Part III – Advanced Probability (Incomplete)

Based on lectures by Dr. S. Andres
Notes taken by Bhavik Mehta

Michaelmas 2019

Contents

Lecture 1

1 Conditional Expectations

Lecture 2 Take a **probability space** $(\Omega, \mathcal{F}, \mathbb{P})$, meaning $\mathcal{F}$ is a σ-algebra and $\mathbb{P}$ is a probability measure, with $\mathbb{P}(\Omega) = 1$. We use the term '**almost surely**' (or a.s.) to mean almost everywhere.

Take X to be a random variable, i.e. $X : \Omega \to \mathbb{R}$ which is $\mathcal{F}$-measurable and write

$$\mathbb{E}[X] = \int X \, d\mathbb{P}$$

for the **expectation** of X. We write also

$$\mathbb{E}[X\mathbb{1}_A] = \int_A X \, d\mathbb{P}$$

for $A \in \mathcal{F}$.

Definition 1.1. Let $B \in \mathcal{F}$ with $\mathbb{P}[B] > 0$. We know

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]},$$

the **conditional probability** of A given B. Similarly,

$$\mathbb{E}[X|B] = \frac{\mathbb{E}[X\mathbb{1}_B]}{\mathbb{P}[B]}$$

the **conditional expectation** of X given B.

There is a significant restriction to this definition: that $\mathbb{P}[B] > 0$. By the end of this lecture, we will generalise this definition to any σ-algebra of events, rather than just one.

Aim. Improve the prediction of X if additional information (given as a sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$) is available.

1.1 Discrete case

Take $B_1, B_2, \ldots \in \mathcal{F}$ a disjoint decomposition of Ω. We take

$$\mathcal{G} = \sigma(B_1, B_2, \ldots) = \left\{ \bigcup_{i \in J} B_i : J \subseteq \mathbb{N} \right\} \subseteq \mathcal{F}.$$

That is, the 'extra information' of $\mathcal{G}$ is that we know which of the disjoint events B_i we fall into.

Then,

$$\mathbb{E}[X|\mathcal{G}](\omega) := \sum_{i:\mathbb{P}[B_i]>0} \mathbb{E}[X|B_i]\mathbb{1}_{B_i}(\omega)$$

is the conditional expectation of X given $\mathcal{G}$.

It is easy to see that $\mathbb{E}[X|\mathcal{G}]$ is a $\mathcal{G}$-measurable random variable, and

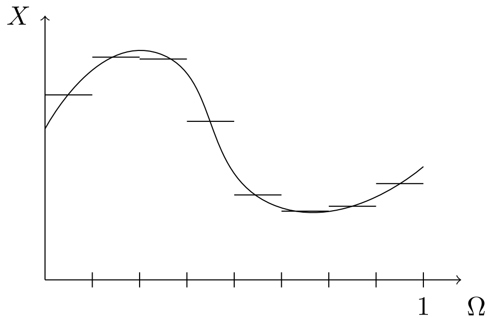
$$\mathbb{E}[\mathbb{1}_A X] = \mathbb{E}[\mathbb{1}_A \mathbb{E}[X|\mathcal{G}]] \quad \forall A \in \mathcal{G}.$$

Example.

(i) Take now $\Omega = (0,1]$, and $\mathcal{F} = \mathcal{B}(\Omega)$, and $\mathbb{P}$ to be Lebesgue measure. Use X as shown below, and use

$$\mathcal{G} = \sigma\left(\left(\tfrac{k}{m}, \tfrac{k+1}{m}\right] : k = 0, \dots, m-1\right).$$

In the picture, we take $m = 8$, and the conditional expectation $\mathbb{E}(X|\mathcal{G})$ is shown.



