Conditional Expectations relative to σ-algebras

Wednesday, November 16, 2011
11:10 AM

Sponsored by the National Grid Foundation nationalgrid The power of action.

Def: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a prob. space, $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra, and $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$

The conditional expectation of X given $\mathcal{G}$: $\mathbb{E}[X|\mathcal{G}]$ is a random variable such that

(i) $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$ measurable

(ii) $\int_A \mathbb{E}[X|\mathcal{G}] \, d\mathbb{P} = \int_A X \, d\mathbb{P}$ for all $A \in \mathcal{G}$

Remark: if $\mathcal{G} = \sigma(Y)$ for some random variable Y we define $\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)]$

examples:

(1) $\mathcal{G} = \mathcal{F} \Rightarrow \mathbb{E}[X|\mathcal{G}] = X$

(2) $\mathcal{G} = \{\phi, \Omega\} \Rightarrow \mathbb{E}[X|\mathcal{G}] = \mathbb{E}X$

i) $\mathbb{E}X$ is a constant, thus it is $\mathcal{G}$ measurable

ii) if $A \in \mathcal{G}$ then $A = \Omega$ or ϕ

$$\int_\Omega \mathbb{E}X \, dP = \mathbb{E}X \cdot \int_\Omega 1 \cdot dP = \mathbb{E}X = \int_\Omega X \, dP$$

(3) let $D \in \mathcal{F}$: $\mathcal{G} = \sigma(1_D) = \{\phi, D, D^c, \Omega\}$

$\Rightarrow \mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X|1_D]$ is a random variable measurable with respect to $\mathcal{G}$.

$\rightsquigarrow \mathbb{E}[X|\mathcal{G}]$ should be constant on D and on D^c

Recall : if σ-algebra $\mathcal{G}$ is generated by a finite or countably infinite partition of Ω then

[X is $\mathcal{G}$ measurable iff X is constant on each element of this partition]

$$E[X \mid 1_D] = \begin{cases} c_1 & \text{on } D \\ c_2 & \text{on } D^c \end{cases}$$

How to find c_1, c_2?

$$\left.\begin{aligned} \int_D E[X \mid 1_D]\, dP &= \int_D X\, dP \\ \int_D E[X \mid 1_D]\, dP &= \int_D c_1\, dP = c_1 \cdot P(D) \end{aligned}\right\} \Rightarrow c_1 = \frac{\int_D X\, dP}{P(D)}$$

- similarly $c_2 = \dfrac{\int_{D^c} X\, dP}{P(D^c)}$

- in particular, if $\Omega = [0,1]$; $\mathcal{F} = \mathcal{B}_{[0,1]}$; $\mathbb{P} = m_{[0,1]}$ and $D = [0, \frac{1}{3}]$; $X(\omega) = \omega$, then

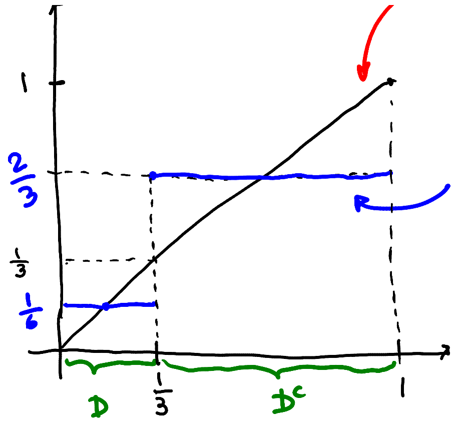
$$\rightsquigarrow E[X \mid 1_D](\omega) = \frac{1}{\mathbb{P}([0,\frac{1}{3}])} \cdot \int_0^{1/3} \omega\, d\omega = 3 \cdot \frac{\omega^2}{2}\Big|_0^{1/3} = \frac{1}{6}$$

if $\omega \in [0, \frac{1}{3}]$

$$\rightsquigarrow E[X \mid 1_D](\omega) = \frac{1}{P([\frac{1}{3}, 1])} \int_{1/3}^{1} \omega\, d\omega = \frac{3}{2} \cdot \frac{\omega^2}{2}\Big|_{1/3}^{1} = \frac{2}{3}$$

if $\omega \in (\frac{1}{3}, 1]$

$X(\omega)$



$E[X \mid 1_D] = f(1_D)$ where

$$f(x) = \begin{cases} 1/6 & \text{if } x = 1 \\ 2/3 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

$\Omega = [0,1]$

(4) this example can be generalized to the case when $\mathcal{G}$ is generated by a finite or countably infinite partition. $\iff$ $\mathcal{G} = \sigma(Y)$ where Y is discrete r.v.

