Feynman Liang
STAT 205A
Lecture #16
10/18/2016

1 General properties of Conditional Expectation

1.1 Idea

Mimic general properties of ordinary expectations

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}X_1 + \mathbb{E}X_2 \qquad \mathbb{E}(cX) = c\mathbb{E}X \tag{1.1}$$

but with $\mathcal{G}$-measurable RVs playing the role of constants c.

1.2 Some basic properties of CE

Let $X : (\Omega, \mathcal{F}P) \to \mathbb{R}$, $\mathbb{E}|X| < \infty$, $\mathcal{G} \subset \mathcal{F}$. $\mathbb{E}[X \mid \mathcal{G}]$ is *the* RV Z such that

(a) Z is $\mathcal{G}$-measurable

(b) $\mathbb{E}[Z1_G] = \mathbb{E}[X1_G] \quad \forall G \in \mathcal{G}$.

Lemma 1.1. *For $Z = \mathbb{E}[X \mid \mathcal{G}]$, $\mathbb{E}[VZ] = \mathbb{E}(VX]$, we have*

(a) $\mathbb{E}[X_1 + X_2 \mid \mathcal{G}] = \mathbb{E}[X_1 \mid \mathcal{G}] + \mathbb{E}[X_2 \mid \mathcal{G}]$

(b) $\mathbb{E}[VX \mid \mathcal{G}] = V\mathbb{E}[X \mid \mathcal{G}]$ *for bounded $\mathcal{G}$-measurable V.*

(c) If $0 \le X_n \uparrow X$ *a.s., then* $\mathbb{E}[X_n \mid \mathcal{G}] \uparrow \mathbb{E}[X \mid Y]$ *a.s.*

(d) If $X \ge 0$ *a.s., then* $\mathbb{E}[X \mid \mathcal{G}] \ge 0$ *a.s.*

(e) $|\mathbb{E}[X \mid \mathcal{G}]| \le \mathbb{E}[|X| \mid \mathcal{G}]$ *a.s.*

(f) $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}X$

(g) If X is $\mathcal{G}$-measurbale, then $\mathbb{E}[X \mid \mathcal{G}] = X$
If $\mathcal{G}$ is trivial, then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}X$.

(h) ***Tower Property****: If $\mathcal{G} \subset \mathcal{H}$ then* $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{H} \mid \mathcal{G}]]$.

Proof. (a) Write $Z_i = \mathbb{E}[X_i \mid \mathcal{G}]$.
Need to show $Z := Z_1 + Z_2 = \mathbb{E}[X_1 + X_2 \mid \mathcal{G}]$
Z is $\mathcal{G}$-measurable because Z_i are $\mathcal{G}$-measurable.

$$\underbrace{\mathbb{E}[Z1_G]}_{=\mathbb{E}[X_1 1_G]+\mathbb{E}[X_2 1_G]} \qquad \underbrace{(X_1 + X_2)1_G}_{=\mathbb{E}[X_1 1_G]+\mathbb{E}[X_2 1_G]} \qquad \forall G \in \mathcal{G} \tag{1.2}$$

(b) Define $Z = V\mathbb{E}[X \mid \mathcal{G}]$. To show $Z = \mathbb{E}[VX \mid \mathcal{G}]$, need to show Z, V, and $\mathbb{E}[X \mid \mathcal{G}]$ are $\mathcal{G}$-measurable.

Z is $\mathcal{G}$-measurable by Lemma applied to $V1_G$. TODO: Check

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] \underbrace{V1_G}_{\mathcal{G}\text{-meas}} = \mathbb{E}[X \underbrace{V1_G}_{\mathcal{G}\text{-meas}}] \quad \forall G \in \mathcal{G} \tag{1.3}$$

(c) Exercise.

(d) Exercise.

(e) Exercise.

(f) $G = \Omega$ in def.

(g) By definition.

$\mathcal{G}$ trivial $\implies$ $\mathbb{E}[X \mid \mathcal{G}]$ constant $\implies$ $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}X$

(h) Write $Z = \mathbb{E}[X \mid \mathcal{G}]$. Need to check $\mathbb{E}[Z1_G] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{H}]1_G]$. But LHS $= \mathbb{E}[X1_G]$ by definition of Z, and RHS $= \mathbb{E}[X1_G]$ by definition of $\mathbb{E}[X \mid H]$ and $G \in \mathcal{G} \implies G \in \mathcal{H}$.

□

(L^2 setting): Now assume $\mathbb{E}X^2 < \infty$.

- $X \mapsto \mathbb{E}[X \mid \mathcal{G}]$ is the orthogonal projection in Hilbert space
- Cauchy-Schwarz $\mathbb{E}|VX| \leq \sqrt{(\mathbb{E}X)^2(\mathbb{E}V)^2} < \infty$

From Lemma

$$\mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}]) \mid V] = 0 \tag{1.4}$$

for V $\mathcal{G}$-measurable and $\mathbb{E}V^2 < \infty$. This gives

Lemma 1.2. *$X - \mathbb{E}[X \mid \mathcal{G}]$ and V are orthogonal $\forall V$ $\mathcal{G}$-measurable.*

Recall $\mathrm{Var}(X) = \mathbb{E}[X - \mathbb{E}[X]]^2$.

