MEASURE THEORY LECTURES

MANJUNATH KRISHNAPUR

1. About the subject and the course

1. What would you say is the length of the following subsets of $\mathbb{R}$? $[0,1]$, $[0,1]\cup(3,8]$, $\mathbb{Z}$, $\{0\}\cup\{\frac{1}{n} : n\ge 1\}$, $\bigcup_{n\ge 1}[\frac{1}{2n+1},\frac{1}{2n}]$, $\mathbb{Q}\cap[0,1]$, Cantor's $\frac{1}{3}$-set. Perhaps some are clear, some are not.

2. In high school we learned the areas of square, rectangle, parallelogram, triangle, trapezium, and many such regular shapes. After learning integral Calculus, we learned how to compute areas of many more irregular objects. Still, one can ask about other shapes for which the answer may not be clear. $\mathbb{Q}^2$, Sierpinski gasket, Sierpinski carpet, etc.

3. From the theory of Riemann integral in many variables, we are abe to calculate the volumes of many shapes in $\mathbb{R}^3$, and by extension, in $\mathbb{R}^d$. Can do the same on the surface of a sphere $\mathbb{S}^{d-1}$, on the torus or more generally on Riemannian manifolds.

4. One extension of this is to ask about volumes of subsets in infinite dimensional spaces such as ℓ^2 and $C[0,1]$. For example, what is the volume of the ball of radius r in these spaces (you may normalize the unit ball to have volume 1)? Does such a notion exist?

5. Another extension is to possibly irregular spaces such as the Cantor set or Sierpinski gasket. For example, if K is the standard $\frac{1}{3}$-Cantor set, we probably agree that $K\cap[0,1/3]$ has half the length of K and that $K\cap[2/9,4/9]$ has $1/9$ of the total length of K. But what about general subsets of K? Similarly, can we measure areas *within* (to be distinguished from the question of the area of the Sierpinski gasket itself) the Sierpinski gasket?

6. When we learned Riemann integration, we saw that integration of functions is more general than computation of areas/lengths/volumes, as the latter are just integrals of indicator functions of subsets. But in another sense, the notion of areas/volumes subsumes integration, since the integral of a (positive) function is the area under its graph. Thus the problem of measuring areas and integrating functions are inseparable.

7. In this course, we shall learn about measures, which is the technical name for very general notions of volume. As in the previous point, it is intimately tied up with an integration theory, known as Lebesgue integral. These notions of measure and Lebesgue integral are vastly more

Date: 15/January/2021.

general than the earlier notions of areas and Riemann integral. In fact they are so general that for almost all of us, it will suffice for a lifetime.

8. Two reasons why this is a core course: For probability and for functional analysis and PDE.

9. When we have a notion of volume of subsets of a set X, if the total volume is 1, we can also talk about "picking a point at random from X", meaning that the chance that the chosen point falls inside a subset A is $\text{vol}(A)$ (e.g., consider the problem of throwing a dart at a dartboard). Thus, the notion of volume is also relevant to the notion of probability. In fact, measure theory turns out to provide a framework for probability theory.

10. Consider the set of Riemann integrable functions on $[0, 1]$ with the metric $d(f, g) = \int_0^1 |f(x) - g(x)|dx$. There is an issue that this is not quite a metric (distance can be zero, for example if f and g differ at finitely many points) but that can be taken care of by quotienting out the functions at zero distance from the zero function. The more serious problem is that even then, this space is not complete. In other words, it has holes that must be filled. Lebesgue's theory of integration ends up doing this! Because of that, it turns out to be particularly good for talking about existence of solutions to various problems on function spaces, e.g., differential equations. This is because one can usually construct approximate solutions, and completeness allows one to find a limiting candidate for the solution (analogy: solving for $x^2 = 2$ in rationals versus reals).

2. Jordan measurable sets in $\mathbb{R}$

11. We investigate the question of assigning a "length" to a subset $A \subseteq \mathbb{R}$. As and when we define the length, it will be denoted $\lambda(A)$.

12. We agree that an interval $I = [a, b]$ (or for that matter $I = (a, b)$ or $I = [a, b)$ or $I = (a, b]$) must be assigned length $|I| := b - a$. If $I_1, \ldots, I_k$ are pairwise disjoint intervals, we also agree that $A = I_1 \sqcup \ldots \sqcup I_k$ must be assigned length $|I_1| + \ldots + |I_k|$. Actually a little care is needed to see that this is well-defined, since a given set may have multiple representations as a finite union of intervals (for example, $[0, 2]$ can also be written as $[0, 1) \sqcup [1, 2]$). For the moment let us assume that it is not an issue, and proceed.

13. To proceed, we need to be more explicit about what properties we require of any notion of length. No one will dispute that it must have monotonicity under inclusion: If $A \subseteq B$, then A cannot have more length than B. This is satisfied when A, B are elementary sets as above.

14. Now for a general $A \subseteq \mathbb{R}$, taking inspiration from Riemann integration theory (where we sandwich the region under the graph of a function between graphs of step functions), define

$$\lambda^{\#}(A) = \inf\{|I_1| + \ldots + |I_k| : k \geq 0, I_j \text{ are p.w. disjoint intervals such that } I_1 \cup \ldots \cup I_k \supseteq A\},$$

$$\lambda_{\#}(A) = \sup\{|I_1| + \ldots + |I_k| : k \geq 0, I_j \text{ are p.w. disjoint intervals such that } I_1 \cup \ldots \cup I_k \subseteq A\}.$$

The monotonicity principle tells us that any number we assign as the length of A must be between $\lambda_{\#}(A)$ and $\lambda^{\#}(A)$. Hence, if $\lambda_{\#}(A) = \lambda^{\#}(A)$, we have no ambiguity and call this common number the length of A and denote it by $\lambda(A)$. Such sets are said to be *Jordan measurable*.

15. An equivalent way to state Jordan measurability is: Given $\epsilon > 0$, there exists elementary sets (i.e., a set that is a finite union of p.w. disjoint intervals) $E_1 \subseteq A \subseteq E_2$ such that $\lambda(E_2) - \lambda(E_1) < \epsilon$.

16. Intervals are Jordan measurable and $\lambda(I) = |I|$ (this needs proof!). But $\mathbb{Q}$ is not Jordan measurable, indeed, $\lambda^{\#}(\mathbb{Q}) = \infty$ and $\lambda_{\#}(\mathbb{Q}) = 0$. The Cantor set K is Jordan measurable and $\lambda(K) = 0$. Indeed, the 2^n intervals of length 3^{-n} each that form the nth stage of the construction give a cover for K proving that $\lambda^{\#}(K) \leq 2^n 3^{-n}$. As this is true for all n, we get $\lambda^{\#}(K) = 0$. Also, K contains no open interval, showing that $\lambda_{\#}(K) = 0$. Construct other Cantor-like sets (deleting a different proportion in the middle of each interval at stage n) and show that there are some that are Jordan measurable, some that are not.

17. Overall, many sets are not Jordan measurable and we have not assigned a length to such sets. Can we not simply take $\lambda^{\#}(A)$ to be the definition of the length of a set A? That would define length for all sets, agrees with $\lambda(A)$ for Jordan measurable A (in particular for elementary sets), and clearly has monotonicity under inclusion.

18. But it misses out on another property of lengths that we implicitly carry in our minds, *finite additivity*: If A and B are disjoint then $\lambda(A \sqcup B) = \lambda(A) + \lambda(B)$ (at least if all three lengths are well-defined). Observe that this property (together with positivity of lengths) subsumes monotonicity, since $A \subseteq B$ means that $A = B \sqcup (A \setminus B)$.

19. Finite additivity is violated by $\lambda^{\#}$. Indeed, if $A = \mathbb{Q} \cap [0,1]$ and $B = \mathbb{Q}^c \cap [0,1]$, then $A \sqcup B = [0,1]$, but $\lambda^{\#}(A) = \lambda^{\#}(B) = \lambda^{\#}([0,1]) = 1$. Hence it appears inevitable that we must give up the idea of defining length for all sets. Even accepting that, we shall now see a more general notion by which a much larger class of sets can be assigned lengths. For instance, in this new notion, $\mathbb{Q}$ will have a length, so will all Cantor-like sets, and in fact it will take some effort to show that there is a set that does not!

3. Lebesgue outer measure

20. Define the *Lebesgue outer measure* of a set $A \subseteq \mathbb{R}$ by

$$\lambda^*(A) = \inf \left\{ \sum_{k=1}^{\infty} |I_k| : I_k \text{ is of the form } (a_k, b_k] \text{ and } \bigcup_k I_k \supseteq A \right\}.$$

21. We allow countable covers in defining λ^* as opposed to only finite covers in defining $\lambda^{\#}$. To see the difference this makes, enumerate rationals as $r_1, r_2, \ldots$, and use the intervals $I_k = (r_k -$

$\epsilon 2^{-k-2}, r_k + \epsilon 2^{-k-2})$ to get $\lambda^*(\mathbb{Q}) \le \epsilon$. Hence $\lambda^*(\mathbb{Q}) = 0$. Contrast with $\lambda^{\#}(\mathbb{Q}) = \infty$. Countability is a running theme in the whole subject of measure theory.

22. There are also some unimportant differences, for example I_k need not be pairwise disjoint (one can remove that condition in the definition of $\lambda^{\#}$ too). Another difference is the use of lorc (left-open, right-closed) intervals. It makes no difference if we allow all intervals (or restrict to such intervals in the definition of $\lambda^{\#}$ too). The collection of lorc intervals will be technically more convenient later, hence we introduced them now.

23. The outer measure does have some desirable properties.

(1) (Monotonicity) $\lambda^*(A) \le \lambda^*(B)$ if $A \subseteq B$.

(2) (Countable subadditivity) $\lambda^*(\bigcup_n A_n) \le \sum_n \lambda^*(A_n)$ for any A_n.

(3) (Outer regularity) Given $A \subseteq \mathbb{R}$ and $\epsilon > 0$, there is an open set U such that $\lambda^*(U) < \lambda^*(A) + \epsilon$.

Monotonicity is trivial. Countable subadditivity follows from the fact that if $\{I_{n,j} : j \ge 1\}$ is a countable cover for A_n, then $\{I_{n,j} : n, j \ge 1\}$ is a countable cover for $\cup_n A_n$. To see outer regularity, get a countable cover $\{(a_j, b_j]\}$ such that $\sum_j (b_j - a_j) < \lambda^*(A) + \epsilon/2$ and then observe that $U = \cup_j (a_j, b_j + \epsilon 2^{-j-1})\}$ does the needful.

24. Still, λ^* does not solve all our problems. For example, there still exists (for now, accept this without proof) a set $A \subseteq [0,1]$ such that $\lambda^*(A) = \lambda^*([0,1] \setminus A) = 1$, which contradicts finite additivity since $\lambda^*([0,1]) \le 1$ (equality holds as shown later). How is this better than $\lambda^{\#}$ - there the same violation was achieved by $A = \mathbb{Q} \cap [0,1]$? The difference is that the violating set A is so much harder to construct and so pathological, that we won't mind not assigning it any length.

25. Construction of A: Regard $([0,1), +)$ as a group where "+" denotes addition modulo 1. Let $\alpha = 1/\sqrt{2}$ (or any irrational number) and consider the cyclic subgroup $G = \mathbb{Z}\alpha$ and $H = 2\mathbb{Z}\alpha$ and $H' = H + \alpha = (2\mathbb{Z}+1)\alpha$. Then, G, H, H' are all dense in $[0,1)$ and $G = H \sqcup H'$.

But G has uncountable index in $[0,1)$, and we "construct" (invoke axiom of choice, to be precise) a set $B \subseteq [0,1)$ that has exactly one element from each coset of G. Let $A = B + H = \{b + h : b \in B,\ h \in H\}$ so that $A^c = B + H'$. We claim that $\lambda^*(A) = \lambda^*(A^c) = 1$. Proof of this claim is postponed.

26. Note that A and A^c in this example are closely enmeshed. That is actually necessary. In fact, if A and B are at positive distance (i.e., $|x - y| > \delta$ for all $x \in A$ and $y \in B$), then $\lambda^*(A \sqcup B) = \lambda^*(A) + \lambda^*(B)$. To see this, observe that in the definition of outer measure, we may restrict to countable covers by intervals of length less than $\epsilon/3$ (why?). Then, if $\{I_n\}$ is such a cover, we can split it into disjoint collections $\{I_n : I_n \cap A = \emptyset\}$ and $\{I_n : I_n \cap B = \emptyset\}$ that cover B and A, respectively. From this, deduce that $\lambda^*(A+B) \geq \lambda^*(A) + \lambda^*(B)$.

27. The upshot of all this is that λ^* does not behave like "length" should. But it looks really close to behaving like one. The way out is to restrict the class of subsets for which we define length.

28. We have seen that $\lambda^{\#}$ and λ^* both fail to assign length to all subsets of $\mathbb{R}$, although we have asserted that λ^* is better and acceptable. Instead of taking special candidates like, wouldn't it be more logical to ask if there is *any* $\lambda^\dagger : 2^{\mathbb{R}} \mapsto [0, \infty]$ such that $\lambda^\dagger(A \sqcup B) = \lambda^\dagger(A) + \lambda^\dagger(B)$ and $\lambda^\dagger(I) = |I|$? In fact we can even ask for translation invariance, $\lambda^\dagger(A + x) = \lambda^\dagger(A)$ for all $A \subseteq \mathbb{R}$ and $x \in \mathbb{R}$. The answer is *yes*, there exists such a function!

29. Why don't we use it then? The problem is when we go to higher dimensions. Even on subsets of $\mathbb{R}^2$, there is a finitely additive function $\lambda^\dagger$ that has translation and rotation invariance and assigns area $|I_1| \times |I_2|$ to $I_2 \times I_2$. But for $d \geq 3$, there is no such function on subsets of $\mathbb{R}^d$! This makes it undesirable, for lengths, areas, volumes are not unrelated concepts living in different dimensions, but very closely connected to one another. For example, we would like the volume of $A \times I$ to be $mbarea(A) \times \text{length}(I)$.

30. The fact that there is no (non-zero) finitely additive, translation invariant notion of vo-Carathéodorylume on $\mathbb{R}^3$ is due the *Banach-Tarski paradox*. It asserts that there are sets $A_1, \ldots, A_5 \subseteq \mathbb{R}^3$ and Euclidean motions $T_1, \ldots, T_5 : \mathbb{R}^3 \mapsto \mathbb{R}^3$ (translations and rotations, i.e., $T(x) = Ax + b$ for some orthogonal matrix A and $b \in \mathbb{R}^3$) such that with $B_i = T_I(A_i)$, we have $A_1 \sqcup \ldots \sqcup A_5 = B(0,1)$ and $B_1 \sqcup \ldots \sqcup B_5 = B(0,2)$. If there was translation-invariant, finitely additive notion of volume $\lambda^\dagger$, when we would get $\lambda^\dagger(B(0,1)) = \lambda^\dagger(B(0,2))$.

4. Lebesgue measurable sets and Lebesgue measure

31. *Definition:* A set $A \subseteq \mathbb{R}$ is said to be *Lebesgue measurable* if it satisfies the *Carathéodory cut condition:* $\lambda^*(A \cap E) + \lambda^*(A^c \cap E) = \lambda^*(E)$ for all $E \subseteq \mathbb{R}$. The length (or *Lebesgue measure*) of a Lebesgue measurable set is defined to be its outer measure.

32. The definition is not very intuitive. One may think of it as the minimal fix to the problem that λ^* is not finitely additive. Here are other equivalent ways of defining Lebesgue measurable sets (the equivalence is not obvious and we must prove it later, but let us put them out there so that you can build the right mental picture from the start) that may look better motivated. Here A is assumed to be bounded.

(1) Say that A is measurable if it satisfies the cut-condition for an intervals E that contains A.

(2) Say that A is measurable if given $\epsilon > 0$, there is an elementary set A_ϵ (i.e., a finite disjoint union of intervals) such that $\lambda^*(A \Delta A_\epsilon) < \epsilon$.

(3) Say that A is measurable if given $\epsilon > 0$, there is an open set $U \supseteq A$ such that $\lambda^*(U \setminus A) < \epsilon$.

A general A is then said to be measurable if $A \cap [-n, n]$ is measurable for all n

33. It is also to be noted that to show that A is Lebesgue measurable, it suffices to show that $\lambda^*(A \cap E) + \lambda^*(A^c \cap E) \leq \lambda^*(E)$ for all $E \subseteq \mathbb{R}$. The other way is obvious by subadditivity of outer measure.

34. Let us analyse the class of all Lebesgue measurable sets, after which various sets of interest will be easily seen to be included. There is no alternate characterization of Lebesgue measurable sets. What we can do best is to know that many basic sets like intervals are in this collection and that the entire collection is closed under many set operations. To state the conclusions, we first introduce two of the fundamental notions in measure theory: sigma-algebra and measure.

35. Sigma algebra: A collection $\mathcal{F}$ of subsets of a non-empty set X is said to be a sigma-algebra if it contains the empty set, is closed under complements and countable unions.

36. Measure: Let $(X, \mathcal{F})$ be a measurable space (i.e., $\mathcal{F}$ is a sigma-algebra of subsets of X). Then $\mu : \mathcal{F} \mapsto [0, \infty]$ is called a *measure* if it is countably additive: $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$ whenever A_n are pairwise disjoint elements of $\mathcal{F}$.

37. Simple examples of sigma-algebas: Let X be any non-empty set and let $\mathcal{F} = \{\emptyset, X\}$ and $\mathcal{G} = 2^X$. Both $\mathcal{F}$ and $\mathcal{G}$ are sigma-algebras. Another one is $\mathcal{H} = \{A \subseteq X : A \text{ or } A^c \text{ is countable}\}$. Of course, if X is countable $\mathcal{H} = \mathcal{G}$, but not so if X is uncountable.

