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

Lecture 2: January 12

Lecturer: Sourav Chatterjee *Scribes: Kenneth Tay*

2.1 Conditional Expectation for L^1 Random Variables

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

We want to define the random variable $\mathbb{E}[X \mid \mathcal{G}]$ with the property that for all $\mathcal{G}$-measurable random variables Y,

$$\mathbb{E}[XY] = \mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}]Y\right]. \tag{2.1}$$

Last lecture, we defined this when $X \in L^2(\Omega, \mathcal{F}, P)$. In that case, $\mathbb{E}[X \mid \mathcal{G}]$ was defined to be the orthogonal projection of X onto the subspace $L^2(\Omega, \mathcal{G}, P)$, and Equation 2.1 was proved for all $Y \in L^2(\Omega, \mathcal{G}, P)$.

Now, we want to extend this concept to L^1 random variables. (We want the conditional expectation to be well-defined whenever the expectation is well-defined.)

Definition 2.1 *Suppose X is a random variable such that $\mathbb{E}|X| < \infty$. Then $\mathbb{E}[X \mid \mathcal{G}]$ is the unique $\mathcal{G}$-measurable random variable with finite expectation such that for all $A \in \mathcal{G}$,*

$$\mathbb{E}[X; A] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]; A], \tag{2.2}$$

where $\mathbb{E}[X; A] := \mathbb{E}[X 1_A]$.

Note that, instead of looking at all $\mathcal{G}$-measurable random variables, we just look at indicator functions. This is because in general, $X \in L^1$ and $Y \in L^1$ does not imply that $XY \in L^1$.

As with the L^2 case, we have to ask: **does such a random variable exist, and is it unique?**

2.1.1 Uniqueness

Observe that if Y is a non-negative random variable such that $\mathbb{E}Y = 0$, then $Y = 0$ a.s. As a consequence:

Proposition 2.2 *If Y_1 and Y_2 are integrable $\mathcal{G}$-measurable random variables such that $\mathbb{E}[Y_1; A] = \mathbb{E}[Y_2; A]$ for all $A \in \mathcal{G}$, then $Y_1 = Y_2$ a.s.*

Proof: Let $A = \{Y_1 \geq Y_2\}$. Then the given condition implies $\mathbb{E}\left[(Y_1 - Y_2)1_{\{Y_1 \geq Y_2\}}\right] = 0$. This is a non-negative random variable, so $(Y_1 - Y_2)1_{\{Y_1 \geq Y_2\}} = 0$ a.s., which in turn implies that $Y_1 \leq Y_2$ a.s.

Similarly, we can show that $Y_2 \leq Y_1$ a.s., hence $Y_1 = Y_2$ a.s. ∎

The proposition implies that $\mathbb{E}[X \mid \mathcal{G}]$, if it exists, must be unique.

2.1.2 Existence

First, suppose that $X \geq 0$ a.s. Let $\{X_n\}$ be a sequence of simple non-negative random variables increasing to X. Since all simple functions are in L^2, $\mathbb{E}[X_n \mid \mathcal{G}]$ is well-defined for all n.

Lemma 2.3 *Let $Y \in L^2(\Omega, \mathcal{F}, P)$. If $Y \geq 0$ a.s., then $\mathbb{E}[Y \mid \mathcal{G}] \geq 0$ a.s.*

(Note: The above is true more generally for all Y for which conditional expectation is defined.)

Proof:

$$\begin{aligned} & Y \geq 0 \text{ a.s.} \\ \Rightarrow \quad & \mathbb{E}[Y; A] \geq 0 \text{ for all } A \in \mathcal{G} \\ \Rightarrow \quad & \mathbb{E}[\mathbb{E}[Y \mid \mathcal{G}]; A] \geq 0 \text{ for all } A \in \mathcal{G}. \end{aligned}$$

Taking $A = \{\mathbb{E}[Y \mid \mathcal{G}] < 0\}$, we see that the above implies implies $\mathbb{E}[Y \mid \mathcal{G}] \geq 0$ a.s. ■

Thus, $\mathbb{E}[X_n \mid \mathcal{G}] \geq 0$ a.s. and is an increasing sequence of non-negative random variables (by linearity of conditional expectation). Thus, a pointwise limit exists. Call it Y, i.e. $Y = \lim \mathbb{E}[X_n \mid \mathcal{G}]$.

