NOTES 1: MEASURE THEORETIC FOUNDATIONS OF PROBABILITY THEORY

CHANDRA NAIR

Disclaimer: These notes are created from a variety of great references in the field, most notably, Varadhan's lecture notes and Dembo's lecture notes.

1. Preliminaries

1.1. **Motivation.** *Basic notion*: One wants to assign chances to outcomes or groups of outcomes of an experiment.

If the number of possible outcomes are finite, say $\omega_1, ..., \omega_n$, then a natural way to assign chances or probabilities is to assign for each outcome a number $\mathrm{P}(\omega_i) = p_i$, where $0 \leq p_i \leq 1$ and $\sum_{i=1}^n p_i = 1$. One can extend this line of thought reasonably to even experiments with *countable* number of outcomes.

However things become a bit complicated when the number of outcomes become uncountable. There is no reasonable way to assign positive probabilities to an uncountable number of outcomes, and still make their sum to be 1. To see this consider the following collection of events $\mathcal{E}_k = \{w : \mathrm{P}(w) \geq \frac{1}{k}\}$. Clearly $|\mathcal{E}_k| \leq k$. Let $\mathcal{E} = \cup_k \mathcal{E}_k = \{w : \mathrm{P}(w) > 0\}$. However, it is clear that the number of elements in $\mathcal{E}$ is countable.

A Remedy: Instead of defining probabilities for individual outcomes, one defines probabilities for collections of outcomes (or *events*).

Let $\mathcal{A}$ denote the collection of events for which probabilities are assigned. Then one would like $\mathcal{A}$ to have the following properties:

- if $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$ ("or" event), and $A \cap B \in \mathcal{A}$ ("and" event)
- $\varnothing, \Omega \in \mathcal{A}$
- if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$

Such a collection $\mathcal{A}$ is called an *algebra* or a *field*. We may also require the set of events for which we wish to assign probabilities to have the following additional property:

- If $A_i \in \mathcal{F}$ then $\cup_i A_i \in \mathcal{F}$.

An algebra or a field with (closed under countable unions) is called a *sigma*-algebra or a *sigma*-field.

Aside: Why can't we then assign probabilities to all subsets of outcomes, i.e. to the power set 2^Ω. Let us take $\Omega = \mathbb{R}^3$. Assume that we want the measure to be constructed to

Date: September 9, 2019.

yield the "volume" of the set. There is a rather deep result that says that we must make one of the following concessions:

(1) The volume of a set might change when it is rotated.
(2) The volume of the union of two disjoint sets might be different from the sum of their volumes.
(3) Some sets might be tagged "non-measurable", and one would need to check whether a set is "measurable" before talking about its volume.
(4) The axioms of ZFC (Zermelo-Fraenkel set theory with the axiom of Choice) might have to be altered.

Most probabilists choose to accept (3), i.e. to tag certain sets as "unmeasurable". See https://en.wikipedia.org/wiki/Non-measurable_set

In any case, in this class we will talk about the

1.2. **Definitions.**

Definition 1.1. A σ-algebra is a collection $\mathcal{F}$ of events $A \subset \Omega$ such that

- $\varnothing, \Omega \in \mathcal{F}$
- $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
- $A_i \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is called a *measurable space*.

Definition 1.2. A *measure* μ on the measurable space $(\Omega, \mathcal{F})$ is a mapping $\mu : \mathcal{F} \mapsto [0, \infty]$ satisfying:

(1) $\mu(\varnothing) = 0$.
(2) if $A_i \in \mathcal{F}$ is a countable collection of pairwise disjoint sets, then $\mu(\sqcup_i A_i) = \sum_i \mu(A_i)$.

Definition 1.3. A *probability measure* is a measure that satisfies $\mathrm{P}(\Omega) = 1$.

Exercise 1.1. Show that the following holds for a probability measure:

(1) $\mathrm{P}(\cup_i A_i) \leq \sum_i \mathrm{P}(A_i)$ (sub-additivity)
(2) if $A_i \uparrow A$, then $\mathrm{P}(A_i) \uparrow \mathrm{P}(A)$ (monotone up-limits)
(3) if $A_i \downarrow A$, then $\mathrm{P}(A_i) \downarrow \mathrm{P}(A)$ (monotone down-limits)

Exercise 1.2. Show that $(1), (2)$ holds for general measures, while (3) may not.

