Feynman Liang
STAT 205A
Lecture #17
10/20/2016

1 Conditional independence/expectation

Jensen's inequality: $E\phi(X) \geq \phi(\mathbb{E}X)$ if ϕ convex, $\mathbb{E}|X| < \infty$, $\mathbb{E}|\phi(X)| < \infty$

Conditional Jensen's inequality: $\mathbb{E}[\phi(X) \mid \mathcal{G}] \geq \phi(\mathbb{E}[X \mid \mathcal{G}])$ a.s.

Recall in MT, independence is property of $\mathcal{G}_1$, $\mathcal{G}_2$. Random variables X and Y are independent

$\iff$ $\sigma(X)$ and $\sigma(Y)$ are independent

$\iff$ $\mathbb{E}[h_1(X_1)h_2(X_2)] = (\mathbb{E}h_1(X_1)) \times (\mathbb{E}h_2(X_2))\ \forall h_i : S_i \to \mathbb{R}$ bounded meas.

$\iff$ $\mathbb{E}[h_1(X_1) \mid X_2] = Eh_1(X_1)$ a.s. $\forall h_1$

Undergrad version: Given discrete RV V, define $P(X_1 = x_1 \mid V = v)$, $P(X_2 = x_2 \mid V = v)$. Then construct (X_1, X_2, V) such that $P(X_1 = x_2, X_2 = x_2 \mid V = v) = P(X_1 =_2| V = v) \times P(X_2 = x_2 \mid V = v)$.

MT version: X_1 and X_2, with σ-fields $\mathcal{H}_1, \mathcal{H}_2$, are *conditionally independent* given $\mathcal{G}$ means

$$\mathbb{E}[\underbrace{h_1(X_1)}_{\text{bdd } \mathcal{H}_1\text{-meas RV}} h_2(X_2) \mid \mathcal{G}] = \mathbb{E}[h_1(X_1) \mid \mathcal{G}] \times \mathbb{E}[h_2(X_2) \mid \mathcal{G}]\ \forall h_i \tag{1.1}$$

Homework later: This is equivalent to

$$\mathbb{E}[h_1(X_1) \mid \mathcal{G}, X_2] = \mathbb{E}[h_1(X) \mid \mathcal{G}] \text{ a.s. } \forall h_1 \tag{1.2}$$

Once you know $\mathcal{G}$, knowing also X_2 gives no *extra* information about X_1.

1.0.1 Relation between conditional probability and conditional expectation

Undergrad: Conditional probabilities and expectations are related in the following way:

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)} \tag{1.3}$$

$$\mathbb{E}[h(Y) \mid X = x] = \sum_y h(y)P(Y = y \mid X = x) \tag{1.4}$$

Graduate: (Conditional probability) $(X, Y) : (\Omega, \mathcal{F}, P) \to S_1 \times S_2$ get kernel Q from S_1 to S_2. $Q(x, B)$ means $P(Y \in B \mid X = x)$.

(Conditional expecation) $W : (\Omega, \mathcal{F}, P) \to \mathbb{R}$, $\mathbb{E}|W| < \infty$, $\mathcal{G} \subset \mathcal{F}$, $\mathbb{E}[W \mid \mathcal{G}] = Z$, specified by $\mathbb{E}[Z1_G] = \mathbb{E}[W1_G]\ \forall G \in \mathcal{G}$.

Where did the connection between the two go?

Write $I : (\Omega, \mathcal{F}) \to (\Omega, \mathcal{G})$ identity function, $(I, Y) : \Omega \to (\Omega, \mathcal{G}) \times (S_2, \mathcal{S}_2)$, $\alpha(\omega, B)$ kernel associated with (I, Y), $\alpha(\omega, B)$ means $P(Y \in B \mid \mathcal{G})(\omega)$. This is called the *regular conditional distribution for Y given $\mathcal{G}$*.

Write $W = h(Y)$, $h : S_{(Y)} \to \mathbb{R}$, $\mathcal{G} = \sigma(X)$.

What is $\mathbb{E}[h(Y) \mid X = x]$ in MT?

$$\mathbb{E}[h(Y) \mid \mathcal{G}](\omega) = \int h(y)\alpha(\omega, dy) \tag{1.5}$$

Proof is a homework exercise.

2 Martingales

A σ-field $\mathcal{G}$ is a collection of events. $A \in \mathcal{G}$ means A is an event.

For RV X, say X *is $\mathcal{G}$-measurable* to mean $\sigma(X) \subset \mathcal{G}$.

2.1 General setup

Definition 2.1. For a probability space $(\Omega, \mathcal{F}, P)$, a sequence of nested sub-σ-fields $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}$ is called a *filtration*.

We interpret $\mathcal{F}_n$ as the "information known at time n."

