STAT 7200

Introduction to Advanced Probability
Lecture 15

Taylor R. Brown

"A First Look at Rigorous Probability Theory" (Jeffrey Rosenthal) Section 12.1

Motivation

Let μ be a probabilty measure on R. We have seen examples of probability measures for discrete random variables, continuous random variables, and discrete/continuous mixture random variables. Are these the only types?

Goal: prove Lebesgue's Decomposition. Write any measure as the sum of three components: discrete, absolutely continuous and singular.

Roadmap: Definitions $\rightarrow$ Hahn's Decomposition theorem $\rightarrow$ Lebesgue's Decomposition Theorem (and Radon-Nikodym theorem as a corollary!)

Definitions

2.1: Definition: Measures dominating other measures

Let μ and λ be two measures. We say μ is dominated by λ if for any $B \in \mathcal{B}$, $\lambda(B) = 0 \implies \mu(B) = 0$. It is written as $\mu \ll \lambda$.

2.2: Definition: Absolutely Continuous

Let μ and λ be two measures. We say μ is absolutely continuous with respect to λ if for any $B \in \mathcal{B}$, $\mu(B) = \int_B f d\lambda$ for some $f \geq 0$.

2.3: Definition: signed measure

Let $(\Omega, \mathcal{F})$ be a measurable space. Then $\phi : \mathcal{F} \to \mathrm{R}$ is a signed measure if 1.) $\phi(\emptyset) = 0$, and 2.) ϕ is countably additive.

Decomposing Ω

2.4: Lemma 12.1.4: Hahn's Decomposition

Let ϕ be a signed measure on $(\Omega, \mathcal{F})$. Then there exists a two-set partition of Ω: $A^+, A^- \in \mathcal{F}$ such that whenever $E \subseteq A^+$, we have $\phi(E) \geq 0$, and whenever $F \subseteq A^-$, we have $\phi(F) \leq 0$.

Hahn Decomposition: Proof

Define

$$\alpha = \sup \{\phi(A) : A \in \mathcal{F}\}.$$

We want to find A^+ such that $\phi(A^+) = \alpha$.

Choose $A_1, A_2, \ldots \in \mathcal{F}$ such that $\phi(A_i) \to \alpha$.
Then define

$$\mathcal{G}_n = \left\{ \bigcap_{k=1}^{n} A'_k : A'_k = A_k \text{ or } A'_k = [\cup_{i=1}^{\infty} A_i] \setminus A_k \right\}.$$

Then define

$$C_n = \bigcup_{S \in \mathcal{G}_n \text{ and } \phi(S) \geq 0} S$$

and set

$$A^+ = \limsup_n C_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} C_k$$

Hahn Decomposition: Proof

For any n, $\phi(C_n) \geq \phi(A_n)$.

Next,

$$\phi\left(C_m \cup \cdots \cup C_{n-1} \cup C_n\right) \geq \phi\left(C_m \cup \cdots \cup C_{n-1}\right)$$

therefore

$$\phi\left(C_m \cup \cdots \cup C_{n-1} \cup C_n\right) \geq \phi\left(C_m\right) \geq \phi\left(A_m\right).$$

Taking the limit on both sides, by continuity of ϕ:

$$\phi\left(\bigcup_{n=m}^{\infty} C_n\right) \geq \phi\left(A_m\right)$$

Finally

$$\phi(A^+) = \phi(\limsup_m C_m) = \lim_{m\to\infty} \phi\left(\bigcup_{k\geq m}^{\infty} C_k\right) \geq \lim_{m\to\infty} \phi(A_m) = \alpha.$$

Hahn Decomposition: Proof

We had $\phi(A^+) \geq \alpha$ from the previous slide. By definition of sup, $\phi(A^+) \leq \alpha$, so $\phi(A^+) = \alpha$.

Take $E \subseteq A^+$. Assume to the contrary that $\phi(E) < 0$. Then $\phi(A^+ \setminus E) = \phi(A^+) - \phi(E) > \phi(A^+) = \sup\{\phi(A) : A \in \mathcal{F}\}$, which contradicts the definition of supremum.

Define $A^- = \Omega \setminus A^+$. Again, we can show that for any $E \subseteq A^-$, we have $\phi(E) \leq 0$. Assume to the contrary that $\phi(E) > 0$. Then, by additivity of ϕ: $\phi(A^+ \cup E) = \phi(A^+) + \phi(E) > \phi(A^+)$. Again, this is a contradiction, so $\phi(E) \leq 0$.

Decomposing Probability Measures

Before we talk about the theorem that decomposes probability measures, we need another definition:

2.5: Definition: Singular Measures

Let μ, ν be two positive measures defined on some probability space $(\Omega, \mathcal{F})$. μ is **singular** with respect to ν, written $\mu \perp \nu$, if there exists an $S \in \mathcal{F}$ such that $\mu(S) = 0$ and $\nu(S^c) = 0$.