38. Examples of measures (on the above three sigma-algebras): Any $\mu : \mathcal{F} \mapsto [0, \infty]$ is a measure on $\mathcal{F}$. On $\mathcal{G}$ it is hard to define any interesting measure, but here is a class of them: Let $\{x_1, x_2, \ldots\} \subseteq X$ be a countable set and let $\alpha : \mathbb{N} \mapsto [0, \infty]$ be any function. Set $\mu(A) = \sum_{k : x_k \in A} \alpha_k$. On $\mathcal{H}$, there is a measure that gives 0 to all countable sets and 1 to all sets whose complement is countable.

39. The examples of sigma-algebras given above are either too small ($\mathcal{F}$ and $\mathcal{H}$) or too big ($\mathcal{G}$). Real examples that actually are used fall in between and are almost never very explicit. And measures on these sigma-algebras are not given - they have to be constructed with some effort.

40. Since sigma-algebras are defined by closure properties, arbitrary intersections of sigma-algebras are sigma-algebras. Therefore, for any $S \subseteq 2^X$, there is a smallest sigma-algebra that contains S (namely the intersection of all sigma-algebras containing it). It is denotes $\sigma(S)$ and is called the sigma-algebra generated by S. Caution: There is *no* simple way to write an element in $\sigma(S)$ in terms of elements of S using countable set operations.

41. If X is a metric space, the sigma-algebra generated by the collection of open sets is called the *Borel sigma-algebra* of X and denoted $\mathcal{B}_X$. On $\mathbb{R}$, check that the sigma-algebra generated by all lorc intervals is the same as $\mathcal{B}_{\mathbb{R}}$.

42. Theorem (Lebesgue). The collection $\mathcal{L}$ of Lebesgue measurable sets in $\mathbb{R}$ forms a sigma-algebra that contains the Borel sigma-algebra $\mathcal{B}$. The outer measure λ^* is a measure when restricted to $\mathcal{L}$.

43. The restriction of λ^* to $\mathcal{L}$ (or to $\mathcal{B}$) will be denoted λ and it is called *Lebesgue measure*. Later when we construct it on $\mathbb{R}^d$, we include a subscript d when there is danger of ambiguity. The existence of Lebesgue measure is the starting point of measure theory.

44. Why did we ask for countable additivity of measures? Would it not have been sufficient and natural to ask for finite additivity? Indeed, there is no motivation I can give now to justify the demand (it is not intuitive to me at all). But we shall give a sort of justification later. For now, let us only say that a rich theory is obtained by asking for countable additivity. Many mathematicians have tried to stay with finite additivity, but a century of experience has not turned up anything that is comparable in richness to the theory with countable additivity.

5. Proof of the existence of Lebesgue measure

45. The proof of Lebesgue's theorem is carried out in several steps, with the proofs of the two claims (that $\mathcal{L}$ is a sigma algebra and that λ^* is a measure on $\mathcal{L}$) closely intertwined.

46. $\mathcal{L}$ is closed under complements: From the symmetry of A and A^c in the cut condition.

47. $\mathcal{L}$ is closed under finite intersections: Indeed, if $A, B \in \mathcal{L}$ and $E \subseteq \mathbb{R}$, then

$$\lambda^*(E) \geq \lambda^*(E \cap A) + \lambda^*(E \cap A^c) \geq \lambda^*(E \cap A \cap B) + \lambda^*(E \cap A \cap B^c) + \lambda^*(E \cap A^c).$$

By subadditivity, the last two add to at least $\lambda^*(E \cap (A \cap B)^c))$, since $(A \cap B^c) \cup A^c = (A \cap B)^c$. Therefore, $\lambda^*(E) \geq \lambda^*(E \cap (A \cap B)) + \lambda^*(E \cap (A \cap B)^c)$. Thus $A \cap B \in \mathcal{L}$.

48. $\mathcal{L}$ is closed under finite unions: $A \cup B = (A^c \cap B^c)^c$.

49. λ^* is finitely additive on $\mathcal{L}$: If $A, B \in \mathcal{L}$ are disjoint and $E \subseteq \mathbb{R}$, then thet cut condition for $E \cap (A \sqcup B)$ shows that $\lambda^*(E \cap (A \sqcup B)) = \lambda^*(E \cap A) + \lambda^*(E \cap B)$. When $E = \emptyset$, this says $\lambda^*(A \sqcup B) = \lambda^*(B)$, which is finite additivity on $\mathcal{L}$. But the statement with general E (which need not be in $\mathcal{L}$) is stronger and used below.

50. Suppose $A_1, A_2, \ldots \in \mathcal{L}$ and $B_n = A_1 \cup \ldots \cup A_n$. Then $B_n \in \mathcal{L}$ and $B_n \uparrow B = \cup_n A_n$. If $E \subseteq \mathbb{R}$, then $\lambda^*(E) \geq \lambda^*(E \cap B_n) + \lambda^*(E \cap B_n^c)$ for every n. As $B_n^c \supseteq B^c$, we have $\lambda^*(E) \geq \lambda^*(E \cap B^c) + \lambda^*(E \cap B_n)$. Thus it suffices to show that $\lim_{n\to\infty} \lambda^*(E \cap B_n) \geq \lambda^*(E \cap B)$. By monotonicity, the limit on the left exists and does not exceed the right side. But the stronger form of finite additivity shown above implies that $\lambda^*(E\cap B_n) = \sum_{k=1}^n \lambda^*(E\cap(A_k\setminus A_{k-1}))$ where $A_0 = \emptyset$. Therefore $\lim_{n\to\infty} \lambda^*(E \cap B_n) \geq \sum_{k\geq 1} \lambda^*(E \cap (A_k \setminus A_{k-1}))$. But countable subadditivity shows that the right hand side is at least $\lambda^*(E \cap B)$. Thus, $B \in \mathcal{L}$.

51. If $A_k \in \mathcal{L}$ are pairwise disjoint and $A = \sqcup A_k$, the previous step shows (as $A_k \setminus A_{k-1} = A_k$) that $\lambda^*(E) = \lambda^*(E \cap A^c) + \sum_{k\geq 1} \lambda^*(E \cap A_k)$ for any $E \subseteq \mathbb{R}$. In particular, take $E = A$ to get countable additivity of λ^* on $\mathcal{L}$.

52. To summarize, we have shown that $\mathcal{L}$ is a sigma-algebra and that λ^* is a measure on it. For all we know, $\mathcal{L}$ could be the trivial sigma algebra $\{\emptyset, \mathbb{R}\}$! We next show that $\mathcal{L}$ contains intervals, and hence it is at least as large as the Borel sigma algebra.

53. $\mathcal{L}$ contains intervals: For definiteness, let $I = [a, b]$ (all other cases are similar) and let $I_\delta = [a - \delta, b + \delta]$ for $\delta > 0$. If $E \subseteq \mathbb{R}$, then $E \cap I$ and $E \cap I_\delta^c$ are at positive distance from each other, hence with $E_\delta = (E\cap I)\cup(E\cap I_\delta^c)$, we have $\lambda^*(E_\delta) = \lambda^*(E\cap I)+\lambda^*(E\cap I_\delta^c)$. But $E\Delta E_\delta$ is contained inside $[a-\delta, a]\cup[b, b+\delta]$ and hence $\lambda^*(E) \geq \lambda^*(E_\delta)-2\delta$ and similarly $\lambda^*(E\cap I_\delta^c) \geq \lambda^*(E\cap I^c)-2\delta$. Therefore, $\lambda^*(E) \geq \lambda^*(E \cap I) + \lambda^*(E \cap I^c) - 4\delta$. As δ is arbitrary, $I \in \mathcal{L}$.

54. $\lambda^*(I) = |I|$: Again, let us take a closed interval $I = [a, b]$, for definiteness. If $\{I_n\}$ is an open cover by lorc intervals so that $\sum_n |I_n| < \lambda^*(I)+\epsilon$, then enlarge I_n by $\epsilon 2^{-n-1}$ to an open interval J_n. By compactness of I, choose a finite subcover $J_1, \ldots, J_m$, where $J_i = (a_i, b_i)$ are ordered without loss of generality so that $a_1 < a_2 < \ldots < a_m$. Then $b_i < a_{i+1}$ and hence

$$\lambda^*(I) \geq \sum_{k=1}^{m} |J_k| - 3\epsilon \geq b_m - a_m + \sum_{k=1}^{m-1} (a_{k+1} - a_k) - 3\epsilon = b_m - a_1 - 3\epsilon$$

which is more than $|I| - 3\epsilon$ (since $a_1 \leq a$ and $b_m \geq b$). Thus $\lambda^*(I) \geq b - a$, the other way inequality being obvious (take the cover with $I_1 = I$ and $I_n = \emptyset$ for $n \geq 2$).

55. We try to understand the structure of the proof of existence of Lebesgue measure and transfer it to a more abstract setting without distracting features.

56. Question. Let X be a non-empty set and let $\mathcal{S}$ be a collection of subsets of X that generates the sigma-algebra $\mathcal{F}$. Given a function $\mu : \mathcal{S} \mapsto [0, \infty]$, does it extend to a measure on $\mathcal{F}$? Is the extension unique?

57. Clearly, this is the prototype of the problem of Lebesgue measure - where we knew what the length of intervals must be, and we extended it to the Lebesgue sigma-algebra.

58. In general, one may expect that extension requires some consistency conditions. For example., if there are $A, B \in \mathcal{S}$ such that $A \cap B = \emptyset$ and $A \cup B \in \mathcal{S}$, but $\mu(A \cup B) \neq \mu(A) + \mu(B)$, then there is no way to extend μ as a measure to $\mathcal{F}$.

59. Perhaps surprisingly, even uniqueness is false! Let $X = \{1, 2, 3\}$ and $\mathcal{S} = \{\{1, 2\}, \{2, 3\}\}$ and $\mu(\{1, 2\}) = \mu(\{2, 3\}) = 1$. There are two measures ν and θ that extend μ to $\mathcal{F} = 2^X$ defined by $\nu(A) = \mathbf{1}_{A \ni 1} + \mathbf{1}_{A \ni 3}$ and $\theta(A) = \mathbf{1}_{A \ni 2}$.

Contrast this with the fact that if two linear functionals on a vector space agree on a generating set, then must be equal. The difference is another reminder that the word 'generated sigma-algebra' has only an external definition, not an internal one (Note: If S is a collection of vectors in a vector space V, then span(S) can be described externally as the intersection of all subspaces that contain S or internally as the collection of all finite linear combinations of elements of S).

60. To understand this and answer the question positively, various kinds of collections of subsets of X are used. We shall probably only use the following (in addition to sigma algebras, which are more restrictive than all these): (I) Algebra: Closed under complements, finite unions. (II) Monotone class: Closed under increasing unions and decreasing intersections. (III) π-system: Closed under intersections. (IV) λ-system: Closed under increasing unions, proper differences.

61. Algebras, monotone classes, π and λ systems, rings, etc., are defined by closure properties, and hence arbitrary intersections of these are collections of the same kind. Hence, we can talk about the algebra (or monotone class or ...) generated by a collection of subsets of X.

62. Monotone class theorem. The monotone class generated by an algebra is a sigma-algebra. Hence $\sigma(\mathcal{A}) = \mathcal{M}(\mathcal{A})$ for any algebra $\mathcal{A}$.

63. *Proof of the monotone class theorem*: Let $\mathcal{M}$ be the monotone class generated by an algebra $\mathcal{A}$. Let $\mathcal{M}_0 = \{A \in \mathcal{M} : A^c \in \mathcal{M}\}$. Then $\mathcal{M}_0$ is a monotone class (if $\mathcal{M}_0 \ni A_n \uparrow A$ then $A \in \mathcal{M}$ and $\mathcal{M}_0 \ni A_n^c \downarrow A^c$, hence $A \in \mathcal{M}_0$, similarly for decreasing limits). Also $\mathcal{M}_0 \supseteq \mathcal{A}$ (as $\mathcal{A}$ is closed under complements), hence $\mathcal{M} = \mathcal{M}_0$ showing that $\mathcal{M}$ is closed under complements.

Now fix $A \in \mathcal{A}$ and let $\mathcal{M}_A = \{B \in \mathcal{M} : A \cup B \in \mathcal{M}\}$. Argue that $\mathcal{M}_A$ is a monotone class that contains $\mathcal{A}$, hence equal to $\mathcal{M}$. Thus if $A \in \mathcal{A}$ and $B \in \mathcal{M}$, then $A \cup B \in \mathcal{M}$. Next fix $A \in \mathcal{M}$ and consider $\mathcal{M}_A$ defined exactly the same way. By what we have shown, $\mathcal{M}_A$ contains $\mathcal{A}$ and by the same proof as before, $\mathcal{M}_A$ is a monotone class. Thus, $\mathcal{M}_A = \mathcal{M}$ for $A \in \mathcal{M}$, implying that $\mathcal{M}$ is closed under finite unions.

If $A_n \in \mathcal{M}$, then $B_n = A_1 \cup \ldots \cup A_n$ are in $\mathcal{M}$ and increase to $\cup_n A_n$. By monotone class property, $\cup_n A_n \in \mathcal{M}$. Thus $\mathcal{M}$ is closed under countable unions. This completes the proof that $\mathcal{M}$ is a sigma-algebra.

64. The point of monotone classes is that measures behave well under monotone limits, but not under arbitrary countable unions/intersections. Indeed, if $(X, \mathcal{F}, \mu)$ is a measure space, then countable additivity is equivalent to the statement that if $A_n \uparrow A$ (all in $\mathcal{F}$), then $\mu(A_n) \uparrow \mu(A)$.

Measures also behave well under decreasing intersections, but with a caveat. If $A_n \downarrow A$ and $\mu(A_n) < \infty$ for some n, then $\mu(A_n) \downarrow \mu(A)$. To see the necessity of the condition of finiteness, let μ be counting measure on $\mathbb{Z}$ and $A_n = \{n, n+1, \ldots\} \downarrow \emptyset$. Then $\mu(A_n) = \infty$ for all n and $\mu(A) = 0$.

65. As an application, we prove that if μ, ν are finite measures on a sigma-algebra $\mathcal{F}$ generated by an algebra $\mathcal{A}$, then $\mu = \nu$.

Indeed, the collection $\{A \in \mathcal{F} : \mu(A) = \nu(A)\}$ is easily seen to be a monotone class (finiteness of the measures is used to get closure under decreasing intersections) and contains $\mathcal{A}$ by assumption, hence equal to $\sigma(\mathcal{A}) = \mathcal{F}$.

66. As in this example, finite measures often have special properties. Quite often they can be extended to sigma-finite measures. μ on $(X, \mathcal{F})$ is said to be sigma-finite if there exist $X_n \in \mathcal{F}$ such that $\sqcup_n X_n = X$ and $\mu(X_n) < \infty$ for all n.

67. As an exercise, show that if μ, ν are two sigma-finite measures on $\mathcal{F} = \sigma(\mathcal{A})$ where $\mathcal{A}$ is an algebra, then if μ, ν agree on $\mathcal{A}$, they agree on $\mathcal{F}$ (Caution: It is not given that the sets X_n are in $\mathcal{A}$).

68. Instead of algebras and monotone classes, one can work with π and λ systems. In this course you may entirely avoid the latter.

69. Sierpinski-Dynkin $\pi - \lambda$ theorem. The λ-system generated by a π-system is the same as the sigma-algebra generated by the π-system.

70. Corollary. If $\mathcal{S}$ is an π-system and generates the sigma-agebra $\mathcal{F}$, then any two measures on $\mathcal{F}$ that agree on $\mathcal{S}$, are equal. In particular, this is true if $\mathcal{S}$ is an algebra.

71. We omit the proofs as exercises (arguments similar to the proof of monotone class theorem).

72. Now we come to the existence question. Suppose $\mu : \mathcal{S} \mapsto [0, \infty]$ is given. To be able to extend it as a measure to $\mathcal{F} = \sigma(\mathcal{S})$, clearly it must be countably additive on $\mathcal{S}$. That is, if $A_n \in \mathcal{S}$

are pairwise disjoint and it so happens that $\sqcup_n A_n \in \mathcal{S}$, then $\mu(\sqcup_n A_n) = \sum_n \mu(A_n)$. It turns out that when $\mathcal{S}$ is an algebra, this consistency condition is also sufficient to guarantee extension!

73. Carathéodory's extension theorem. Let $\mathcal{F}$ be a sigma-algebra that is generated by an algebra $\mathcal{A}$. If $\mu : \mathcal{A} \mapsto [0, \infty]$ is countably additive on $\mathcal{A}$, then it extends to a measure on $\mathcal{F}$ in a unique way.

74. The earlier extension theorem that we gave is a special case of this, where $\mathcal{A} = \{I_1 \sqcup \ldots \sqcup I_k : k \geq 0, I_r \text{ are lorc intervals}\}$ and $\lambda : \mathcal{A} \mapsto [0, \infty]$ is defiined by $\lambda(I_1 \sqcup \ldots \sqcup I_k) = \lambda(I_1) + \ldots + \lambda(I_k)$. In the earlier proof, we did not explicitly check countable additivity of λ on $\mathcal{A}$, but what we did can be rephrased that way (and we shall do this checking later).

75. One might notice that in that extension theorem, the measure extended to $\mathcal{L}$, a larger sigma-algebra than $\mathcal{F} = \sigma(\mathcal{A})$ (in that example $\mathcal{F} = \mathcal{B}(\mathbb{R})$). In fact the proof given next shows that it is also true in general. But we are often happy enough to restrict the measure to $\mathcal{F}$. It has the advantage that $\mathcal{F}$ depends only on $\mathcal{A}$, but the larger sigma algebra depends on μ too.