Proposition 2.4 *Y satisfies Equation 2.2, and hence is the conditional expectation of X given $\mathcal{G}$.*

Proof: By the Monotone Convergence Theorem,

$$\begin{aligned} \mathbb{E}[X; A] &= \lim \mathbb{E}[X_n; A], \\ \mathbb{E}[Y; A] &= \lim \mathbb{E}[\mathbb{E}[X_n \mid \mathcal{G}]; A]. \end{aligned}$$

For each n, the 2 quantities on the RHSes are the same (since $X_n \in L^2$), hence the 2 quantities on the LHSes must be the same. ■

Finally, consider a general random variable X with finite expectation. Let X^+ and X^- be the positive and negative parts of X. Then we define

$$\mathbb{E}[X \mid \mathcal{G}] := \mathbb{E}[X^+ \mid \mathcal{G}] - \mathbb{E}[X^- \mid \mathcal{G}]. \tag{2.3}$$

Why is the above well-defined? Since $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$, both $\mathbb{E}X^+$ and $\mathbb{E}X^-$ are finite. Taking A to be the full space, we have

$$\mathbb{E}[X^+] = \mathbb{E}[X^+; \Omega] = \mathbb{E}[\mathbb{E}[X^+ \mid \mathcal{G}]; \Omega] = \mathbb{E}[\mathbb{E}[X^+ \mid \mathcal{G}]].$$

This implies that $\mathbb{E}[X^+ \mid \mathcal{G}]$ is a non-negative random variable with finite expectation, and so it is finite a.s. The same applies for $\mathbb{E}[X^- \mid \mathcal{G}]$, so the RHS of Equation 2.3 is well-defined.

2.2 Properties of Conditional Expectation

Here are some important properties of conditional expectation:

- **Linearity:**
$$\mathbb{E}[aX + bY \mid \mathcal{G}] = a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}].$$

- **Tower property:** If $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$, then

$$\mathbb{E}[\mathbb{E}(X \mid \mathcal{G}_2) \mid \mathcal{G}_1] = \mathbb{E}[X \mid \mathcal{G}_1] \text{ a.s.}$$

- If X and Y are 2 random variables and Y is $\mathcal{G}$-integrable and $X \in L^1$, $XY \in L^1$, then

$$\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}].$$

  Proof:(Sketch) For all $A \in \mathcal{G}$,

$$\begin{aligned}\mathbb{E}[XY; A] &= \mathbb{E}[X(Y1_A)] \\ &= \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]Y1_A] \\ &= \mathbb{E}[Y\mathbb{E}[X \mid \mathcal{G}]; A].\end{aligned}$$

  By Proposition 2.2, we must have $\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}]$.

  (Note that the second equality follows directly by definition for $X \in L^2$, but not so for $X \in L^1$. A little more work needs to be done involving the dominated convergence theorem.) ∎

- If $\sigma(X)$ and $\mathcal{G}$ are independent σ-algebras, then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}X$ a.s.
- If $\mathcal{H}$ is independent of $\sigma(X, \mathcal{G})$, then $\mathbb{E}[X \mid \sigma(\mathcal{H}, \mathcal{G})] = \mathbb{E}[X \mid \mathcal{G}]$.
- **Jensen's Inequality:** If $\phi : \mathbb{R} \to \mathbb{R}$ is convex, and $\mathbb{E}|\phi(X)| < \infty$, $\mathbb{E}|X| < \infty$, then

$$\mathbb{E}[\phi(X) \mid \mathcal{G}] \geq \phi(\mathbb{E}[X \mid \mathcal{G}]) \text{ a.s.}$$

  Proof: Just imitate the unconditional version of Jensen's inequality. ∎

2.3 Connection with Naïve Notion of Conditional Expectation

Example: Let (X, Y) be a pair of real-valued random variables with joint density $f_{X,Y}(x, y)$. We say that the conditional density of Y given $X = x$ is

$$f_{Y|X=x}(y) := \frac{f_{X,Y}(x,y)}{f_X(x)} := \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy}.$$