(ii) Take a random variable $Z : \Omega \to \{z_1, z_2, \dots\} \subseteq \mathbb{R}$, and use $\mathcal{G} = \sigma(Z) = \sigma(\{Z = z_i\} : i = 1, 2, \dots)$. Then,

$$\begin{aligned}\mathbb{E}[X|Z] &:= \mathbb{E}[X|\sigma(Z)] \\ &= \sum_{i:\mathbb{P}[Z=z_i]>0} \mathbb{E}[X|Z = z_i]\mathbb{1}_{\{Z=z_i\}}.\end{aligned}$$

This is not satisfactory quite yet: if Z has an absolutely continuous distribution (eg $\mathcal{N}(0,1)$), i.e. $\mathcal{P}[Z = z] = 0$ for every z, then $\mathbb{E}[X|Z]$ is not defined yet!

1.2 General case

Definition 1.2. Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. A random variable Y is called (a version of) the **conditional expectation** of X given $\mathcal{G}$ if

(i) Y is $\mathcal{G}$-measurable

(ii) $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$ for all $A \in \mathcal{G}$.

We notate $Y = \mathbb{E}[X|\mathcal{G}]$.

Remark 1.3.

(a) We took $X \in L^1$, but this can be changed to $X \geq 0$ throughout.

(b) If $\mathcal{G} = \sigma(\mathcal{C})$ for some $\mathcal{C} \subseteq \mathcal{F}$ which is a π**-system** (i.e. stable under intersections), it suffices to check (ii) for all $A \in \mathcal{C}$.

(c) If $\mathcal{G} = \sigma(Z)$ where Z is a random variable, we write $\mathbb{E}[X|Z] := \mathbb{E}[X|\sigma(Z)]$. This is $\sigma(Z)$ measurable by (i), so it's of the form $f(Z)$ for some function f. It's then common to define $\mathbb{E}[X|Z = z] = f(z)$.

Theorem 1.4 (Existence and uniqueness)**.** Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra.

(i) $E[X|\mathcal{G}]$ exists

(ii) Any two versions of $\mathbb{E}[X|\mathcal{G}]$ coincide $\mathbb{P}$-almost surely.

Proof.

(ii) Uniqueness. Let Y be as in Definition 1.2, and let Y' satisfy Definition 1.2(i) and (ii) for some $X' \in L^1$ with $X \le X'$ almost surely. Let $Z = (Y - Y')\mathbb{1}_A$ with $A := \{Y \ge Y'\} \in \mathcal{G}$.

$$\mathbb{E}[Y\mathbb{1}_A] = E[X\mathbb{1}_A] \le \mathbb{E}[X'\mathbb{1}_A] = \mathbb{E}[Y'\mathbb{1}_A] < \infty$$

and note that $\mathbb{E}[X'\mathbb{1}_A] < \infty$, so $\mathbb{E}[Y'\mathbb{1}_A] < \infty$.

By definition of Z, this means $\mathbb{E}[Z] \le 0$. But $Z \ge 0$ almost surely, so $Z = 0$ a.s. therefore $Y \le Y'$ a.s. (This shows monotonicity of conditional expectation.) If $X = X'$, we can run the same argument to show that $Y = Y'$ almost surely (using $A = \{Y > Y'\}$ and $A = \{Y < Y'\}$, we see both sets are measure zero).

(i) Existence. Step 1: Assume first $X \in L^2(\mathcal{F})$. Since $L^2(\mathcal{G})$ is a complete subspace of $L^2(\mathcal{F})$, X has an orthogonal projection Y on $L^2(\mathcal{G})$, i.e. there is $Y \in L^2(\mathcal{G})$ such that $\mathbb{E}[(X-Y)Z] = 0$ for every $Z \in L^2(\mathcal{G})$. Choosing $Z = \mathbb{1}_A$ for $A \in \mathcal{G}$ we get $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$ so Y satisfies the conditions of Definition 1.2.