7. Proof of the Carathéodory extension theorem

76. Let $\mathcal{A} \subseteq 2^X$ and $\mu : \mathcal{A} \mapsto [0, \infty]$ and $\mathcal{F} = \sigma(\mathcal{A})$. At the moment, no assumptions on $\mathcal{A}$ or μ.

77. Define the *outer measure* $\mu^* : 2^X \mapsto [0, \infty]$ by

$$\mu^*(A) = \inf\left\{\sum_{n=1}^{\infty} \mu(A_n) : A_n \in \mathcal{A},\ \cup_n A_n \supseteq A\right\}.$$

It satisfies (1) Monotonicity: $\mu^*(A) \leq \mu^*(B)$ if $A \subseteq B$, (2) Countable subadditivity: $\mu^*(\cup_n A_n) \leq \sum_n \mu^*(A_n)$, (3) $\mu^*(\emptyset) = 0$. In general, an *outer measure* is a non-negative function on 2^X satisfying these three properties. It need not arise from a μ as here.

78. The key step is Carathéodory's cut condition: Let $\mathcal{L}_{\mu^*}$ be the collection of all $A \subseteq X$ for which $\mu^*(E) = \mu^*(A \cap E) + \mu^*(A^c \cap E)$ for all $E \subseteq X$. Sets in $\mathcal{L}_{\mu^*}$ are said to be μ^*-measurable. The non-trivial part of the cut condition is $\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$, the other way inequality follows from subadditivity.

79. The proof of Carathéodory's extension theorem can be broken into two parts.

(1) If μ^* is any outer measure, then $\mathcal{L}_{\mu^*}$ is a sigma algebra and μ^* is a measure on it.

(2) If μ^* arises from a countably additive μ on an algebra $\mathcal{A}$, then $\mu^* = \mu$ on $\mathcal{A}$ and $\mathcal{L}_{\mu^*} \supseteq \mathcal{A}$.

80. For a proof of the first statement, Paragraphs 46-51 can be copied verbatim, with $\mathcal{L}_{\mu^*}$ in place of $\mathcal{L}$ and μ^* in place of λ^*. Steps 46-49 show that $\mathcal{L}$ is an algebra on which μ is finitely additive, and the next two steps upgrade algebra to sigma-algebra and finite additivity to countably additivity.

81. To prove the second statement, we fist prove that $\mu^* = \mu$ on $\mathcal{A}$. Let $A \in \mathcal{A}$. Clearly $\mu^*(A) \leq \mu(A)$ as the singleton $\{A\}$ is a cover for A. To see equality, consider any countable cover

$\{A_n\}$ for A with $A_n \in \mathcal{A}$. Let $B_n := (A_n \cap A) \setminus \cup_{k=1}^{n-1}(A_k \cap A)$. Then B_n are pairwise disjoint, $B_n \in \mathcal{A}$ and $\sqcup_n B_n = A$. Hence $\sum_{k=1}^{n} \mu(B_k) = \mu(B_1 \sqcup \ldots \sqcup B_n) \uparrow \mu(A)$ by countable additivity of μ on $\mathcal{A}$. Therefore $\mu(A) = \sum_n \mu(B_n)$. But $\mu(B_n) \le \mu(A_n)$ for all n, from which it follows that $\sum_n \mu(A_n) \ge \mu(A)$. As this is true for any cover, $\mu^*(A) \ge \mu(A)$.

82. It only remains to show that $\mathcal{A} \subseteq \mathcal{L}_{\mu^*}$. Let $A \in \mathcal{A}$ and $E \subseteq X$. Let $\{B_n\} \subseteq \mathcal{A}$ be a cover for E satisfying $\sum_n \mu(B_n) \le \mu^*(E) + \epsilon$. As μ is finitely additive on $\mathcal{A}$, we see that $\mu(B_n) = \mu(B_n \cap A) + \mu(B_n \cap A^c)$. When summed over n, the left side is bounded above by $\mu^*(E) + \epsilon$ while countable subaditivity implies that the two sums on the right side are bounded below by $\mu^*(E \cap A)$ and $\mu^*(E \cap A^c)$, respectively. Thus, $\mu^*(E) + \epsilon \ge \mu^*(E \cap A) + \mu^*(E \cap A^c)$. As ϵ is arbitrary, $A \in \mathcal{L}_{\mu^*}$.

8. Checking countable additivity on an algebra

83. If $\mu : \mathcal{A} \mapsto [0, \infty]$ is finitely additive and finite ($\mu(X) < \infty$), then the condition of countable additivity is equivalent to showing that $\mu(A_n) \downarrow 0$ if $A_n \in \mathcal{A}$ and $A_n \downarrow \emptyset$.

If μ is sigma-finite in the sense that there are $X_n \in \mathcal{A}$ such that $X = \sqcup_n X_n$ and $\mu(X_n) < \infty$, then we can restrict to any X_n with $\mathcal{A}_n = \{A \cap X_n : A \in \mathcal{A}\}$ and $\mu_n = \mu\big|_{\mathcal{A}_n}$ and apply the above criterion. It may also be noted that $\sigma(\mathcal{A}) = \{A = \sqcup_n A_n : A_n \in \sigma(\mathcal{A}_n)\}$.

84. Even this appears complicated to check. For example, in the standard example of the algebra generated by lorc intervals in $\mathbb{R}$, we need to take $A_n = I_{n,1} \sqcup \ldots \sqcup I_{n,m_n}$ that decreases to $\emptyset$ (as $\mathbb{R}$ is sigma finite, as explained earlier, we may assume $A_n \subseteq [-M, M]$ for some M for all n). The number of intervals (m_n) can increase without bound. A more convenient way to check countable additivity would be nice.

85. Claim: Let $\mathcal{A}$ be an algebra of subsets of a metric space X. Let $\mu : \mathcal{A} \mapsto [0, \infty]$ be finitely additive. Assume: (1) For any $A \in \mathcal{A}$ and $\epsilon > 0$, there exists $C \in \mathcal{A}$ and compact K such that $C \subseteq K \subseteq A$ and $\mu(C) > \mu(A) - \epsilon$. (2) For $A \in \mathcal{A}$ with $\mu(A) = \infty$, there exist $\mathcal{A} \ni C_n \subseteq A$ such that $\infty > \mu(C_n) \uparrow \infty$. Then μ is countably additive on $\mathcal{A}$.

86. It would have been simpler to say that there must be a compact $K \subseteq A$ such that $\mu(K) > \mu(A) - \epsilon$, but in standard examples like the algebra of lorc intervals on $\mathbb{R}$, compact sets are not in the algebra.

Those who care about generalities may observe that the proof below works just as well if the class of compact sets is replaced by a *compact class* (recall: $\mathcal{K}$ is a compact class if $A_n \in \mathcal{K}$ and $\cap_n A_n = \emptyset$, then $\cap_{n \le N} A_n = \emptyset$ for some N).

87. Proof of the claim: Let $A_n, A \in \mathcal{A}$ and $A_n \uparrow A$. Let $B_n = A \setminus A_n$ so that $\mathcal{A} \ni B_n \downarrow \emptyset$.

If $\mu(A) < \infty$, find compact K_n and $C_n \in \mathcal{A}$ such that $C_n \subseteq K_n \subseteq B_n$ and $\mu(C_n) \ge \mu(B_n) - \frac{\epsilon}{2^n}$. Let $K'_n = K_1 \cap \ldots \cap K_n$ and $C'_n = C_1 \cap \ldots \cap C_n$, so that we still have K'_n compact, $C'_n \in \mathcal{A}$ and $C'_n \subseteq K'_n \subseteq B_n$. Further, $\mu(B_n \setminus C'_n) \le \sum_{j=1}^{n} \mu(B_n \setminus C_j) \le \sum_{j=1}^{n} \mu(B_j \setminus C_j) \le \epsilon$, which means that $\mu(B_n) \le \mu(C'_n) + \epsilon$. But $B_n \downarrow \emptyset$, hence $K'_n \downarrow \emptyset$ and by finite intersection property, $K'_n = \emptyset$ for large n which also forces $C'_n = \emptyset$ for large n. Thus, $\limsup \mu(B_n) \le \epsilon$, i.e., $\mu(B_n) \downarrow 0$. Equivalently, $\mu(A_n) \uparrow \mu(A)$.

If $\mu(A) = \infty$, then find $\mathcal{A} \ni C_k \subseteq A$ such that $\mu(C_k) < \infty$ and $\mu(C_k) \uparrow \infty$. Then $A_n \cap C_k \uparrow C_k$ for any k, as $n \to \infty$. Therefore, $\mu(A_n \cap C_k) \uparrow \mu(C_k)$ as $n \to \infty$. This shows that $\liminf \mu(A_n) \ge \mu(C_k)$ for any k, which implies that $\mu(A_n) \uparrow \infty$.

88. Recall that from any outer measure μ^* one gets a measure on $\mathcal{L}_{\mu^*}$. The whole point of the countable additivity of μ on an algebra $\mathcal{A}$ was to ensure that $\mathcal{L}_{\mu^*}$ is not too small (and that μ^* equals μ on $\mathcal{A}$). Here is another way to ensure this. Observe that the condition is on μ^*.

89. Claim: Suppose (X, d) is a metric space and let μ^* be an outer measure on 2^X. Suppose we have the additivity property $\mu^*(A \sqcup B) = \mu^*(A) + \mu^*(B)$ whenever $d(A, B) := \inf\{d(x, y) : x \in A,\ y \in B\} > 0$. Then, $\mathcal{L}_{\mu^*} \supseteq \mathcal{B}(X)$.

90. To prove the claim, it suffices to show that any closed $A \subseteq X$ satisfies the cut condition. For fixed $E \subseteq X$ we wish to check that $\mu^*(E) \ge \mu^*(E \cap A) + \mu^*(E \cap A^c)$. Assume $\mu^*(E) < \infty$, otherwise this is trivially true.

The idea is to write $A^c = \sqcup_{n \ge 0} B_n$, where $B_n = \{x \in X : \frac{1}{n+1} \le d(x, A) < \frac{1}{n}\}$ (an "annulus"). Let $C_n = B_0 \sqcup \ldots \sqcup B_{n-1}$. Then $d(C_n, A) \ge \frac{1}{n}$, hence $\mu^*(E) \ge \mu^*(E \cap A) + \mu^*(E \cap C_n)$. If we can argue that $\mu^*(E \cap C_n) \uparrow \mu^*(E \cap A^c)$, then the cut condition is verified.

By subadditivity, $\mu^*(E \cap A^c) - \mu^*(E \cap C_n)$ is bounded from above by $\mu^*(E \cap (B_n \sqcup B_{n+1} \sqcup \ldots)) \le \sum_{j \ge n} \mu^*(E \cap B_j)$. The tail of a convergent series converges to 0, hence it suffices to show that $\sum_j \mu^*(E \cap B_j) < \infty$. Now, $d(B_j, B_{j+2}) \ge \frac{1}{j+1} - \frac{1}{j+2} > 0$ (this does not work for B_j and B_{j+1}). Hence, by the metric property of μ^*, for any $m \ge 1$ we have

$$\sum_{j=1}^{m} \mu^*(E \cap B_{2j}) = \mu^*(E \cap (B_2 \sqcup B_4 \sqcup \ldots \sqcup B_{2m})) \le \mu^*(E)$$

which is finite. Therefore, $\sum_j \mu^*(E \cap B_{2j}) < \infty$. Similarly, $\sum_{j \ge 1} \mu^*(E \cap B_{2j-1}) < \infty$.

91. The metric condition of an outer measure is often quite easy to check, and quickly gives a Borel measure. Two books that take this approach are of Stein and Shakarchi and of Wheeden and Zygmund. Bogachev's book has the criterion with compact classes and much more.

9. Lebesgue measure on $\mathbb{R}^d$

92. Now consider $\mathbb{R}^d$ with the π-system $\mathcal{S}$ of lorc rectangles $R = I_1 \times \ldots \times I_d$ where I_q are lorc intervals in $\mathbb{R}$ and set $|R| = |I_1| \times \ldots \times |I_d|$, which agrees with our usual notion of volume.

93. The algebra generated by $\mathcal{S}$ is $\mathcal{A} = \{R_1 \sqcup \ldots \sqcup R_k : k \ge 0,\ R_i \in \mathcal{S}\}$ and we naturally extend the notion of volume to $\mathcal{A}$ by defining $\lambda(R_1 \sqcup \ldots \sqcup R_k) = \lambda(R_1) + \ldots + \lambda(R_k)$. One must check that this is a legitimate definition, since there are multiple ways to write a set as a union of disjoint rectangles. For example, $(0,1] \times (0,2] = (0,1] \times (0,1] \sqcup (0,1] \times (1,2]$.

94. If $A = R'_1 \sqcup \ldots \sqcup R'_m$ and $A = R_1 \sqcup \ldots \sqcup R_n$, then $R'_i = \sqcup_{j=1}^n (R'_i \cap R_j)$ for any $i \le m$. Therefore to show that $\sum_{i=1}^m |R'_i| = \sum_{j=1}^n |R_j|$, it suffices to show that $|R'_i| = \sum_{j=1}^n |R'_i \cap R_j|$. In words, if an lorc rectangle is a union of disjoint lorc rectangles, then its area is the sum of the areas of the smaller rectangles. We refer to Stein and Shakarchi, Lemma 1.1 for a proof.

95. Finite additivity of λ on $\mathcal{A}$ follows easily: If $A = R_1 \sqcup \ldots \sqcup R_n$ and $B = R'_1 \sqcup \ldots \sqcup R'_m$ and $A \cap B = \emptyset$, then $A \sqcup B = R_1 \sqcup \ldots \sqcup R_n \sqcup R'_1 \sqcup \ldots \sqcup R'_m$.

96. Thus, to check countable additivity on the algebra, we only need to check the conditions in the previous section. Let $R = \times_{q=1}^d (a_q, b_q]$. If R has an infinite side, then we may take $C_n = R \cap (-n,n]^d \in \mathcal{A}$ to get $\infty > \mu(C_n) \uparrow \infty$. If R has finite sides, we may take $C = \times_{q=1}^d (a_q + \delta, b_q] \in \mathcal{A}$ and $K = \times_{q=1}^d [a_q + \delta, b_q]$ compact. Then $C \subseteq K \subseteq A$ and if δ is small enough, we also have $\lambda(C) \ge \lambda(A) - \epsilon$. This verifies the conditions for a single rectangle. For a general element in the algebra, write it as a finite union of rectangles and do the same inside each component rectangle.

97. Conclusion: There exists a sigma-algebra $\mathcal{L}_d \supseteq \mathcal{B}(\mathbb{R}^d)$ on $\mathbb{R}^d$ on which there is a measure λ_d that satisfies $\lambda_d(R) = |R|$ for any rectangle R. This is known as the Lebesgue measure on $\mathbb{R}^d$ and elements of $\mathcal{L}_d$ are said to be Lebesgue measurable.

10. Regularity of Lebesgue measure and related matters

98. Let λ_d (we drop the subscript often) denote the Lebesgue measure on the Lebesgue sigma-algebra $\mathcal{L}_d$ that is larger than the Borel sigma-algebra $\mathcal{B}(\mathbb{R}^d)$. What are the relationships between these sigma-algebras? How do we understand the sets in them if they are not explicitly described?

99. Outer regularity: $\lambda_d(A) = \inf\{\lambda_d(G) : A \subseteq G \text{ open}\}$ for any $A \in \mathcal{L}_d$.

Trivially $\lambda_d(A) \le \lambda_d(G)$ for any $G \supseteq A$. To see the inequality the other way, if $\lambda_d(A) < \infty$ (otherwise take $G = \mathbb{R}$), find a countable covering of A by lorc rectangles so that $\sum_n |R_n| \le \lambda_d(A) + \epsilon$. Enlarge each R_n to an open rectangle $\tilde{R}_n$ so that $|\tilde{R}_n| \le |R_n| + \epsilon 2^{-n}$. Then $G := \cup \tilde{R}_n$ is open, contains A and $\lambda_d(G) \le \sum_n |\tilde{R}_n| < \sum_n |R_n| + \epsilon \le \lambda_d(A) + 2\epsilon$.

100. Inner regularity: $\lambda_d(A) = \sup\{\lambda_d(K) : A \supseteq K \text{ compact}\}$ for any $A \in \mathcal{L}_d$.

Let $Q_n = [-n,n]^d$. As $\lambda_d(A \cap Q_n) \uparrow \lambda_d(A)$, we only need to show this for bounded A. If $A \subseteq Q_N$, use outer regularity to find an open set $G \supseteq Q_N \setminus A$ such that $\lambda_d(G) < \lambda_d(Q_N \setminus A) + \epsilon$

which is $\lambda_d(Q_N) - \lambda_d(A) + \epsilon$. Then $K := Q_N \setminus G$ is a compact set contained inside A. Further, $\lambda_d(K) = \lambda_d(Q_N) - \lambda_d(G) \geq \lambda_d(A) - \epsilon$.

101. The cardinality of $\mathcal{B}(\mathbb{R}^d)$ is the same as $\mathbb{R}$ while the cardinality of $\mathcal{L}_d$ is that of $2^{\mathbb{R}}$. We do not justify the first statement here, but the second one easily follows from the next point. For now, note the consequence, $\mathcal{B}_d$ is strictly smaller than $\mathcal{L}_d$.

102. If $\lambda^*(A) = 0$, we claim that $A \in \mathcal{L}_d$. As a consequence, all subsets of A are also measurable. If A is the standard Cantor set (which has zero measure), then A has the same cardinality as $\mathbb{R}$ and hence the collection of subsets of A is itself equinumerous with $2^{\mathbb{R}}$. Of course $\mathcal{L}_d \subseteq 2^{\mathbb{R}}$, hence the other way is clear. This shows that $\mathcal{L}_d$ has the same cardinality as $2^{\mathbb{R}}$.