- More precisely, let $D_1, D_2, \dots$ form a partition of Ω : $D_i \in \mathcal{F}$, $\mathbb{P}(D_i) > 0$

$\rightsquigarrow$ let $\mathcal{G} = \sigma(D_1, D_2, \dots)$, $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ then

$$E[X \mid \mathcal{G}](\omega) = \frac{\int_{D_i} X \, dP}{P(D_i)} \quad \text{for } \omega \in D_i$$

- if $\mathcal{G} = \sigma(Y)$, where Y is discrete then

$$E[X \mid Y](\omega) = \frac{\int_{\{Y = y_i\}} X \, dP}{P(Y = y_i)} \quad \text{for } \omega : Y(\omega) = y_i$$

- let's get back to the general def.

Questions regarding $E[X \mid \mathcal{G}]$: Existence ; Uniqueness

$\rightsquigarrow$ the answers come from Radon-Nikodym th.

① Let $X \geq 0$, st. $EX \neq 0$; define $Q : \mathcal{G} \to [0,1]$ s.t.

$\int X \, dP$

$$Q(A) = \frac{\int_A X\, dP}{\int_\Omega X\, dP} \qquad \text{for all } A \in \mathcal{G}$$

$\Longrightarrow$ Q is a prob. measure and $Q \ll P$.
(easy to check)

- By R-N th. there exists a r.v. $Z(\omega)$ which is $\mathcal{G}$ measurable, such that

$$Q(A) = \int_A Z(\omega)\, dP(\omega)$$

thus:
$$\int_A Z(\omega)\, dP(\omega) = \int_A \frac{X(\omega)}{EX}\, dP(\omega) \qquad (\forall)\, A \in \mathcal{G}$$

def of $E[x|\mathcal{G}]$

$\Rightarrow$ $Z(\omega) \cdot EX$ satisfies (i) & (ii)
since $Z(\omega)$ is unique $\Rightarrow$ $E[X|\mathcal{G}]$ exists and is unique.

(2) In general: $X = X^+ - X^-$ so we use the existence & uniqueness for X^+, X^-

$$E[X|\mathcal{G}] = E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}]$$

Corollary:
- If $X \geq 0$ then $E[X|\mathcal{G}] \geq 0$ a.s.
- If $X \geq Y \geq 0$ then $E[X|\mathcal{G}] \geq E[Y|\mathcal{G}]$ a.s.

MCT
- If $X_n \geq 0$, $X_n \uparrow X$ a.s. then $E[X_n|\mathcal{G}] \uparrow E[X|\mathcal{G}]$ a.s.

DCT
- If $\lim_{n\to\infty} X_n = X$ a.s and $|X_n| \leq Y$ a.s, $Y \in L^1$

then $\lim E[X_n|\mathcal{G}] = E[X|\mathcal{G}]$ a.s.

then $\lim_{n\to\infty} \mathbb{E}[X_n | \mathcal{G}] = \mathbb{E}[X|\mathcal{G}]$ a.s.

Proof of (MCT):

- let $Y_n = \mathbb{E}[X_n | \mathcal{G}]$; $Y = \lim_{n\to\infty} Y_n$ (a.s.)

  Y_n is $\mathcal{G}$ measurable $\Rightarrow$ Y is $\mathcal{G}$ measurable also

- Moreover, for all $A \in \mathcal{G}$ we have

  $1_A Y_n \uparrow 1_A Y$ a.s. as $n \to \infty$

- Monotone Convergence Th. gives :

$$\left.\begin{aligned} &\mathbb{E}\, 1_A Y_n = \int_A Y_n \, d\mathbb{P} \underset{(ii)}{=} \int_A X_n \, d\mathbb{P} \xrightarrow{MCT} \int_A X \, d\mathbb{P} \\ &\text{also } \mathbb{E}\, 1_A Y_n \longrightarrow \mathbb{E}\, 1_A Y = \int_A Y \, d\mathbb{P} \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow Y = \mathbb{E}[X|\mathcal{G}] \text{ a.s.}$$

Theorem: Let $X_1, X_2, X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$ σ-alg.

