d} X$ with $|X_i| < M$ for all i. Th there exists a sequence $\{Y_n\}$ with $|Y_i| < M$ almost surely and

$$X_n \stackrel{d}{=} Y_n \xrightarrow{a.s.} Y \stackrel{d}{=} X$$

So

$$\lim \mathbb{E}X_n = \lim \mathbb{E}Y_n = \mathbb{E} \lim Y_n = \mathbb{E}Y = \mathbb{E}X$$

Because convergence in probability implies convergence in distribution, we see that this result also tells us that bounded convergence also works for convergence in probability. (DOES dominated convergence work in general f convergence in probability?)

This corollary immediately tells us that if $X_n \xrightarrow{d} X$ and g is any bounded continuous function, then $\lim \mathbb{E}g(X_n) = \mathbb{E}g(X)$. Actually, this is one way of characterizing convergence in distribution. EXPLAIN.

How do we know that $X_n \xrightarrow{d} X$ implies $g(X_n) \xrightarrow{d} g(X)$ for bounded contin uous g? SHOW THE details.

We've seen that convergence in distribution means converge of cdfs at continuity points, but there's no end to the ways in which probability measur can be compared or said to converge. We'll survey a number of these other ways of quantifying the "difference" between distributions in section REF. (I KOLMOGOROV distance just the right thing for capturing convergence in distribution? Or is it stronger since it's uniform convergence? AND I DON'T think it's excluding discontinuity points, so that may be another difference.)