To justify the claim, let $E \subseteq \mathbb{R}^d$. Then $\lambda^*(A \cap E) = 0$ while $\lambda^*(A^c \cap E) \leq \lambda^*(E)$, by monotonicity. Therefore, $\lambda^*(A \cap E) + \lambda^*(A^c \cap E) \leq \lambda^*(E)$, showing that A satisfies the cut-condition.

103. Relationship between Lebesgue and Borel sigma-algebras: Let $A \subseteq \mathbb{R}^d$. Then $A \in \mathcal{L}_d$ if and only if there are $B, C \in \mathcal{B}(\mathbb{R}^d)$ such that $B \subseteq A \subseteq C$ and $\lambda_d(B) = \lambda_d(C)$.

If there are such Borel sets B, C, then $\lambda(C \setminus B) = \lambda(C) - \lambda(B) = 0$. Hence, $A \setminus B$ (a subset of $C \setminus B$) is also measurable and so is $A = B \sqcup (A \setminus B)$.

Conversely, if $A \in \mathcal{L}_d$, then using regularity, find $K_n \subseteq A \subseteq G_n$ such that G_n is open, K_n is compact and $\lambda(G_n) \downarrow \lambda(A)$ and $\lambda(K_n) \uparrow \lambda(A)$. Now set $B = \cup_n K_n$ and $C = \cap_n G_n$. Clearly $B \subseteq A \subseteq C$ and B, C are Borel sets. Further, $\lambda(K_n) \leq \lambda(B) \leq \lambda(A) \leq \lambda(C) \leq \lambda(C_n)$ for all n. Let $n \to \infty$ to see that $\lambda(C) = \lambda(B)$.

104. Another way to say this is that given a measurable set A, there is a Borel set B such that $\lambda(A \Delta B) = 0$. In that sense, there is no big loss in working with Borel sets than measurable sets.

105. Ultimately one only understands intervals/rectangles. To understand Borel sets, one must relate them to rectangles. Here are three connections that are useful in different ways.

(1) If $A \in \mathcal{L}_d$ and $\lambda_d(A) < \infty$, then for any $\epsilon > 0$ there is an elementary set $B_\epsilon = R_1 \sqcup \ldots \sqcup R_n$ such that $\lambda(A \Delta B_\epsilon) < \epsilon$.

(2) If $A \in \mathcal{L}_d$ and $\lambda_d(A) > 0$, then for any $\epsilon > 0$ there is a rectangle R such that $\lambda_d(A \cap R) \geq (1 - \epsilon)\lambda_d(R)$.

(3) If $A \in \mathcal{L}_d$ and $\lambda_d(A) > 0$, then $A - A$ contains a neighbourhood of the origin.

106. To prove the first statement, find a covering of A by lorc rectangles $\{R_n\}$ so that $\sum_n |R_n| \leq \lambda(A) + \epsilon$. Find large enough N so that $\sum_{n>N} |R_n| < \epsilon$ and set $B = R_1 \cup \ldots \cup R_N$. Then $\lambda(A \Delta B) < 2\epsilon$

and B is an elementary set (even if the R_i are not disjoint, we can find a different set of pairwise disjoint rectangles by taking intersections).

107. For the second statement, again find a covering $\{R_n\}$ for A so that $\sum_n |R_n| \le (1-\epsilon)^{-1}\lambda(A)$ (valid even if $\lambda(A) = \infty$). But then $\lambda(A) = \sum_n \lambda(A \cap R_n)$, which forces that there must be at least one n for which $|R_n| \le (1-\epsilon)^{-1}\lambda(A \cap R_n)$.

108. The third statement is known as Steinhaus' lemma. Observe that $x \in A - A$ if and only if $A \cap (A + x) \neq \emptyset$. A set cannot be empty if it has positive measure. To show that $A \cap A + x$ has positive emasure (if x is close enough to the origin), find R as in the second step so that $\lambda(A') \ge 0.9|R|$ where $A' = A \cap R$. But then $\lambda(A' \cap (A' + x)) \ge \lambda(A') + \lambda(A' + x) - \lambda(A' \cup (A' + x))$ which is at least $1.8|R| - |R \cup (R + x)|$. But if R has sides $\ell_1, \ldots \ell_d$, then $R \cup (R + x)$ is contained in a rectangle of lengths $\ell_i + x_i$. Hence if x is small enough, $|R \cup (R + x)| \le 1.1 \times |R|$. Therefore, for such x, $\lambda(A' \cap (A' + x)) > 0$, showing that $A \cap (A + x) \neq \emptyset$.

11. Non-measurable sets

109. For the purposes of the course, it suffices to read the first construction.

110. Non-measurable set: Consider the group $G = [0, 1)$ with addition modulo 1 (thus $0.3 + 0.8 = 0.1$). Consider the subgroup $H = \mathbb{Q} \cap [0, 1)$ and create a set A having one element from each coset of H in G. This is possible if one assumes the *axiom of choice*, as we do.

111. We claim that $\sqcup_{r \in H}(A + r) = [0, 1)$.

If $x \in (A+r) \cap (A+s)$, then $x - r \in A$ and $x - s \in A$, but $x - r$ and $x - s$ are in the same coset of H (since their difference is $r - s$ which is in H). Hence we must have $x - r = x - s$ or equivalently that $r = s$. This shows that $(A + r) \cap (A + s) = \emptyset$ for $r \neq s$.

Next, if $x \in [0, 1)$, there is some $a \in A$ in the same coset of H as x, which means that $x = a + r$ for some $r \in H$. Thus, the union of $A + r$, $r \in H$, is the whole of $[0, 1)$.

112. Now if A is measurable, then so is $A + r$ and $\lambda(A + r) = \lambda(A)$ (observe that this is true despite the meaning of '+' as addition modulo 1). But then either $\lambda(A) > 0$ or $\lambda(A) = 0$. In either case, countable additivity requirement $\lambda([0, 1)) = \sum_{r \in H} \lambda(A + r)$ is violated. The contradiction shows that A cannot be measurable.

113. We now show the stronger statement that there is some A, B such that $A \sqcup B = [0, 1)$ and such that $\lambda^*(A) = \lambda^*(B) = 1$. This of course shows that A and B must be non-measurable, but more strongly it shows that λ^* is not even finitely additive (a claim we had made but not justified so far).

114. The proof is similar to the previous one, except that we change the sub-group to $H = \{n\alpha : n \in \mathbb{Z}\}$, where α is a fixed irrational, e.g., $\alpha = 1/\sqrt{2}$. Here again $n\alpha$ is interpreted modulo

1. We leave it as an exercise to check that H is dense in $[0,1)$, and that $n\alpha \neq n\beta$ (mod 1) if $n \neq m$. Because of this, the map $n \mapsto n\alpha$ is an isomorphism from $\mathbb{Z}$ onto H.

115. The difference from rationals is that this subgroup has finite index subgroups. In particular, let $H' = \{2n\alpha : n \in \mathbb{Z}\}$. Then H' is a subgroup of H with two cosets, H' and $H'' = H' + \alpha$. Observe that both H' and H'' are dense in $[0,1)$ (for similar reasons why H is).

116. As before, create a set A by picking one element from each coset of H in G. Then $[0,1) = \sqcup_{r\in H}(A+r) = B \sqcup C$ where $B = \sqcup_{r\in H'}(A+r)$ and $C = \sqcup_{r\in H''}(A+r)$. We claim that $\lambda^*(B) = \lambda^*(C) = 1$, which finishes the proof.

117. Any element of $B - B$ is of the form $z = a + r - a' - s$ where $a, a' \in A$ and $r, s \in H'$. If $a = a'$, then this element is $r - s$ which is in H' (as H') is a subgroup and in particular, $z \notin H''$. But if $a \neq a'$, then $a - a' \notin H$ and $r - s \in H'$, hence again $z \notin H''$ (in fact $z \notin H$). In short, we have proved that $(B - B) \cap H'' = \emptyset$. As H'" is dense, this shows that $B - B$ does not contain any interval. Similarly $C - C$ also does not contain any interval.

118. Now if $\lambda^*(B) < 1$, then find intervals $\{I_n\}$ such that $\cup_n I_n \supseteq B$ and $\sum_n |I_n| < 1$. Then $X = (\cup_n I_n)^c$ (complement inside $[0,1)$) is a measurable set of positive measure and further, $X \subseteq C$. But then $C - C \supseteq X - X$, and by Steinhaus' lemma the latter contains an interval around 0, contradicting that $C - C$ does not contain any interval. Thus we must have $\lambda^*(B) = 1$ and similarly $\lambda^*(C) = 1$.

119. A third construction of a nonmeasurable set due to Sierpinski is outlined in the problem set. There, the idea is to (a) regard $\mathbb{R}$ as a vector space over $\mathbb{Q}$, (b) pick a basis B that is contained inside the standard Cantor set, (c) define $E_0 = B \sqcup (-B) \sqcup \{0\}$ and $E_n = E_{n-1} - E_{n-1}$ for $n \geq 1$.

Then argue that there is a first n for which $\lambda^*(E_n) > 0$ and show that this particular E_n is not measurable (again using Steinhaus' lemma).

12. Radon measures on $\mathbb{R}^d$

120. What about other Borel measures on $\mathbb{R}$? For simplicity let us consider a finite measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then we can associate to μ a function $F_\mu : \mathbb{R}^d \mapsto \mathbb{R}$ by defining $F_\mu(x) = \mu(R_x)$ where $R_x = (-\infty, x_1] \times \ldots \times (-\infty, x_d]$. It is called the *cumulative distribution function* of μ.

121. F_μ is (A) increasing in each co-ordinate, (B) is right-continuous, (C) $F_\mu(x) \to \mu(\mathbb{R}^d)$ if $\min\{x_1, \ldots, x_d\} \to \infty$ and $F_\mu(x) \to 0$ if $\min\{x_1, \ldots, x_d\} \to -\infty$.

These properties follow from the definition of a measure. For example, if $x_n \to x$ from the right (meaning that each co-ordinate of x_n decreases to the corresponding co-ordinate of x), then

$R_{x_n} \downarrow R_x$. Therefore, $\mu(R_{x_n}) \downarrow \mu(R_x)$ (as we assumed that μ is a finite measure). This shows that F_μ is continuous from the right. Other properties are similar and left as exercise.

122. Conversely let $F : \mathbb{R}^d \mapsto [0, \infty)$ be any function that is increasing in each co-ordinate, is right-continuous, $F_\mu(x) \to 0$ if $\min\{x_1, \ldots, x_d\} \to -\infty$ and $\lim F_\mu(x) < \infty$ as $\min\{x_1, \ldots, x_d\} \uparrow \infty$. Then we claim that $F = F_\mu$ for a unique finite Borel measure μ on $\mathbb{R}^d$.

123. The uniqueness is clear from the fact that rectangles form a π-system that generates the Borel sigma-algebra, hence two measures that agree on rectangles must be equal. For existence, start by defining the measure of any lorc rectangle using F (explained below) and extend it naturally to the algebra $\mathcal{A}$. This function has countable additivity (needs checking, the compact class criterion that we gave earlier helps in reducing the work involved in checking), and therefore extends to a measure μ on $\mathcal{B}(\mathbb{R}^d)$. Of course F is the cumulative distribution function of μ.

124. How to define μ on rectangles? In one dimension, if $R = (a, b]$, then we define $\mu(R) = F(b) - F(a)$. In $d = 2$, if $R = (a_1, b_1] \times (a_2, b_2]$, we define $\mu(R) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$. Similarly, in higher dimensions, the formula (coming from inclusion-exclusion) is $F(R) = \sum \pm F(a_1^{\pm}, \ldots a_d^{\pm})$ where $R = (a_1^-, a_1^+] \times \ldots \times (a_d^-, a_d^+]$ and the sum is over all 2^d choices of $\pm$, and the sign in front of $F(a_1^{\epsilon_1}, \ldots a_d^{\epsilon_d})$ is negative is the product of $\epsilon_1, \ldots, \epsilon_d$.

125. Thus finite Borel measures on $\mathbb{R}^d$ and cumulative distribution functions are in one-one correspondence with each other. As CDFs are easier to understand, this gives us a good understanding of the collection of all finite Borel measures on $\mathbb{R}^d$. One can extend the notion of CDF to infinite measures too, provided they are Radon (i.e., $\mu(K) < \infty$ for all compact K), but we do not bother about that here (just as one example, we can take the function $F(x) = x$ on $\mathbb{R}$ as the CDF of Lebesgue measure, since $\mu(a, b] = F(b) - F(a)$ for all $a < b$). Instead let us see a few examples.

126. Suppose $F(x) = 0$ for $x \le 0$, $F(x) = x$ for $0 \le x \le 1$ and $F(x) = 1$ for $x \ge 1$. It is easy to see that the corresponding measure is $\mu = \lambda\big|_{[0,1]}$, Lebesgue measure restricted to $[0, 1]$. By this we just mean that $\mu(A) = \lambda(A \cap [0, 1])$ for $A \in \mathcal{B}_{\mathbb{R}}$.

127. Let $S = \{a_1, a_2, \ldots\}$ be any countable set and let $p_i > 0$ be such that $\sum_i p_i < \infty$. Define $F(x) = \sum_{i: a_i \le x} p_i$. This is the CDF of the "discrete measure" $\mu = \sum_i p_i \delta_{a_i}$. That is, $\mu(A) = \sum_{i: A \ni a_i} p_i$. Observe that the CDF here has jumps of magnitude p_i at location a_i. It is not left continuous at these points.

128. Let $f : \mathbb{R} \mapsto [0, \infty)$ be a continuous (or even piecewise continuous) function such that $\int_{-\infty}^{\infty} f(x)dx < \infty$. One example is $f(x) = e^{-x^2/2}$. Then define $F(x) = \int_{-\infty}^{x} f(t)dt$. Here we are using that (improper?) Riemann integral. It is easy to see that F is increasing (because $f > 0$) and continuous (in fact, by the fundamental theorem of Calculus, F is differentiable and $F' = f$). Also $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to \int_{-\infty}^{\infty} f(t)dt$ as $x \to +\infty$. Thus F is a continuous CDF, and corresponds to a finite Borel measure μ.

This gives us innumerable examples of measures on $\mathbb{R}$.

129. Let K be the standard Cantor set and let K_n be the n-th level (consisting of a union of 2^n intervals of length 3^{-n} each). Let $\mu_n = (3/2)^n \lambda\big|_{K_n}$ (the Lebesgue measure on K_n rescaled to have total mass 1). Write down the first few F_{μ_n} and show that they converge uniformly (this was done in detail in class) by checking the Cauchy criterion (for example, check that $\sup_{x\in\mathbb{R}} |F_{\mu_n}(x) - F_{\mu_{n+1}}(x)| \le 2^{-n}$). Let F denote the limiting function. Clearly F is non-decreasing and continuous, and further, $F(x) = 0$ for $x \le 0$ and $F(x) = 1$ for $x \ge 1$. Thus, there is a finite Borel measure μ with CDF F. The measure μ with distribution function F is called the Cantor measure.

130. The Cantor measure is a somewhat curious object that bears thinking about. Observe that as F is continuous, μ has no atoms. We now claim that $\mu(K^c) = 0$, i.e., μ "sits on" the Cantor set. To see this, observe that $K^c = \cup_n K_n^c$ is a disjoint union of the intervals $(-\infty, 0), (1, \infty), (\frac{1}{3}, \frac{2}{3})$, $(\frac{1}{9}, \frac{2}{9}), (\frac{7}{9}, \frac{8}{9})$,... Take any one of these intervals, say $J = ((2k-1)3^{-n}, 2k3^{-n})$. Then F_m is constant on J for $m \ge n$, hence F is constant on J. This shows that $\mu(J) = 0$. Thus, $\mu(K^c) = 0$ by countable additivity.

Thus, μ allows us to measure the relative sizes of sets inside the Cantor set.

131. A remark may clarify some of the above examples. If $(X, \mathcal{F}, \mu)$ is any measure space and $B \in \mathcal{F}$, then we can define a new measure space $(B, \mathcal{G}, \nu)$ by defining $\mathcal{G} = \{A \in \mathcal{F} : A \subseteq B\}$ (we may also write $\mathcal{G} = \{A \cap B : A \in \mathcal{F}\}$) and $\nu(A) = \mu(A)$ for $A \in \mathcal{G}$. We leave it as a simple thing to check that this is indeed a measure space. However, ν is a non-trivial measure if and only if $\mu(B) > 0$. In that case, we say that ν is the restriction of μ to B and write $\nu = \mu\big|_B$.

This is what we did to construct Lebesgue measure on $[0, 1]$ and also to define the measures μ_n in the construction of Cantor measure (μ_n was just a rescaling of $\lambda\big|_{K_n}$). However, we cannot define Cantor measure directly like this, as $\lambda(K) = 0$, hence the more indirect approach via appropriate limits. For those with inclination towards probability, this is akin to the difficulty of conditioning on zero probability events.

13. Measurable functions

132. Let $(X, \mathcal{F})$ and $(Y, \mathcal{G})$ be measurable spaces. A function $T : X \mapsto Y$ is said to be *measurable* (w.r.t. $\mathcal{F}$ and $\mathcal{G}$, although the sigma algebras will be usually not explicitly mentioned unless necessary) if $T^{-1}B := \{x \in X : T(x) \in B\} \in \mathcal{F}$ for any $B \in \mathcal{G}$.

133. Observe that T^{-1} need not exist as a function. The meaning of $T^{-1}(B)$ for $B \subseteq Y$ is that of the inverse-image of B. It is easy to check that $T^{-1}(B^c) = (T^{-1}(B))^c$ (the complements are in

Y and X, respectively), $T^{-1}(\bigcup_n B_n) = \bigcup_n T^{-1}(B_n)$. The last point is valid even for uncountable unions.