(A) Linearity: $\mathbb{E}[c_1 X_1 + c_2 X_2 | \mathcal{G}] = c_1 \mathbb{E}[X_1|\mathcal{G}] + c_2 \mathbb{E}[X_2|\mathcal{G}]$

(B) "Taking out what is known":

assume X_1 is $\mathcal{G}$ measurable, $X_1 X_2 \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, then

$$\mathbb{E}[X_1 X_2 | \mathcal{G}] = X_1 \cdot \mathbb{E}[X_2 | \mathcal{G}]$$

(C) "Tower Property" : if $\mathcal{H} \subset \mathcal{G}$, then

(**) $$\mathbb{E}\big[\mathbb{E}[X|\mathcal{G}] \,\big|\, \mathcal{H}\big] = \mathbb{E}[X|\mathcal{H}]$$

* $E\big[E[X|\mathcal{H}]\,|\,\mathcal{G}\big] = E[X|\mathcal{H}]$

(D) If X is independent of $\mathcal{G}$ then: $E[X|\mathcal{G}] = EX$

(E) Conditional Jensen's inequality: $\varphi(\cdot)$ convex

$$\varphi\big(E[X|\mathcal{G}]\big) \le E\big[\varphi(X)\,|\,\mathcal{G}\big]$$

(F)
- $\| E(X|\mathcal{G}) \|_p \le \|X\|_p$ for $p \geq 1$
- $X_n \xrightarrow{L^2} X \Rightarrow E(X_n|\mathcal{G}) \xrightarrow{L^2} E(X|\mathcal{G})$

(G) Repeated Conditioning: assume

$\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{G}_2 \subset \cdots \subset \mathcal{G}_\infty = \sigma\left(\bigcup_{n=1}^{\infty} \mathcal{G}_n\right)$

$X \in L^p$ with $p \geq 1$. Then

$$E[X|\mathcal{G}_n] \longrightarrow E[X|\mathcal{G}_\infty] \quad \text{a.s.}$$

$$E[X|\mathcal{G}_n] \longrightarrow E[X|\mathcal{G}_\infty] \quad \text{in } L^p$$

Proofs: (A) follows from the uniqueness of the cond. expectation

(B) standard arguement:

Step 1: $X_1 = 1_D$, $D \in \mathcal{G} \Rightarrow$

$E(X_1 X_2 \,|\, \mathcal{G}) = E(1_D X_2 \,|\, \mathcal{G})$ and for all $A \in \mathcal{G}$

$$\int_A E(1_D X_2 | \mathcal{G})\, dP = \int_A 1_D X_2 \, dP = \int_{A\cap D} X_2 \, dP = \int_{A \cap D} E(X_2|\mathcal{G})\, dP$$

$$= \int_A 1_D\, E(X_2|\mathcal{G})\, dP$$

- since both $E(1_D X_2|\mathcal{G})$ and $1_D E(X_2|\mathcal{G})$ are $\mathcal{G}$ meas.

and both satisfy (ii) by the uniqueness of the and both satisfy (ii), by the uniqueness of the conditional expectation we have

$$E(1_D X_2 | \mathcal{G}) = 1_D E(X_2 | \mathcal{G}) \quad \text{a.s}$$

Step2: extends the equality by linearity to the case where X_1 is a linear combination of indicator functions of some sets in $\mathcal{G}$. (X_1 = simple r.v $\mathcal{G}$ meas.)

Step3: extends to all $X_1 \geq 0$ by MCT

Step4: extends to all $X_1 \in L^1(\Omega, \mathcal{G}, P)$ by

$$X_1 = X_1^+ - X_1^-$$

(c) (*) obvious

(**)

- for any $G \in \mathcal{G}$: $\int_G E(X|\mathcal{G})\, dP = \int_G X\, dP$
- for any $H \in \mathcal{H}$: $\int_H E(X|\mathcal{H})\, dP = \int_H X\, dP$
- but $\mathcal{H} \subseteq \mathcal{G} \Rightarrow H \in \mathcal{G}$

$\Rightarrow$

$$\int_H E(X|\mathcal{G})\, dP = \int_H E(X|\mathcal{H})\, dP \quad (\forall)\ H \in \mathcal{H}$$

$$\Rightarrow E(E(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{H}) \quad \text{a.s.}$$

