PROBABILITY THEORY: LECTURE NOTES 5

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.

7. Signed Measure

Definition 1. Let $(\Omega, \mathcal{F})$ be a measurable space. A *finite signed-measure* is a mapping of $\mathcal{F}$ to $\mathbb{R}$ satisfying: if $\{A_i\}$ are pairwise disjoint; then $\lambda(\cup_i A_i) = \sum_i \lambda(A_i)$.

Note that $\lambda(\emptyset) = 0$.

Lemma 1. *Let λ be a finite signed measure, then* $\sup_{A \subseteq \mathcal{F}} |\lambda(A)| < \infty$.

Proof. For any set A, let $\lambda_+(A) := \sup_{B \subset A} |\lambda(B)|$. If there exists any set A such that $\lambda_+(A)$ and $\lambda_+(A^c)$ are both finite, we are done. Assume otherwise, i.e. for every set A, at least one of $\lambda_+(A)$ or $\lambda_+(A^c)$ is infinite.

Take a set A such that $\lambda_+(A) = \infty$. Let $A_1 \subset A$ such that $|\lambda(A_1)| > n(1+|\lambda(A)|)$ and $\lambda_+(A_1) = \infty$. Such A_1 exists because: for any $B \subseteq A$, if $|\lambda(B)| \geq k(1+|\lambda(A)|)$, then

$$|\lambda(A \setminus B)| \geq |\lambda(B)| - |\lambda(A)| \geq k + (k-1)|\lambda(A)| \geq (k-1)(1 + |\lambda(A)|).$$

Further $\lambda_+(A) \leq \lambda_+(B) + \lambda_+(A \setminus B)$.

Repeat. Take $A_2 \subset A_1$ such that $|\lambda(A_2)| > n(1 + |\lambda(A_1)|) > n^2(1 + |\lambda(A)|)$ and $\lambda_+(A_2) = \infty$.

Therefore by induction, we have a decreasing sequence of sets $A_n \downarrow A_*$. Hence by countable additivity $\lambda(A_n) \to \lambda(A_*)$; however $|\lambda(A_n)| \to \infty$, contradicting finiteness of $\lambda(A_*)$. □

Definition 2. A set A is called *totally positive*, if for every $B \subseteq A$ we have $\lambda(B) \geq 0$.

Lemma 2. *If $\lambda(A) = l \geq 0$, then these exists a totally positive set $A_+ \subseteq A$ such that $\lambda(A_+) \geq l$.*

Proof. Let $m = \inf_{B \subseteq A} \lambda(B)$. w.l.o.g. $m < 0$ (else $m = 0$ and set $A_+ = A$). Let B_1 satisfy $\lambda(B_1) < \frac{m}{2}$. Set $A_1 = A \setminus B_1$. Clearly $\lambda(A_1) \geq l$ and $m_1 = \inf_{B \subseteq A_1} \lambda(B) > \frac{m}{2}$.

repeat. Take $B_2 \subseteq A_1$ such that $\lambda(B_1) < \frac{m_1}{2}$, and set $A_2 = A_1 \setminus B_2$. Note that $m_2 = \inf_{B \subseteq A_2} \lambda(B) > \frac{m_1}{2} > \frac{m}{4}$. Further $\lambda(A_2) \geq l$.

Proceeding similarly we have that $A_n \downarrow A_*$ such that $m_k = \inf_{B \subseteq A_2} \lambda(B) > \frac{m}{2^k}$ $\forall k$ and $\lambda(A_k) \geq l$. Hence $\lambda(A_*) \geq l$ and $m_* = \inf_{B \subseteq A_*} \lambda(B) \geq 0$. This A_* is the required totally positive subset of A. □

Note that the class of totally positive sets is closed under countable unions. Further any subset of a totally positive set is also totally positive.

Date: November 4, 2019.