134. If $(X, \mathcal{F})$, $(Y, \mathcal{G})$ and $(Z, \mathcal{H})$ are measurable spaces and $T : X \mapsto Y$ and $S : Y \mapsto Z$ are measurable, then so is $S \circ T : X \mapsto Z$. This is because $(S \circ T)^{-1}(C) = T^{-1}(S^{-1}C)$.

135. Making $\mathcal{F}$ larger makes it easier for T to be measurable. In particular, if $\mathcal{F} = 2^X$, then every function $T : X \mapsto Y$ is measurable. Similarly, making $\mathcal{G}$ smaller helps T to be measurable. In particular, if $\mathcal{G} = \{\emptyset, Y\}$, then every $T : X \mapsto Y$ is measurable.

136. Let $\sigma(T) := \{T^{-1}B : B \in \mathcal{G}\}$. Then $\sigma(T)$ is a sigma-algebra on X (from the earlier observation about inverse-images of complements and unions) and is called the *sigma algebra generated by* T. In terms of this sigma-algebra, T is measurable w.r.t. $\mathcal{F}$ and $\mathcal{G}$ if and only if $\sigma(T) \subseteq \mathcal{F}$. In other words, $\sigma(T)$ is the smallest sigma-algebra on X that makes T a measurable function.

137. The collection $\mathcal{G}_0 := \{B \subseteq Y : T^{-1}B \in \mathcal{F}\}$ is a sigma-algebra (easy to check) and hence so is $\mathcal{G}_0 \cap \mathcal{G} = \{B \in \mathcal{G} : T^{-1}B \in \mathcal{F}\}$. Therefore, if $\mathcal{S}$ is a collection (of subsets of Y) that generates $\mathcal{G}$, and $T^{-1}B \in \mathcal{F}$ for $B \in \mathcal{S}$, then $\mathcal{G}_0 \supseteq \mathcal{S}$ and hence $\mathcal{G}_0 \supseteq \mathcal{G}$. In other words, to check measurability of T, it suffices to check the condition $T^{-1}B \in \mathcal{F}$ only for $B \in \mathcal{S}$. This is helpful in practise.

138. Most important for us will be measurable functions into $\mathbb{R}$, in which case the latter is always assumed to be endowed with the Borel sigma-algebra. In short, $T : X \mapsto \mathbb{R}$ is measurable if $T^{-1}B \in \mathcal{F}$ for any $B \in \mathcal{B}_{\mathbb{R}}$. Observe that if X itself is $\mathbb{R}$, we may take $\mathcal{F}$ to be either the Borel or the Lebesgue sigma-algebra, but on the target space it is always the Borel sigma-algebra.

139. For $A \subseteq X$, its indicator function is $1_A : X \mapsto \mathbb{R}$ defined as $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ if $x \notin A$. Then, $\mathbf{1}_A$ is a measurable function if and only if A is a measurable set (i.e., $A \in \mathcal{F}$). Thus, if we identify sets with their indicator functions, then measurable functions are a generalization of measurable sets. Another way to say the same is that an indicator function is a binary measurement (an answer to a yes-no question) while a general function is a finer measurement.

140. If $T : \mathbb{R} \mapsto \mathbb{R}$ is continuous, then $T^{-1}G$ is open if G is open. As open sets generate the Borel sigma-algebra, it follows that $T : \mathbb{R} \mapsto \mathbb{R}$ is measurable (w.r.t. Borel sigma algebra on both the domain and co-domain). Same if $T : X \mapsto \mathbb{R}$ if X is a metric space endowed with the Borel sigma-algebra.

141. Of course, we can take a larger sigma-algebra such as the Lebesgue sigma-algebra on the domain without losing measurability. However, if we take Lebesgue sigma-algebra on both the domain and co-domain, then there are continuous functions that are not measurable! This is not obvious, but an example is the inverse of the strictly increasing function $x \mapsto (x + F(x))/2$ from $[0, 1] \mapsto [0, 1]$ (one can extend it to $\mathbb{R}$, but the essence of the matter is the same), where F is the CDF of the Cantor measure. We relegate the details to the problem set.

142. As $\{(-\infty, t] : t \in \mathbb{R}\}$ generates $\mathcal{B}_{\mathbb{R}}$, we see that $T : X \mapsto \mathbb{R}$ is measurable if and only if $\{T \leq t\} \in \mathcal{F}$ for all $t \in \mathbb{R}$. Here $\{T \leq t\}$ is a short-form for $\{x \in X : T(x) \leq t\}$. For exactly analogous reasons, each of the following conditions is equivalent to measurability of T: (a) $\{T < t\} \in \mathcal{F}$ for any $t \in \mathbb{R}$, (b) $\{T \geq t\} \in \mathcal{F}$ for any $t \in \mathbb{R}$, (c) $\{T > t\} \in \mathcal{F}$ for any $t \in \mathbb{R}$.

143. It follows immediately from this that if $T : \mathbb{R} \mapsto \mathbb{R}$ is increasing, then it is measurable (with Borel sigma algebra on both domain and co-domain). This is because $\{T \leq t\}$ is necessarily an interval (of the form $(-\infty, t]$ or $(-\infty, t)$ or $\emptyset$ or $\mathbb{R}$). Of course, decreasing functions are also measurable.

144. We leave it as an exercise to check that if $T : \mathbb{R} \mapsto \mathbb{R}$ is left-continuous, then it is measurable (same for right-continuous, of course). Similarly, upper semi-continuous and lower semi-continuous functions are measurable.

145. If $A \subseteq \mathbb{R}$ is a non-measurable set, then $1_A : \mathbb{R} \mapsto \mathbb{R}$ is not a measurable function (even if we take the Lebesgue sigma algebra on the domain). But as the difficulty of constructing non-measurable sets suggests, it is not easy to construct non-measurable functions. This is further demonstrated by the fact that the collection of measurable functions is closed under various (countable) operations.

146. Let $(X, \mathcal{F})$ be a measurable space. If $T_n : X \mapsto \mathbb{R}$ are measurable, then so are

$$\sup_n T_n, \quad \inf_n T_n, \quad \limsup_n T_n, \quad \liminf T_n, \quad \lim_n T_n, \quad \sum_n a_n T_n \ \ (\text{where } a_n \in \mathbb{R})$$

provided they exist and are finite (otherwise, they are not functions from X to $\mathbb{R}$). Countable operations include finite ones, hence $\max\{T_1, T_2\}$, $\min\{T_1, T_2\}$, $a_1 T_1 + a_2 T_2$, are measurable.

147. Proofs of the claim: If $T = \sup_n T_n$ is finite, then $\{T \leq t\} = \cap_n \{T_n \leq t\}$ is in $\mathcal{F}$. Thus $\sup_n T_n$ is measurable. Similarly $\{\inf_n T_n \geq t\} = \cap_n \{T_n \geq t\}$. If $T = \limsup_n T_n$, then $\{T < t\} = \cap_{j \geq 1} \cup_{n \geq 1} \cap_{k \geq n} \{T_k < t - \frac{1}{j}\}$ (pay attention to the strict inequality - the equality is false if we write $\leq t$ instead of $< t$) showing that $\limsup_n T_n$ is measurable. Similarly $\liminf_n T_n$ is measurable. If $\lim_n T_n$ exists, it is same as $\limsup_n T_n$, so it is measurable. We can write $\sum_n a_n T_n$ as a limit of finite sums, hence it suffices to show that $T_1 + T_2$ is measurable. To see that, we write $\{T_1 + T_2 < t\} = \bigcap_{s \in \mathbb{Q}} \{T_1 < s\} \cap \{T_2 < t - s\}$ (again, the strict inequality is crucial).

148. So far we have not had to bring in measure in discussing measurable functions. When we are working in a measure space $(X, \mathcal{F}, \mu)$, we shall see that it often does not matter what happens on a set of zero measure. For example, if $\sup_n T_n(x) < \infty$ for $x \in X \setminus A$ where $\mu(A) = 0$, then it is just as good as if $\sup_n T_n(x) < \infty$ for all x.

149. In fact it is so common that most of our statements look like "Property P holds for a.e. x" or "P holds for x a.s. $[\mu]$", where a.e (almost every) and a.s. (almost surely) indicate that the given property is true for all x outside of a set of zero measure w.r.t. μ. For example, $f = g$ *a.e.* means that $\mu\{f \neq g\} = 0$; $f_n \to f$ *a.s.* means that $\mu\{x : f_n(x) \not\to f(x) \text{ as } n \to \infty\} = 0$ and so on.

150. Returning to the fact that we don't mind if one of our functions takes infinite values on zero measure sets, and that does occur often when taking limits or supremum etc., it is convenient to consider measurable functions taking values in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. But what is the Borel sigma-algebra on $\overline{\mathbb{R}}$? The metric on $\overline{\mathbb{R}}$ is $d(x,y) = |x-y|/(1+|x-y|)$ where $|x - \infty| = \infty$ if $x \neq +\infty$ and $|x-(-\infty)| = \infty$ if $x \neq -\infty$ and the convention is the $\infty/\infty = 1$. Another way to say this is that $\overline{\mathbb{R}}$ is homeomorphic to $[-1,1]$ with the usual metric (for homeomorphism we have many choices, e.g., $x \mapsto x/(1+|x|)$ or $x \mapsto \frac{2}{\pi}\arctan(x)$ with the obvious definitions for $x = \pm\infty$).

151. The open intervals in $\overline{\mathbb{R}}$ include all open intervals in $\mathbb{R}$ as well as the sets $[-\infty, t)$ and $(t, \infty]$ and $[-\infty, \infty]$. These open intervals generate the Borel sigma-algebra of $\overline{\mathbb{R}}$. Hence, to check that $T : X \mapsto \overline{\mathbb{R}}$ is measurable, it suffices to check that $\{T < t\} \in \mathcal{F}$ for any $t \in \mathbb{R}$. Then $\{T = -\infty\} = \cap_n \{T < -n\}$ and $\{T = +\infty\} = (\cup_n \{T < n\})^c$ are also measurable, and then it is easy to see that the pre-images of other open sets are measurable too (e.g., $\{-\infty < T < t\} = \{T < t\} \setminus \{T = -\infty\}$).

Just as with $\mathbb{R}$-valued functions, instead of $\{T < t\} \in \mathcal{F}$ for $t \in \mathbb{R}$, one can instead check that $\{T \le t\} \in \mathcal{F}$ or that $\{T \ge t\} \in \mathcal{F}$ or that $\{T > t\} \in \mathcal{F}$.

152. A function taking values in $\mathbb{R}$ can also be viewed as taking values in $\overline{\mathbb{R}}$, and whether or not it is measurable does not change which view point we take. Henceforth, whenever we say real-valued measurable function, we shall mean $\overline{\mathbb{R}}$-valued measurable function, unless we say otherwise.

14. PUSH-FORWARD OF MEASURES BY MEASURABLE FUNCTIONS

153. Let $T : X \mapsto Y$ be a measurable function. For this section, the sigma algebra on X, Y, Z will be $\mathcal{F}, \mathcal{G}, \mathcal{H}$ respectively, and if any of them is a metric space (in particular a subset of $\mathbb{R}^d$), then the corresponding sigma-algebra will be the Borel sigma algebra. So far measures did not enter the picture.

154. Let μ be a measure on $(X, \mathcal{F})$. Then define $\nu : \mathcal{G} \mapsto [0, \infty]$ by $\nu(B) = \mu(T^{-1}B)$. If B_n are pairwise disjoint elements of $\mathcal{G}$, then $T^{-1}B_n$ are pairwise disjoint and elements of $\mathcal{F}$ (as T is measurable), hence $\nu(\sqcup_n B_n) = \mu(\sqcup_n T^{-1}B_n) = \sum_n \mu(T^{-1}B_n) = \sum_n \nu(B_n)$, showing that ν is a measure on $(Y, \mathcal{G})$.

155. ν is called the *push-forward* of μ under T, and denoted $\nu = \mu \circ T^{-1}$. In general, there is no such thing as a *pull-back* measure. Convince yourself why. If we construct one measure, we can get many more by pushing it forward under various measurable mappings into various spaces. In fact, it turns out that pretty much every measure of interest is push-forward of Lebesgue measure on $\mathbb{R}$.

156. That would mean that there is no need to go through the Carathéodory extension again and again, it is enough to have done it once to construct Lebesgue measure. This is not entirely correct, as it is sometimes not easy to find the measurable function with which to push-forward λ_1 to get a specified target measure.

157. Here is an example that may seem surprising. Define $B_n : [0,1] \mapsto [0,1]$ by $B_n(x) = \lfloor 2^n x \rfloor$ (this is the nth digit in the binary expansion of x, i.e., $x = \sum_n B_n(x)2^{-n}$). Then define $T_1(x) = \sum_{n\geq 1} B_{2n}(x)2^{-n}$ and $T_2(x) = \sum_{n\geq 1} B_{2n-1}(x)2^{-n}$ and set $T(x) = (T_1(x), T_2(x))$. Then $T : [0,1] \mapsto [0,1]^2$ and we claim that $\lambda_1 \circ T^{-1} = \lambda_2$ (here λ_d denotes Lebesgue measure restricted to $[0,1]^d$). In a very similar way, $T : [0,1] \mapsto [0,1]^d$ defined by $T(x) = (T_1(x), \ldots, T_d(x))$ where $T_i(x) = \sum_{n\geq 0} B_{nd+i}(x)2^{-n}$, is measurable and $\lambda_1 \circ T^{-1} = \lambda_d$.

158. The point of this example is that dimension is irrelevant in measure theory. All the measure spaces $([0,1]^d, \mathcal{B}([0,1]^d), \lambda_d)$ are *isomorphic* in the sense that there are measurable transformations from $[0,1]^k$ to $[0,1]^\ell$ that pushes forward λ_k to λ_ℓ. It is no surprise to anyone for $k \geq \ell$, but that it is true for $k < \ell$ is. Contrast it with theorems in Topology/Geometry class that assert that $\mathbb{R}$ is not homeomorphic/diffeomorphic to $\mathbb{R}^2$ or with the fact that $\mathbb{R}$ and $\mathbb{R}^2$ are not isomorphic as vector spaces.

159. The proof of the claim above is not difficult. Let us restrict to $d = 2$. Let $B = (k2^{-n}, (k+1)2^{-n}] \times (\ell 2^{-n}, (\ell+1)2^{-n}]$ for some $0 \leq k, \ell \leq 2^n - 1$. What is $T^{-1}(B)$?

15. Lebesgue integration

160. Let $(X, \mathcal{F}, \mu)$ be a measure space. The goal is to define a notion of integral for a large class of functions $f : X \mapsto \mathbb{R}$. For the measure space $(\mathbb{R}, \mathcal{L}, \lambda)$, it will turn out to be a generalisation of the Riemann integral (i.e., Riemann integrable functions will turn out to be Lebesgue integrable and the values of the integral in the two notions agree). The Lebesgue integral will turn out to be good in the sense that we shall never need any more general notion.

161. What are the requirements out of the integral that we are about to construct? The integral of a $I(f)$ of a real-valued function f (if defined) will be a real number. The association of integral to a function must be linear ($I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$) and positive ($I(f) \geq 0$ if $f \geq 0$ pointwise). The dependence of the integral on the measure is captured in the requirement that $I(\mathbf{1}_A) = \mu(A)$ for $A \in \mathcal{F}$.

162. We do not make any effort to justify these requirements but mention them to keep the end-goal in sight. However, note that the Riemann integral also has linearity and positivity. Later we shall see that in fact any linear and postive functional on a large class of functions is an integral with respect to a measure!

163. A function $f : X \mapsto \mathbb{R}$ is said to be *simple* if it has finite range. If $a_1, \ldots, a_n$ are the distinct elements in the range, and $A_k = f^{-1}\{a_k\}$, then $\{A_1, \ldots, A_n\}$ is a partition of X (i.e., $X = A_1 \sqcup A_2 \sqcup \ldots \sqcup A_n$) and $f = a_1\mathbf{1}_{A_1} + \ldots + a_n\mathbf{1}_{A_n}$. Observe that f is measurable if and only if $A_k \in \mathcal{F}$ for each k.

It is worth noting that any finite linear combination of indicators, $g = b_1\mathbf{1}_{B_1} + \ldots + b_m\mathbf{1}_{B_m}$ is a simple function. Its canonical form as above is $g = \sum_S a_S \mathbf{1}_{A_S}$ where the sum is over subsets of $\{1, 2, \ldots, m\}$ and $a_S = \sum_{i\in S} b_i$ and $A_S = \cap_{i\in S} B_i \cap_{i\notin S} B_i^c$. The empty sum is 0, and the terms with $A_S = \emptyset$ are to be dropped from the sum. For example, if $m = 2$, then the canonical form of g is $b_1\mathbf{1}_{B_1\cap B_2^c} + b_2\mathbf{1}_{B_1^c\cap B_2} + (b_1+b_2)\mathbf{1}_{B_1\cap B_2} + 0\mathbf{1}_{B_1^c\cap B_2^c}$.

164. The definition of integral will proceed in three steps.

(1) If $f : X \mapsto \mathbb{R}_+$ is non-negative, simple and measurable with canonical representation $f = a_1\mathbf{1}_{A_1} + \ldots + a_n\mathbf{1}_{A_n}$, then define $I_1(f) := a_1\mu(A_1) + \ldots + a_n\mu(A_n)$. Then $I_1(f) \in [0, +\infty]$. Note that measures of sets can be infinite, which is why we don't define for general simple functions (we may encounter $\infty - \infty$).

(2) If $f : X \mapsto \mathbb{R}_+$ is non-negative and measurable, define $I_2(f) = \sup\{I_1(\varphi) : 0 \le \varphi \le f,\ \varphi \text{ simple, measurable}\}$.

