STATS 310B: Theory of Probability II **Winter 2016/17**

Lecture 1: January 10

Lecturer: Sourav Chatterjee *Scribes: Kenneth Tay*

1.1 Conditional Probability

Set-up: We have a probability space $(\Omega, \mathcal{F}, P)$ and a sub σ-algebra $\mathcal{G} \subseteq \mathcal{F}$.

Definition 1.1 *For an event $A \in \mathcal{F}$, define the* ***conditional probability of*** *A* ***given*** *$\mathcal{G}$, written $P(A \mid \mathcal{G})$, to be a $\mathcal{G}$-measurable random variable such that for all $B \in \mathcal{G}$,*

$$P(A \cap B) = \mathbb{E}\left[P(A \mid \mathcal{G})1_B\right] = \int_B P(A \mid \mathcal{G}) dP.$$

Why does such a random variable exist, and if it does, is it unique? We will spend the rest of the lecture on this.

(Note: One reason for the complexity of the theory of conditional expectation is the issue of how to condition on events of probability 0. For example, see the Borel-Kolmogorov paradox.)

1.1.1 Conditional Expectation for L^2 Random Variables

Definition 1.2 $L^2(\Omega, \mathcal{F}, P) := \left\{ \text{all } \mathcal{F}\text{-measurable random variables } X \text{ such that } \mathbb{E}X^2 < \infty \right\}$.

We will first define conditional expectation for L^2 random variables (this lecture), then extend the notion to L^1 random variables (next lecture). Before that, we will work through a few lemmas that we will need.

Lemma 1.3 *L^2 is complete, i.e. Cauchy sequences converge.*

Proof: For any $X \in L^2$, let $\|X\| := \sqrt{\mathbb{E}X^2}$. If $\{X_n\}$ is a Cauchy sequence, then $\|X_n - X_m\| \to 0$ as $n, m \to \infty$.

Find a subsequence $n_1, n_2, \ldots$ such that $\|X_{n_{i+1}} - X_{n_i}\| \leq 2^{-i}$ for all i. By the Monotone Convergence Theorem and since the L^2 norm dominates the L^1 norm for probability spaces,

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^\infty |X_{n_{i+1}} - X_{n_i}|\right) &= \sum_{i=1}^\infty \mathbb{E}|X_{n_{i+1}} - X_{n_i}| \\
&\leq \sum_{i=1}^\infty \|X_{n_{i+1}} - X_{n_i}\| \\
&< \infty.
\end{aligned}$$

We thus conclude that $\displaystyle\sum_{i=1}^\infty |X_{n_{i+1}} - X_{n_i}|$ is finite a.s. Since $X_{n_i} - X_{n_1} = \displaystyle\sum_{j=1}^{i-1} X_{n_{j+1}} - X_{n_j}$, we can further conclude that $\displaystyle\lim_{i \to \infty} X_{n_i}$ exists a.s.

Call this limit Y. Because

$$\begin{aligned}
\|X_{n_i}\| - \|X_{n_1}\| &= \sum_{j=1}^{i-1} \|X_{n_{j+1}}\| - \|X_{n_j}\|, \\
\Big| \|X_{n_i}\| - \|X_{n_1}\| \Big| &\leq \sum_{j=1}^{i-1} \Big| \|X_{n_{j+1}}\| - \|X_{n_j}\| \Big| \\
&\leq \sum_{j=1}^{i-1} \|X_{n_{j+1}} - X_{n_j}\|,
\end{aligned}$$

which converges to a finite limit, we can apply Fatou's Lemma to obtain

$$\mathbb{E}Y^2 \leq \liminf \mathbb{E}X_{n_i}^2 = \liminf \|X_{n_i}\| < \infty,$$

i.e. $Y \in L^2$. Since if $j > i$,

$$\begin{aligned}
\|X_{n_i} - X_{n_j}\| &\leq \|X_{n_i} - X_{n_{i+1}}\| + \cdots + \|X_{n_{j-1}} - X_{n_j}\| \\
&\leq 2^{-i} + 2^{-(i+1)} + \cdots + 2^{-j} \\
&\leq 2^{-(i-1)},
\end{aligned}$$

we can apply Fatou's Lemma again to obtain

$$\mathbb{E}(X_{n_i} - Y)^2 \leq \liminf \mathbb{E}(X_{n_i} - X_{n_j})^2 = 0,$$

i.e. $X_{n_i} \xrightarrow{L^2} Y$.