Lemma 3. *There is a partition of Ω into a totally positive set Ω_+ and a totally negative set Ω_-.*

Proof. Let $m = \lambda_+(\Omega) := \sup_{A \subseteq \Omega} \lambda(A)$. Assume $m > 0$ (else set $\Omega_+ = \emptyset$). Then pick a totally positive A_1 such that $\lambda(A_1) > \frac{m}{2}$, and set $B_1 = \Omega \setminus A_1$. It is clear that $m_1 = \lambda_+(B_1) \leq \frac{m}{2}$. Now pick a totally positive $A_2 \subset B_1$ such that $\lambda(A_2) \geq \frac{m_1}{2}$, and set $B_2 = B_1 \setminus A_2$. Note that $m_2 = \lambda_+(B_2) \leq \frac{m_1}{2} \leq \frac{m}{4}$.

Define $\Omega_+ = \cup_i A_i$, and let $B_i \downarrow \Omega_-$. Note that Ω_+ and Ω_- partition Ω and Ω_+ is a totally positive set and Ω_- is a totally negative set. □

Remark: Note that this partitioning is not necessarily unique; however any two partitions differ by a set that is both totally positive and totally negative (hence it and all its subsets have measure 0). Let us call the sets that are totally positive and totally negative to be totally zero sets.

Therefore every finite signed measure λ induces two non-negative finite measures defined by: $\mu_+(A) = \lambda(A \cap \Omega_+)$, and $\mu_-(A) = -\lambda(A \cap \Omega_-)$. Further $\lambda(A) = \mu_+(A) - \mu_-(A)$.

Let $f(\omega)$ be an integrable function. Then define

$$\lambda(A) := \int f 1_A d\mu.$$

Observe that λ is a signed measure. (Countable additivity follows from dominated convergence theorem.) Further if $\mu(A) = 0$ then $\lambda(A) = 0$. To see this, let $\mu(A) = 0$ and note that

$$0 = n\mu(A) \geq \int (|f| \wedge n) 1_A d\mu \uparrow \int |f| 1_A d\mu \geq |\lambda(A)|.$$

Definition 3. A measure λ is said to be *absolutely continuous* with respect to another measure μ, denoted by $\lambda \ll \mu$, if $\mu(A) = 0$ implies $\lambda(A) = 0$.

Theorem 1 (Radon-Nikodym)**.** *Let λ be a finite-signed-measure on $(\Omega, \mathcal{F})$ and μ be a non-negative measure on $(\Omega, \mathcal{G}), \mathcal{G} \supseteq \mathcal{F}$, such that $\mu(\Omega) < \infty$. If $\lambda \ll \mu$, then there exists a integrable function f, measurable w.r.t. $\mathcal{F}$ such that*

$$\lambda(A) = \int f 1_A d\mu, \quad \forall A \in \mathcal{F}.$$

Proof. For every $q \in \mathbb{Q}$, define the finite-signed-measure on $(\Omega, \mathcal{F})$

$$\lambda_q(A) = \lambda(A) - q\mu(A).$$

Let Ω_+^q be a totally positive partition of Ω induced by λ_q. Further by suitably discarding totally zero sets, we can have Ω_+^q to be a decreasing sequence of sets in q. (Argue why?). Once we have this collection, define

$$f(\omega) = \sup\{q : \omega \in \Omega_+^q\}.$$

To show the measurability of f, note that

$$\{\omega : f(\omega) > x\} = \cup_{q > x} \Omega_+^q.$$

Nest step is to show finiteness of $f(\omega)$ almost everywhere. Let $A = \cap_q \Omega_+^q$. Since $\lambda_q(A) \geq 0 \ \forall q$, we have $\lambda(A) \geq q\mu(A) \ \forall q$, which can happen only if $\mu(A) = 0$ (by finiteness of λ), and by absolute continuity $\lambda(A) = 0$ as well.

Suppose there is a set A such that $A \cap \Omega_q^+ = \emptyset$ for all $q \in Q$. This implies that $A \subseteq (\Omega_q)^c$ which is essentially same as Ω_q^-. Hence $\lambda(A) - q\mu(A) \leq 0$ for all q. This can only happen again only if $\mu(A) = 0$ and hence $\lambda(A) = 0$. Thus $f(\omega)$ is finite on a set of measure 1.