Intuitively, the two measures put probability on different places, and these two places partition the sample space.

Decomposing Probability Measures

2.6: Lemma 12.1.1: Lebesgue's Decomposition

Any probability measure μ on R can uniquely be decomposed as

$$\mu = \mu_{\text{disc}} + \mu_{\text{ac}} + \mu_{\text{sing}},$$

where

1. $\mu_{\text{disc}}(\mathrm{R}) = \sum_{x \in \mathrm{R}} \mu_{\text{disc}}(\{x\})$
2. $\mu_{\text{ac}}(A) = \int_A f d\lambda$ for any $A \in \mathcal{B}$, where f is a nonnegative, Borel-measurable function f, and λ is the Lebesgue measure on R, and
3. $\mu_{\text{sing}}(\{x\}) = 0$ for all $x \in \mathrm{R}$ but $\exists S \subseteq \mathrm{R}$ such that $\lambda(S) = 0$ and $\mu_{\text{sing}}(S^c) = 0$.

Lebesgue Decomposition: Proof

The first part is easy. We can just define

$$\mu_{\text{disc}}(A) := \sum_{x \in A} \mu(\{x\})$$

for any $A \in \mathcal{B}$.

In this way, $\mu - \mu_{\text{disc}}$ has no discete part:

$$\sum_{x \in \mathrm{R}} [\mu - \mu_{\text{disc}}](\{x\}) = 0.$$

Without loss of generality, write μ for $\mu - \mu_{\text{disc}}$.

Lebesgue Decomposition: Proof

Now we construct the ac part by finding a density $f \geq 0$ such that

$$\mu_{\text{ac}}(A) = \int_A f d\lambda$$

for any $A \in \mathcal{B}$. Before we do that, we he have to define a **candidate density**.

2.7: Definition: candidate density

$g : \mathrm{R} \to \mathrm{R}^+$ is a **candidate density** if, for all $E \in \mathcal{B}$:

$$\mu(E) \geq \int_E g d\lambda$$

Lebesgue Decomposition: Proof

If g_1 and g_2 are candidate densities, then so is $\max\{g_1, g_2\}$ because

$$\begin{aligned}
\int_E \max\{g_1, g_2\} d\lambda &= \int_{E \cap \{g_1 \le g_2\}} g_2 d\lambda + \int_{E \cap \{g_1 > g_2\}} g_1 d\lambda \\
&\le \mu(E \cap \{g_1 \le g_2\}) + \mu(E \cap \{g_1 > g_2\}) \\
&= \mu(E)
\end{aligned}$$

Also, if $h_n \nearrow h$ pointwise, and each h_n is a candidate density, then so is h because

$$\int_E h d\lambda = \lim_{n \to \infty} \int_E h_n d\lambda \le \mu(E).$$

So, for any arbitrary collection of candidate densities $\{g_n\}$, $\lim_{n \to \infty} \max\{g_1, \ldots, g_n\} = \sup_n g_n$ is a candidate density.

Lebesgue Decomposition: Proof

Define

$$\beta = \sup \left\{ \int_{\mathsf{R}} g d\lambda : g \text{ is a candidate density} \right\}.$$

For each $n \in \mathbb{N}$, select g_n such that $\beta - n^{-1} < \int_{\mathsf{R}} g_n d\lambda$.

Finally, choose $f = \sup_{n \geq 1} g_n$. Clearly $\int_{\mathsf{R}} f d\lambda = \beta$.

Last, define

$$\mu_{\text{ac}}(A) = \int_A f d\lambda,$$

for any $A \in \mathcal{B}$

Lebesgue Decomposition: Proof

Finally, define $\mu_{\text{sing}} = \mu - \mu_{\text{ac}}(A)$. We have to show that $\mu_{\text{sing}} \perp \lambda$.

For each $n \in \mathbb{N}$, define $\phi_n = \mu_{\text{sing}} - n^{-1}\lambda$.

Let A_n^+, A_n^- be the Hahn decomposition for ϕ_n, and call $M = \bigcup_{n=1}^{\infty} A_n^+$.

$\cap_{n=1}^{\infty} A_n^- = M^c \subseteq A_n^-$, so, for all $n \in \mathbb{N}$, $0 \leq \mu_{\text{sing}}(M^c) \leq n^{-1}\lambda(M^c)$. So $\mu_{\text{sing}}(M^c) = 0$.

Now we have to show that $\lambda(M) = 0$.