Step 2: Assume $X \ge 0$. Then $X_n = X \wedge n \in L^2(\mathcal{F})$ and $0 \le X_n \nearrow X$ as $n \to \infty$. By Step 1, we can find $Y_n \in L^2(\mathcal{G})$ such that $\mathbb{E}[X_n\mathbb{1}_A] = \mathbb{E}[Y_n\mathbb{1}_A]$ for all $A \in \mathcal{G}$ and $0 \le Y_n \le Y_{n+1}$ almost surely (from the proof of (ii)). Let $Y_\infty = \lim_n Y_n \mathbb{1}_{\Omega_0}$ with

$$\Omega_0 = \{\omega \in \Omega : 0 \le Y_n(\omega) \le Y_{n+1}(\omega)\ \forall n\}.$$

Then Y_∞ is a non-negative random variable, is $\mathcal{G}$-measurable as a limit of $\mathcal{G}$-measurable r.v.s and by monotone convergence $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y_\infty\mathbb{1}_A]$ for every $A \in \mathcal{G}$. Taking $A = \Omega$, $\mathbb{E}[Y_\infty] = \mathbb{E}[X] < \infty$, since $X \in L_1$. So $Y_\infty < \infty$ almost surely and $Y := Y_\infty \mathbb{1}_{\{Y_\infty < \infty\}}$ satisfies Definition 1.2(i) and (ii).

Step 3: For general $X \in L^1$, apply Step 2 on X^+ and X^- to obtain Y^+ and Y^-. Then $Y = Y^+ - Y^-$ satisfies the conditions of Definition 1.2. $\square$

Lecture 3 **Example** (Conditional density functions)**.** Let U and V be random variables with a joint density function $f_{U,V}$ in $\mathbb{R}^2$. Then

$$f_U(u) = \int_{\mathbb{R}} f_{U,V}(u,v)\,dv$$

is the density of U, and

$$f_{V|U}(v|u) = \begin{cases} \frac{f_{U,V}(u,v)}{f_U(u)} & \text{if } f_U(u) > 0 \\ 0 & \text{else} \end{cases}$$

is the conditional density of V given U.

Assume $X = h(V) \in L^1$. Then $\mathbb{E}[X|U] = g(U)$ with $g(u) = \int_{\mathbb{R}} h(v) f_{V|U}(v|u)\,dv$. Indeed, since every $A \in \sigma(U)$ takes the form $A = \{U \in B\}$ for some $B \in \mathcal{B}(\mathbb{R})$.

$$\begin{aligned} \mathbb{E}[X\mathbb{1}_A] &= \int_{\mathbb{R}^2} h(v)\mathbb{1}_B(u) f_{U,V}(u,v)\,du\,dv \\ &= \int_{\mathbb{R}} \underbrace{\left(\int_{\mathbb{R}} h(v) f_{V|U}(v|u)\,dv\right)}_{g(u)} f_U(u)\mathbb{1}_B(u)\,du \\ &= \mathbb{E}[g(U)\mathbb{1}_{\{U\in B\}}] = \mathbb{E}[g(U)\mathbb{1}_A]. \end{aligned}$$

1.3 Properties of conditional expectation

Let $X \in L^1$, $\mathcal{G} \subseteq \mathcal{F}$ σ-algebras.

(i) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$ (proof: use $A = \Omega$ in Definition 1.2(ii))

(ii) If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] = X$ a.s. (proof: X satisfies the conditions of Definition 1.2)

(iii) If X is independent of $\mathcal{G}$ (i.e. $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$ for all $A \in \mathcal{G}$ and $B \in \sigma(X)$) then $\mathbb{E}[X|G] = \mathbb{E}[X]$ a.s. Proof: $\mathbb{E}[X]$ is constant and thus $\mathcal{G}$-measurable. For $A \in \mathcal{G}$

$$\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[X]\mathbb{E}[\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X]\mathbb{1}_A]$$

by independence then linearity.

(iv) If $X \geq 0$ almost surely then $\mathbb{E}[X|\mathcal{G}] \geq 0$ almost surely. (proof: see Theorem 1.4(ii)).

(v) $\mathbb{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha\mathbb{E}[X|\mathcal{G}] + \beta\mathbb{E}[Y|\mathcal{G}]$ almost surely for $Y \in L^1$ and $\alpha, \beta \in \mathbb{R}$.

