A Little Bit of Measure Theory

Lecture 5: Conditional Expectations

Jyotirmoy Bhattacharya

July 22, 2020

Some reminders

- Probability space $(\Omega, \mathcal{F}, \mathbb{P})$: Ω is the set of states of the world, $\mathcal{F}$ is a σ-algebra on it and $\mathbb{P}$ a measure on $\mathcal{F}$ with $\mathbb{P}(\Omega) = 1$.
- A function X is measurable with respect to a σ-algebra $\mathcal{F}$ if for every λ
$$\{x \colon f(x) \leq \lambda\} \in \mathcal{F}$$
- A random variable is a measurable function on a probability space.

Integration on a set

Definition

Given a measurable function f and a measurable set G we have

$$\int_G f \, d\mu = \int f \cdot 1_G \, d\mu$$

Less than full information

- Let Ω be the set of all states of the world.
- An agent who does not know the full state of the world may still be able to answer yes/no questions about subsets of Ω.
- Eg., if you have observed only the first of a sequence of tosses, you can still answer whether or not $\omega \in (T??\ldots)$.

σ-algebras as information structures

- Assume that the set of events about which the agent can answer yes/no questions is a σ-algebra. (Convincing?)

σ-algebras as information structures

- Assume that the set of events about which the agent can answer yes/no questions is a σ-algebra. (Convincing?)
- If $\mathcal{A} \subset \mathcal{B}$ then which σ-algebra represents greater information $\mathcal{A}$ or $\mathcal{B}$?

σ-algebras as information structures

- Assume that the set of events about which the agent can answer yes/no questions is a σ-algebra. (Convincing?)
- If $\mathcal{A} \subset \mathcal{B}$ then which σ-algebra represents greater information $\mathcal{A}$ or $\mathcal{B}$?
- Answer: $\mathcal{B}$.
- The least information $\{\Omega, \emptyset\}$.
- Why not just restrict Ω to some Ω' (say $\{T, H\}$ if you just know the result of the first-toss)?

σ-algebra generated by a set of RVs

Definition

Given a family of random variables X_α, for $\alpha \in A$ in some index set A, the σ-algebra generated by these random variables, denoted $\sigma(X_\alpha)$, is the smallest σ-algebra with respect to which all these random variables are measurable.

Interpretation

This is the σ-algebra denoting knowledge about the X_α and nothing more. So the events we can answer yes/no questions about are those pertaining to the values of members of X_α extended to a σ-algebra.

Measurability wrt σ-algebra

Definition

A random X is measurable with respect to a σ-algebra $\mathcal{G}$ if for every λ

$$\{\omega \colon X(\omega) \leq \lambda\} \in \mathcal{G}$$

Interpretation

X is a random variable about whose values we can answer questions based on the information in $\mathcal{G}$.

Linking the two concepts

Theorem

If $Y(\omega)$ *is measurable with respect to* $\sigma(X_1, \ldots, X_n)$ *then is a measurable function* Φ *such that*

$$Y(\omega) = \Phi(X_1(\omega), \ldots, X_n(\omega))$$

and vice versa.

A simpler approach: partitions

- Define $\omega_1 \sim \omega_2$ if the agent cannot distinguish between ω_1 and ω_2. Say $\omega_1 \sim \omega_2$ if $X(\omega_1) = X(\omega_2)$
- Represent information by the partition of Ω generated by this equivalence relation or by the equivalence relation itself.

Partitions will not always do

- Suppose we want to capture the infomation contained in a random variable X.
- We take $\omega_1 \sim \omega_2$ if $X(\omega_1) = X(\omega_2)$.
- The σ-algebra we get from the partition is

$$\{\omega \colon X(\omega) \in A\}$$

  for sets A for which either A or A^c is countable.
- These are too few sets if X takes on an uncountable set of values.

Towards conditional expectation

Suppose we have a random variable X defined on some σ-algebra $\mathcal{F}$. But we only have information given by some sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$. What is the best approximation we can get about X based on our information?

Towards conditional expectation (contd.)

Theorem

Given a random variable X, $\mathbb{E}[X]$ is the best predictor of X in expected squared error terms. That is, given any number c we have

$$\mathbb{E}[(X - \mathbb{E}[X])^2] \leq \mathbb{E}[(X - c)^2]$$

Proof.

$$\begin{aligned}
(X - c)^2 &= (X - \mathbb{E}[X] + \mathbb{E}[X] - c)^2 \\
&= (X - \mathbb{E}[X])^2 + 2(X - \mathbb{E}[X])(\mathbb{E}[X] - c) + (\mathbb{E}[X] - c)^2 \\
\mathbb{E}[(X - c)^2] &= \mathbb{E}[(X - \mathbb{E}[X])^2] + 2(\mathbb{E}[X] - c)\mathbb{E}[(X - \mathbb{E}[X])] \\
&+ (\mathbb{E}[X] - c)^2 \\
&= \mathbb{E}[(X - \mathbb{E}[X])^2] + (\mathbb{E}[X] - c)^2 \\
&\geq \mathbb{E}[(X - \mathbb{E}[X])^2]
\end{aligned}$$