For any two real number $a < b$, consider the set

$$I_{[a,b]} \subseteq \{\omega : f(\omega) \in [a,b]\}.$$

Therefore $I_{[a,b]} \subseteq \Omega_a^+ \ \forall q \leq a$ and $I_{[a,b]} \subseteq (\Omega_q)^- \ \forall q \geq b$. This implies that $\lambda(I_{[a,b]}) - q\mu(I_{[a,b]}) \geq 0 \ \forall q \leq a$ and $\lambda(I_{[a,b]}) - q\mu(I_{[a,b]}) \leq 0 \ \forall q \geq b$. Hence $a\mu(I_{[a,b]}) \leq \lambda(I_{[a,b]}) \leq b\mu(I_{[a,b]})$.

Consider a grid and set

$$I_n = \{\omega : nh \leq f(\omega) \leq (n+1)h\}, \ -\infty \leq n \leq \infty.$$

Now, note that for all $A \in \mathcal{F}$ and every n

$$\begin{aligned} \lambda(A \cap I_n) - h\mu(A \cap I_n) \leq nh\mu(A \cap I_n) \leq \int_{A \cap I_n} f d\mu \\ \leq (n+1)h\mu(A \cap I_n) \leq \lambda(A \cap I_n) + h\mu(A \cap I_n). \end{aligned} \tag{7.1}$$

Take $A_+ = \{\omega : f(\omega) \geq 0\}$.

Summing over n (and using countable additivity and monotone convergence theorem) we obtain

$$\lambda(A_+) - h\mu(A_+) \leq \int_{A_+} f d\mu = \int_\Omega f_+ d\mu \leq \lambda(A_+) + h\mu(A_+).$$

This shows that $\int f_+ < \infty$. Similarly $\int f_- < \infty$; thus f is integrable.

Now take a generic A and sum (7.1) over n and (and using countable additivity and dominated convergence theorem) we get

$$\lambda(A) - h\mu(A) \leq \int_A f d\mu \leq \lambda(A) + h\mu(A).$$

Taking $h \to 0$ completes the proof. □

Remark 1. Note that the Radon-Nikodym derivative is essentially unique. To see this, suppose f, d were two derivatives, let $A_\epsilon = \{\omega : f(\omega) - g(\omega) \geq \epsilon\}$. Then since $\mu(A_\epsilon) = \int_{A_\epsilon} f d\mu = \int_{A_\epsilon} g d\mu$, which on the other hand $\int_{A_\epsilon}(f-g)d\mu \geq \epsilon\mu(A_\epsilon)$ we must have $\mu(A_\epsilon) = 0$. This shows that the functions match almost surely.

8. Conditional Expectation

Proposition 1. *Let f be an integrable function defined on $(\Omega, \mathcal{F}, \mu)$. Let $\mathcal{G} \subset \mathcal{F}$. Then there is a $\mathcal{G}$-measurable and integrable function g such that*

$$\int_A f d\mu = \int_A g d\mu \quad \forall g \in \mathcal{G}.$$

Further if g_1 and g_2 are two such functions, then $g_1 = g_2$ almost surely.

Proof. Define a finite-signed-measure on $(\Omega, \mathcal{G})$ by

$$\lambda(A) = \int_A f d\mu \ \forall A \in \mathcal{G}.$$

Note that $\lambda \ll \mu$. Hence by Radon-Nikodym theorem, there exists an integrable function $\hat{g} = \frac{d\lambda}{d\mu}$ that is $\mathcal{G}$-measurable such that

$$\lambda(A) = \int_A \hat{g} d\mu \ \forall A \in \mathcal{G}.$$

Setting $g = \hat{g}$ completes the first part.