(3) If $f : X \mapsto \mathbb{R}$ is measurable, let $f_+ = f \vee 0$ and $f_- = (-f)_+ = -(f \wedge 0)$. Both f_+ and f_- and non-negative and measurable, $f_+ - f_- = f$ and $f_+ + f_- = |f|$. If $I_2(f_+) < \infty$ and $I_2(f_-) < \infty$, then we say that f is integrable and define $I_3(f) = I_2(f_+) - I_2(f_-)$.

    If exactly one of $I_2(f_+)$ and $I_2(f_-)$ is equal to $+\infty$, then we define $I_3(f)$ to be $\pm\infty$ accordingly, but we do not say that f is integrable.

If f is non-negative, simple, measurable, then $I_1(f) = I_2(f) = I_3(f)$ and if f is non-negative and measurable, then $I_2(f) = I_3(f)$ (this needs justification which will be provided later). Hence it is legitimate to use a common notation for I_1, I_2, I_3. The standard notation is $\int_X f d\mu$ (also written as $\int_X f(t)d\mu(t)$ with a dummy variable t that has no meaning) and read as the (Lebesgue) integral of f w.r.t. μ. Our first goal is the following theorem.

165. Theorem: The space of integrable functions is a vector space. Further, (1) $\int_X (\alpha f + \beta g)d\mu = \alpha \int_X f d\mu + \beta \int_X g d\mu$ for any f, g integrable and $\alpha, \beta \in \mathbb{R}$. (2) $\int_X f d\mu \ge 0$ if $f \ge 0$ pointwise and is measurable, and the inequality is strict unless $f = 0\ a.s.[\mu]$.

Further, f is integrable if and only if $|f|$ is integrable and then $\left|\int_X f d\mu\right| \le \int_X |f| d\mu$.

In words, the integral is linear and positive. A consequence of these is monotonicity: $\int_X f d\mu \le \int_X g d\mu$ if $f \le g$ pointwise and both are integrable.

166. To prove this theorem, we must prove similar properties at each step - for I_1, then for I_2 and then for I_3. The part that is most non-trivial is the additivity f I_2: If f, g are non-negative measurable, so if $h = f + g$. From the definition of I_2 it is easy to deduce that $I_2(h) \ge I_2(f) + I_2(g)$. The other way inequality is what requires work and will also lead us to one of the appealing features of Lebesgue integration theory, namely the clean limit theorems that allow easy interchange of limits and integrals (which leads to similar theorems about interchange of two integrals, of a sum with an integral, of differentiation under an integral and so on).

167. Proposition 1: Let f, g be non-negative, simple, measurable. Then so is $\alpha f + \beta g$ for $\alpha, \beta \geq 0$, and $I_1(\alpha f + \beta g) = \alpha I_1(f) + \beta I_1(g)$. If $f \leq g$ pointwise then $I_1(f) \leq I_1(g)$ with equality if and only if $f = g \; a.s.[\mu]$.

168. Before proving the proposition, it is worth observing that if $g = b_1 \mathbf{1}_{B_1} + \ldots + b_m \mathbf{1}_{B_m}$ is a simple function (not necessarily in canonical form), then it is true that $I_1(g) = b_1 \mu(B_1) + \ldots + b_m \mu(B_m)$. Indeed, we have seen that the canonical form of g is $\sum_S \left(\sum_{i \in S} b_i\right) \mathbf{1}_{A_S}$ where $A_S = \cap_{i \in S} B_i \cap_{i \notin S} B_i^c$. Hence,

$$I_1(g) \;=\; \sum_S \left(\sum_{i \in S} b_i\right) \mu(A_S) \;=\; \sum_{k=1}^m b_k \sum_{S : S \ni k} \mu(A_S) \;=\; \sum_{k=1}^m b_k \mu(B_k)$$

since $\sqcup_{S:S \ni k} A_S = B_k$.

169. *Proof of Proposition 1:* Therefore, if $f = a_1 \mathbf{1}_{A_1} + \ldots + a_n \mathbf{1}_{A_n}$ and $g = b_1 \mathbf{1}_{B_1} + \ldots + b_m \mathbf{1}_{B_m}$ where $A_i, B_j \in \mathcal{F}$ and $a_i, b_j \geq 0$, then $\alpha f + \beta g = \alpha a_1 \mathbf{1}_{A_1} + \ldots + \alpha a_n \mathbf{1}_{A_n} + \beta b_1 \mathbf{1}_{B_1} + \ldots + \beta b_m \mathbf{1}_{B_m}$. From the previous step, we can write $I_1(f), I_1(g), I_1(h)$ without having to worry about their canonical forms and see that $I_1(h) = \alpha I_1(f) + \beta I_1(g)$.

Next, if $f \leq g$, then we can find a common partition and represent them as $f = \alpha_1 \mathbf{1}_{C_1} + \ldots + \alpha_k \mathbf{1}_{C_k}$ and $g = \beta_1 \mathbf{1}_{C_1} + \ldots + \beta_k \mathbf{1}_{C_k}$. Then $f \leq g$ imples that $\alpha_j \leq \beta_j$ for each j. Consequently $I_1(f) = \sum_i \alpha_i \mu(C_i) \leq \sum_i \beta_i \mu(B_i) = I_1(g)$. Equality holds if and only if $\alpha_i = \beta_i$ for all i for which $\mu(C_i) > 0$. This is the same as $f = g \; a.s.[\mu]$.

170. MCT for simple functions: Suppose φ, φ_n, $n \geq 1$, are non-negative simple measurable functions on $(X, \mathcal{F}, \mu)$ such that $\varphi_n \uparrow \varphi$ pointwise. Then $I_1(\varphi_n) \uparrow I_1(\varphi)$.

171. *Proof:* By monotonicity we see that $I_1(\varphi_n)$ is increasing in n and $I_1(\varphi_n) \leq I_1(\varphi)$ for all n. Therefore $\lim_n I_1(\varphi_n)$ exists and is at most $I_1(\varphi)$. Let $\varphi = a_0 \mathbf{1}_{A_0} + a_1 \mathbf{1}_{A_1} + \ldots + a_m \mathbf{1}_{A_m}$ where $\{A_k\}$ form a measurable partition and $a_j > 0$ for $j \geq 1$. Let $A_{n,k} = \{\varphi_n > (1-\epsilon) a_k\}$ for $1 \leq k \leq n$. As $\varphi_n \uparrow \varphi$, it follows that $A_{n,k} \uparrow A_k$ for each $1 \leq k \leq m$ and hence $\mu(A_{n,k}) \uparrow \mu(A_k)$. Then $\varphi_n \geq \psi_n := (1-\epsilon)[a_1 \mathbf{1}_{A_{n,1}} + \ldots + a_m \mathbf{1}_{A_{n,m}}]$. Both φ_n, ψ_n are non-negative simple functions, hence by monotonicity $I_1(\varphi_n) \geq I_1(\psi_n) = (1-\epsilon)[a_1 \mu(A_{n_1} + \ldots a_m \mu(A_{n,m})]$ and the right hand quantity increases to $(1-\epsilon)I_1(\varphi)$ as $n \to \infty$. Thus $\lim_n I_1(\varphi_n) \geq (1-\epsilon)I_1(\varphi)$ for any $\epsilon > 0$, hence $\lim_n I_1(\varphi_n) \geq I_1(\varphi)$. ■

172. Proposition 2: Let f, g be non-negative, simple, measurable. Then so is $\alpha f + \beta g$ for $\alpha, \beta \geq 0$, and $I_2(\alpha f + \beta g) = \alpha I_2(f) + \beta I_2(g)$. If $f \leq g$ point wise then $I_2(f) \leq I_2(g)$ with equality if and only if $f = g\ a.s.[\mu]$. Further, $I_2(f) = I_1(f)$ if f is simple, non-negative and measurable.

173. The last part is a direct consequence of monotonicity of I_1. Indeed, if f is simple, and $0 \leq \varphi \leq f$, then $I_1(\varphi) \leq I_1(f)$, which shows that $I_2(f) \leq I_1(f)$. Further, the set of simple functions over which we take supremum contains f too, hence $I_1(f) \leq I_2(f)$. Thus, $I_2(f) = I_1(f)$.

174. The second part is easy. If $f \leq g$, then the supremum in the definition of $I_2(g)$ is over a larger class of non-negative simple measurable functions than the supremum in the definition of $I_2(f)$, hence $I_2(g) \geq I_2(f)$. If $A_\delta = \{g > f + \delta\}$ has $\mu(A_\delta) > 0$, then take any simple, non-negative measurable $\varphi \leq f$ and observe that $\psi = \varphi + \delta \mathbf{1}_{A_\delta} \leq g$, hence $I_2(g) \geq I_1(\varphi) + \delta\mu(A_\delta) > I_2(f)$. Thus, to have $I_2(f) = I_2(g)$, we must have $\mu(A_\delta) = 0$ for all $\delta > 0$. Take intersection over $\delta = 1/j$, $j \geq 1$, to get $\mu\{g > f\} = 0$.-

175. To prove the first part, it is obvious that $h = \alpha f + \beta g$ is non-negative and measurable. Let $0 \leq \varphi \leq f$ and $0 \leq \psi \leq g$ be simple and measurable such that $I_1(\varphi) \geq I_2(g) - \epsilon$ and $I_1(\psi) \geq I_2(g) - \epsilon$. Then, $0 \leq \alpha\varphi + \beta\psi \leq h$ is also simple and measurable, hence $I_2(h) \geq I_1(\alpha\varphi + \beta\psi) = \alpha I_1(\varphi) + \beta I_1(\psi) \geq \alpha I_1(f) + \beta I_1(g) - 2\epsilon$ (by Proposition 1). Thus $I_2(h) \geq \alpha I_1(f) + \beta I_1(g)$.

To prove the other way inequality suppose we can find simple, non-negative, measurable φ_n (respectively ψ_n) that increase to f (respectively g) point wise and set $\hat{\varphi}_n = \varphi_n \wedge \varphi$ and $\hat{\psi}_n = \psi_n \wedge \psi$ (both are again simple and measurable). Then $\hat{\varphi}_n \uparrow \varphi$ and $\hat{\psi}_n \uparrow \psi$ point wise. By the MCT for simple functions, $I_1(\hat{\varphi}_n) \uparrow I_1(\varphi)$ and $I_1(\hat{\psi}_n) \uparrow I_1(\psi)$ and $I_1(\alpha\hat{\varphi}_n + \beta\hat{\psi}_n) \uparrow I_1(\alpha\varphi + \beta\psi) \geq (1-\epsilon)[I_2(f) + \beta I_2(g)]$. But $\alpha\hat{\varphi}_n + \beta\hat{\psi}_n \leq h$, hence $I_2(h) \geq (1-\epsilon)[\alpha I_1(f) + \beta I_1(g)]$ for any $\epsilon > 0$. This proves the first part.

176. It still remains to show that for a given measurable $f \geq 0$, there exists simple, measurable $\varphi_n \geq 0$ such that $\varphi_n \uparrow f$ point wise. An explicit example is $\varphi_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbf{1}_{\frac{k}{2^n} \leq f < \frac{k+1}{2^n}} + n\mathbf{1}_{f \geq n}$.

In words, we partition $[0, \infty)$ into $[0, 2^{-n})$, $[2^{-n}, 2 \times 2^{-n})$, $\ldots [n - 2^{-n}, n)$ and $[n, \infty)$, and if the value of f falls in an interval, φ_n will take value equal to the lower end-point of that interval. Observe that each partition is a refinement of the previous one, which ensures that $\varphi_n \leq \varphi_{n+1}$. This works even if f takes the value $+\infty$.

177. Theorem: The space of integrable functions is a vector space. Further, (1) $I_3(\alpha f + \beta g) = \alpha I_3(f) + \beta I_3(g)$ for any f, g integrable and $\alpha, \beta \in \mathbb{R}$. (2) $I_3(f) \geq 0$ if $f \geq 0$ point wise and is measurable, and the inequality is strict unless $f = 0$ $a.s.[\mu]$. Further, $I_3(f) = I_2(f)$ if f is non-negative and measurable.

178. *Proof:* The last part is again obvious, since $f \geq 0$ implies that $f_+ = f$ and $f_- = 0$. The second part is a restatement of the last part of Proposition 2 (after we know that $I_3(f) = I_2(f)$ for $f \geq 0$).

For the remaining part, since $|f| = f_+ + f_-$, we see that a measurable f is integrable if and only if $|f|$ (a non-negative measurable function) has $I_2(|f|) < \infty$. Since $|\alpha f + \beta g| \leq |\alpha||f| + |\beta||g|$, this shows that if f and g are integrable, so is $\alpha f + \beta g$.

For the first part, let us first consider scalar multiplication. If $\alpha \geq 0$, then $(\alpha f)_\pm = \alpha f_\pm$ while if $\alpha < 0$ then $(\alpha f)_\pm = -\alpha f_\mp$. From this, it is easy to see that $I_3(\alpha f) = \alpha I_3(f)$. Now consider $h = f + g$. Write $I_3(f) = I_2(f_+) - I_2(f_-)$ and $I_3(g) = I_2(g_+) - I_2(g_-)$ and $I_3(h) = I_2(h_+) - I_2(h_-)$. Rearrange $h = f + g$ as $h_+ + f_- + g_- = h_- + f_+ + g_+$ and use Proposition 2 to deduce that $I_2(h_+) + I_2(f_-) + I_2(g_-) = I_2(h_-) + I_2(f_+) + I_2(g_+)$. Rearrange again to get $I_3(h) = I_3(f) + I_3(g)$. This proves the linearity. ■

179. This completes the construction and the most basic properties of Lebesgue integral. Henceforth we shall use the notation $\int_X f(t)d\mu(t)$ or $\int_X f d\mu$. One more property that we have not mentioned is that for integrable f,

$$|\int_X f d\mu| \leq \int_X |f| d\mu.$$

The proof is that the right side quantity is the sum of two non-negative numbers $\int_X f_+ d\mu$ and $\int_X f_- d\mu$, while the left side quantity is the absolute value of the difference between the same numbers.

16. Some remarks on Lebesgue integral

180. So far our measurable functions were taking values in $\mathbb{R}$. What if they take values in $\overline{\mathbb{R}}$. As remarked already, if f is takes values in $[0, \infty]$, then $I_2(f)$ can be defined exactly as before. But if $\mu\{f = +\infty\} > 0$, then necessarily $\int_X f d\mu = +\infty$ (the converse is not true). If we want a measurable function to be integrable, then we shall need $\mu\{f_+ = \infty\} = 0$ and $\mu\{f_- = \infty\} = 0$ or equivalently, $\mu\{f = \pm\infty\} = 0$. This is why, even though we allow the values $\pm\infty$, in most situations we only allow that on a set of zero measure. The convenience is that when we take limits of functions, we may get infinite values on sets of zero measure.

181. Turning it around, we can say that if a non-negative measurable function is integrable, then it is finite almost surely. This looks trivial but is useful in situations like the following: Suppose $f_n \geq 0$ are measurable and $\sum_n \int_X f_n d\mu < \infty$. Then $\sum_n f_n$ converges to a finite value $a.s.[\mu]$. This is because $f := \sum_n f_n$ is a non-negative measurable function taking values in $[0, +\infty]$ whose integral is finite. Hence the sum is finite $a.s.$ (convergence of the series is easy as partial sums are increasing, finiteness is the question).

In particular, if $\sum_k \mu(A_k) < \infty$, then $\sum_k \mathbf{1}_{A_k} < \infty \; a.s.[\mu]$. But $\sum_k \mathbf{1}_k(x) < \infty$ if and only if x belongs to only finitely many of the A_ks. Hence, $\mu(\limsup A_k) = 0$, where $\limsup A_k = \cap_k \cup_{n\geq k} A_n$ is the set of x that belong to infinitely many of the A_ns. This is known as *first Borel-Cantelli lemma*.

182. A quantitative extension of the previous remark is the simple but useful *Markov inequality:* If $f \geq 0$ is measurable, then $\int_X f d\mu \geq t\mu\{f \geq t\}$ for any $t > 0$. This follows by monotonicity of the integral applied to $f \geq t\mathbf{1}_{f\geq t}$.

183. Extending the previous point, if f, g are two measurable functions such that $f = g \; a.s.[\mu]$, then $\int_X |f - g| d\mu = 0$ and in particular, if one is integrable then so is the other and their integrals are equal. For all intents and purposes, they are the same function. This is made precise by declaring an equivalence relationship $f \sim g$ if $f = g \; a.s.[\mu]$. That it is reflexive ($f \sim f$) and symmetric ($f \sim g \implies g \sim f$) is clear. Transitivity follows because $\mu\{f \neq h\} \leq \mu\{f \neq g\} + \mu\{g \neq h\}$. The equivalence classes are often denoted $[f]$, where f is any representative of the class.

184. The operations of addition, scalar multiplication, product and almost-sure limits along sequences carry over to the equivalence classes. For this all one needs to check is that if $f_n \sim g_n$ then $\lim_n f_n = \lim_n g_n \; a.s.[\mu]$ (if the limits exist almost everywhere), that $\alpha f_1 + \beta f_2 \sim \alpha g_1 + \beta g_2$ $a.s.[\mu]$ and that $f_1 f_2 = g_1 g_2 \; a.s.[\mu]$. Thus, one can write $[f] + [g]$ (same as $[f + g]$), $[f] \times [g]$ (same as $[fg]$) and $\limsup_n [f_n]$ and $\liminf [f_n]$ (same as $[\limsup f_n]$ and $[\liminf f_n]$). But after a while one gets tired of writing brackets (we tell ourselves that we are picturing them mentally) and we refer to equivalence classes as functions, choose representatives at abandon and so on.