(Hint: To see a counter example to (3), define $B_i = p_i \mathbb{N}$ (here p_i is the ith prime) and set $A_i = \mathbb{N} \setminus \cup_{j=1}^{i} B_i$. Use counting measure as the measure and show that $\mu(A_i) = \infty, \forall i$, while $A = \{1\}$, and hence $\mu(A) = 1$.)

Remark 1.1. If A_i is a collection such that $\mathrm{P}(A_i) = 0, \forall i$ then $\mathrm{P}(\cup_i A_i) = 0$ (from subadditivity).

Exercise 1.3. Consider events $\{A_n\}$ in a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ that are almost pairwise disjoint, i.e. $\mathrm{P}(A_n \cap A_m) = 0$ whenever $n \neq m$. Show that

$$\mathrm{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathrm{P}(A_n).$$

The triple $(\Omega, \mathcal{F}, \mathrm{P})$ is called a *probability space*.

Definition 1.4. $(\Omega, \mathcal{F}, \mathrm{P})$ is said to be *non-atomic* if $\forall A$ s.t. $\mathrm{P}(A) > 0, \exists B \subset A, B \in \mathcal{F}$ s.t. $\mathrm{P}(A) > \mathrm{P}(B) > 0$.

Exercise 1.4. If $(\Omega, \mathcal{F}, \mathrm{P})$ is non-atomic and $\mathrm{P}(A) > 0$ then show that

(1) $\forall \epsilon > 0, \exists B \subset A$ s.t. $0 < \mathrm{P}(B) < \epsilon$.
(2) If $0 < a < \mathrm{P}(A)$ then there exists $B \subset A$ s.t. $\mathrm{P}(B) = a$.

Consider a collection of events $\mathcal{A}$. We wish to *extend* the collection to a sigma field. Or in other words is there a smallest σ-field that contains $\mathcal{A}$. The answer is Yes. To see this, the following exercise is useful.

Exercise 1.5. Let $\mathcal{F}_\alpha$ be an arbitrary collection of σ-fields. Then $\cap_\alpha \mathcal{F}_\alpha$ is also a σ-field.

From the above exercise the intersection of all σ-algebras that contain $\mathcal{A}$ is a σ-algebra and is clearly the smallest σ-algebra that contains $\mathcal{A}$. This is denoted as $\sigma(\mathcal{A})$, and called the σ-algebra generated by $\mathcal{A}$.

Supposed Ω is a topological space (i.e. equipped with the notion of open sets). Then the σ-algebra generated by the open sets is called the *Borel σ-algebra*.

Exercise 1.6. Given a collection of sets $\mathcal{A}$, let $\sigma(\mathcal{A})$ denote the smallest σ-field containing the elements of $\mathcal{A}$. Verify the following alternate definitions for Borel σ-field $\mathcal{B}_R$ of reals (i.e. show that all the following σ-fields are identical):

- $\sigma(\{(a, b) : a < b \in \mathbb{R}\})$
- $\sigma(\{[a, b] : a < b \in \mathbb{R}\})$
- $\sigma(\{(-\infty, b) : b \in \mathbb{R}\}$
- $\sigma(\{(-\infty, b) : b \in \mathbb{Q}\})$
- $\sigma(\{\mathcal{O} \subset \mathbb{R}$ is open$\})$.

1.3. **Existance and construction of measures.**

Theorem 1.1. *(Caratheodory Extension Theorem)*
Any countably additive probability measure P on a field $\mathcal{A}$ extends uniquely as a countably additive probability measure on $\mathcal{F} = \sigma(\mathcal{A})$.

Proof. The proof consists of various steps: Here we outline the main steps

1. Define a quantity $\mathrm{P}^*(B), B \in \mathcal{F}$ (an outer measure) as follows:

$$\mathrm{P}^*(B) = \inf_{A_i \in \mathcal{A}, \cup_i A_i \supseteq B} \sum_i \mathrm{P}(A_i).$$

(S.t. without loss of generality can assume A_i to be disjoint.)

