Lecture 20 : Conditional Probability

STAT205 *Lecturer: Jim Pitman* Scribe: *Tye Lidman* `<tlid@berkeley.edu>`

These notes were partially adapted from the notes of Charles C. Fowlkes from previous years. Reference: [1], Section 4.1.

20.1 Review

Recall that for a probability space $(\Omega, \mathcal{F}, P)$, a real r.v. $X \in L^1(\Omega, \mathcal{F}, P)$ (i.e. $\mathbb{E}|X| < \infty$), and sub-σ-field $\mathcal{G} \subset \mathcal{F}$, $\mathbb{E}(X|\mathcal{G})$ is the (a.s.) unique r.v. $\hat{X}$ satisfying:

1. $\hat{X} \in \mathcal{G}$; and
2. $\hat{X}$ integrates like X over $\mathcal{G}$-sets; i.e.

$$\mathbb{E}(\hat{X}1_G) = \mathbb{E}(X1_G) \text{ for all } G \in \mathcal{G}. \tag{20.1}$$

Note that you never have to check condition (20.1) above for all $G \in \mathcal{G}$ because for any pair of random variables X and Y which are integrable, $\{G \mid \mathbb{E}(X1_G) = \mathbb{E}(Y1_G)\}$ is a λ system (provided $\Omega \in \mathcal{G}$), so you only have to check (20.1) for G in a π-system which generates $\mathcal{G}$.

20.1 extends immediately to

$$\mathbb{E}(\hat{X}Z) = \mathbb{E}(XZ) \text{ for } Z \in \mathcal{G} \text{ bounded} \tag{20.2}$$

by the usual argument of taking linear combinations of indicators and passing to the limit via the dominated convergence theorem, since $c|X|$ is a dominating variable (where c is the bound on $|Z|$). This gives $\mathbb{E}(XZ) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})Z)$ since Z and $\mathbb{E}(X|\mathcal{G})$ are both $\mathcal{G}$-measurable, so $\hat{X}$ is a kind of projection of X onto $\mathcal{G}$-measurable things.

For $X \in L^2(\Omega, \mathcal{F}, P)$, $\hat{X}$ is the *orthogonal projection* of X onto $L^2(\Omega, F, P)$ (up to quibbles about equivalence relations of $\mathcal{F}$ and $\mathcal{G}$). That is, (20.2) says $\mathbb{E}((X - \hat{X})Z) = 0$ for all $Z \in L^2(\Omega, \mathcal{F}, P)$. This is the definition of an orthogonal projection in a Hilbert space.

Last time we proved the existence of $\mathbb{E}(X|\mathcal{G})$ for $\mathcal{G}$ that are countably generated and $X \in L^2$ by letting $\hat{X} = \lim_{n\to\infty} \mathbb{E}(X|\mathcal{G}_n)$, with $\mathcal{G}_n \uparrow \mathcal{G}$. What about $X \in L^1(\Omega, \mathcal{F}, P)$?

20.2 Properties of Conditional Expectation

We record some basic properties of $\mathbb{E}(\cdot|\mathcal{G})$ as an operator, $X \longmapsto \mathbb{E}(X|\mathcal{G})$:

- $\mathbb{E}(\cdot|\mathcal{G})$ is positive:
$$Y \geq 0 \Rightarrow \mathbb{E}(Y|\mathcal{G}) \geq 0$$
- $\mathbb{E}(\cdot|\mathcal{G})$ is linear:
$$\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$$
- $\mathbb{E}(\cdot|\mathcal{G})$ is a projection:
$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{G}) = \mathbb{E}(X|\mathcal{G})$$
- $\mathbb{E}(\cdot|\mathcal{G})$ is continuous with norm 1 in any of the usual $\mathbf{L}^p$ spaces for $p \geq 1$:
$$\|\mathbb{E}(X|\mathcal{G})\|_p \leq \|X\|_p$$
and
$$X_n \xrightarrow{\mathbf{L}^2} X \text{ implies } \mathbb{E}(X_n|\mathcal{G}) \xrightarrow{\mathbf{L}^2} \mathbb{E}(X|\mathcal{G})$$
- Conditional expectation has the tower property. If $\mathcal{H} \subset \mathcal{G}$ then:
$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(\mathbb{E}(X|\mathcal{H})\mathcal{G}) = \mathbb{E}(X|\mathcal{H})$$
- $\mathbb{E}(\cdot|\mathcal{G})$ commutes with multiplication by $\mathcal{G}$-measurable variables:
$$\mathbb{E}(XY|\mathcal{G}) = E(X|\mathcal{G})Y \text{ for } \mathbb{E}|XY| < \infty \text{ and } Y \in \mathcal{G}$$
- $\mathbb{E}(\cdot|\mathcal{G})$ respects monotone convergence:
$$0 \leq X_n \uparrow X \text{ implies } \mathbb{E}(X_n|\mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G})$$
- If ϕ is convex and $\mathbb{E}|\phi(X)| < \infty$ then a conditional form of Jensen's inequality holds:
$$\phi(\mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(\phi(X)|\mathcal{G})$$
- As before, $\mathbb{E}(\cdot|\mathcal{G})$ is an orthogonal projection in $\mathbf{L}^2$:
$$\mathbb{E}((X - \hat{X})Z) = 0 \text{ for all } Z \in L^2(\Omega, \mathcal{F}, \mathbb{P})$$
- *Repeated Conditioning*: for $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \ldots$, $\mathcal{G}_\infty = \sigma(\cup \mathcal{G}_i)$ and $X \in \mathbf{L}^p$ with $p \geq 1$,
$$\mathbb{E}(X|\mathcal{G}_n) \xrightarrow{a.s.} \mathbb{E}(X|\mathcal{G}_\infty)$$
$$\mathbb{E}(X|\mathcal{G}_n) \xrightarrow{\mathbf{L}^p} \mathbb{E}(X|\mathcal{G}_\infty)$$