Lebesgue Decomposition: Proof

Now we have to show that $\lambda(M) = 0$. Assume to the contrary that $\lambda(M) = \lambda\left(\bigcup_{n=1}^{\infty} A_n^+\right) > 0$.

There exists $n \in \mathbb{N}$ such that $\lambda(A_n^+) > 0$. For any $E \subseteq A_n^+$, we have $\mu_{\text{sing}}(E) \geq n^{-1}\lambda(E)$.

$g = f + n^{-1}1_{A_n^+}$ is a candidate density because, for any $D \in \mathcal{F}$

$$
\begin{aligned}
\int_D g d\lambda &= \int_D f d\lambda + \frac{1}{n}\int 1_{A_n^+} 1_D d\lambda \\
&= \mu_{\text{ac}}(D) + \frac{1}{n}\lambda\left(A_n^+ \cap D\right) \\
&\leq \mu_{\text{ac}}(D) + \mu_{\text{sing}}(A_n^+ \cap D) \\
&\leq \mu_{\text{ac}}(D) + \mu_{\text{sing}}(D) \\
&= \mu(D)
\end{aligned}
$$

Unfortunately, this is a contradiction, though: $\int_{\mathsf{R}} g d\lambda = \beta + \frac{1}{n}\lambda(A_n^+) > \beta$.

Lebesgue Decomposition: Proof

We showed that $\mu_{\text{sing}} \perp \lambda$. Now we show uniqueness.

Suppose there are two decompositions $\mu = \mu_{\text{sing}} + \mu_{\text{ac}} = \nu_{\text{sing}} + \nu_{\text{ac}}$.

Looking at the singular parts, there exists S_1 and S_2 such that $\lambda(S_1) = \lambda(S_2) = 0$ and $\mu_{\text{sing}}(S_1^C) = \nu_{\text{sing}}(S_2^C) = 0$.

Looking at the absolutely continuous parts, there exists f, g such that $\mu_{\text{ac}}(A) = \int_A f d\lambda$ and $\nu_{\text{ac}}(A) = \int_A g d\lambda$.

We will show $\lambda(\{g = f\}) = 1$ by showing $\lambda(\{g > f\}) = 0$ and $\lambda(\{g < f\}) = 0$. To show $\lambda(\{g > f\}) = 0$ we will show $\lambda(\{g > f\} \cap S) = 0$ and $\lambda(\{g > f\} \cap S^c) = 0$, where $S = S_1 \cup S_2$.

Lebesgue Decomposition: Proof

First we show $\lambda(\{g > f\} \cap S^c) = 0$. Again, define $S = S_1 \cup S_2$. Also define $B = S^c \cap \{g > f\}$.

$g > f$ on B, and we have $\int_B (g - f) d\lambda = \nu_{\text{ac}}(B) - \mu_{\text{ac}}(B) = \nu_{\text{ac}}(B) + \nu_{\text{sing}}(B) - \mu_{\text{ac}}(B) - \mu_{\text{sing}}(B) = \mu(B) - \mu(B) = 0$. Together these mean $\lambda(B) = 0$.

Also, $\lambda(S_1 \cup S_2) \leq \lambda(S_1) + \lambda(S_2) = 0$. So $\lambda(\{g > f\}) = 0$.

Similarly, $\lambda(\{g < f\}) = 0$. So $g = f$ λ-a.s. So $\mu_{\text{ac}} = \nu_{\text{ac}}$. So $\mu_{\text{sing}} = \nu_{\text{sing}}$.

The Radon-Nikodym Theorem

2.8: Corollary 12.1.2: Radon-Nikodym Theorem

A Borel probability measure μ is dominated by λ if and only if there exists $f \geq 0$ such that

$$\mu(A) = \int_A f d\lambda$$

for any $A \in \mathcal{B}$.

The Radon-Nikodym Theorem: Proof

Let μ be a Borel probability measure. Then

$$\mu = \mu_{\text{sing}} + \mu_{\text{ac}} + \mu_{\text{disc}}.$$

Suppose μ is absolutely continuous with respect to λ, then $\mu_{\text{sing}}(A) = \mu_{\text{ac}}(A) = 0$ for any $A \in \mathcal{B}$. Clearly, if $\lambda(B) = 0$, then $\mu(B) = 0$ as well.

Suppose $\mu \ll \lambda$. For any $x \in \mathrm{R}$, $\lambda(\{x\}) = 0$. Also, if S is such that $\lambda(S) = 0$ (and $\mu_{\text{sing}}(S^c) = 0$), then $\mu(S) = 0$, and specifically $\mu_{\text{sing}}(S) = 0$. So $\mu_{\text{sing}}(\Omega) = \mu_{\text{sing}}(S) + \mu_{\text{sing}}(S^c) = 0$. The only part left over of the three is the absolutely continuous part.