185. We have only integrated functions over the whole space. If $A \in \mathcal{F}$, we define the integral over A as $\int_A f d\mu = \int_X f \mathbf{1}_A d\mu$. It is a good exercise to check that this is the same as integrating $g = f\big|_A$ over the whole set in the measure space $(A, \mathcal{G}, \nu)$ where $\mathcal{G} = \{B \cap A : B \in \mathcal{F}\}$ and $\nu = \mu\big|_A$.

186. An interesting property is that if $f \geq 0$, then $\mu_f(A) = \int_A f d\mu$ defines a new measure on $(X, \mathcal{F})$. All one needs to check is countable additivity. But if A_n are pairwise disjoint and their union is A, then $\mathbf{1}_A = \sum_n \mathbf{1}_{A_n}$ and by MCT (applied to the sequence of partial sums), $\int_X f\mathbf{1}_A d\mu = \sum_n \int_X f \mathbf{1}_{A_n} d\mu$. This is another way to generate new measures from old. In terminology that we shall introduce later, *μ_f is absolutely continuous to μ with Radon-Nikodym derivative f.*

17. Limit theorems

187. As remarked earlier, a striking and convenient feature of Lebesgue integration theory is the ease with which limits and integrals can be interchanged. There are three primary theorems that we shall have occasion to use repeatedly. Throughout, we fix a measure space $(X, \mathcal{F}, \mu)$ and $\overline{\mathbb{R}}$-valued functions on it.

188. Monotone convergence theorem (MCT): Suppose f_n, f are non-negative measurable functions such that $f_n \uparrow f$ $a.s.[\mu]$. Then $\int_X f_n d\mu \uparrow \int_X f d\mu$ (valid also when $\int_X f d\mu = +\infty$).

189. Fatou's lemma: Suppose f_n are non-negative measurable functions. Then $\int_X (\liminf f_n) d\mu \leq \liminf_n \int_X f_n d\mu$.

190. Dominated convergence theorem (DCT): Suppose $f_n \to f$ $a.s.[\mu]$ and $|f_n| \leq g$ $a.s.$ for all n for an integrable g. Then $\int_X |f_n - f| d\mu \to 0$ and $\int_X f_n d\mu \to \int_X f d\mu$.

191. Is non-negativity necessary in MCT? As $f_n \uparrow f$. If $f_n \uparrow f$, then $g_n = f_n - f_1$ are non-negative and increase to $g = f - f_1$. Applying MCT, we get $\int (f_n - f_1) d\mu \uparrow \int (f - f_1) d\mu$. Using linearity, $\int f_n d\mu - \int f_1 d\mu \uparrow \int f d\mu - \int f_1 d\mu$, which implies that $\int f_n d\mu \uparrow \int f d\mu$. Where is the flaw in this argument?

There is none, provided $\int f_1 d\mu > -\infty$, or $\int f_N d\mu > -\infty$ for some N (then we can start the sequence at f_N). Consider $f_n = -\mathbf{1}_{[n,\infty)}$ which increases to 0 as $n \to \infty$, but $\int f_n d\lambda = -\infty$ for all n and $\int f d\lambda = 0$. This shows that MCT can be violated.

192. Is the domination condition necessary in DCT? The example of $f_n = n\mathbf{1}_{[0,\frac{1}{n}]}$ on $(\mathbb{R}, \mathcal{L}, \lambda)$ (clearly $f_n \to 0$ $a.s.[\lambda]$) shows that some condition is needed along with almost sure convergence to conclude convergence of integrals. This example also helps to avoid the most common mistake I have seen in applying DCT. All too often someone states that if $f_n \to f$ $a.s.[\mu]$ and f is integrable, then $\int f_n d\mu \to \int f d\mu$. False!

However, the domination condition is not necessary. In fact, there is a necessary and sufficient condition (finiteness of μ needed?) which, together with $f_n \to f$ $a.s.[\mu]$ implies that $\int f_n d\mu \to \int f d\mu$. That condition is *uniform integrability*: Given $\epsilon > 0$, there exists $M < \infty$ such that $\int_X |f_n| \mathbf{1}_{|f_n| > M} d\mu < \epsilon$ for all n. Check that the domination condition implies uniform integrability. Domination is one of the most easily checkable sufficient conditions for uniform integrability.

193. Is non-negativity required in Fatou's lemma? Again the example of $-\mathbf{1}_{[n,\infty)}$ on $(\mathbb{R}, \mathcal{L}, \lambda)$ shows that we cannot just omit it. Another point is that strict inequality can hold in the lemma. The example $f_n = n\mathbf{1}_{[0,\frac{1}{n}]}$ shows this.

194. *Proof of MCT:* We already know this when f_n and f are simple and convergence happens at all x. Let us relax this one by one.

First suppose that f_n are simple and convergence happens at all points of X. Then for any non-negative, simple, measurable φ such that $0 \leq \varphi \leq f$, we have $(\varphi \wedge f_n) \uparrow \varphi$ pointwise and hence $\int_X \varphi = \lim_{n\to\infty} \int_X (\varphi \wedge f_n) d\mu \leq \liminf \int_X f_n d\mu$. Take supremum over

Next suppose that convergence happens at all points of X but no assumption that f_n or f are simple. Find non-negative simple measurable $\varphi_{n,k}$ that increase to f_n as $k \uparrow \infty$. Set $\psi_n = \varphi_{n,1} \vee \ldots \vee \varphi_{n,n}$. Then ψ_n are simple and increase to f point wise (why?). Hence, $\int_X f = \lim_{n\to\infty} \int_X (\varphi \wedge f_n) d\mu \leq \liminf \int_X f_n d\mu$

195. *Proof of Fatou's lemma:* Let $g_k = \inf_{n\geq k} f_n$. Then g_k are non-negative measurable functions that increase to $g = \liminf f_n$. By MCT, $\int g_k d\mu \uparrow \in g d\mu$. But $g_k \leq f_k$ for each k, hence $\liminf_k \int f_k d\mu \geq \int g d\mu$. ■

196. *Proof of DCT:* $g - f_n \geq 0$ and $g - f_n \to g - f$ $a.s.[\mu]$. By Fatou's lemma $\liminf \int (g - f_n) d\mu \geq \int (g - f) d\mu$. Cancel $\int g d\mu$ on both sides (since it is finite by integrability of g) to get $\limsup \int f_n d\mu \leq \int f d\mu$. Similarly from $g + f_n$ (which are non-negative and converge to $g + f$ a.s.) conclude that $\liminf \int f_n d\mu \leq \int f d\mu$. Together these show that $\int f_n d\mu \to \int f d\mu$.

To get the apparently stronger conclusion that $\int |f_n - f| d\mu \to 0$, apply what we have already proved to the sequence $|f_n - f|$ which converges to 0 $a.s.[\mu]$ and is dominated by $2g$ (since $|f_n - f| \leq |f_n| + |f|$ and $f_n \to f$ $a.s.[\mu]$ implies that $|f| \leq g$ too). ■

18. COMPLETENESS OF LEBESGUE SPACES

197. Let $(X, \mathcal{F}, \mu)$ be a measure space. Recall the equivalence relationship $f = g$ $a.s.[\mu]$ introduced earlier. If $f \sim g$, then f is integrable if and only if g is, and then their integrals are equal. Hence we can define an equivalence class to be integrable if one (and hence all) element in it is integrable. The space of equivalence classes of integrable functions is denoted $L^1(X, \mathcal{F}, \mu)$ - it is the first Lebesgue space. We may simply write $L^1(\mu)$ or L^1 if the setting is clear.

198. As noted earlier, the operations of addition, scalar multiplication, products, almost sure limits can be naturally defined on the collection of equivalence classes. In particular, $L^1(\mu)$ is a vector space with the L^1-norm $\|[f]\|_1 := \int_X |f| d\mu$ (well-defined as the choice of representative does not change the value of the integral). It is a norm because of homogeneity $\int |\alpha f| d\mu = |\alpha| \int |f| d\mu$ and triangle inequality $\int |f + g| d\mu \leq \int (|f| + |g|) d\mu = \int |f| d\mu + \int |g| d\mu$. Also, $\int |f| d\mu = 0$ if and only if $f = 0$ $a.s.[\mu]$, which means that $[f] = 0$ (this is the reason why we moved to equivalence classes, otherwise we would only get a pseudo-norm).

199. Completeness of L^1: The space $L^1(X, \mathcal{F}, \mu)$ is complete under the L^1 norm.

200. *Proof:* Pick a Cauchy sequence in $L^1(\mu)$ and representatives f_n in them. The Cauchy property says that $\int_X |f_n - f_m| d\mu \to 0$ as $m, n \to \infty$. In particular, we can find $N_1 < N_2 < \ldots$ such that $\int_X |f_n - f_m| d\mu \leq 2^{-k}$ if $m, n \geq N_k$. As 2^{-k} is summable, it follows that the series $f_{N_1} + (f_{N_2} - f_{N_1}) + (f_{N_3} - f_{N_2}) + \ldots$ converges absolutely almost surely. Thus, $f_{N_k} \to f$ $a.s.[\mu]$ for some finite f that is integrable. Further, $\int_X |f_{N_k} - f| d\mu \leq \sum_{j\geq k} \int_X |f_{N_j} - f_{N_{j+1}}| d\mu \leq 2^{-k+1}$.

Therefore, $\|[f_{N_k}] - [f]\|_1 \to 0$. In a Cauchy sequence, if a subsequence converges then the entire sequence does, thus $\|[f_n] - [f]\|_1 \to 0$. This shows the completeness of $L^1(\mu)$. ■

201. Other Lebesgue spaces. For $1 \le p < \infty$, define $\|f\|_p = \left(\int_X |f|^p d\mu\right)^{1/p}$ and for $p = \infty$ we define $\|f\|_\infty = \inf\{t \ge 0 : \mu\{f > t\} = 0\}$. Then, for $1 \le p \le \infty$, the vector space $L^p(X, \mathcal{F}, \mu) := \{[f] : \|f\|_p < \infty\}$. It is endowed with the L^p norm $\|[f]\|_p = \|f\|_p$ (the choice of representative does not matter). It is clear that $\|\alpha[f]\|_p = |\alpha|\|[f]\|_p$ and that $\|[f]\|_p = 0$ if and only if $[f] = 0$. Triangle inequality is not obvious and will be proved later (*Minkowski's inequality*). That is where $p \ge 1$ is required. One example that may already be familiar is when $X = \mathbb{N}$ and μ is the counting measure ($\mathcal{F}$ is the power set), in which case $L^p(\mu) = \ell^p = \{x = (x_1, x_2, \ldots) : x_n \in \mathbb{R},\ \sum_n |x_n|^p < \infty\}$.

202. L^p spaces are important, but the most important and intuitive of them are L^1, L^2 and L^∞. We have already seen L^1.

203. The space L^∞ is easy to work with. The elements are equivalence classes of measurable that contain a bounded function, with the norm being the smallest of the sup-norms of elements in the equivalence class. Triangle inequality and completeness of L^∞ are easy to prove (similar to the way one proves that sup-norm is a complete norm on $C[0, 1]$).

204. The space L^2 is the most special of all. Observe that $\|f\|_2^2 = \langle f, f\rangle$ where $\langle f, g\rangle := \int_X fg d\mu$. This is a (pseudo) inner product, which becomes a genuine inner product at the level of equivalence classes. It is also complete, which makes it a *Hilbert space*.

205. Assuming that the L^p-norm is a norm, let us prove the completeness. Take a Cauchy sequence of equivalence classes in $L^p(\mu)$ and representatives f_n in them. Then find a subsequence $\{N_k\}$ such that $\|f_n - f_m\|_p < 2^{-2k}$ for $m, n \ge N_k$. Then the set $A_k := \{|f_{N_k} - f_{N_{k+1}}| \ge 2^{-k}\}$ has measure $\mu(A_k)2^{-kp} \le \int_X |f_{N_k} - f_{N_{k+1}}|^p d\mu \le 2^{-2kp}$, whence $\mu(A_k) \le 2^{-kp}$. As this is summable, we see that $|f_{N_k} - f_{N_{k+1}}| \ge 2^{-k}$ for all large k, for a.e. x $[\mu]$. Therefore, $f_{N_1} + \sum_{k\ge1}(f_{N_{k+1}} - f_{N_k})$ converges absolutely, $a.s.[\mu]$. Call the limit f (at points where the limit does not exist, define $f(x) = 0$) so that $f_{N_k} \to f$ $a.s.[\mu]$. Further, $\|f - f_{N_k}\|_p \le \sum_{j\ge k} \|f_{N_{j+1}} - f_{N_j}\|$, which shows that $f_{N_k} \to f$ in L^p. As the whole sequence is Cauchy in L^p, this implies that the sequence converges in L^p to f (more precisely that $[f_n]$ converges to $[f]$ in L^p).

206. We are yet to prove Minkowski's inequality. We first prove Hölder's inequality.

207. **Hölder's inequality:** Let $f, g \ge 0$ be measurable functions on $(X, \mathcal{F}, \mu)$. Then for any $1 \le p, q \le \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, we have $\int fg d\mu \le \left(\int f^p d\mu\right)^{1/p} \left(\int g^q d\mu\right)^{1/q}$, with equality iff $f^p = g^q$ $a.s.[\mu]$. As a corollary[1], if $f \in L^p$ and $g \in L^q$, then $fg \in L^1$ and $\|fg\|_1 \le \|f\|_p\|g\|_q$.

208. The key inequality is that $\frac{a^p}{p} + \frac{b^q}{q} \ge ab$ for $a, b \ge 0$. Apply this with $a = f(x)$, $b = g(x)$ and integrate over x w.r.t. μ to get $\frac{1}{p}\|f\|_p^p + \frac{1}{q}\|g\|_q^q \ge \|fg\|_1$. Replace f by $f/\|f\|_p$ and g by $g/\|g\|_q$ (if

[1]Strictly speaking one should write $[f], [g], [fg]$ in this statement, but it is usual to be loose with language and say that a function is in L^p etc.

one of $\|f\|_p$ or $\|g\|_q$ is 0 or ∞, the statement is obvious) to get $\frac{\|fg\|_1}{\|f\|_p\|g\|_q} \leq \frac{1}{p} + \frac{1}{q} = 1$. This proves the first statement. Applying it to $|f|$ and $|g|$ gives the second.

To prove the inequality used, consider the graph $y = x^{p-1}$, which is the same as $x = y^{q-1}$, where $x, y > 0$. Observe that $\frac{a^p}{p}$ is the area of $\{(x,y) : 0 \leq x \leq a,\ 0 \leq y \leq x^{p-1}\}$ and $\frac{b^q}{q}$ is the area of $\{(x,y) : 0 \leq y \leq b,\ 0 \leq x \leq y^{q-1}\}$. The union of these regions covers the rectangle $0 \leq x \leq a$, $0 \leq y \leq b$, which has area ab.

It is clear that equality holds if and only if $b = a^{p-1}$ or equivalently $a = b^{q-1}$ or equivalently $a^p = b^q$. Thus, in Hölder's inequality for non-negative functions, equality holds if and only if $f^p = g^q\ a.s.[\mu]$. For general f, g, equality holds if and only if $|f|^p = |g|^q$ and fg has a constant sign, $a.s.[\mu]$. ∎

209. Minkowski's inequality: Let $f, g \geq 0$ are measurable, then for any $1 \leq p \leq \infty$, we have $\|f+g\|_p \leq \|f\|_p + \|g\|_p$. In particular, $\|\cdot\|_p$ defines a norm on $L^p(\mu)$.

210. *Proof:* The cases $p = 1$ and $p = \infty$ are easy to verify directly. Hence assume $1 < p < \infty$ and let $1 < q < \infty$ be the conjugate exponent satisfying $\frac{1}{p} + \frac{1}{q} = 1$. Write $\int (f+g)^p d\mu$ as $\int f(f+g)^{p-1} d\mu + \int g(f+g)^{p-1} d\mu$. Apply Hölder's inequality with exponents p and q to the first integral to get $\int f(f+g)^{p-1} d\mu \leq (\int f^p d\mu)^{\frac{1}{p}} (\int (f+g)^p d\mu)^{\frac{1}{q}}$. Similarly $\int g(f+g)^{p-1} d\mu \leq (\int g^p d\mu)^{\frac{1}{p}} (\int (f+g)^p d\mu)^{\frac{1}{q}}$. Adding the two, we get $\int (f+g)^p d\mu \leq (\int (f+g)^p d\mu)^{\frac{1}{q}} ((\int f^p d\mu)^{\frac{1}{p}} + (\int g^p d\mu)^{\frac{1}{p}})$. As $1 - \frac{1}{q} = 1$, we get $(\int (f+g)^p d\mu)^{\frac{1}{p}} \leq (\int f^p d\mu)^{\frac{1}{p}} + (\int g^p d\mu)^{\frac{1}{p}})$. Equality holds if and onlu if $f^p = (f+g)^{(p-1)q}$ and $g^p = (f+g)^{(p-1)q}\ a.s.[\mu]$, which is the same as $f = g\ a.s.[\mu]$.

All this holds even if the integrals are infinite. Now if $f, g \in L^p$, then apply the obtained inequality to $|f|$ and $|g|$ to get the second statement of the theorem. Equality holds if and only if $|f| = |g|$ and $|f+g| = |f| + |g|\ a.s.[\mu]$, which is the same as saying that f and g are collinear. ∎

19. Measure from integral

211. Sticking to $\mathbb{R}$, we constructed Lebesgue measure and then Lebesgue integral. It is also possible to go in the other direction - start with a notion of integral on a large enough class of functions and then construct the measure. What class of functions? The notion of measurability depends on the sigma algebra we take, but we probably agree that any notion of integral must apply to compactly supported continuous functions. Hence let us start with an integral on this class. But what is an integral? We ask only fo linearity and positivity.