Definition 1.3. The *conditional variance*

$$\mathrm{Var}(X \mid \mathcal{G}) = \mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])^2 \mid \mathcal{G}] \tag{1.5}$$

Lemma 1.4 (Bias-variance decomposition)**.** *If Y is $\mathcal{G}$-measurable, $\mathbb{E}Y^2 < \infty$*

$$\mathbb{E}[(X - Y)^2 \mid \mathcal{G}] = Var(X \mid \mathcal{G}) + (\mathbb{E}[X \mid \mathcal{G}] - Y)^2 \tag{1.6}$$

Proof.

$$\mathbb{E}[(X-Y)^2 \mid \mathcal{G}] = \mathbb{E}[X^2 - 2XY + Y^2 \mid \mathcal{G}] \tag{1.7}$$
$$= \mathbb{E}[X^2 \mid \mathcal{G}] - 2Y\mathbb{E}[X \mid \mathcal{G}] + Y^2 \tag{1.8}$$
$$= (\mathbb{E}[X^2 \mid \mathcal{G}] - \mathbb{E}[X \mid \mathcal{G}]^2) + (\mathbb{E}[X \mid \mathcal{G}]^2 - 2Y\mathbb{E}[X \mid \mathcal{G}] + Y^2) \tag{1.9}$$
$$= \mathrm{Var}(X \mid \mathcal{G}) + (\mathbb{E}[X \mid \mathcal{G}] - Y)^2 \tag{1.10}$$

□

Lemma 1.5. *$Var(X) = \mathbb{E}Var(X \mid \mathcal{G}) + Var\mathbb{E}[X \mid \mathcal{G}]$*

Proof. Replace X by $X-$ changes no term, so wlog assume $\mathbb{E}X = 0$.

$$\mathrm{Var}(X) = \mathbb{E}[X^2] = \mathbb{E}[\mathbb{E}[X^2 \mid \mathcal{G}]] \tag{1.11}$$
$$\mathbb{E}[X^2 \mid \mathcal{G}] = \mathbb{E}\left[\left(\underbrace{(X - \mathbb{E}[X \mid \mathcal{G}])}_{\to a} + \underbrace{\mathbb{E}[X \mid \mathcal{G}]}_{\to b}\right)^2 \mid \mathcal{G}\right], \qquad \mathbb{E}[ab \mid \mathcal{G}] = 0 \tag{1.12}$$
$$= \mathbb{E}[a^2 \mid \mathcal{G}] + b^2 \tag{1.13}$$
$$= \mathrm{Var}(X \mid \mathcal{G}) + (\mathbb{E}[X \mid \mathcal{G}])^2, \qquad \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}X = 0 \tag{1.14}$$
$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid \mathcal{G}) + (\mathbb{E}[X \mid \mathcal{G}])^2] \tag{1.15}$$
$$= \mathbb{E}\mathrm{Var}(X \mid \mathcal{G}) + \underbrace{\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] - 0]^2}_{=\mathrm{Var}\mathbb{E}[X \mid \mathcal{G}]} \tag{1.16}$$

□

Lemma 1.6 (Connection with independence)**.** *A S-valued RV X is independent of $\mathcal{G}$ $\iff$ $\mathbb{E}[h(X) \mid \mathcal{G}] = \mathbb{E}h(X)$ $\forall$ bounded measurable $h : S \to \mathbb{R}$.*

Proof. $\Rightarrow$. NTS $\mathbb{E}[\mathbb{E}h(x)1_G] = \mathbb{E}[h(X)1_G]$ $\forall G \in \mathcal{G}$. But $\mathbb{E}h(X)$ is a consstant so LHS $= (\mathbb{E}h(X))(\mathbb{E}1_G)$ and by independence RHS $= (\mathbb{E}h(X))(\mathbb{E}1_G)$

$\Leftarrow$. Take $h = 1_B$ for $B \subset S$. From the same argument

$$\mathbb{E}[h(X)1_G] = \mathbb{E}[h(x)]\mathbb{E}[1_G] \tag{1.17}$$
$$= P(X \in B, G) = P(X \in B)P(G) \tag{1.18}$$

Holds $\forall B, G \implies X$ and $\mathcal{G}$ independent. □

2 Background to conditional independence

Recall

Definition 2.1. *X, Y independent* $\iff$ $\mathbb{E}(h_1(X)h_2(Y)) = (\mathbb{E}h_1(X))(\mathbb{E}h_2(Y))$ for all bounded meas. h_1, h_2

Example 2.2. Bayes (X_i) conditionally independent given Θ

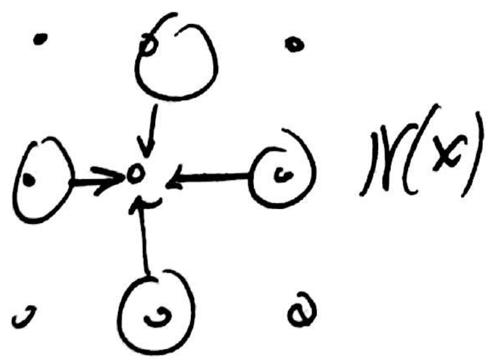
(i) Random Θ, values in $\{\text{PMs on } \mathbb{R}\} = \mathcal{P}(\mathbb{R})$

(ii) Conditional on $\Theta = \theta \in \mathcal{P}(\mathbb{R})$ take $X_1, X_2, X_3, \cdots$ IID(θ).

Simple Markov property for $(X_n, n \geq 0)$ Past $X_{0:(n-1)}$ and future X_{n+1} conditionally independnet given present X_n.

$$P(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \cdots, X_0 = x_0) = P(X_{n+1} = x_{n+1} \mid X_n = x_n) \tag{2.1}$$

Locally dependent : Given $(W_{\tilde{x}}, \tilde{x} = (x_1, x_2) \in \mathbb{Z}^2)$.



Idea: $W_{\tilde{x}}$ depends only on $W_{\tilde{y}} : \tilde{y} \in N(\tilde{x})$ and not on the other Ws.

Formally: TODO: Conditionally indep given ...