Show that

(1) Subadditivity: $\mathrm{P}^*(\cup_i B_i) \leq \sum_i \mathrm{P}^*(B_i)$. (Hint: A collection of "good" covers of B_i is only one possible cover for $\cup_i B_i$.)
(2) if $A \in \mathcal{A}$ then $\mathrm{P}^*(A) \leq \mathrm{P}(A)$ (trivial)
(3) if $A \in \mathcal{A}$ then $\mathrm{P}^*(A) \geq \mathrm{P}(A)$ (Hint: take a good cover of A and use countable additivity of P on $\mathcal{A}$.)

2. Define a set $E \in \mathcal{F}$ to be measurable if for all $B \in \mathcal{F}$

$$\mathrm{P}^*(B) \geq \mathrm{P}^*(B \cap E) + \mathrm{P}^*(B \cap E^c)$$

(clearly from subadditivity, this forces an equality.) Let $\mathcal{E}$ be the set of all measurable sets. Show that $\mathcal{E}$ is a σ-field and that P^* is a countably additive probability measure on $\mathcal{E}$. Finally show that $\mathcal{E} \supseteq \mathcal{A}$ and hence $\mathcal{E} \supseteq \sigma(\mathcal{A})$.

Outline of proof: Clearly $\mathrm{P}^*(\varnothing) = 0$, hence $\varnothing \in \mathcal{E}$. It is also clear that if $E \in \mathcal{E}$ then $E^c \in \mathcal{E}$. Now suppose $E_1, E_2 \in \mathcal{E}$. Then observe that

$$\begin{aligned}
\mathrm{P}^*(B) &= \mathrm{P}^*(B \cap E_1) + \mathrm{P}^*(B \cap E_1^c) \\
&= \mathrm{P}^*(B \cap E_1) + \mathrm{P}^*(B \cap E_1^c \cap E_2) + \mathrm{P}^*(B \cap E_1^c \cap E_2^c) \\
&\geq \mathrm{P}^*(B \cap (E_1 \cup E_2)) + \mathrm{P}^*(B \cap E_1^c \cap E_2^c).
\end{aligned}$$

Here the equalities follow from the measurability of E_1, E_2, and the inequality follows from sub-additivity of $\mathrm{P}^*(\cdot)$. This implies finite unions are measurable. We can conclude that $\mathcal{E}$ is an algebra.

Let E_i be a pairwise disjoint collection. Clearly $G_n = \cup_{i=1}^n E_i$ is measurable. Let $G = \cup_i E_i$. Therefore

$$\begin{aligned}
\mathrm{P}^*(B) &= \mathrm{P}^*(B \cap G_n) + \mathrm{P}^*(B \cap G_n^c) \\
&\geq \mathrm{P}^*(B \cap G_n) + \mathrm{P}^*(B \cap G^c) \\
&= \mathrm{P}^*(B \cap G_n \cap E_n) + \mathrm{P}^*(B \cap G_n \cap E_n^c) + \mathrm{P}^*(B \cap G^c) \\
&= \mathrm{P}^*(B \cap E_n) + \mathrm{P}^*(B \cap G_{n-1}) + \mathrm{P}^*(B \cap G^c) \\
&= \sum_{i=1}^n \mathrm{P}^*(B \cap E_i) + \mathrm{P}^*(B \cap G^c).
\end{aligned}$$

Taking limits $\mathrm{P}^*(B) \geq \sum_{i=1}^{\infty} \mathrm{P}^*(B \cap E_i) + \mathrm{P}^*(B \cap G^c) \geq \mathrm{P}^*(B \cap G) + \mathrm{P}^*(B \cap G^c)$, where the last inequality follows from sub-additvity. Hence $G = \cup_i E_i$ is measurable or $\mathcal{E}$ is an σ-algebra.

To show that $\mathrm{P}^*(\cdot)$ is a countably additive probability measure on $\mathcal{E}$, note the following. Let E_i be a pairwise disjoint collection and let $G = \cup_i E_i$. Then

$$\begin{aligned}\mathrm{P}^*(E_1 \cup E_2) &= \mathrm{P}^*((E_1 \cup E_2) \cap E_1) + \mathrm{P}^*((E_1 \cup E_2) \cap E_1^c) \\ &= \mathrm{P}^*(E_1) + \mathrm{P}^*(E_2).\end{aligned}$$

Hence $\mathrm{P}^*(\cdot)$ is a finitely additive probability measure on $\mathcal{E}$. Countable additivity is a simple consequence of subadditivity and the following:

$$\mathrm{P}^*(G) \geq \mathrm{P}^*(G_n) = \sum_{i=1}^{n} \mathrm{P}^*(E_i).$$

Hence $\mathrm{P}^*(G) = \sum_{i=1}^{\infty} \mathrm{P}^*(E_i)$.