Now let $A_\epsilon = \{\omega : g_1(\omega) - g_2(\omega) > \epsilon\}$. Since A_ϵ is $\mathcal{G}$ measurable for every $\epsilon > 0$, we must have

$$\int_{A_\epsilon} g_1 d\mu = \int_{A_\epsilon} g_2 d\mu \quad \implies \mu(A_\epsilon) = 0.$$

Now $A_+ = \{\omega : g_1(\omega) - g_2(\omega) > 0\} = \cup_n A_{1/n}$ and hence $\mu(A_+) = 0$. Similarly, by symmetry, we see that $\mu(A_-) = 0$ where $A_- = \{\omega : g_2(\omega) - g_1(\omega) > 0\} = \cup_n A'_{1/n}$. □

Thus we denote any $g(\omega)$ that satisfies the proposition as the *conditional expectation* $E(f|\mathcal{G})$.

Theorem 2 (Properties of Conditional Expectation)**.** *Let f be integrable and $(\Omega, \mathcal{F})$-measurable. Let $\mathcal{G} \subset \mathcal{F}$. The following properties hold:*

(1) $\mathrm{E}(f) = \mathrm{E}(\mathrm{E}(f|\mathcal{G}))$.

(2) $f \geq 0$ *implies that* $\mathrm{E}(f|\mathcal{G}) \geq 0$ *a.s.*

(3) If f_1, f_2 be integrable and $(\Omega, \mathcal{F})$-measurable then

$$\mathrm{E}(af_1 + bf_2|\mathcal{G}) = a\,\mathrm{E}(f_1|\mathcal{G}) + b\,\mathrm{E}(f_2|\mathcal{G}) \text{ a.s.}$$

(4) $\mathrm{E}(|f|) \geq \mathrm{E}(|\,\mathrm{E}(f|\mathcal{G})|)$

(5) If h is bounded and $\mathcal{G}$-measurable, then

$$\mathrm{E}(hf|\mathcal{G}) = h\,\mathrm{E}(f|\mathcal{G}) \text{ a.s.}$$

(6) If $\mathcal{G}_1 \subseteq \mathcal{G}$ then

$$\mathrm{E}(f|\mathcal{G}_1) = \mathrm{E}(\mathrm{E}(f|\mathcal{G})|\mathcal{G}_1) \text{ a.s.}$$

(7) (Jensen's inequality) If Φ is a convex function then

$$\mathrm{E}(\Phi(f)|\mathcal{G}) \geq \Phi(\mathrm{E}(f|\mathcal{G})) \text{ a.s.}$$

Proof. The proofs are rather straightforward

(1) Take $A = \Omega$ and apply definition.

(2) For $\epsilon > 0$, let $A_\epsilon^- = \{\omega : \mathrm{E}(f|\mathcal{G}) \leq -\epsilon\}$. Then argue that $\mu(A_\epsilon^-) = 0$. Take $\epsilon_n = \frac{1}{n} \downarrow 0$.

(3) Follows from definition and linearity of expectation.

(4) Let $A_+ = \{\omega : \mathrm{E}(f|\mathcal{G}) \geq 0\}$. Since $A_+ \in \mathcal{G}$, by definition,

$$\int_{A_+} \mathrm{E}(f|\mathcal{G}) d\mu = \int_{A_+} f d\mu \leq \int_{A_+} |f| d\mu.$$

Similarly we can consider $A_- = \{\omega : \mathrm{E}(f|\mathcal{G}) < 0\}$. Again

$$-\int_{A_-} \mathrm{E}(f|\mathcal{G}) d\mu = -\int_{A_-} f d\mu \leq \int_{A_-} |f| d\mu.$$

Adding them yields the desired result.

(5) If $h = 1_A$ for $A \in \mathcal{G}$, this is immediate from definition. Linearity extends it to simple functions. Let h_n be simple and $h_n \to h$ uniformly. Let $g_n = \mathrm{E}(h_n f|\mathcal{G})$ and $g = \mathrm{E}(hf|\mathcal{G})$. Note that from 4)

$$\int |g_n - g| dP \leq \int (|h_n - h||f|) dP \to 0.$$

Therefore for any fixed $A \in \mathcal{G}$

$$\int_A g dP \leftarrow \int_A g_n dP = \int_A h_n f dP \rightarrow \int_A h f dP.$$

(6) This follows from definitions easily. (Check!)
(7) The key is again to write a convex function as the pointwise supremum of a set of affine functions. For each such affine function we have

$$\Phi(f) \geq a_\alpha f + b_\alpha \quad \implies \mathrm{E}(\Phi(f)|\mathcal{G}) \geq a_\alpha \mathrm{E}(f|\mathcal{G}) + h_\alpha.$$

Now taking supremum over the class of affine functions that yields Φ implies the result.