Definition 2.2. A sequence $(X_n)_{n \geq 0}$ is *adapted* to $(\mathcal{F}_n)$ means $X_n \in \mathcal{F}_n$ $\forall n$.

Definition 2.3. A $\mathbb{R}$-valued process $(X_n)_{0 \leq n < \infty}$ is a *martingale* (MG) if

(a) $\mathbb{E}|X_n| < \infty$ $\forall n$

(b) (X_n) is adapted to $(\mathcal{F}_n)$

(c) $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$, $0 \leq n < \infty$

- *sub-martingale*: $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n$, $0 \leq n < \infty$, (X_n below i.e. sub X_{n+1})
- *super-martingale*: $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n$, $0 \leq n < \infty$

Typical uses of the theory:

- Complicated (Y_n)
- We look for h such that $h(Y_n)$ is a MG
- Take $\mathcal{F}_n = \sigma(Y_0, Y_1, \cdots, Y_n)$
- $X_n = h(Y_n)$, (X_n) *is* adapted to $(\mathcal{F}_n)$

Convention: If we define X_n and say "X_n is a MG", we are taking

$$\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n) \tag{2.1}$$

this is called the *natural filtration*.

Example 2.4. Let $\xi_1, \xi_2, \cdots$ be *independent* RVs, $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$ the natural filtration.

(1) If $\mathbb{E}|\xi_i| < \infty$ and $\mathbb{E}\xi_i = 0\ \forall i$, then $S_n = \sum_{i=1}^n \xi_i$ is a MG. To check this, note

$$\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[\underbrace{S_n}_{\in \mathcal{F}_n} + \xi_{n+1} \mid \mathcal{F}_n] = S_n + \mathbb{E}[\underbrace{\xi_{n+1} \mid \mathcal{F}_n}_{\text{indep}}] = S_n + \mathbb{E}\xi_{n+1} = S_n \tag{2.2}$$

(2) As in (1), suppose also $\sigma_i^2 = \mathbb{E}\xi_i^2 < \infty$. Then $Q_n = S_n^2 - \sum_{i=1}^n \sigma_i^2$ is a MG.

$$Q_{n+1} - Q_n = S_{n+1}^2 - S_n^2 - \sigma_{n+1}^2 = 2S_n\xi_{n+1} + \xi_{n+1}^2 - \sigma_{n+1}^2 \tag{2.3}$$

$$\mathbb{E}[Q_{n+1} - Q_n \mid \mathcal{F}_n] = \mathbb{E}[2\underbrace{S_n}_{\in \mathcal{F}_n} \xi_{n+1} \mid \mathcal{F}_n] + \mathbb{E}[\underbrace{\xi_{n+1}^2 \mid \mathcal{F}_n}_{\text{indep}}] - \sigma_{n+1}^2 \tag{2.4}$$

$$= 2S_n \underbrace{\mathbb{E}[S_{n+1} \mid \mathcal{F}_n]}_{=0} = 0 \tag{2.5}$$

(3) Suppose (xi_i) independent, $\mathbb{E}\xi_i = 1$, then $M_n = \prod_{i=1}^n \xi_i$ is a MG.

$$M_{n+1} = \underbrace{M_n}_{\in \mathcal{F}_n} \xi_{n+1} \tag{2.6}$$

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[M_n\xi_{n+1} \mid \mathcal{F}_n] = M_n\mathbb{E}[\xi_{n+1} \mid \mathcal{F}_n] = M_n\mathbb{E}[\xi_{n+1}] = M_n \cdot 1 \tag{2.7}$$

(4) Take (ξ_i) iid. Take density functions f and $g > 0$. Define $L_n = \prod_{i=1}^n \frac{g(\xi_i)}{f(\xi_i)}$.

(a) If (ξ_i) have density f then $\forall g$, (L_n) is a MG.

$$L_n = \prod_{i=1}^n Y_i \qquad Y_i = \frac{g(\xi_i)}{f(\xi_i)} \tag{2.8}$$

$$\mathbb{E}Y_i = \int \frac{g(y)}{f(y)} f(y)dy \tag{2.9}$$

$$= \int g(y)dy = 1 \tag{2.10}$$

(b) If (ξ_i) has density g then, provided $\mathbb{E}L_n < \infty$, (L_n) is a sub-MG.

$$(a) \implies (1/L_n) \text{ is a sub MG} \tag{2.11}$$

$$\text{conditional Jensen} \implies 1/L_n = \mathbb{E}(1/L_n \mid \mathcal{F}_n) \geq \frac{1}{\mathbb{E}[L_{n+1} \mid \mathcal{F}_n]} \implies \mathbb{E}[L_{n+1} \mid \mathcal{F}_n] \geq L_n \text{ su} \tag{2.12}$$