Analysis of Probabilistic Systems

Bootcamp Lecture 2: Measure and Integration

Prakash Panangaden[1]

[1]School of Computer Science
McGill University

Fall 2016, Simons Institute

Outline

What is measure theory?

- We want to assign a “size” to sets so that we can use it for quantitative purposes, like integration or probability.
- We could count the number of points, but this is useless for a continuous space.
- We want to generalize the notion of “length” or “area.”
- What is the “length” of the rational numbers between 0 and 1?
- We want a consistent way of assigning sizes to these and (all?) other sets.

What are measurable sets ?

- Alas! Not all sets can be given a sensible notion of size that generalizes the notion of length of an interval.
- We take a family of sets satisfying “reasonable” axioms and deem them to be “measurable.”
- **Countable** unions are the key.

Measurable spaces

A **measurable space** (X, Σ) is a set X together with a family Σ of subsets of X, called a σ**-algebra** or σ**-field**, satisfying the following axioms:

- $\emptyset \in \Sigma$,
- $A \in \Sigma$ implies that $A^c \in \Sigma$, and
- if $\{A_i \in \Sigma | i \in I\}$ is a *countable* family then $\cup_{i \in I} A_i \in \Sigma$.

Basic facts

- The intersection of any collection of σ-algebras on a set is another σ-algebra.
- Thus, given any family of sets $\mathcal{F}$ there is a least σ-algebra containing $\mathcal{F}$: the σ-algebra *generated* by $\mathcal{F}$; written $\sigma(\mathcal{F})$.
- Measurable sets are complicated beasts, we often want to work with the sets of family of simpler sets that generate the σ-algebra.

Examples

- X a set, $\Sigma = \{X, \emptyset\}$
- $\sigma = \mathcal{P}(X)$, the power set.
- The σ-algebra generated by intervals in $\mathbb{R}$ is called the *Borel* algebra. For any topological space the σ-algebra generated by the opens (or the closed sets) is called its Borel algebra.
- There is a larger σ-algebra containing the Borel sets called the Lebesgue σ-algebra; more later.
- Fix a finite set A; $A^\infty =$ finite and infinite sequences of elements from A. Define, for $x \in A^*$, $x\uparrow = \{y : x \leq y\}$. The σ-algebra generated by the $x\uparrow$ is very commonly used to study discrete-step processes.

Monotone classes

- Notation: If $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \ldots$ and $A = \bigcup_n A_n$ we write $A_n \uparrow A$; similarly $A_n \downarrow A$.
- If $\mathcal{M}$ is a collection of sets and is closed under up and down arrows it is called a **monotone class**.
- Arbitrary intersections of monotone classes form a monotone class; hence we have the monotone class *generated* by a family of sets.
- A collection of sets closed under complements and finite unions is called a **field** of sets.
- Any σ-algebra is a monotone class and if a monotone class is also a field it is a σ-algebra.
- **Theorem**: If $\mathcal{F}$ is a field of sets then the monotone class that it generates is the same as the σ-algebra that it generates.

π and λ systems

- A π-system is a family of sets closed under finite intersections.
- The open intervals of $\mathbf{R}$ form a π-system. It generates the Borel sets.
- Slogan: π-systems are usually easy to describe but they can generate complicated σ-algebras. Try to use a generating π-system.
- A λ-**system** over X is a family of subsets of X containing X, closed under complements and closed under countable unions of pairwise disjoint sets.
- **Prop**: If Ω is a π-system and a λ-system it is a σ-algebra.
- If Ω is a π-system and Λ is a λ-system and $\Omega \subseteq \Lambda$ then $\sigma(\Omega) \subseteq \Lambda$. Dynkin's $\lambda - \pi$ theorem.

Measurable functions

- $f : (X, \Sigma) \to (Y, \Omega)$ is *measurable* if for every $B \in \Omega$, $f^{-1}(B) \in \Sigma$.
- Just like the definition of continuous in topology.
- Why is this the definition? Why backwards?
- $x \in f^{-1}(B)$ if and only if $f(x) \in B$.
- No such statement for the forward image.
- Exactly the same reason why we give the Hoare triple for the assignment statement in terms of preconditions.
- Older books (Halmos) give a more general definition that is not compositional.
- Measurable spaces and measurable functions form a category.

Examples

- If $A \subset X$ is a measurable set, $\mathbf{1}_A(x) = 1$ if $x \in A$ and 0 otherwise is called the *indicator* or *characteristic* function of A and is measurable.
- The sum and product of real-valued measurable functions is measurable.
- If we take *finite* linear combinations of indicators we get *simple* functions: measurable functions with finite range.

Convergence properties

- If $\{f_i : \mathbf{R} \longrightarrow \mathbf{R}\}_{i \in \mathbf{N}}$ converges pointwise to f and all the f_i are measurable then so is f.
- Stark difference with continuity.
- If $f : (X, \Sigma) \longrightarrow (\mathbb{R}, \mathcal{B})$ is non-negative and measurable then there is a sequence of non-negative *simple* functions s_i such that $s_i \leq s_{i+1} \leq f$ and the s_i converge pointwise to f.
- The secret of integration.