Proposition 2.5 *Let $\mathcal{G}$ be the σ-algebra generated by X. For any function $Z = h(X, Y)$ of X and Y, $\mathbb{E}[Z \mid \mathcal{G}] := \mathbb{E}[Z \mid X] = g(X)$, where $g(x) = \mathbb{E}[Z \mid X = x]$.*

Proof:

$$\begin{aligned}g(x) &= \mathbb{E}[Z \mid X = x] \\ &= \int_{-\infty}^{\infty} h(x,y) f_{Y|X=x}(y)dy \\ &= \int_{-\infty}^{\infty} h(x,y) \frac{f_{X,Y}(x,y)}{f_X(x)}dy.\end{aligned}$$

We have to show that for any $\mathcal{G}$-measurable W, $\mathbb{E}[ZW] = \mathbb{E}[g(X)W]$. Since W is $\mathcal{G}$-measurable, it may be written as $W = u(X)$. Then

$$
\begin{aligned}
\mathbb{E}[ZW] &= \mathbb{E}[h(X,Y)u(X)] \\
&= \int\int h(x,y)u(x)f_{X,Y}(x,y)dxdy, \\
\mathbb{E}[g(X)W] &= \int g(x)u(x)f_X(x)dx \\
&= \int\int h(x,y)\frac{f_{X,Y}(x,y)}{f_X(x)}u(x)f_X(x)dydx,
\end{aligned}
$$

and the two expressions are the same. ■

2.3.1 Resolving the Borel-Kolmogorov Paradox

Think about narrow bands of longitudes or latitudes. For longitudes, we can't just look at narrow bands and take the limit, hoping to get the right conditional distribution. This is because there is "too much mass" at the poles, and this irregularity does not disappear even when you take the limit. (This however works for bands of latitudes.)

The solution is not to define conditional distributions on each longitude individually, but to define them altogether so that the averaging properties "work out right".

Concretely, let X be a point chosen uniformly on a sphere. Let L_1 and L_2 be the latitude and longitude of X respectively. Then we need

$$\int \mathbb{E}[f(X) \mid L_1 = l]f_{L_1}(l)dl = \mathbb{E}[f(X)]$$

for any f. The way to guarantee this is to the treat the conditional distribution of X given $L_1 = l$ as uniform on that latitude.

Looking at longitudes, we also need

$$\int \mathbb{E}[f(X) \mid L_2 = l]f_{L_2}(l)dl = \mathbb{E}[f(X)].$$

This works out only if we take the conditional distribution of $X \mid L_2 = l$ as the appropriate non-uniform distribution. (If uniform, too much mass is given to the poles.)

2.4 Martingales

Definition 2.6 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A **filtration** on this space is an increasing sequence of sub σ-algebras $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$*

Definition 2.7 *A sequence of random variables $Z_1, Z_2, \ldots$ is **adapted** to a filtration $\{\mathcal{F}_n\}$ if for all n, Z_n is $\mathcal{F}_n$-measurable.*

Definition 2.8 *An adapted sequence is called a **martingale** if:*

1. $\mathbb{E}|Z_n| < \infty$ *for all n, and*

2. $\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] = Z_n$ *a.s. for all* n*.*

Martingales were invented by Joseph Doob in the 1950s.

Example of a martingale: Let $Y_1, Y_2, \ldots$ be i.i.d. mean 0 random variables with finite expectation. Let $Z_n = \sum_{i=1}^{n} Y_i$, $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. Then

$$\begin{aligned}
\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[Y_{n+1} + Z_n \mid \mathcal{F}_n] \\
&= \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] + \mathbb{E}[Z_n \mid \mathcal{F}_n] \\
&= \mathbb{E}Y_{n+1} + Z_n \mathbb{E}[1 \mid \mathcal{F}_n] \\
&= 0 + Z_n.
\end{aligned}$$

Thus $\{Z_n\}$ is a martingale.