(D) X is independent of $\mathcal{G} \iff (\forall)\ A \in \mathcal{G}$, X and 1_A are independent

$$\Rightarrow E(X 1_A) = EX \cdot E1_A$$

- for an arbitrary $A \in \mathcal{G}$

$$\int_A E(X|\mathcal{G})\,dP = \int_A X\,dP = \int_\Omega X\cdot 1_A\,dP = E\left[X\cdot 1_A\right]$$

$$= EX \cdot E 1_A = EX \cdot P(A) = \int_A EX\,dP$$

$$\Rightarrow \int_A \left[E(X|\mathcal{G}) - E(X)\right] dP = 0 \quad (\forall)\ A\in\mathcal{G}$$

- since $E(X|\mathcal{G}) - EX$ is $\mathcal{G}$ measurable $\Rightarrow$ it has to be equal to 0. a.s

(E) proof is omitted.

(F) easy consequence of Jensen's ineq.

(G) proof is omitted.

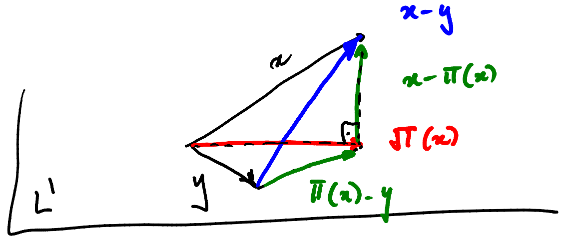
<u>Theorem</u>: let $X\in L^2(\Omega,\mathcal{F},P)$, $\mathcal{G}\subset\mathcal{F}$ σ-algebra, $Y\in L^1(\Omega,\mathcal{G},P)$

$$\Rightarrow E\left(X - E[X|\mathcal{G}]\right)^2 \le E(X-Y)^2$$

$\rightsquigarrow$ $E(X|\mathcal{G})$ is the best approximation to X in L^2 sense among all $\mathcal{G}$ measurable and square integrable r.v.

Analogy with: $\|x-y\|^2 = \|x-\pi(x)\|^2 + \|\pi(x)-y\|^2$

for $(\forall)$ $y\in L$



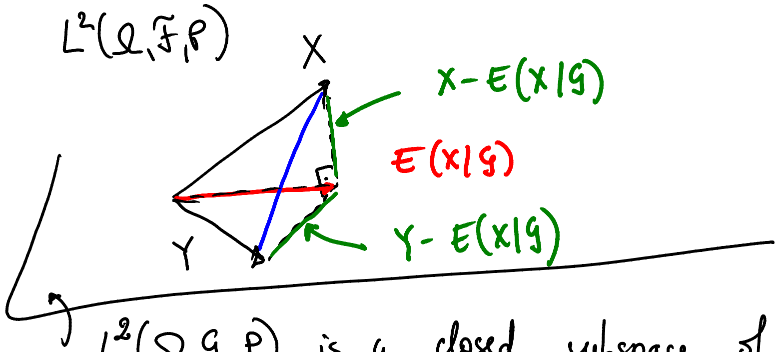
$\pi(x)$ orthogonal projection of x onto L'

$\|x-y\|^2 \ge \|x-\pi(x)\|^2$ for all $y\in L$

$$\|x-y\|^2 \geq \|x-\pi(x)\|^2 \quad \text{for all} \quad y \in L$$

$$\Rightarrow \|x-\pi(x)\|^2 = \min_{y \in L} \|x-y\|^2$$

- in terms of conditional expectation



$L^2(\Omega,\mathcal{G},P)$ is a closed subspace of $L^2(\Omega,\mathcal{F},P)$

$\langle X,Y\rangle := E(XY) \rightsquigarrow$ scalar product in L^2

$E(X|\mathcal{G}) - Y \perp X - E(X|\mathcal{G})$

which means $E\Big[\big(E(X|\mathcal{G})-Y\big)\underline{\big(X-E(X|\mathcal{G})\big)}\Big] = 0$

$$E\Big[X\cdot E(X|\mathcal{G}) - [E(X|\mathcal{G})]^2 - XY + Y E(X|\mathcal{G})\Big] =$$

$$\underbrace{E\Big[X\cdot E(X|\mathcal{G}) - (E(X|\mathcal{G}))^2\Big]}_{=0} \underbrace{- EXY + E\big(YE(X|\mathcal{G})\big)}_{=0}$$