To show that $\mathcal{A} \subseteq \mathcal{E}$, for any B take a "good" cover of B, and show that this induces a cover on $B \cap A$ and $B \cap A^c$. Hence show that

$$\mathrm{P}^*(B) \geq \mathrm{P}^*(B \cap A) + \mathrm{P}^*(B \cap A^c).$$

Thus $\mathcal{A} \subseteq \mathcal{E}$ and since $\mathcal{E}$ is a sigma-algebra, we also have $\sigma(\mathcal{A}) \subseteq \mathcal{E}$.

3. To show uniqueness, show the following: Let $\mathcal{M} = \{B : p_1(B) = p_2(B)\}$ be the collection of all sets in which the two countably additive probability extensions agree. Then it is clear that $\mathcal{M}$ is a monotone class (i.e. contains monotone limits). As with σ-algebras, arbitrary intersection of monotone class is also a monotone class, hence it makes sense to talk of Monotone class *generated* by a collection of sets. The proof is then completed by showing the following proposition. □

Remark 1.2. An example of a monotone class on the reals is the collection $\{(0,1], (2,3]\}$. This is clearly not closed under unions or intersections. However increasing unions can only be formed by taking identical elements and hence it is a monotone class.

Proposition 1.2 (Monotone Class Theorem)**.** *Let $\mathcal{A}$ be an algebra. Let $\mathcal{M}(\mathcal{A})$ be the monotone class generated by $\mathcal{A}$. Then $\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A})$.*

Proof. Since $\sigma(\mathcal{A})$ is closed under monotone limits, it is clear that $\sigma(\mathcal{A}) \supseteq \mathcal{M}(\mathcal{A})$. Therefore suffices to show that $\mathcal{M}(\mathcal{A})$ is a σ-algebra. Since $\varnothing \in \mathcal{A}$ we have $\varnothing \in \mathcal{M}(\mathcal{A})$.

Let $\mathcal{M}_0 = \{B : B \in \mathcal{M}(\mathcal{A}) \text{ such that } B^c \in \mathcal{M}(\mathcal{A})\}$. Clearly $\mathcal{A} \subseteq \mathcal{M}_0$. Since $\mathcal{M}(\mathcal{A})$ is closed under monotone limits, it is clear that $\mathcal{M}_0$ is also a monotone class, hence $\mathcal{M}_0 = \mathcal{M}(\mathcal{A})$. Thus $B \in \mathcal{M}(\mathcal{A})$ implies that $B^c \in \mathcal{M}(\mathcal{A})$.

To complete the proof, it suffices to show that $\mathcal{M}(\mathcal{A})$ is a field. Towards this, fix $A \in \mathcal{A}$, and define

$$\mathcal{M}_A = \{B \in \mathcal{M} : A \cap B \in \mathcal{M}\}.$$

Clearly $\mathcal{A} \subseteq \mathcal{M}_A$ and $\mathcal{M}_A$ is a monotone class. Hence $\mathcal{M}_A = \mathcal{M}(\mathcal{A})$.

Now fix $C \in \mathcal{M}(\mathcal{A})$ and define

$$\mathcal{M}_C = \{B \in \mathcal{M} : C \cap B \in \mathcal{M}\}.$$

Clearly $\mathcal{A} \subseteq \mathcal{M}_C$ and $\mathcal{M}_C$ is a monotone class. Hence $\mathcal{M}_C = \mathcal{M}(\mathcal{A})$.

Therefore $\mathcal{M}(\mathcal{A})$ is closed under finite unions and intersections, hence an algebra. Now closure under monotone limits follows from definition implying that $\mathcal{M}(\mathcal{A})$ is a σ-algebra. □

Thus Caratheordory's extension theorem reduces the burden of construction of measures on σ-algebras to those on algebras.

