

STATS310B – Lecture 1

Sourav Chatterjee
Scribed by Michael Howes

01/04/22

Contents

1	Announcements	1
2	Conditional expectation	1
2.1	Motivation	1
2.2	Definition, uniqueness and existence	2

1 Announcements

An email will be sent out with administrative information. The course will cover two main topics: conditional expectation and martingales. There will also be some applications to Markov chains and ergodic theory.

2 Conditional expectation

2.1 Motivation

The quantities $P(A|X = x)$ and $E[Y|X = x]$ are clear when X is a discrete random variable. In this case we have

$$P(A|X = x) = \frac{P(A \cap \{X = x\})}{P(X = x)},$$

when $P(X = x) > 0$. It is unclear how we can define $P(A|X = x)$ when $P(X = x) = 0$, but this is important when X is a continuous random variable. We can normally interpret $P(A|X = x)$ as

$$\lim_{\varepsilon \rightarrow 0} P(A|X \in (x - \varepsilon, x + \varepsilon)).$$

But this immediately leads to problems. Notably, when does the limit exist? Do we have to put restrictions on A or X ?, etc. There is also a big problem called the Borel–Kolmogorov paradox. One version of the paradox is as follows (see [wikipedia](#) for another example).

Example 1. Let (X, Y) be uniformly distributed on the unit disk in \mathbb{R}^2 . What is the distribution of Y given $X = 0$? That is what is the distribution of Y , given that (X, Y) is on the y -axis? The answer is that Y should be uniform on $[-1, 1]$. Indeed, if we take limits of the conditional distribution given $X \in (-\varepsilon, \varepsilon)$, then we do get the uniform distribution.

But the event $\{X = 0\}$ can be written in a different way. Let (R, Θ) be (X, Y) in polar coordinates with $\Theta \in [0, 2\pi)$. Then $\{X = 0\} = \{\Theta = \pi/2\} \cup \{\Theta = 3\pi/2\}$. So we can think of the conditional distribution of Y given $X = 0$ as the limit of the conditional distribution of Y given $\Theta \in (\pi/2 - \varepsilon, \pi/2 + \varepsilon) \cup (3\pi/2 - \varepsilon, 3\pi/2 + \varepsilon)$. But this limiting distribution is not uniform on $[-1, 1]$. The limiting

distribution is proportional to $|y|$. This is because we are now approximating $\{X = 0\}$ as two skinny wedges rather than a vertical strip.

The upshot of this example is as follows. Conditioning on events of probability zero is not well-defined as the limit of conditioning on approximating events of positive probability. The limit may depend on the choice of approximating events.

Here is one resolution to the paradox. We cannot view the conditional distribution of Y given $X = 0$ as the same object as the conditional distribution of Y given $\Theta \in \{\pi/2, 3\pi/2\}$. What remains true is that we can still compute $P(Y \in A|X = x)$ by using a limiting procedure. Fixing A we can define a function $f(x) = P(Y \in A|X = x)$. We can interpret $P(Y \in A|X = 0)$ as $f(0)$. We can also define a function $g(\theta) = P(Y \in A|\Theta = \theta)$, but these are different functions. Then we can plug X or Θ into f and g to get random variables. The random variable $f(X)$ is what's written as $P(Y \in A|X)$. Thus, conditional probabilities are random variables.

In rigorous conditional probability, we define conditional expectations of a random variable Y given a σ -algebra \mathcal{G} instead of an event. The conditional expectation is a random variable and is written as $E[Y|\mathcal{G}]$. In our examples \mathcal{G} was $\sigma(X)$.

2.2 Definition, uniqueness and existence

We will now give the rigorous definition of conditional expectation.

Definition 1. Let (Ω, \mathcal{F}, P) be a probability space. Let X be a real valued integrable random variable defined on (Ω, \mathcal{F}) and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub- σ -algebra of \mathcal{F} . An integrable random variable Y is called $E[X|\mathcal{G}]$ if

- (1) The random variable Y is \mathcal{G} -measurable.
- (2) For all $B \in \mathcal{G}$, $E[Y\mathbf{1}_B] = E[X\mathbf{1}_B]$, where $\mathbf{1}_B$ denotes the indicator function of B .

Lemma 1. If $E[X|\mathcal{G}]$ exists, then it is unique almost surely.

Before we prove the above lemma, let's consider an example.

Example 2. Again consider (X, Y) uniformly distributed on the unit disk. Let $\mathcal{G} = \sigma(X)$, what is $E[Y^2|\mathcal{G}]$? (Note: usually we will write $E[Y|X]$ to denote $E[Y|\sigma(X)]$).

Let's first guess and then verify. Given $X = x$, we believe $Y \sim \text{Unif}[-\sqrt{1-x^2}, \sqrt{1-x^2}]$. Thus,

$$\begin{aligned} E[Y^2|X = x] &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{2\sqrt{1-x^2}} y^2 dy \\ &= \frac{1}{2\sqrt{1-x^2}} \left[\frac{1}{3} y^3 \right]_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \\ &= \frac{1}{3} (1 - x^2). \end{aligned}$$

So, our guess is $E[Y^2|\mathcal{G}] = \frac{1}{3}(1 - X^2)$. Let's check the two properties. The random variable $Z = \frac{1}{3}(1 - X^2)$ is a continuous function of X and hence \mathcal{G} -measurable. The random variable Z is also bounded and hence integrable. Lastly, if $B \in \mathcal{G}$, then $B = \{X \in A\}$ for some Borel set A . It follows

that

$$\begin{aligned}
 E[Y^2 \mathbf{1}_B] &= E[Y^2 \mathbf{1}_{X \in A}] \\
 &= \frac{1}{\pi} \int_{-1}^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y^2 \mathbf{1}_A(x) dy dx \\
 &= \frac{1}{\pi} \int_A \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y^2 dy dx \\
 &= \int_A \frac{2\sqrt{1-x^2}}{\pi} \frac{1}{3} (1-x^2) dx \\
 &= E \left[\mathbf{1}_{X \in A} \frac{1}{3} (1-X^2) \right] \\
 &= E \left[\mathbf{1}_B \frac{1}{3} (1-X^2) \right]
 \end{aligned}$$

the last line holds because $\frac{2\sqrt{1-x^2}}{\pi}$ is the marginal density of X .