212. The considerations extend to any locally compact Hausdorff space. If not familiar with general topology, assume that it is a metric space where closed balls are compact. All open sets in

$\mathbb{R}^d$ are examples. But $C[0,1]$ is not a locally compact space (why?). Let $C_c(X)$ denote the space of continuous $f : X \mapsto \mathbb{R}$ such that $f = 0$ outside a compact set.

213. **Riesz's representation theorem:** Let X be a locally compact Hausdorff space. Let $L : C_c(X) \mapsto \mathbb{R}$ be a functional that is linear ($L(\alpha f + \beta g) = \alpha L(f) + \beta L(g)$ for $\alpha, \beta \in \mathbb{R}$ and $f, g \in C_c(X)$) and positive ($L(f) \geq 0$ if $f \geq 0$ point wise). Then, there exists a unique regular Borel measure μ on $(X, \mathcal{B}_X)$ such that $L(f) = \int_X f d\mu$ for all $f \in C_c(X)$.

214. Recall that a Borel measure μ is regular if it is Radon ($\mu(K) < \infty$ for all compact K), outer regular ($\mu(A) = \inf\{\mu(G) : G \supseteq A \text{ is open}\}$ for all $A \in \mathcal{B}_X$) and inner regular ($\mu(A) = \sup\{\mu(K) : K \subseteq A \text{ is compact}\}$ for all $A \in \mathcal{B}_X$ with $\mu(A) < \infty$). If X is sigma-compact, then we may drop the condition of $\mu(A) < \infty$ in inner regularity.

215. Riesz's representation theorem says that an integral is nothing except a positive, linear functional on $C_c(X)$. If X is compact, then $C_c(X)$ is just $C(X)$.

216. When we defined measure as a countably additive function on a sigma-algebra, we never justified the requirement of *countable* additivity. Even our intuitive notion of volumes would ask only for finite additivity. Indeed, countable additivity was initially not universally accepted as the right framework, especially in probability. However, the theory with countably additivity is far richer than just finite additivity, and has gradually become the norm.

Riesz's representation theorem can be seen as a justification for countable additivity. The question of what linear functions on $C_c(\mathbb{R})$ are positive, has no countability in it. But the answer turns out to be integration w.r.t. (countably additive) measures. Thus it brings out countable additivity without secretly putting it into the question itself!

217. Some examples: Let $X = \mathbb{R}$ and $L_1(f) = \int_{\mathbb{R}} f(x)dx$ (Riemann integral, say), $L_2(f) = f(1)$, $L_3(f) = \sum_{n\in\mathbb{Z}} w_n f(n)$ where $w_n > 0$ for all n. It is clear that all three are positive linear functionals. The measure corresponding to them (as per Riesz's representation theorem) are $\mu_1 = \lambda_1$ (Lebesgue measure on $\mathbb{R}$), $\mu_2 = \delta_1$ and $\mu_3 = \sum_{n\in\mathbb{Z}} w_n \delta_n$.

218. Let us discuss the approach to Riesz's representation theorem in the context of $X = I = [0,1]$ and $L(f) = \int_0^1 f(x)dx$ for $f \in C(I)$ (here we may take Riemann integral, to make the point that this could have been defined before Lebesgue integration theory). The one way to construct measures that we have seen is to construct an outer measure then use the Carathéodory cut condition. For Lebesgue measure, we constructed the outer measure starting with lengths of intervals; now we must do that using integrals of continuous functions.

219. A first thought would be to define $\mu^{\#}(A) := \inf\{L(f) : f \in C(I),\ f \geq \mathbf{1}_A\}$ for $A \subseteq I$. But then $\mu^{\#}(\mathbb{Q} \cap [0,1]) = 1$, as $f \geq 1$ on $\mathbb{Q} \cap [0,1]$ and continuity of f force that $f \geq 1$ on $[0,1]$. But we know that we must get μ to be the Lebesgue measure, hence there is a problem with this approach. This is reminiscent of the problem with Jordan measurable sets. Indeed, it is not hard to show that $\mu^{\#}(A) = \lambda^{\#}(A)$ where the outer Jordan measure $\lambda^{\#}(A) = \inf \sum_{j=1}^{k} |I_j|$, where the infimum is over all finite covers of A by intervals.

220. *Proof that* $\mu^{\#} = \lambda^{\#}$*:* Given an interval cover $I_1, \ldots, I_m$ of A, pick $f_i \in C(I)$ such that $f_i = 1$ on I_i and $f_i = 0$ outside an ϵ-neigbourhood of I_i and $0 \le f_i \le 1$ everywhere. Then $f := f_1 + \ldots + f_m \ge \mathbf{1}_A$ and $L(f) = L(f_1) + \ldots + L(f_m) \le |I_1| + \ldots + |I_m| + m\epsilon$. Taking infima, we see that $\mu^{\#}(A) \le \lambda^{\#}(A)$.

Conversely, given $f \in C(I)$ such that $f \ge \mathbf{1}_A$, observe that $\{f > 1-\epsilon\}$ is a finite union of finitely many open intervals $I_1, \ldots, I_m$ (why?). Then $|I_1| + \ldots + |I_m| \le (1-\delta)^{-1} \int_0^1 f(x)dx$. Again taking infima, we see that $\lambda^{\#}(A) \le \mu^{\#}(A)$.

221. We now move back to a more general situation but more restricted than in the statement of Riesz's theorem. **Henceforth, assume that X is a compact metric space**. Given L is a positive linear functional on $C(X)$, define for $A \subseteq X$.

$$\mu^*(A) = \inf \left\{ \sum_{i=1}^{\infty} L(f_i) : f_i \in C(X),\ \sum_{i=1}^{\infty} f_i \ge \mathbf{1}_A \right\}.$$

The choice $f_1 = 1$ and $f_i = 0$ for $i \ge 2$ shows that the infimum is over a non-empty set.

222. *Proof that μ^* is an outer measure:* Clearly $0 \le \mu^*(A) \le L(1)$. Monotonicity is also obvious: If $A \subseteq B$ then $\mu^*(A) \le \mu^*(B)$. To check countable subadditivity, let A_n be subsets and $A = \cup_n A_n$. Find $f_{n,k}$, $k \ge 1$, such that $\sum_n L(f_{n,k}) \le \mu^*(A_n) + \epsilon 2^{-n}$ and then the whole collection $f_{n,k}$, $n, k \ge 1$, satisfies $\sum_{n,k} L(f_{n,k}) \le \sum_n \mu^*(A_n) + \epsilon$ and $\sum_{n,k} f_{n,k} \ge \mathbf{1}_A$. Hence, $\mu^*(A) \le \epsilon + \sum_n \mu^*(A_n)$, showing subadditivity of μ^*.

223. *Metric property of μ^*:* Let $A, B \subseteq X$ with $d(A,B) > 0$. Consider $\varphi_{A,\epsilon}(x) := 1 - (\frac{1}{\epsilon} d(A,x) \wedge 1)$, a continuous function that is 1 on A and 0 outside the ϵ-neighbouhood of A and between 0 and 1 everywhere. Given any $f \ge \mathbf{1}_{A \sqcup B}$, let $g = f\varphi_A$ and $h = f\varphi_B$ where $2\epsilon < d(A,B)$. Then $g, h \in C(X)$, have disjoint supports, $g \ge \mathbf{1}_A$ and $h \ge \mathbf{1}_B$. Disjointness of supports shows that $f \ge g + h$, hence $L(f) = L(g) + L(h) \ge \mu^*(A) + \mu^*(B)$. Taking infimum over f we get $\mu^*(A \sqcup B) \ge \mu^*(A) + \mu^*(B)$. The other way inequality follows from subadditivity. Thus μ^* is a metric outer measure.

224.

225. Recall that a metric outer measure is a measure when restricted to the Borel sigma algebra. Let this measure be denoted μ. All that remains is to prove the regularity of μ. To see the outer regularity, let $A \subseteq X$ and find $f \in C(X)$ such that $\mu(A) > L(f) - \epsilon$. Let $G = \{f > 1-\delta\}$, an open set that contains A. Observe that $(1-\delta)^{-1} f \ge \mathbf{1}_G$ and hence $\mu(G) \le (1-\delta)^{-1} L(f) < (1-\delta)^{-1}(\mu(A) + \epsilon)$. As ϵ and δ are arbitrary positive numbers this shows that $\inf\{\mu(G) : G \supseteq A \text{ is open}\} \le \mu(A)$. The other way inequality is trivially true. Thus μ is outer regular. In case of compact space X, by going to complements, this also gives inner regularity. This completes the existence part of Riesz's representation theorem.

To prove the uniqueness, observe that if A is closed, then the sequence of functions $\varphi_n = 1 - (\frac{1}{\epsilon}d(A,x) \wedge 1)$ decrease to $\mathbf{1}_A$. Hence, by monotone convergence theorem, $\mu(A) = \lim L(\varphi_n)$. Thus, two measures that represent L must agree on all closed sets. Since they form a π-system that generates the Borel sigma algebra, the measures must be equal. ■

226. This completes the proof for compact metric spaces. The metric condition is not too restrictive for actual applications, but compactness is. For example, the case of $\mathbb{R}^d$ is not covered. Where did we use the compactness of X in the proof above? Suppose X is a locally compact metric space, we may define μ^* exactly as above, except that the infimum may be over an empty set (which is defined to be $+\infty$). The proof that μ^* is an outer measure goes through. When it comes to the metric property of μ^*, the function φ_A is in $C_c(X)$ only if $\overline{A}$ is compact. Thus, the metric property $\mu^*(A \sqcup B) = \mu^*(A) + \mu^*(B)$ holds when A and B are separated and their closures are compact. It must be possible to extend this to all sets or may be this suffices to show that μ^* is a measure when restricted to $\mathcal{B}_X$. I'll think about this at some point of time and update the notes, but for the course, we just stay with the proof for the compact metric case.

20. Radon-Nikodym theorem

227. Let $(X, \mathcal{F}, \mu)$ be a measure space. If f is a non-negative measurable function on X, then $\nu_f : \mathcal{F} \mapsto [0, \infty]$ defined by $\nu_f(A) = \int_A f d\mu$, defines another measure on $\mathcal{F}$ (this was an exercise - a simple application of MCT). When this is the case, we say that f is the *Radon-Nikodym* derivative of ν with respect to μ. The uniqueness is clear, hence we say "the" derivative. It is customary to write $d\nu = f d\mu$.

228. Now we address the converse question: Given a measure ν on $\mathcal{F}$, does ν have a Radon-Nikodym derivative w.r.t. μ, i.e., is there a function f such that $\nu = \nu_f$?

229. A necessary condition: If $\mu(A) = 0$ then $\nu_f(A) = 0$, because $f\mathbf{1}_A = 0 \; a.s.[\mu]$, which implies that $\int_A f d\mu = 0$. Hence, if there is even a single $A \in \mathcal{F}$ such that $\nu(A) > 0 = \mu(A)$, then ν cannot have a RN-derivative w.r.t. μ.

230. *Definition:* Let μ, ν be measures on $(X, \mathcal{F})$. We say that ν is *absolutely continuous* to μ and write $\nu \ll \mu$ if $\nu(A) = 0$ whenever $\mu(A) = 0$. If $\nu \ll \mu$ and $\mu \ll \nu$, we say that μ and ν are *mutually absolutely continuous*. We say that ν and μ are *singular* and write $\mu \perp \nu$ if there exists a set $A \in \mathcal{F}$ such that $\mu(A) = 0$ and $\nu(A^c) = 0$. Singularity is a symmetric relationship.

231. Examples: If ν is a measure on $\mathbb{R}^d$ that has an atom then it is not absolutely continuous to λ_d. Lebesgue measure on a proper subspace of $\mathbb{R}^d$ is singular to Lebesgue measure on $\mathbb{R}^d$.

The Cantor measure and Lebesgue measure on the line are singular to each other. On $(\mathbb{R}, \mathcal{B})$, if $\mu = \lambda\big|_{[0,1]}$, then $\mu \ll \lambda$ but λ is not absolutely continuous to μ.

232. The discussion earlier says that if ν has a Radon-Nikodym derivative w.r.t. μ, then it is necessary that $\nu \ll \mu$. It turns out that this is the only obstacle! Well, we must restrict to sigma-finite measures though.

233. Radon-Nikodym theorem: Let μ, ν be σ-finite measures on $(X, \mathcal{F})$. If $\nu \ll \mu$, then it has a Radon-Nikodym derivative w.r.t. μ.

234. If we prove the theorem for finite measures μ, ν, then the result for the sigma-finite case can be deduced as follows: Find a measurable partition $\{X_n\}$ of X such that $\mu(X_n) < \infty$ and $\nu(X_n) < \infty$. Consider the finite measures $\mu_n = \mu\big|_{X_n}$ and $\nu_n = \nu\big|_{X_n}$ on $(X_n, \mathcal{F}_n)$, where $\mathcal{F}_n = \{B \in \mathcal{F} : B \subseteq X_n\}$. Then $\nu_n \ll \mu_n$, write $d\nu_n = f_n d\mu_n$, where $f_n \in L^1(X_n, \mathcal{F}_n, \mu_n)$. Then define $f : X \mapsto \mathbb{R}$ as $f = \sum_n f_n \mathbf{1}_{X_n}$. For any $A \in \mathcal{F}$, write $A = \sqcup_n (A \cap X_n)$ and hence

$$\nu(A) = \sum_n \nu_n(A \cap X_n) = \sum_n \int_{X_n} f_n d\mu_n = \int_X f d\mu,$$

showing that $d\nu = f d\mu$.

235. *Proof of RN theorem for finite measures:* Consider $\mathcal{S} := \{f : X \mapsto \overline{\mathbb{R}}_+ : \text{measurable}, \int_A f d\mu \leq \nu(A)\ \forall A \in \mathcal{F}\}$. As $\mathcal{S}$ contains the zero function, it is non-empty. Let $J(f) := \int_X f d\mu$ for $f \in \mathcal{S}$. We wish to show that J attains its supremum on $\mathcal{S}$. Assuming this, let $g \in \mathcal{S}$ be a maximiser of J over $\mathcal{S}$. We claim that g is the RN-derivative of ν w.r.t. μ. If not, then $\int_X g d\mu < (1-\epsilon)\nu(X)$ for some $\epsilon > 0$. We make the claim (and this is the one that uses the hypothesis of absolute continuity) that there exists $A \in \mathcal{F}$ such that $\mu(A) > 0$ and $\int_C g d\mu \leq (1-\epsilon)\nu(C)$ for all measurable $C \subseteq A$. Once this claim is granted, $h = g + \epsilon \mathbf{1}_A \in \mathcal{S}$ and $J(h) > J(g)$, giving a contradiction. Modulo the two claims, the proof is complete. ■

236. *Proof that J attains its supremum on $\mathcal{S}$:* First we observe that $\mathcal{S}$ is closed under taking finite maxima. Indeed, if $f_1, f_2 \in \mathcal{S}$, then partition X into $X_1 = \{f_1 \geq f_2\}$ and $X_2 = \{f_2 > f_1\}$, so that $\int_A (f_1 \vee f_2) d\mu = \int_{A \cap X_1} f_1 d\mu + \int_{A \cap X_2} f_2 d\mu$. The summands are bounded by $\nu(A \cap X_1)$ and $\nu(A \cap X_2)$, because $f_i \in \mathcal{S}$. Adding up, we see that $\int_A g d\mu \leq \nu(A)$, for any $A \in \mathcal{F}$. That is, $g \in \mathcal{S}$.

Now find $f_n \in \mathcal{S}$ such that $J(f_n) \uparrow \sup_{\mathcal{S}} J$. Then $g_n := f_1 \vee f_2 \vee \ldots \vee f_n \in \mathcal{S}$ and g_n are point-wise increasing, say to g. By MCT, it follows that $g \in \mathcal{S}$ and also that $J(g) = \lim J(g_n) = \sup_{\mathcal{S}} J$. ■

237. *Proof of the existence of the set A:* To remove distracting features, define the measures $d\theta = g d\mu$ and $\tau = (1-\epsilon)\nu$. Now define $a_1 = \sup\{\theta(B) - \tau(B) : B \subseteq X\}$ and observe that $a_1 > 0$ since $\theta(X) > \tau(X)$ by assumption. Find $A_1 \subseteq X$ such that $\theta(A_1) - \tau(A_1) \geq 0.9a_1$. Then define $a_2 = \sup\{\theta(B) - \tau(B) : B \subseteq A_1\}$ and find $A_2 \subseteq A$ such that $\theta(A_2) - \tau(A_2) \geq 0.9a_2$. Continue to define a_n and A_n (if $a_n = 0$ then set $A_k = 0$ for $k \geq n$). Then A_n are pairwise disjoint and

$\sqcup_n A_n \subseteq A$, hence $0.9 \sum_n a_n \leq \sum_n \theta(A_n) \leq 1$, showing that $a_n \to 0$. Therefore, if $B = A \setminus \sqcup_n A_n$, then for any $C \subseteq B$, we must have $\theta(C) \leq \tau(C)$, otherwise, $\theta(C) > \tau(C) + a$ for some $a > 0$ and then a_n would have been more than a for all n. If $\mu(B) = 0$, then by absolute continuity we also have $\nu(B) = 0$, which means that $\theta(B) = \tau(B) = 0$. ■

238. *Second proof of Radon-Nikodym theorem:* Again we assume that μ and ν are finite measures. Let $\theta = \mu + \nu$. Then $\mu(A) \leq \theta(A)$ for all A, hence $L^2(\theta) \subseteq L^2(\mu)$. Further, the linear functional $L(f) = \int_X f d\mu$ defined on $L^2(\theta)$ is continuous. Hence, there exists $g \in L^2(\theta)$ such that $\int_X f d\mu = \int_X f g d\theta$ or equivalently, $\int_X f(1-g) d\mu = \int_X f g d\nu$.