□

8.1. **Conditional Probability.** The goal of this section is to define a *conditional probability* distribution which is defined below.

Definition 4. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{G} \subseteq \mathcal{F}$. A mapping $\mu : \Omega \times \mathcal{F} \to [0,1]$ is called a *regular conditional probability* if it satisfies the following:
(1) For every $A \in \mathcal{F}$, $\mu(\omega, A) = \mathrm{E}(1_A|\mathcal{G})$
(2) For almost every ω, $\mu(\omega, A)$ is a probability measure on $(\Omega, \mathcal{F})$.

It turns out that these regular conditional probabilities need not always exist. However, if the space is "nice", then they do exist. When is a space "nice": If Ω is a Polish space, $\mathcal{F}$ is its Borel σ-algebra, and $\mathcal{G}$ is a countably generated sub σ-algebra of $\mathcal{F}$, then the space is "nice" enough.

Clearly, an initial approach would be to define for every $A \in \mathcal{F}$, $\mu(\omega, A) := \mathrm{E}(1_A|\mathcal{G})$. This would meet condition (1) above, but it is not necessary that for every ω, and for $A \subset B \in \mathcal{F}$ we would have $\mu(\omega, A) \leq \mu(\omega, B)$. Of course, we can throw away a set of measure zero and make the previous inequality hold. However of there are uncountable such pairs of sets, then this would cause issues. On the other hand, if the space is nice and if the σ-algebra is countably generated then we can try to avoid the above issue. We will work with $\Omega = [0,1]$ and $\mathcal{F}$ to be the Borel σ-algebra.

We consider the collection $A_q = (-\infty, q]$ for $q \in Q$, and we define

$$\mu(\omega, A_q) = \mathrm{E}(1_{A_q}|\mathcal{G}) \quad q \in Q.$$

Since this collection is countable, we know that we can obtain a set S with $P(S) = 1$ such that for all $\omega \in S$, $\mu(\omega, A_q)$ is non-decreasing in q, and $\mu(\omega, A_q) = 1$ for all $q \geq 1$ and $\mu(\omega, A_q) = 0$ for all $q < 0$.

Consider the set $A_y = (-\infty, y], y \in \mathbb{R}$ and define

$$\mu(\omega, A_y) = \inf_{q > y} \mu(\omega, A_q).$$

Clearly, the above $F_\omega(y) := \mu(\omega, A_y)$ is a distribution function for every $\omega \in S$. Since this is in natural 1-1 correspondence with a probability measure (see Lebesgue's theorem) on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ we set $\mu(\omega, A)$ to be that probability measure. Hence we are left to show that this choice also satisfies the first condition.

Observe that, by construction, if $A = (-\infty, q], q \in Q$ then the first condition is satisfied. Consider the interval $A_y = (-\infty, y]$ for y in reals. Then for any $q > y$ we have

$$\int 1_{A_q} dP = \int \mu(\omega, A_q) dP.$$

Let $q \downarrow y$, monotone convergence theorem (on both sides) implies that

$$\int 1_{A_y} dP = \int \mu(\omega, A_y) dP.$$

Hence from linearity of conditional expectation, equality holds on the algebra comprised of finite disjoint unions of intervals of the form $(a, b]$. Now consider the collection, $\mathcal{S}$, of all sets for which

$$\int 1_A dP = \int \mu(\omega, A) dP.$$

This is a monotone class, since the limit on the left hand side is argued using monotone convergence theorem. On the left hand side, for every $\omega \in S$ the limit exists and now dominated convergence theorem implies that $\mathcal{S}$ is a monotone class. Hence $\mathcal{S}$ contained the σ-filed generated by the algebra of finite disjoint union of left-open right-closed intervals, completing the proof.