Measures

A **measure** (**probability measure**) μ on a measurable space (X, Σ) is a function from Σ (a set function) to $[0, \infty]$ ($[0, 1]$), such that if $\{A_i | i \in I\}$ is a countable family of pairwise disjoint sets then

$$\mu(\bigcup_{i \in I} A_i) = \sum_{i \in I} \mu(A_i).$$

In particular if I is empty we have

$$\mu(\emptyset) = 0.$$

A set equipped with a σ-algebra and a measure defined on it is called a **measure space**.

A simple example

Fix a set X and a point x of X. We define a measure, in fact a probability measure, on the σ-algebra of all subsets of X as follows. We use the slightly peculiar notation $\delta(x, A)$ to emphasize that x is a parameter in the definition.

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

This measure is called the *Dirac delta measure*. Note that we can fix the set A and view this as the definition of a (measurable) function on X. What we get is the characteristic function of the set A, χ_A.

Monotonicity and Continuity

- Fix (X, Σ, μ); write $A, B, C, \ldots$ for sets in Σ.
- If $A \subseteq B$ then $\mu(A) \leq \mu(B)$.
- If $A_n \uparrow A$ then $\lim_{n \longrightarrow \infty} \mu(A_n) = \mu(A)$.
- If $A_n \downarrow A$ *and* $\mu(A_1)$ *is finite* then $\lim_{n \longrightarrow \infty} \mu(A_n) = \mu(A)$.

Digression: Choquet Capacities

- Consider combining probability and nondeterminism.
- Given (X, Σ), suppose we have a family of measures μ_i. Define $c(A) := \sup_i \mu_i(A)$. Is it a measure?
- No! It is not additive, not even finitely.
- But, it does satisfy monotonicity and *both* continuity properties.
- Such a thing is called a “Choquet capacity.”
- Not all capacities arise in this way.

Outer measure on $\mathbb{R}$

- For *any* subset A of $\mathbb{R}$ we define *the outer measure of* A, $\mu^*(A)$, as the infimum of the total length of any family of intervals covering A.
- The rationals have outer measure zero.
- μ^* is not additive so it does not give a measure defined on all sets.
- It does however satisfy countable subadditivity: $\mu^*(\cup A_i) \leq \sum_i \mu^*(A_i)$.
- We define an outer measure to be a set function satisfying monotonicity and countable subadditivity and defined on *all* sets.

From outer measure to measure

- Let X be a set and μ^* an other measure defined on it.
- There are some sets that "split all other sets nicely."
- For some sets A, $\forall E \subseteq X, \mu^*(E) = \mu^*(A \cap E) + \mu^*(A^c \cap E)$. Call the collection of all such sets Σ.
- Define $\mu(A) = \mu^*(A)$ for $A \in \Sigma$.
- (X, Σ, μ) is a measure space.
- The proof uses the $\lambda - \pi$ theorem.
- Applied to $\mathbb{R}$ with the outer measure above we get the Lebesgue measure on the Borel sets.

An extension theorem

- Want to define measures on “nice” sets and *extend* to all the sets in the generated σ-algebra.
- A family of sets $\mathcal{F}$ is called a *semi-ring* if:
  - $\emptyset \in \mathcal{F}$
  - $A, B \in \mathcal{F}$ implies $A \cap B \in \mathcal{F}$
  - $A \subset B$ implies there are finitely many pairwise disjoint sets $C_1, \ldots, C_k$, all in $\mathcal{F}$, such that $B \setminus A = \cup_{i=1}^{k} C_i$.
- If μ defined on a semi-ring satisfies:
  - $\mu(\emptyset) = 0$,
  - μ is finitely additive and
  - μ is countably subadditive,
- then μ extends uniquely to a measure on $\sigma(\mathcal{F})$.

Why I like π-systems

- Given two measures, μ_1, μ_2 on (X, Σ); are they the same?
- Suppose that $\Sigma = \sigma(\mathcal{P})$ where $\mathcal{P}$ is a π-system.
- Then if μ_1, μ_2 agree on $\mathcal{P}$ they will agree on Σ.
- We need to know in advance that μ_1, μ_2 are measures.

A measure on $\{H, T\}^\infty$

- Recall: A a finite set, $A^\infty =$ finite and infinite sequences of elements from A. Define, for $x \in A^*$, $x \uparrow = \{y : x \leq y\}$. Here $A = \{H, T\}$.
- We have $\Pr(H), \Pr(T)$. Want to define a measure on sets of $H - T$ (possibly infinite) sequences.
- Define, $\Pr(a_1 \ldots a_n) = \Pi_i \Pr(a_i)$.
- The sets of the form $x \uparrow$ form a semi-ring so Pr extends to a measure on the generated σ-algebra.