Since $X_{n_i} \to Y$ in L^2 and X_n is Cauchy in L^2, it follows that $X_n \to Y$ in L^2, as required. ∎

Lemma 1.4 *Let $\mathcal{C} \subseteq L^2(\Omega, \mathcal{F}, P)$ be non-empty, closed and convex. Then there exists a unique $X \in \mathcal{C}$ such that $\|X\| = \inf\{\|Z\| : Z \in \mathcal{C}\}$.*

(Note: This is a general fact about Hilbert spaces.)

Proof: Recall the parallelogram identity:

$$\|X_n - X_m\|^2 + \|X_n + X_m\|^2 = 2\|X_n\|^2 + 2\|X_m\|^2,$$

or equivalently,

$$\|X_n - X_m\|^2 = 2\left(\|X_n\|^2 + \|X_m\|^2 - 2\left\| \frac{X_n + X_m}{2} \right\|^2 \right).$$

Let $\lambda := \inf\{\|Z\| : Z \in \mathcal{C}\}$, and pick any sequence $\{X_n\}$ in $\mathcal{C}$ such that $\|X_n\| \to \lambda$. Since $\mathcal{C}$ is convex, we have $\dfrac{X_n + X_m}{2} \in \mathcal{C}$ and so the parallelogram identity yields

$$\|X_n - X_m\|^2 \leq 2\left(\|X_n\|^2 + \|X_m\|^2 - 2\lambda^2 \right).$$

Since $\|X_n\| \to \lambda$, the above implies that $\{X_n\}$ is a Cauchy sequence. Hence, Lemma 1.3 implies that there is a $Y \in L^2(\Omega, \mathcal{F}, P)$ such that $X_n \xrightarrow{L^2} Y$. Since $\mathcal{C}$ is closed by assumption, we have $Y \in \mathcal{C}$. Y must minimize the norm in $\mathcal{C}$ because $X_n \xrightarrow{L^2} Y$ implies $\|X_n\| \to \|Y\|$.

To show uniqueness: Assume that Y and W are 2 random variables which minimize the norm in $\mathcal{C}$. Pick $\{X_n\} = Y, W, Y, W, \ldots$. Then the argument above shows that we must have $Y = W$. ∎

Lemma 1.5 *Let M be a closed linear subspace of L^2, and let X be some element of L^2. Then there exists a unique $Z \in M$ such that $\|X - Z\| = \inf\{\|X - Y\| : Y \in M\}$.*

Moreover, $X - Z \perp Y$ for all $Y \in M$, i.e. $< X - Z, Y > := \mathbb{E}[(X - Z)Y] = 0$.

Moreover, if $W \in M$ satisfies $X - W \perp Y$ for all $Y \in M$, then $W = Z$.

Proof: Let $\mathcal{C} = \{X - Y : Y \in M\}$. Then we can apply Lemma 1.4 to get the existence and uniqueness of Z.

To show that $X - Z \perp Y$ for all $Y \in M$: Define the function $f(c) := \|X - Z + cY\|^2$. Note that $Z - cY$ is an element of $\mathcal{C}$ for all c. By definition of Z, the function f must obtain its minimum at $c = 0$, which implies that $f'(0) = 0$. However,

$$\begin{aligned} f(c) &= \|X - Z\|^2 + 2c < X - Z, Y > + c^2\|Y\|^2, \\ f'(c) &= 2 < X - Z, Y > + 2c\|Y\|^2, \\ f'(0) &= 2 < X - Z, Y >, \end{aligned}$$

i.e. $< X - Z, Y >= 0$, as required.

The last assertion is left as an exercise. ■

We can now define conditional expectation for L^2 random variables:

Definition 1.6 *For $X \in L^2(\Omega, \mathcal{F}, P)$, define the* ***conditional expectation of*** X ***given*** $\mathcal{G}$*, written $\mathbb{E}[X \mid \mathcal{G}]$, as the $\mathcal{G}$-measurable random variable in L^2 such that*

$$\mathbb{E}XZ = \mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}]Z\right]$$

for all $Z \in L^2(\Omega, \mathcal{G}, P)$ (i.e. preserves inner products with X).

Theorem 1.7 $\mathbb{E}[X \mid \mathcal{G}]$ *exists and is unique. It is the orthogonal projection of X onto the space $L^2(\Omega, \mathcal{G}, P)$.*

Proof: Let $M = L^2(\Omega, \mathcal{G}, P)$. This is a closed subspace: L^2-convergent sequences have an a.s.-convergent subsequence (proved in Lemma 1.3), and hence the limit of the sequence is $\mathcal{G}$-measurable.

We can apply Lemma 1.5, and so the projection of X onto this subspace satisfies the criterion for conditional expectation, and is the only $\mathcal{G}$-measurable L^2 random variable satisfying this condition. ■