1.3.1. *Constructing countably additive probability measures on algebras.* We will now see that there is a canonical way of constructing countably additive probability measure on $\mathcal{F}_B$, the borel σ-algebra on real line.

Consider the following collection of intervals: $\mathcal{I} = \{I_{a,b} : -\infty \leq a < b \leq \infty\}$, where $I_{a,b} = (a, b]$ when $b < \infty$ and $I_{a,\infty} = (a, \infty)$.

Exercise 1.7. Show that the class, $\mathcal{A}_B$, of finite disjoint union of members of $\mathcal{I}$ is an algebra.

Assume we are given a function $F(x)$ which is nondecreasing, right-continuous, and satisfies

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} F(x) = 1.$$

Then we can define a finitely additve P by first defining $P(I_{a,b}) = F(b) - F(a)$ for intervals, and then extending it to $\mathcal{A}_B$ by defining it as sum for disjoint unions from $\mathcal{I}$.

We will now prove Lebesgue's theorem which shows when one can extend the finitely additive P to a countably additive P.

Theorem 1.3. *(Lebesgue)*
Let P be a finitely additive probability measure on $\mathcal{A}_B$. P is countably additive on $\mathcal{A}_B$ if and only if $F(x) = P((-\infty, x])$ is right continuous function of x.

Remark 1.3. Essentially this means (using Caratheodory's theorem) that for every right continuous function non-decreasing $F(x)$ that satisfies $F(-\infty) = 0$ and $F(\infty) = 1$, there is a unique countably additive probability measure on $\mathcal{F}_B$; and conversely every countably additive probability measure on $\mathcal{F}_B$ comes from a right continuous function.

Proof. Suppose P is countably additive on $\mathcal{A}$. Then for any x and $\{\epsilon_n\} \downarrow 0$, the collection of intervals $J_n = (x, x + \epsilon_n]$ decreases to $\varnothing$. This means that $F(x + \epsilon_n) - F(x) \downarrow 0$ and this suffices (why? hint: $F(x)$ is non-decreasing).

The tricky part is clearly the reverse direction, i.e. getting countable additivity from right continuous and non-decreasing $F(x)$ with $F(-\infty) = 0$ and $F(\infty) = 1$. Suppose $A_j \in \mathcal{A}, A_j \downarrow \varnothing$. Assume that $\mathrm{P}(A_j) \geq \delta > 0$. (We wish to show a contradiction). Now pick l large enough that $1-F(l)+F(-l) < \frac{\delta}{2}$, and define $B_j = A_j \cap (-l, l]$. Clearly $\mathrm{P}(B_j) \geq \frac{\delta}{2}, \forall j$.

Since B_j is a finite disjoint union of left open right closed intervals, create $C_j \subset B_j$ by moving the left (open) end point of the intervals to the right (i.e. shortening each interval). Clearly this can be done so as to guarantee that

$$\mathrm{P}(B_j \setminus C_j) \leq \frac{\delta}{3 \cdot 2^j},\ \forall j.$$

Define $D_j = closure(C_j)$. Clearly $D_j \subset B_j \subset A_j$.

We know that B_j's are decreasing but C_j's may not be. Therefore define $E_j = \cap_{i=1}^{j} C_i$, and $F_j = \cap_{i=1}^{j} D_i$. Clearly $F_j \supset E_j$ (by construction), and observe that

$$\begin{aligned}\mathrm{P}(E_j) &\geq \mathrm{P}(B_j) - \mathrm{P}(B_j \cap E_j^c) \geq \mathrm{P}(B_j) - \sum_{i=1}^{j} \mathrm{P}(B_j \cap C_i^c) \\ &\geq \mathrm{P}(B_j) - \sum_{i=1}^{j} \mathrm{P}(B_i \cap C_i^c) \geq \frac{\delta}{2} - \frac{\delta}{3} > 0.\end{aligned}$$

Therefore F_j is non-empty. Thus F_j is non-empty, closed, bounded and decreasing. Thus $\cap_j F_j$ cannot be the $\varnothing$ (finite intersection property). However $F_j \subset A_j$ and $A_j \downarrow \varnothing$, leading to a contradiction.(!) $\square$

$F(\cdot)$ is *the distribution function* corresponding to the probability measure P.