Lebesgue integration

- Proceed by working up from “simple” functions.
- Fix (X, Σ, μ). If $A \in \Sigma$ define $\chi_A : X \longrightarrow \mathbb{R}$ by $\chi_A(x) = 1$ if $x \in A$ else 0.
- Natural definition: $\int \chi_A \mathrm{d}\mu = \mu(A)$
- Define *simple* functions as finite linear combinations of characteristic functions: $s = \sum_i r_i \chi_{A_i}$; $r_i \in \mathbb{R}, A_i \in \Sigma$.
- $\int s \mathrm{d}\mu = \sum_i r_i \mu(A_i)$.
- Need to verify that the integral of s does not depend on how it is represented.
- Fact: every positive measurable function is the pointwise limit of a sequence of simple functions.
- For a positive measurable function f we define $\int f \mathrm{d}\mu = \bigvee_{s \leq f} \int s \mathrm{d}\mu$.
- For a general measurable function we split it into positive and negative parts and compute the integrals separately.
- I have skated over some issues about integrability.

Monotone convergence theorem

- Suppose that $f_n : X \longrightarrow \mathbb{R}$ is a sequence of measurable functions such that
- $\forall x \in X, 0 \leq f_1(x) \leq f_2(x) \leq \ldots \leq f_n(x) \leq \ldots < \infty$ and
- $\forall x \in X, \bigvee_n f_n(x) = f(x)$ then
- f is measurable and
- $\int f \mathrm{d}\mu = \bigvee_n \int f_n \mathrm{d}\mu$.
- One uses this theorem to prove that the integral is linear.

The monotone convergence mantra: an example

- Let $T : (X, \Sigma, \mu) \longrightarrow (Y, \Omega, \nu)$ be measurable.
- Let $f : Y \longrightarrow \mathbb{R}$ be measurable.
- Suppose that $\nu = \mu \circ T^{-1}$ then for any $B \in \Omega$,
- $\int_{T^{-1}(B)} f \circ T \mathrm{d}\mu = \int_B f \mathrm{d}\nu$.
- Easy to check that the equation holds for $f = \chi_A$.
- Hence true for f a simple function by linearity of integration.
- Hence true for any positive measurable function by the monotone convergence theorem.
- Hence true for any measurable function by splitting.

The Radon-Nikodym theorem

- Fix a measurable space (X, Σ) and two measures μ, ν.
- We say μ, ν are *mutually singular* if there are disjoint measurable sets A, B with $\mu(X \setminus A) = 0 = \nu(X \setminus B)$. We write $\mu \perp \nu$.
- We say ν is absolutely continuous with respect to μ, written $\nu << \mu$, if $\mu(A) = 0$ implies $\nu(A) = 0$.
- If we define ν by $\nu(A) = \int_A f \mathrm{d}\mu$ for some positive measurable function we will have $\nu << \mu$.
- If both μ, ν are (σ-) finite measures, then ν can be decomposed into $\nu = \nu_a + \nu_s$ with $\nu_a << \mu$ and $s \perp \mu$.
- There is a non-negative measurable function h such that $\nu_a(A) = \int_A h \mathrm{d}\mu$.
- If h' satisfies the same property as h then h, h' differ at most on a set of μ-measure 0.
- h is called the Radon-Nikodym derivative of ν_a with respect to μ and is sometimes written as $\frac{\mathrm{d}\nu_a}{\mathrm{d}\mu}$.

Conditional probability I

- Product space $X \times Y$, joint probability measure P on $X \times Y$; marginals P_X, P_Y.
- Suppose I know that the X coordinate is x, how do I revise my estimate of the probability distribution over Y?
- Fix a measurable subset $A \subseteq X$, there is a measurable function $P_A : Y \longrightarrow [0, 1]$ which satisfies: $\forall B \subseteq Y, P(A \times B) = \int_B P_A(y)\mathrm{d}P_Y$.
- Similarly there is a function P_B such that $\forall A \in \Sigma_X, P(A \times B) = \int P_B(x)\mathrm{d}P_X$.
- How do we know such things exist? Radon-Nikodym! $P(A \times \cdot) << P(X \times \cdot)$.
- I will write $P(x, B)$ and $P(y, A)$.

Conditional probability II

- We have a probability space (X, Σ, P).
- Suppose we have $\Lambda \subset \Sigma$. I tell you for every $B \in \Lambda$ whether the result is in B or not. How do we now estimate probabilities?
- For any $A \in \Sigma$, there is a Λ-measurable function, written $P[A||\Lambda](\cdot)$ such that for any $B \in \Lambda$ we have:
- $P(A \cap B) = \int_B P[A||\Lambda](x) \mathrm{d}P$.

Disintegration

- Back to the product case: I wrote $P(x, B)$.
- For fixed x it is a probability measure. For fixed B it is a measurable function.
- Not quite! For a fixed countable family of measurable sets we get countable additivity *almost everywhere*.
- But there are lots of countable families; we could end up with something that is not a proper measure anywhere!
- We want something stronger than what RNT promises: regular conditional probabilities or *disintegrations*.
- For disintegrations the statements of (2) are true everywhere.
- How do we construct disintegrations? They can be constructed on spaces that are equipped with *metric* structure.
- A Polish space is the topological space underlying a complete separable metric space. On Polish spaces disintegrations can always be constructed.