20.3 Conditioning on Random Variables

Take $\mathcal{G} = \sigma(Y)$ for some random variable Y. Recall that Z is $\sigma(Y)$ is measurable if and only if $Z = \phi(Y)$ for some Borel-measurable function ϕ. We will use the notation $\mathbb{E}(X|\sigma(Y)) = \phi(Y)$. We read this in two ways. $E(X|\sigma(Y))$ is a random variable on $(\Omega, \mathcal{F}, P)$ and $E(X|\sigma(Y))$ is some function of Y, as the notation suggests.

Example 20.1 *(Baby Bivariate Normal)* If X and Y are bivariate normal $(X, Y) \stackrel{d}{=} (X, X\cos\theta + Z\sin\theta)$, with $\cos\theta = \rho$, and X, Z i.i.d. $N(0,1)$, then:

$$\mathbb{E}X = \mathbb{E}Y = 0$$

$$\mathbb{E}X^2 = \mathbb{E}Y^2 = 1$$

$$\mathbb{E}(XY) = \rho = \text{correlation of } X \text{ and } Y$$

$$\mathbb{E}(Y|X) = \rho X$$

$$\mathbb{E}(X|Y) = \rho Y$$

Notice that if $(X', Y') \stackrel{d}{=} (X, Y)$ then $\mathbb{E}(X|Y) = \phi(Y)$, $\mathbb{E}(X'|Y') = \phi(Y')$.

20.4 Relation to Undergraduate Probability

If (X, Y) has joint density $f(x, y)$ with respect to Lebesgue measure $dx\,dy$,

$$\mathbb{P}(X \in dx, Y \in dy) = f(x, y)\,dx\,dy.$$

This means $\mathbb{E}(g(X, Y)) = \int_{\mathbb{R}}\int_{\mathbb{R}} g(x, y)f(x, y)\,dx\,dy$ for all $g \geq 0$ or g bounded. We let the density of X be $\mathbb{P}(X \in dx) = f_X(x)\,dx$ and the density of Y be $\mathbb{P}(Y \in dy) = f_Y(y)\,dy$. Then this means that $f_X(x) = \int f(x, y)\,dy$ and $f_Y(y) = \int f(x, y)\,dx$.

Now for all y with $f_Y(y) > 0$, we can define:

$$f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)} \geq 0, \text{ with } \int f_{X|Y=y}(x)\,dx = 1$$

This is the "formal" conditional density of X given $Y = y$. We can relate this to $\mathbb{E}(X|Y) = \phi(Y)$ by defining $\phi(y)$ as

$$\phi(y) = \begin{cases} \int x\frac{f(x,y)}{f_Y(y)}dx & \text{if } f_Y(y) > 0 \\ 0 & \text{if } f_Y(y) = 0 \end{cases}.$$

We must check that this is measurable with respect to $\sigma(Y)$: this follows from part of Fubini's theorem. Also, we need to check that $\mathbb{E}(\phi(Y)h(Y)) = \mathbb{E}(Xh(Y))$ for h measurable.

In the same setting, we can let $\mathbb{E}(k(X)|Y) = \phi_k(Y)$. $\phi_k(y)$ is defined by simply replacing x with $k(x)$:

$$\phi_k(y) = \begin{cases} \int k(x)\frac{f(x,y)}{f_Y(y)}dy & \text{if } f_Y(y) > 0 \\ 0 & \text{if } f_Y(y) = 0 \end{cases}.$$

In this setup, the expectation of any $k(X)$ given Y can be obtained by integration with respect to a regular conditional distribution, which is a distribution of X that depends on the value of Y. Here, given $Y = y$, we have the density $f_{X|Y=y}(x) = f(x,y)/f_Y(y)$.

References

[1] Richard Durrett. *Probability: theory and examples, 3rd edition.* Thomson Brooks/Cole, 2005.