- now recall Y is $\mathcal{G}$ measurable $\Rightarrow$ ·

$$Y E(X|\mathcal{G}) = E(X\cdot Y|\mathcal{G})$$

and also $E\big(E(XY|\mathcal{G})\big) = E(XY)$

(we have used "Tower" property & "Take out what is known")

- on the other hand

$\mathcal{G}$-measurable (known)

$$E\big(XE(X|\mathcal{G})\big) = E\big(E(X\cdot E(X|\mathcal{G})|\mathcal{G})\big)$$

$$\begin{cases} E\big(X E(X|\mathcal{G})\big) = E\Big(E\big(X\cdot \underbrace{E(X|\mathcal{G})} \,|\, \mathcal{G}\big)\Big) \\ = E\big(E(X|\mathcal{G})\cdot E(X|\mathcal{G})\big) = E\Big(\big(E(X|\mathcal{G})\big)^2\Big) \end{cases}$$

· suppose (X,Y) has a joint density $f(x,y)$
$\mathcal{G} = \sigma(Y)$. How to find $E(X|Y) = ?$

· answer: $f_{X|Y}(x|y) = \begin{cases} \dfrac{f(x,y)}{f_Y(y)} & \text{if } f_Y(y) \neq 0 \\ 0 & \text{if } f_Y(y) = 0 \end{cases}$

conditional density of X given $Y=y$

$$\Rightarrow E\big[X \,|\, Y=y\big] = \int_{-\infty}^{+\infty} x\, f_{X|Y}(x|y)\,dx$$

$$E\big[X|Y\big](\omega) = \int_{-\infty}^{+\infty} x\, f_{X|Y}\big(x|Y(\omega)\big)\,dx$$

random, depends on ω

example: X = Poisson r.v. with parameter λ
S = nr of successes in X independent Bernoulli trials with prob of success p.

Find $E(S|X)$ and $E(X|S)$

solution: :

· given $X=n$: $P(S=k \,|\, X=n) = \binom{n}{k} p^k (1-p)^{n-k}$
(Binomial distr. (n,p))

$$\Rightarrow \quad E(S \mid X=n) = n \cdot p \quad ; \quad E(S \mid X) = X \cdot p$$

- given $S=k$ let's find $E(X \mid S=k)$

$$= \sum_{n=k}^{\infty} n \cdot P(X=n \mid S=k) =$$

$$= \sum_{n=k}^{\infty} n \cdot \frac{P(X=n, S=k)}{P(S=k)} =$$

$$= \sum_{n=k}^{\infty} n \cdot \frac{P(X=n) \cdot P(S=k \mid X=n)}{\sum_{m=k}^{\infty} P(X=m) \cdot P(S=k \mid X=m)}$$

$$= \sum_{n=k}^{\infty} n \; \frac{\cancel{e^{-\lambda}} \frac{\lambda^n}{n!} \cdot \binom{n}{k} \cancel{p^k} (1-p)^{n-k}}{\sum_{m=k}^{\infty} \cancel{e^{-\lambda}} \frac{\lambda^m}{m!} \binom{m}{k} \cancel{p^k} (1-p)^{m-k}}$$

$$= \sum_{n=k}^{\infty} n \cdot \frac{\frac{[(1-p)\cdot\lambda]^{n-k}}{(n-k)!} \cdot \cancel{\frac{\lambda^k}{k!}}}{\sum_{m=k}^{\infty} \frac{[(1-p)\lambda]^{m-k}}{(m-k)!} \cdot \cancel{\frac{\lambda^k}{k!}}}$$

$$= \sum_{n=k}^{\infty} n \cdot \frac{\frac{[(1-p)\lambda]^{n-k}}{(n-k)!}}{e^{(1-p)\lambda}}$$

$$= \sum_{n-k=0}^{\infty} (n-k) \frac{[(1-p)\lambda]^{n-k}}{(n-k)!} e^{-(1-p)\lambda} + \sum_{n-k=0}^{\infty} k \cdot \frac{[(1-p)\lambda]^{n-k}}{(n-k)!} \cdot e^{-(1-p)\lambda}$$

$$= \lambda(1-p) + k$$

$$\Rightarrow \quad E(X \mid S) = \lambda(1-p) + S$$