(vi) If $0 \leq X_n \nearrow X$ almost surely, then $\mathbb{E}[X_n|\mathcal{G}] \nearrow \mathbb{E}[X|\mathcal{G}]$ almost surely. Proof: $\mathbb{E}[X_n|\mathcal{G}] \nearrow Y$ almost surely for some $\mathcal{G}$-measurable Y. For every $A \in \mathcal{G}$,

$$\mathbb{E}[X\mathbb{1}_A] = \lim_n \mathbb{E}[X_n\mathbb{1}_A] = \lim_n \mathbb{E}[\mathbb{E}[X_n|\mathcal{G}]\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$$

so $Y = \mathbb{E}[X|\mathcal{G}]$.

(vii) Fatou. If $X_n \geq 0$ almost surely $\forall n$, then

$$\mathbb{E}[\liminf X_n|\mathcal{G}] \leq \liminf \mathbb{E}[X_n|\mathcal{G}].$$

(Proof as for $\mathbb{E}[\,]$).

(viii) Dominated convergence. If $X_n \to X$ almost surely, and $|X_n| \leq Y$ almost surely $\forall n$ for $Y \in L^1$ then $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$ almost surely. Proof as for $\mathbb{E}[\,]$).

(ix) Jensen's inequality. If $c : \mathbb{R} \to (-\infty, \infty]$ convex, then $c(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[c(X)|\mathcal{G}]$. Proof: c can be written as

$$c(x) = \sup_n (a_n x + b_n) \quad x \in \mathbb{R}$$

so

$$\mathbb{E}[c(X)|\mathcal{G}] \geq a_n\mathbb{E}[X|\mathcal{G}] + b_n$$

for all n. Taking $\sup_n$ on the right gives the claim.

(x) $\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]|^p] \leq \mathbb{E}[|X|^p]$ for $1 \leq p < \infty$. Follows from (ix) and (ii).

(xi) Tower property: If $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ σ-algebras, then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$$

almost surely. Proof: Clearly $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$ is $\mathcal{H}$-measurable. Take $A \in \mathcal{H}$, so $A \in \mathcal{G}$. Then

$$\begin{aligned}\mathbb{E}[X\mathbb{1}_A] &= \mathbb{E}[\mathbb{1}_A\mathbb{E}[X|\mathcal{G}]] \\ &= \mathbb{E}[\mathbb{1}_A\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]].\end{aligned}$$

(xii) Let $Y \in L^1$ be $\mathcal{G}$-measurable, and such that $XY \in L^1$. Then

$$\mathbb{E}[YX|\mathcal{G}] = Y\mathbb{E}[X|G]$$

almost surely. '$\mathcal{G}$-measurable random variables behave like constants'.

Proof: The right hand side of $\mathcal{G}$-measurable. If $Y = \mathbb{1}_B$ for $B \in \mathcal{G}$. Then $\forall A \in \mathcal{G}$,

$$\mathbb{E}[XY\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_{A\cap B}] = \mathbb{E}[(\mathbb{E}[X|\mathcal{G}]\mathbb{1}_B)\mathbb{1}_A].$$

So the claim holds for simple random variables. For general Y, the statement follows by linearity, approximation, etc.

(xiii) If $\sigma(X, \mathcal{G})$ is independent of $\mathcal{H}$ then $\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{G}]$ almost surely. Proof: For $A \in \mathcal{G}$ and $B \in \mathcal{H}$,

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})]\mathbb{1}_{A\cap B}] &= \mathbb{E}[X\mathbb{1}_{A\cap B}] \\
&= \mathbb{E}[X\mathbb{1}_{A\cap B}] \\
&= \mathbb{E}[X\mathbb{1}_A]\mathbb{P}[B] \\
&= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}_A]\mathbb{P}[B] \\
&= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}_{A\cap B}] \\
\implies \mathbb{E}[(\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] - \mathbb{E}[X|\mathcal{G}])\mathbb{1}_{A\cap B}] &= 0
\end{aligned}$$

The set of such intersections $A \cap B$ is a π-system generating $\sigma(\mathcal{G}, \mathcal{H})$, and it is a standard result of measure theory that this implies $\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] - \mathbb{E}[X|\mathcal{G}] = 0$ almost surely (see PM notes, Proposition 3.1.4).

Index