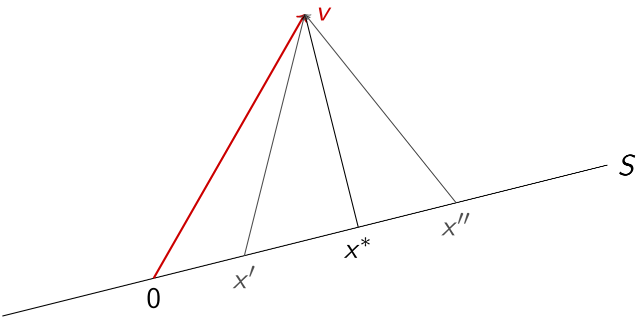
Conditional expectation: what do we want?

Suppose X is a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}[X^2] < \infty$ and a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ we want a $\mathcal{G}$ measurable random variable Z with $\mathbb{E}[Z^2] < \infty$ which minimizes

$$\mathbb{E}[(X - Z)^2]$$

i.e., we want the best $\mathcal{G}$-measurable approximation to X in expected squared error terms.

Projection



Theorem

If x^* *minimizes* $\|v - x\|$ *for* $x \in S$ *then* $(v - x^*) \cdot x = 0$ *for all* $x \in S$.

The orthogonality condition

- If Z fits our requirement then for any $\mathcal{G}$ measurable function Y we must have

$$\int (X - Z) Y \, d\mathbb{P} = 0$$

The orthogonality condition

- If Z fits our requirement then for any $\mathcal{G}$ measurable function Y we must have

$$\int (X - Z) Y \, d\mathbb{P} = 0$$

- In particular for any $G \in \mathcal{G}$ if we take $Y = 1_G$ we will have

$$\int (X - Z) 1_G \, d\mathbb{P} = \int_G (X - Z) \, dP = 0 \Rightarrow \int_G X \, d\mathbb{P} = \int_G Z \, d\mathbb{P}$$

The orthogonality condition

- If Z fits our requirement then for any $\mathcal{G}$ measurable function Y we must have

$$\int (X - Z) Y \, d\mathbb{P} = 0$$

- In particular for any $G \in \mathcal{G}$ if we take $Y = 1_G$ we will have

$$\int (X - Z) 1_G \, d\mathbb{P} = \int_G (X - Z) \, dP = 0 \Rightarrow \int_G X \, d\mathbb{P} = \int_G Z \, d\mathbb{P}$$

- On the other hand if the second condition holds for all $G \in \mathcal{G}$ then the first condition holds for all Y measurable wrt $\mathcal{G}$. Check:
  - Simple function.
  - Non-negative functions.
  - All functions.

Kolmogorov's definition

Definition

Given a random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\mathbb{E}[|X|] < \infty$ and a σ-algebra $\mathcal{G} \subset \mathcal{F}$ we define $\mathbb{E}[X \mid \mathcal{G}]$ to be a $\mathcal{G}$ measurable random variable such that for all $G \in \mathcal{G}$

$$\int_G \mathbb{E}[X \mid \mathcal{G}] \, d\mathbb{P} = \int_G X \, d\mathbb{P}$$

Note: $\mathbb{E}[X \mid \mathcal{G}]$ is a function of ω.

Kolmogorov's definition

Definition

Given a random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\mathbb{E}[|X|] < \infty$ and a σ-algebra $\mathcal{G} \subset \mathcal{F}$ we define $\mathbb{E}[X \mid \mathcal{G}]$ to be a $\mathcal{G}$ measurable random variable such that for all $G \in \mathcal{G}$

$$\int_G \mathbb{E}[X \mid \mathcal{G}] \, d\mathbb{P} = \int_G X \, d\mathbb{P}$$

Note: $\mathbb{E}[X \mid \mathcal{G}]$ is a function of ω.

- We have weakened $\mathbb{E}[X^2] < \infty$ to $\mathbb{E}[|X|] < \infty$
- Exists
- Unique upto sets of measure 0.

Properties of conditional expectations

- Linearity

$$\mathbb{E}[\alpha X + \beta Y \mid \mathcal{G}] = \alpha \mathbb{E}[X \mid \mathcal{G}] + \beta \mathbb{E}[Y \mid \mathcal{G}]$$

- Monotonicity If $X \geq Y$ then $\mathbb{E}[X \mid \mathcal{G}] \geq \mathbb{E}[Y \mid \mathcal{G}]$.
- Monotone convergence If X_n is a monotonically increasing sequence of non-negativing random variables with $\lim_{n\to\infty} X_n = X$ then

$$\lim_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}].$$

Conditional probability

Definition

$$\mathbb{P}(A \mid \mathcal{G}) = \mathbb{E}[1_A \mid \mathcal{G}]$$

Condition probability is a random variable, depends on ω.

What about good old

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Take $\mathcal{G} = \{\Omega, B, B^c, \emptyset\}$