We will now prove Lemma 1.

Proof. Suppose Y_1, Y_2 are \mathcal{G} measurable and

$$E[\mathbf{1}_B Y_1] = E[\mathbf{1}_B Y_2] = E[\mathbf{1}_B X],$$

for all $B \in \mathcal{G}$. Let $B = \{Y_1 > Y_2\}$. Since Y_1, Y_2 are both \mathcal{G} -measurable, $B \in \mathcal{G}$. Thus, by assumption, $E[\mathbf{1}_B(Y_1 - Y_2)] = 0$. Since $\mathbf{1}_B(Y_1 - Y_2)$ is non-negative, this implies $P(\mathbf{1}_B(Y_1 - Y_2) = 0) = 1$ and hence $P(Y_1 \leq Y_2) = 1$. By symmetry, we must have $P(Y_2 \leq Y_1) = 1$ and thus $P(Y_1 = Y_2) = 1$ and $Y_1 = Y_2$, almost surely. \square

If we return to example 2, we can see that $E[Y^2|X]$ is only defined up to a set of measure zero. Indeed, the random variable

$$Z' = \begin{cases} \frac{1}{3}(1 - X^2) & \text{if } X \neq 0, \\ 100 & \text{if } X = 0. \end{cases}$$

is also a candidate for $E[Y^2|X]$. Our next goal is to prove the following theorem.

Theorem 1. *For any probability space (Ω, \mathcal{F}, P) , any integrable random variable X and any sub- σ -algebra \mathcal{G} , the conditional expectation $E[X|\mathcal{G}]$ exists and is unique almost surely.*

We already proved uniqueness, and so we only need to prove existence. We'll start with a lemma where we allow ourselves an additional assumption.

Lemma 2. *If $X \in L^2(\Omega, \mathcal{F}, P)$, then $E[X|\mathcal{G}]$ exists.*

The notation $X \in L^2(\Omega, \mathcal{F}, P)$ means that X is a random variable with respect to (Ω, \mathcal{F}, P) and $E[X^2] < \infty$.

Proof. Assume $X \in L^2$ and let

$$a = \inf_{Y \in L^2(\mathcal{G})} E(X - Y)^2.$$

Then there exist $Y_1, Y_2, \dots \in L^2(\mathcal{G})$ such that

$$\lim_{n \rightarrow \infty} E(X - Y_n)^2 = a.$$

Recall the parallelogram identity, for all $u, v \in \mathbb{R}$,

$$(u - v)^2 + (u + v)^2 = 2u^2 + 2v^2.$$

Now fix any n, m and take $u = X - Y_n$ and $v = X - Y_m$ and so

$$(Y_m - Y_n)^2 + (2X - Y_n - Y_m)^2 = 2(X - Y_n)^2 + 2(X - Y_m)^2.$$

Taking expectations and rearranging, we get

$$E(Y_m - Y_n)^2 = 2E(X - Y_n)^2 + 2E(X - Y_m)^2 - 4E\left(X - \frac{Y_m + Y_n}{2}\right)^2.$$

Now, $\frac{Y_m + Y_n}{2} \in L^2(\mathcal{G})$ and so $E\left(X - \frac{Y_m + Y_n}{2}\right)^2 \geq a$ and thus

$$E(Y_m - Y_n)^2 \leq 2E(X - Y_n)^2 + 2E(X - Y_m)^2 - 4a.$$

As $n, m \rightarrow \infty$, $E(X - Y_n)^2, E(X - Y_m)^2 \rightarrow a$ and so

$$\limsup_{n, m \rightarrow \infty} E(Y_m - Y_n)^2 \leq 0,$$

which implies $\lim_{n, m \rightarrow \infty} \mathbb{E}(Y_m - Y_n)^2 = 0$. Thus, $\{Y_n\}_{n=1}^\infty$ is a Cauchy sequence in $L^2(\mathcal{G})$. The space $L^2(\mathcal{G})$ is a complete Hilbert space and so there exists $Y \in L^2(\mathcal{G})$ such that $Y_n \rightarrow Y$. It follows that $E(X - Y)^2 = a$.

We now wish to show that $Y = E[X|\mathcal{G}]$. Take any $B \in \mathcal{G}$. We wish to show $E[\mathbf{1}_B X] = E[\mathbf{1}_B Y]$. For any $\lambda \in \mathbb{R}$, $Y + \lambda \mathbf{1}_B \in L^2(\mathcal{G})$ and so $E(X - Y - \lambda \mathbf{1}_B)^2 \geq a$ and furthermore

$$E(X - Y - \lambda \mathbf{1}_B)^2 = E(X - Y)^2 - 2\lambda E[(X - Y)\mathbf{1}_B] + \lambda^2 P(B).$$

Since the above quadratic is minimized when $\lambda = 0$, we must have $2E[(X - Y)\mathbf{1}_B] = 0$ as required. \square

Remark 1. It might be helpful to think of $E[X|\mathcal{G}]$ as regressing X on \mathcal{G} . What we were doing above is a form of least squares.

In the next class we will prove the result for integrable X . We will see that most of the work is done. This is because the above lemma implies that $E[X|\mathcal{G}]$ exists for simple X . The next step will be a familiar limiting argument.